The role of guidance in computer-based problem solving for the development of concepts of logic

Tessa Eysink, Sanne Dijkstra, Jan Kuper

Running head: GUIDANCE IN PROBLEM SOLVING

The Role of Guidance in Computer-Based Problem Solving for the Development of Concepts of Logic

Tessa, H. S. Eysink, Sanne Dijkstra, and Jan Kuper

University of Twente, The Netherlands

Faculty of Educational Science and Technology

Department of Instructional Technology

University of Twente

P.O. Box 217

7500 AE, Enschede

The Netherlands

Tel.: +31 53 4893572 / 4893606

Fax: +31 53 4892895

E-mail: Eysink@edte.utwente.nl

Eysink, T.H.S., Dijkstra, S., & Kuper, J. (2002). The role of guidance in computer-based problem solving for the development of concepts of logic. Instructional Science, 30, 307-333. This is the author's version of the article published in Instructional Science, 30 © Springer, 2002. The original publication is available at http://www.springerlink.com/content/k6n8262570178297/

Abstract

The effect of two instructional variables, manipulation of objects and guidance, in learning to use the logical connective, conditional, was investigated. Instructions for 72 first- and second year social science students were varied in the computer-based learning environment Tarski's World, designed for teaching first-order logic (Barwise & Etchemendy, 1992). Guidance, which was operationalised by giving the learners problems that guided them to all different types of basic problem situations that could be derived from the conditional and stimulated them to (physically or mentally) manipulate the geometrical objects, significantly influenced the scores from pre- to post test. Manipulation, which was operationalised by giving the learners a visual representation in which (concrete) geometrical objects could be manipulated, yielded significant results from post- to retention test. The results support the authors' view that guidance in combination with the possibility to manipulate objects in a domain, support the acquisition of knowledge and skills in that domain. Guidance leads to the students seeing all situations that are relevant for the development of new knowledge, which has a positive effect directly after instruction. The possibility to manipulate leads to the students playing around in such a way that they (systematically) experience the results of their actions by the feedback given, having an effect after two weeks of non-instruction.

Keywords: Logical Reasoning, Logic Teaching, Problem Solving, Computer-Based Instruction, Guidance, Manipulation.

Introduction

Various studies have shown that a substantial part of all students in higher education have difficulties learning and using logic. Logic is often seen as boring and difficult (e.g., Goldson et al., 1993; Fung et al., 1994) as well as meaningless (e.g., Barwise & Etchemendy, 1998). Students indicate that they especially have difficulties translating sentences of natural language into sentences of formal language, that they have problems to break down a problem into a series of smaller problems to be solved, and that they have trouble manipulating formal notations and understanding the meaning of this formal language. Fung et al. concluded that these difficulties are often caused by a lack of mathematical background and an unfamiliarity with formal notations and language.

In sum, students generally find it difficult to think in abstract terms and to handle abstract concepts as used in the general, formal and mathematical techniques of logic. This can either be caused by the fact that the cognitive development of the students has not yet reached the stage of development in which abstract concepts can evolve, or by inappropriate instruction. In his theory of cognitive development, Piaget (1970) stated that children reach the formal operational stage in which mathematical abstract concepts can develop at the age of twelve. He said, that at this age, children start to think in abstract terms. They start to generalise from (classes of) real objects to abstract notions. The child learns to think abstractly by deducing abstract concepts from *concrete* objects or situations. Learning of and thinking in abstract, formal terms without this reference to real world objects or situations is difficult, even for a large percentage of adults (Pintrich, 1990). People often develop and use abstract concepts in close connection to concrete objects in their everyday world. This then raises the questions what makes mathematical and logical abstraction so difficult for most people, and what instructional variables play a role in order to improve mathematical and logical skills. In this article, several instructional variables in logic teaching are discussed, including the connection between abstract concepts and concrete objects. A computer-based learning environment in which this connection has been realised is Tarski's World (Barwise & Etchemendy, 1992). In this environment, information from formal logical sentences is combined with information from visual representations of (concrete) geometrical objects. As a result, students can focus on the semantics of logic rather than on the syntax. A typical example of Tarski's World is given in Figure 1.

Insert Figure 1 about here

In order to study the idea that connecting formal sentences to concrete objects would indeed lead to a better understanding of the concepts of logic, Van der Pal and Eysink (1999) designed an instruction in which Tarski's World was used. They compared three conditions: (a) a condition in which situations were described by formal sentences (sentential condition); (b) a condition in which situations were described by visual representations of geometrical objects (geometrical condition); and (c) a condition in which situations were described by formal sentences as well as visual representations of geometrical objects (combined sentential and geometrical condition). The results showed that the last condition yielded the best performances, although an unpublished replication experiment (Eysink, 1998) did not confirm these results. Thus, a clear answer to the question whether the combination of formal sentences and concrete objects yielded better understanding of the concepts of logic could not yet be given. Further analyses of both studies revealed that, in the combined instruction, two instructional variables were confounded. The subject matter was represented as (a) a visualisation in which (b) geometrical objects could be manipulated. Thus, the question arose whether a possible effect was caused by combining the formal sentences with the visual representation of objects, or whether the effect was caused by the combination or interaction of both the visual representation of objects and the possibility to manipulate these objects in this visualisation. Moreover, two claims were made concerning visualisation and manipulation of objects. The first claim concerning visualisation was based on the phenomenon which Johnson-Laird (1989) called 'the mental models of discourse'. When presented sentential descriptions, imaginations or mental models arise automatically on the basis of these descriptions. The mental models make explicit the structure of the situation as it automatically arises from the sentential descriptions instead of making explicit the structure of the situation as described exactly by the sentence. As logic is precise, these slightly different imagined visualisations can cause a greater chance of making errors. Furthermore, a visual representation gives the possibility to visually check the reasoning process in the available visualisation and to easily retain the steps made in this reasoning process. Thus, it was expected that giving a visual representation of geometrical objects would lead to better results than giving a sentential description of the situation.

The second claim on manipulation of objects was derived from Piaget’s theory (1970) that learners need to act in the environment if knowledge development is to ensue. Knowledge is constructed through actions on objects in the environment. The logic-mathematical experience (in contrast to the physical experience) consists of abstracting knowledge by operating on several objects. In addition to characteristics already present, new characteristics are attributed to objects by these operations. Experience, then, refers to the relation between the characteristics attributed to the objects by manipulating them or operating on them. In this sense, knowledge is seen to be extracted from the operations as such and not from the physical features of the object, as is the case with physical experience. For instance, a child can learn the concept of order by the activity of ordering different balls to size. In this case, size is a physical feature the balls already possess, order is a logic-mathematical concept being added by actually operating on the balls. A parallel can be drawn to the logical connectives, conjunction ($\wedge$), disjunction ($\vee$), conditional ($\rightarrow$) and negation ($\neg$). Learners can learn the concepts of these connectives by actually manipulating or operating on the concrete sentences connected to each other by one (or more) of the connectives and they can see whether the combined sentence is true or false. This can be elaborated on further by creating a possibility to manipulate objects in a visual representation as described by sentences. Thus, by introducing real objects that can be manipulated, the development of understanding logical combinations of sentences can be supported. Therefore, it was expected that students having the possibility to manipulate, would profit more from the instructional material than students who did not have this possibility.

In order to study the correctness of these claims, Eysink et al. (2001) designed an experiment in which two instructional variables, *visualisation* and *manipulation* of geometrical objects were investigated. For this purpose, three conditions were administered: (a) a sentential, non-manipulation condition SN; (b) a geometrical, non-manipulation condition GN; and (c) a geometrical, manipulation condition GM. The results showed that not all the claims made were confirmed. First of all, students who were given a visual representation of geometrical objects did not profit more from the instructional material than students given a sentential description. However, this was explained by the fact that students given a sentential description had the possibility to draw their own visual representation, something the students frequently used, especially when problems became more complex. Clearly, the students preferred a visualisation to not having one. Secondly, students who were given the possibility to manipulate did not outperform students not given this possibility. A closer look and further analysis of the data showed that this result could be explained by the fact that students did not fully use the possibilities of manipulation. They played around in such a way that they were confronted with the subject matter they already understood, but that they did not confront themselves with the subject matter put in a new situation. The students were tempted to stay in familiar situations, even when given freedom to explore. They repeatedly checked things they already knew, instead of trying out something new and they headed straight for the solution without straying from their path, even if this could have resulted in a better solution.

The finding that students did not fully use the possibilities of manipulation fits in with results from other studies. Several researchers (e.g. Klayman & Ha, 1987; Dunbar, 1993, Kuhnet al., 1992) have found that learners tend to stay in familiar situations when testing a certain hypothesis, the so-called *confirmation bias*. They mainly conduct experiments that confirm their current hypothesis. They keep repeating the same situations to which they (think they) know the answers. The learners do not think of conducting experiments that will disconfirm their current beliefs. When, nevertheless, confronted with evidence that is inconsistent with their current hypothesis, they just ignore this evidence (e.g. Mynatt et al., 1978; Dunbar, 1993, Chinn & Brewer, 1993). Instead of changing their current theories, learners reject or distort these findings. One way of doing this, is stating that the anomaly in the data must have been caused by error. In order to test this, learners can conduct a series of the same experiments, and only when all these experiments give disconfirming evidence, they will start thinking of a new explanation for this finding.

Apparently, most learners are not able to regulate their own problem solving process. In complex domains the learners can have difficulties getting an overall notion of all aspects that cover the domain. When no feedback is given on the suitability of the working method used by the learner, this can lead to unsystematic behaviour resulting in missing parts of the subject matter, not seeing what is important and what are side-issues, and working without a goal or mission, even if the problem is well-defined. In addition, learners find it difficult to doubt and change existing knowledge and to increase their skills.

A solution to this problem is to regulate the problem solving process by giving *guidance* to learners. Guidance is a form of instruction that supports and facilitates the development of knowledge and skills. In the present study, guiding the learners in their problem solving process will be realised by giving problems in which the learners are directed to all possible problem situations, especially to those in which the development of new knowledge is most likely to occur, that is in situations in which problems are presented that cannot be solved with the learners' existing knowledge. In these situations critical for the development of new knowledge, the learners experience that the feedback they receive is in conflict with their existing ideas about logic. Learners can react on this infirmative feedback in different ways. They can discard or reject the feedback on the error, they can search for the error made in order to explain it, or they can try out the same, similar or other situations to experience the same or new feedback. In the latter case, the learners learn from the feedback they receive after making errors and, sooner or later, they can make inferences that lead to new knowledge or to change and extension of existing knowledge. However, the firmer the knowledge base of existing beliefs, the harder it will be to abandon or modify these beliefs and the clearer the evidence for an alternative theory should be, so the more often the learners should be confronted with the subject matter involved. Without guidance, no supervision or feedback is given on the working method of the learner. This can increase the possibility of working unsystematically without a goal. Learners can get lost, they are not forced to study every possible situation and thus they will not as often be confronted with situations that are critical for the development of new knowledge, as they can decide not to confront themselves with situations they do not know the answer to, and after having been confronted with unexpected feedback the learners can decide to avoid this problem situation. In such situations, guidance of the learner may help to develop new knowledge.

In order to investigate the influence of guidance on learning logic with or without the possibility to manipulate objects, an experiment needs to be designed in which two instructional variables, *manipulation* and *guidance*, will be varied systematically. Manipulation will be operationalised by giving the students the possibility to manipulate geometrical objects in a visual representation. By adding and removing objects, and by changing their features like size or position, students will be able to observe and (physically) experience whether the logical sentences they constructed become true or false in the visualisation and what effect certain changes in the visualisation has on the truth or falsity of the logical sentences. The group of students in the non-manipulation conditions will be given a fixed visual representation of geometrical objects in which they can evaluate whether the logical sentences are true or false.

The guidance variable will be operationalised by giving problems in which the students are directed to all possible problem situations, especially to those problem situations that are in conflict with the students' existing knowledge, so that new knowledge can be constructed. In these problem situations the students will be given tasks to (mentally or physically) manipulate the geometrical objects in the visual representation, so that the students experience what influence every change in the visualisation has on the truth or falsity of the elements of a sentence and what impact this has on the truth or falsity of the complete sentences composed of the elements combined by one or more connectives. By doing this in such a way that the different types of basic problem situations that can be derived from the conditional are all presented, the students can experience what impact every change in the visualisation has on the formal logical sentences concerning their truth or falsity. A typical example of a problem in which the students are guided, is given in Figure 2.

Insert Figure 2 about here

In the upper part of Figure 2, the student has already solved the first two tasks of the problem presented. The natural language sentences have been translated into first-order predicate language and the truth or falsity of the sentences has been checked by the student. The answer of the student appeared to be correct: both sentences are indeed true in the world given, as in both sentences the first part of the sentence as well as the second part is true. The next task of the problem concerns the guidance to a problem situation of which students are assumed to have difficulties with, namely the situation in which the first part of the sentence is false and the situation in which the objects mentioned in the first part do not exist at all. The students are stimulated to change the world in such a way that they are confronted with both problem situations (see for the changed world the lower part of Figure 2). Both sentences are checked again in this new problem situation. The existing everyday life knowledge of the student can easily lead to the conclusion that both sentences have become false by the change made. The feedback given by the computer programme shows the student that this answer is not correct. If the first part of the sentence before the conditional is false or if the objects mentioned in this part do not exist, the complete sentence is true. The guidance to these problem situations and the feedback of the programme on the answers given by the students can cause the students to change existing knowledge or develop new knowledge. When systematically confronting the students with the different kinds of situations that are possible and showing which changes in the truth or falsity of the elements of a sentence are responsible for changes in the truth or falsity of the combined sentence, the students will realise that their initial, everyday life beliefs about logic are not always consistent with the rules logic holds. It is supposed that this newly learnt knowledge will (eventually) change the students' concepts of logic.

The students who are not guided in their problem solving process, will be given problems in which they have the possibilities to create the same problem situations as the students who receive guidance, with the distinction that the students who are not guided, will not especially be directed to critical problem situations in which new knowledge is most likely to develop. Furthermore, students who are not guided, will not be stimulated to (mentally or physically) manipulate the geometrical objects in the visual representation.

Students are only supposed to profit from the possibility to manipulate if this possibility is combined with problems that guide the students through the subject matter to all possible, and especially critical situations. When guided through all different situations, learners receive feedback on every possible situation. Manipulation without guidance is supposed to be less helpful, as students in this case probably will only stay in situations they already know the answer to and they will not confront themselves with all possible situations of the subject matter. The feedback they receive on their actions, therefore, does not include all situations, and especially not those situations they do not know the answer to or situations they are uncertain about, that is situations that are critical to develop new knowledge. Students not guided through the subject matter will profit less from the material presented, independently whether they have the possibility to manipulate or not, as the advantage of manipulation is supposed not to reveal without guidance. In sum, it is supposed that students who are guided will more easily develop new knowledge, especially when they have the possibility to manipulate objects in the world.

Students not guided will profit least from the material presented, independent of the possibility to manipulate.

Method

Participants

The participants were 72 first- and second-year social science students (18 male, 54 female; mean age 19.0 years, *SD* 1.1). They volunteered for the experiment for which they were paid a fee of 50 Dutch guilders (approximately 23 euro). None of the students had any experience in computer programming or logic.

Learning environment

The computer-based learning environment Tarski’s World 4.1 for Windows (Barwise & Etchemendy, 1992) was used. Tarski’s World provides an introduction in first-order logic. In the problems to be solved, a well-defined, simple world of three kinds of geometrical objects (cubes, tetrahedrons and dodecahedrons) was used. Participants could change the size of the objects (small, middle or large) and the position of the objects (to the left of, to the right of, at the back of, in front of, and between). The learning environment consisted of four main components (see Figure 1): (a) the world module in which students could place the objects of a certain size and shape in the proper position; (b) the sentence module with the same possibilities but in formal notation; (c) the keyboard module for constructing sentences in the sentence module; and (d) the inspector module in which sentences from the sentence module could be checked to verify whether they were well-formed, correct and true/false in relation to the world in the world module.

The programme was adapted to fit the experimental design by making four versions that corresponded with the four conditions. In all four conditions, the students had to translate Dutch sentences into first-order predicate language. These sentences had to be checked in a world. In the manipulation conditions, the geometrical objects in the world could be manipulated and new worlds could be constructed. In the non-manipulation conditions, the given worlds consisted of geometrical objects that could not be manipulated. In the guidance conditions, problems were given that led the students systematically to all different types of basic problem situations that can be derived from the conditional, with the emphasis on those problem situations that were supposed to be difficult and that would lead to the development of new knowledge. Additionally, the problems stimulated the students in (mentally or physically) manipulating the geometrical objects in the world, so that changes from the existing problem situation to a new situation and their effects on the truth or falsity of the logical sentences become clear. In the non-guidance conditions, the students were not guided in their problem solving process to explore what the effect was of the different kinds of connectives, including the conditional, on the truth or falsity of complex sentences combined by these connectives and they were not stimulated to (mentally or physically) manipulate the geometrical objects in the world. They were given problems in which they had the possibilities to create the same problem situations as the students who were guided, with the distinction that the students who were not guided were not especially directed to critical problem situations. Furthermore, students who were not guided were not stimulated to (mentally or physically) manipulate the geometrical objects in the visual representation.

The following changes in Tarski's World were made: (a) the menu bar was made invisible, so that students were not able to give commands themselves; (b) the programme was translated from English into Dutch, so language could not interfere with the results; (c) worlds and sentences were automatically loaded and saved when starting and finishing a task; and (d) in the non-manipulation conditions worlds were given which could not be manipulated by the students.

The instruction accompanying Tarski's World was provided in the browser of Netscape Communicator 4.06. The changes in the browser were as follows: (a) the menu options were disabled, so that students could not navigate completely freely in the browser nor surf on the internet; (b) the browser was linked to Tarski's World, so commands in one programme resulted in actions in the other programme.

Learning materials

The subject matter involved logic, in particular the language of *first-order logic*. Logic is the science of (both human and machine) reasoning, which tries to discover conditions by which conclusions are justified and correct. In order to reach precision, contemporary logic is presented in a formal, mathematical way. The (formalised) language of first-order logic contains *names* (*a*, *b*, …, *x*, *y*, …) to denote individual objects, and *predicates* to express properties of objects and relations between these objects. For example, *Large* is a predicate and *Large*(*a*) says that object *a* is large. Likewise, *Larger*(*a*, *b*) says that object *a* is larger than object *b*, and *Between*(*x*, *y*, *z*) expresses that object *x* is positioned between the objects *y* and *z*. Expressions of this form are called (*elementary*) *propositions*. They can be combined into more complex propositions by the following *connectives*: *negation* ($\neg$, not), *conjunction* ($\wedge$, and), *disjunction* ($\vee$, or), and *conditional* ($\rightarrow$, if … then …, sometimes called *implication*). For example, the formula *Larger*(*a*, *b*) $\rightarrow$ *Smaller*(*b*, *a*) says that "*if a* is larger than *b*, *then b* is smaller than *a*".

Propositions express states of affairs about the world and they can either be *true* or *false*. The connectives are *truth functional*, that is, the truth or falsehood of a complex proposition is completely determined by the truth or falsehood of the propositions of which it is composed, as described by the *truth tables* given in Table 1.

Insert Table 1 about here

So, for example, if p is true, then $\neg p$ is false. And also, if p is true and q is false, then $p \wedge q$ is false, but $p \vee q$ is true. Notice that $p \rightarrow q$ is only false if p is true and q false. In all other cases $p \rightarrow q$ is true.

As mentioned above, the formal language contains names to denote objects. There are two kinds of names: constants (a, b, …) denote specific objects (it is assumed that one knows which objects are denoted), whereas variables (x, y, …) denote arbitrary objects (one does not know which objects are denoted). In combination with variables two *quantifiers* can be used, the universal quantifier ($\forall$, for all) and the *existential* quantifier ($\exists$, there is at least one). For example, $\forall x\ Large(x)$ says that *all* objects x are large, whereas $\exists x\ Large(x)$ expresses that there is *at least one* object x which is large. A more complicated example is $\forall x\ (Cube(x) \rightarrow Large(x))$. Literally, this formula says that “every object x is large, if it is a cube”. In everyday language this is normally said as “every cube is large”, or “all cubes are large”. In this formula all occurrences of the variable x are said to be within the *scope* of the quantifier, and all occurrences of x are *bound* by the quantifier. If a variable is not bound by a quantifier, it is called *free*. If a formula contains a free variable, one cannot know whether this formula is true or false, since it is not known which object is denoted by this variable. For example, we do not know whether $Large(x)$ is true or false, since we do not know which object is denoted by x. On the other hand, when x is bound by a quantifier, as in $\forall x\ Large(x)$, we can know whether this formula is true or false, as x

ranges over all objects now, and for each individual object it can be checked whether it is large or not. In general, one can determine whether a formula is true or false whenever that formula does not contain free variables. Such formulas are called *sentences*.

During the verbal instruction and general training, students received information about this basic knowledge of logic. During the exercises of the experimental phase, the *conditional* was emphasised.

Tests, problems and questionnaires

To measure the students' knowledge of the meaning of the conditional ($p \rightarrow q$), transfer tests were administered. These tests were divided into three parts with a total of 27 items. The first part, four puzzles consisting of 14 items, tested the knowledge of the students on the conditional within the domain of Tarski's World. The students were given a certain situation which was either a world with several geometrical objects or a sentential description. In addition, several statements in natural language were given. The students' task was to tell whether these statements were true or false in the given situation or whether it was valid or not for certain conclusions to be drawn from the statements. Figure 3 shows a typical example of the items given in the first part of the transfer test.

Insert Figure 3 about here

The second part, consisting of nine items, tested the knowledge of the students on the conditional within an everyday life setting. The students were given a statement and a conclusion about an everyday life situation and they had to decide whether the conclusion was true or false or whether you could not tell from the information given. Figure 4 shows a typical example of the items given in the second part of the transfer test.

Insert Figure 4 about here

The third part of the test consisted of four items testing the conditional. These items involved the Wason Selection Task (1966). Figure 5 shows a typical example of an item of this task, as used in the experiment. One abstract Wason (card)task, two concrete, non-arbitrary Wason tasks and one near-transfer Wason task in a Tarski's World setting were used in the third part of the transfer test.

Insert Figure 5 about here

The students had to complete three comparable versions of the transfer test, namely a pre-, post- and retention test. These tests were designed to measure the knowledge gained after the various instructions. All students received the same tests.

During the experiment, students were given four different types of problems: (a) investigation problems type I: problems in which the students had to investigate all basic combinations of truth values of several self-made sentences in a self-made or given world; (b) investigation problems type II: problems in which the students (mentally or physically) had to investigate all basic combinations of truth values of one self-made sentence in several self-made or given worlds; (c) production problems: problems in which the students had to produce a world in which given sentences were true or false or in which the students had to produce sentences that were true or false in a given, fixed world; (d) modification problems: problems in which the students had to modify (explicitly or implicitly) a given world, in such a way that given sentences would become true or false.

Furthermore, two questionnaires were administered. The first concerned the former education of the students on mathematics, programming and logic. The second questionnaire was an evaluation of the instruction in combination with Tarski's World.

All tests, problems to be solved and questionnaires were administered on the computer.

Log files

Two log files were generated during the experiment. The first log file logged all the actions of the students while working in Tarski's World. It logged the status of the sentences and the matching world every time the students checked this combination on Well Formed Formula (WFF), sentence and/or truth. The second log file logged all the actions of the students while working in the browser. This, among others, concerned the answers of students on two questionnaires, the answers of students on the transfer tests, and the time registration.

From these log files, the mean number of checks per student could be computed. As the subject matter concerned the *conditional* with the general format $p \rightarrow q$, these checks could be divided into the four possible truth-falsity combinations ($1 \rightarrow 1$, $1 \rightarrow 0$, $0 \rightarrow 1$, $0 \rightarrow 0$, in which 1 = true, and 0 = false).

Design and procedure

Experimental conditions. A full-factorial design of four conditions was administered: (a) a manipulation / guidance condition Man+Guid+; (b) a manipulation / non-guidance condition Man+Guid-; (c) a non-manipulation / guidance condition Man-Guid+; and (d) a non-manipulation / non-guidance condition Man-Guid-. The subjects were randomly assigned to one of the four conditions. In all conditions, the students were given Dutch sentences, which they had to translate in first order predicate language. In the Man+Guid+ condition, the students had to construct geometrical worlds in which they had to check the sentences to be true or false. The problems presented led them through the subject matter to all different types of basic problem situations that can be derived from the conditional, especially to those problem situations in which the students were supposed to develop new knowledge. Moreover, the students were specifically directed towards manipulating certain objects in order to see what effect this change had on the truth values of the sentences. In the Man+Guid- condition, the students had to construct geometrical worlds in which they had to check whether the sentences were true or false. The students were given problems that instructed them to manipulate objects in the world to see what happened with the truth values of the sentences. By doing this the students were free to explore and manipulate whatever they wanted without guiding them in this process. In the Man-Guid+ condition, the students had to check whether the sentences were true or false in a given, fixed geometrical world. Problems presented led them through the subject matter to all different types of basic problem situations that can be derived from the conditional, especially to those problem situations in which the students were supposed to develop new knowledge. Moreover, the students were specifically directed towards imagining what would happen with the truth values of the sentences, if they changed the world in a certain way. The students were also given feedback on this. In the Man-Guid- condition, the students had to check whether the sentences were true or false in a given, fixed geometrical world. The students were given problems in which they were instructed to construct sentences that could be checked in the given world. By giving this possibility the students were free to explore whatever they wanted without being guided in this process and without the possibility to manipulate objects.

Procedure. The experiment was held in three consecutive sessions; the first (pre test, instruction and general training) and the second (exercises and post test) on two successive days and the third (retention test) two weeks later.

The first session started with an introduction after which the students had to complete a questionnaire about their previous education in mathematics and logic. This questionnaire was followed by a pre test consisting of three parts with a total of 27 items measuring the knowledge of the subjects of several aspects of the conditional. Successively, the subjects received a verbal instruction of 30 minutes in which they had an introductory course into first-order logic, as used in Tarski's World. This instruction gave the subjects an idea of what logic can be used for, what Tarski's World can do, what logic operators and quantifiers are available, how these operators and quantifiers can be used and what truth and falsity means. This, together with some examples was the knowledge the students were equipped with. After the instruction, the students received a general training of about two hours depending on the condition to which they were assigned. During the training, the students had to complete eleven assignments in which they learnt to work with Tarski's World and with the logic operators. For this, model progression was used, an idea introduced by White and Frederiksen (1990). One of the general principles of model progression is to structure the rich information source and to keep the environment manageable by not introducing too many ideas at one time. Model progression entails starting with a simplified version of a model and gradually offering more complex versions of the model. In this case, the model was the field of predicate logic. The concepts were introduced in the following order: (a) predicates and constants, (b) connectives and parentheses, (c) quantifiers and variables, and (d) conditional. If the students had any questions, assistance was given by one of the experimenters present.

The second session started with twelve exercises that had to be solved by the students in the four conditions. The first two problems were presented to refresh the knowledge acquired the day before. Consequently, ten exercises addressed the conditional. In all the exercises students were asked to translate Dutch sentences into first-order logic and to check the truth of the sentences in the (given or self-constructed) world of geometrical objects. After the students had completed the exercises, a post test was administered. In this post test, the students were again tested on their knowledge of the conditional. The post test consisted of the same type of items as used in the pre test. The second questionnaire in which the instruction and Tarski's World was evaluated, was also administered during this second session. Two weeks after the experiment, the students had to return for the retention test. This test consisted of comparable items as were used in the pre- and post test.

Results

Reliability

The reliability of pre-, post- and retention test, as measured with Cronbach's α, was $\alpha = .43$; $\alpha = .74$; and $\alpha = .78$ respectively. Deleting items from the test did not lead to significant higher reliabilities.

Pre-, post- and retention tests

Table 2 shows the means and standard deviations on the pre-, post- and retention test for the four conditions Man+Guid+, Man+Guid-, Man-Guid+ and Man-Guid-. The maximum score was 27. A significant mean effect was found from pre- to post- to retention test ($F(2, 67) = 21.71$, $p < .001$) and an interaction was found between the three tests and the four conditions (F(6,

136) = 5.02, $p < .001$). The scores of the students in condition Man+Guid+ increased significantly from pre- to post test ($F(1, 17) = 18.35$, $p < .001$) and from post- to retention test ($F(1, 17) = 9.03$, $p < .01$). The scores of the students in condition Man-Guid+ increased significantly from pre- to post test ($F(1, 17) = 19.17$, $p < .001$), but decreased significantly from post- to retention test ($F(1, 17) = 5.28$, $p < .05$). Scores on pre-, post- and retention tests did not increase significantly for condition Man+Guid- ($F(2, 34) = 1.54$, $p > .05$) and condition Man-Guid- ($F(2, 34) = 1.01$, $p > .05$).

Insert Table 2 about here

The scores of the students were divided into the scores on the three different parts of the transfer tests (Tarski's World items, everyday life items and Wason tasks, see Figure 6). In part 1, the Tarski's World items, a significant mean effect was found from pre- to post- to retention test ($F(2, 67) = 13.29$, $p < .001$) and an interaction was found between the three tests and the four conditions ($F(6, 136) = 2.91$, $p < .05$). The scores on pre- to post test increased significantly for the two guidance conditions, Man+Guid+ ($F(1, 17) = 13.73$, $p < .01$) and Man-Guid+ ($F(1, 17) = 13.91$, $p < .01$). All other differences were non-significant.

In part 2, the everyday life items, again a significant mean effect was found from pre- to post- to retention test ($F(2, 67) = 13.38$, $p < .001$) and an interaction was found between the three tests and the four conditions ($F(6, 136) = 4.34$, $p < .001$). Scores on pre- to post test increased significantly for the two guidance conditions, Man+Guid+ ($F(1, 17) = 11.98$, $p < .01$) and Man-Guid+ ($F(1, 17) = 18.98$, $p < .001$). Scores on post- to retention test again increased for condition Man+Guid+ ($F(1, 17) = 8.12$, $p < .05$), but decreased for condition Man-Guid+ ($F(1, 17) = 5.39$, *p* < .05). The scores in the two non-guidance conditions, Man+Guid- and Man-Guid-, did not differ from pre- to post- to retention test.

In part 3, the Wason tasks, a significant mean effect was found from pre- to post- to retention test ($F(2, 67) = 3.60$, $p < .05$) and an interaction was found between the three tests and the four conditions ($F(6, 136) = 2.19$, $p < .05$) caused by a significant increase of scores from pre- to post- to retention test for condition Man+Guid+. No other differences were found for the other conditions between scores on pre- to post- to retention tests.

Insert Figure 6 about here

Process data

In Table 3, a summary is given of the mean number of checks of students working on the experimental exercises in Tarski's World. As can be seen, students in condition Man+Guid+ used significantly more checks on all sentences than students in the condition Man+Guid- ($F(1, 34) = 12.07$, $p < .001$), students in the condition Man-Guid+ ($F(1, 34) = 14.67$, $p < .001$) and students in the condition Man-Guid- ($F(1, 34) = 93.50$, $p < .001$). Students in condition Man-Guid- used significantly fewer checks on all sentences than students in the condition Man+Guid- ($F(1, 34) = 38.13$, $p < .001$) and students in the condition Man-Guid+ ($F(1, 34) = 97.57$, $p < .001$).

Furthermore, students in the non-guidance conditions, Man+Guid- and Man-Guid-, checked most in the situations to which they already knew the answer ($1 \rightarrow 1$ and $1 \rightarrow 0$). In unfamiliar situations ($0 \rightarrow 1$ and $0 \rightarrow 0$), they checked significantly less (a ratio of 3.0 to 1 and 4.8 to 1 resp.). Students in the guidance conditions, Man+Guid+ and Man-Guid+, spread their checks equally over all possible situations (a ratio of 1.3 to 1 and 1.0 to 1 resp.).

Insert Table 3 about here

Discussion

Reliability

The fact that the reliability of the pre test is lower than the reliabilities of the post- and retention test is explained by the small number of correct answers on the pre test (M = 14.07, SD = 2.52). Apparently, the pre test was difficult, so that the students may have been guessing when answering the items, which has a negative influence on the reliability of the test.

Overall Learning Results

The overall scores on the pre-, post- and retention tests clearly show that students who received problems in which they were directed to situations in which new knowledge could be developed (i.e. students in the conditions Man+Guid+ and Man-Guid+), profited from this instructional support. These students were able to solve significantly more logic problems correctly on the post- and retention tests in comparison to the pre test. Students having the possibility to manipulate even increased their scores from post- to retention test, whereas the scores of the students having no possibilities to manipulate decreased again after two weeks, but were still higher than the scores on the retention test of the students who were not guided. The latter students (i.e. students in the conditions Man+Guid- and Man-Guid-) did not develop new knowledge concerning the conditional, whether they had the possibility to manipulate or not. Having the possibility to manipulate is shown to be only effective when combined with guidance. The process data confirm the hypothesis that this is caused by the students not using the possibilities of manipulation when they are given freedom to explore whatever they want. Only guidance will induce students to explore all possible situations, so they also experience what happens with the truth values of the sentences in situations they do not know the answer to or they are uncertain about. The results show that guidance in combination with the possibility to manipulate objects in a domain supports the development of new knowledge and skills in that domain. Guidance leads to the students seeing all possible situations including the ones in which new knowledge can be developed, which has positive effects directly after instruction. The possibility to manipulate leads to the students playing around in such a way that they (systematically) experience the results of their actions by the feedback given on the same problems in different worlds. This 'seeing with your own eyes' and (physically) experiencing what happens with the truth values of the logical expressions when the world is changed is assumed to lead to making inferences, to lead to a better memorisation of the subject matter and to a continuing reflection on the logic problems in the interval between administering the post- and retention test. The students needed time for integrating their newly developed knowledge with their existing knowledge, as the positive effect of this variable appears after two weeks of non-instruction.

Tarski’s World puzzles: near-transfer

The scores of the students on the items in the first part of the transfer tests, the Tarski’s World puzzles, show that near-transfer is possible if (and only if) the students are given guidance, independently of the possibility to manipulate. The scores of the students who were guided (i.e.

scores of the students in the conditions Man+Guid+ and Man-Guid+) increased significantly from pre- to post test, showing that, in near-transfer tasks, students directly profit from this form of guided instruction. In addition, the scores of the students who were guided, remained at the same level from post- to retention test, showing that, in near-transfer tasks, the subject matter learnt is well integrated into the students' knowledge base to correctly solve near-transfer tasks after two weeks of non-instruction, but that the subject matter is not reflected upon anymore after instruction. A possible cause for this finding is that, during the near-transfer test, the students recognise the subject matter from the material presented during instruction, but after instruction they are not confronted anymore with similar situations. Furthermore, the scores of the students in the non-manipulation conditions did not differ from the scores in the comparable manipulation conditions. Apparently, manipulation did not influence the scores of the students on the near-transfer tests.

Everyday-life puzzles: medium-transfer

The scores of the students on the items in the second part of the transfer tests, the everyday-life puzzles, show that medium-transfer is possible if (and only if) the students are given guidance and they show that this transfer continues after an instruction-free period when combined with manipulation. The increase of the scores from pre- to post test of the students who were guided (i.e. students in the conditions Man+Guid+ and Man-Guid+) show that students also directly profit from this form of guided instruction in medium-transfer tasks. However, from post- to retention test this guided instruction interacts with the possibility to manipulate. Students having this possibility continue reflecting upon the subject matter in the instruction-free period. They start to apply their newly developed knowledge in everyday life situations, which results in increasing scores from post- to retention test on the everyday-life puzzles in the experiment. The scores of students who were guided but who did not have the possibility to manipulate decreased from post- to retention test. The first possible explanation for this is that the students did not reflect upon the subject matter in the instruction-free period and that they did not completely retain all knowledge acquired, because they did not have the possibility to manipulate and thus could not (physically) experience their actions. The second explanation is that the students did reflect upon the subject matter in the instruction-free period, but that they did not fully retain the newly learnt knowledge caused by the lack of possibilities to manipulate. This may then have led to confusion resulting from an interplay between the newly acquired knowledge and the existing everyday life knowledge coming forward again in everyday life situations. As the authors cannot explain why the possibility to manipulate would lead to reflection and the lack of the possibility to manipulate would not lead to reflection, they assume that the second explanation is the most probable one.

Wason tasks: far-transfer

The scores of the students on the items in the third part of the transfer tests, the Wason tasks, show that far-transfer is still difficult, even after instruction in which the students are guided in their problem solving process. The scores of the students in all conditions are low, although there is a trend from post- to retention test of increasing scores of the students in the Man+Guid+ condition. It is possible that students in this condition continued reflecting on these puzzles, but that integration of their newly developed knowledge was not yet completely established for far-transfer tasks.

Main findings and conclusion

Students who were not guided during instruction were not able to control their problem solving process themselves. They stayed in familiar situations and did not confront themselves with situations they (thought they) did not know the answer to or they were uncertain about. The result was that the students did not develop new, or change existing knowledge concerning the conditional.

Students who were guided through the subject matter were confronted with all different types of basic problem situations that could be derived from the conditional, especially those situations that were supposed to lead to the development of new knowledge. The students could infer the behaviour of the conditional resulting in confirming existing everyday life knowledge that was correct and learning and developing new knowledge. Apparently, the students were able to use this knowledge directly after instruction when solving near-transfer and middle-transfer problems. Far-transfer problems still yielded difficulties. It is assumed that, in the two-week instruction-free period, the students tried to cognitively elaborate on their newly developed knowledge resulting in different outcomes. First of all, the students were most probably not confronted anymore with near-transfer problems. Therefore, they did not reflect on these problems, but the knowledge resulting from these problems became integrated well enough to yield similar scores on the retention test compared to the post test. Secondly, the students were still confronted with the middle-transfer problems in their everyday life. Students who, in addition to being guided, had the possibility to manipulate during instruction retained their newly learnt knowledge correctly and continued to reflect on this knowledge. Students who did not have this possibility might not have retained the newly learnt knowledge correctly, leading to confusion between this newly acquired knowledge and the incorrect everyday life knowledge, resulting in lower scores on the retention test. Thirdly, students were still not able to solve the far-transfer problems although it seems that the students in the guided, manipulation condition also reflected on these problems leading to a small increase of scores on the Wason tasks.

The findings of this study give reasons to reflect on the design of learning environments. Questions arise, such as, how "open" should the problem solving situations be? To what extent can learners be left on their own? To what extent do they need at least some kind of instructional support or guidance? Can learners be held responsible at all for their own development of knowledge and skills? Is the extent of required guidance age-bound? What influence do these open learning environments have on the quality and quantity of the subject matter learnt? And what about the effects in the long run? All these (and other) questions are important for the design of learning environments and for the design of instruction. The results of the present study can be used as a start in an attempt to answer some of them, of course with the restriction in mind that no social interaction was possible during the experiment. This study shows that giving learners complete freedom to explore in an open problem solving environment does not easily lead to new knowledge development. In such open situations, learners continue to use their (often incorrect) existing knowledge and they do not get any feedback on the way they approach the subject matter. Consequently, they do not confront themselves with all different types of basic problem situations and thus do not receive any feedback on these critical situations in which new knowledge could be developed. Therefore, learners need to be guided in their problem solving processes. In addition, guidance should stimulate learners to operate on objects so that the correctness of their existing knowledge is (physically) experienced in the world given. This results in making inferences about what was correct and what was incorrect leading to changes of incorrect knowledge or development of new knowledge and integrating this with correct existing knowledge, leading to better long-term effects.

References

Barwise, J., & Etchemendy, J. (1992). *The language of first-order logic: Including the Microsoft Windows program Tarski's World 4.0 for use with IBM-compatible computers.* Stanford, CA: CSLI.

Barwise, J., & Etchemendy. J. (1998). Computers, visualization, and the nature of reasoning. In T. W. Bynum & J. H. Moor (Eds.), *The digital phoenix: How computers are changing philosophy* (pp. 93-116). London: Blackwell.

Bruner, J.S. (1961). The act of discovery. *Harvard Educational Review, 31*(1), 21-32.

Chinn, A.C. & Brewer, W.F. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1), 1-49.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17,* 397-434.

Eysink, T.H.S., Dijkstra, S., Kuper, J. (2001). Cognitive processes in solving variants of computer-based problems used in logic teaching. *Computers in Human Behavior, 17*(1), 1-19.

Fung, P., O'Shea, T., Goldson, D., Reeves, S., & Bornat, R. (1994). Why computer science students find formal reasoning frightening. *Journal of Computer Assisted Learning, 10,* 240-250.

Goldson, D., Reeves, S., & Bornat, R. (1993). A review of several programs for the teaching of logic. *The Computer Journal*, *36*(4), 373-386.

Johnson-Laird, P. N. (1989). Mental Models. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 469-499). Cambridge, MA: The MIT Press.

Jonassen, D.H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology: Research and Development*, *39* (3), 5-14.

Klayman, J. & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Kuhn, D., Schauble, L. & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*(4), 285-327.

Mynatt, C.R., Doherty, M.E. & Tweney, R.D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology, 30*, 395-406.

Piaget, J. (1970). *Epistémologie génétique.* [Genetic epistemology]. Paris: Presses Universitaires de France.

Pintrich, P. R. (1990). Implications of psychological research on student learning and college teaching for teacher education. In W. R. Houston, M. Haberman & J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 826-857). New York: MacMillan.

Van der Pal, J., & Eysink, T. (1999). Balancing situativity and formality: The importance of relating a formal language to interactive graphics in logic instruction. *Learning and Instruction, 9*, 327-341.

Table 1

Truth tables of the four connectives negation (¬), conjunction (∧), disjunction (∨), and conditional (→)

Elementary Propositions

p	$\neg p$
0	1
1	0

Complex Propositions

p	q	$p \wedge q$	$p \vee q$	$p \rightarrow q$
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	1

Note. p and q denote arbitrary propositions. 0 = false; 1 = true.

Table 2

Means and Standard Deviations for Each Condition on Pre-, Post- and Retention Tests

	Condition			
Test	Man+Guid+	Man+Guid-	Man-Guid+	Man-Guid-
Pre test				
M	13.22	14.17	14.11	14.50
SD	2.18	3.37	2.14	1.98
Post test				
M	17.50	14.49	18.22	15.50
SD	3.90	4.17	4.15	3.17
Retention test				
M	19.22	15.33	17.33	15.11
SD	4.07	4.43	4.21	3.31

Note. Maximum score = 27. Man+Guid+ = manipulation, guidance condition; Man+Guid- = manipulation, non-guidance condition; Man-Guid+ = non-manipulation, guidance condition; Man-Guid- = non-manipulation, non-guidance condition.

Table 3

Summary of Process Data of Students Working in Tarski's World

	Condition			
	Man+Guid+	Man+Guid-	Man-Guid-	Man-Guid-
# checks				
all sentences[a]	73.7	55.4	56.9	31.4
$1 \rightarrow 1$[b]	25.3	24.7	18.8	14.4
$1 \rightarrow 0$[b]	16.7	17.6	10.2	11.6
$0 \rightarrow 1$[b]	13.2	4.7	10.2	2.3
$0 \rightarrow 0$[b]	13.7	7.4	10.9	2.6
0 (na) $\rightarrow$ x[b]	4.7	1.1	6.8	0.5

Note. Man+Guid+ = manipulation, guidance condition; Man+Guid- = manipulation, non-guidance condition; Man-Guid+ = non-manipulation, guidance condition; Man-Guid- = non-manipulation, non-guidance condition.

[a] mean number of checks over all sentences during the experimental exercises.

[b] mean number of checks per student made on five possible checks of the conditional ($p \rightarrow q$) during the experimental exercises (1= true; 0 = false; 0 (na) = empty set; x = unknown truth value).

Figure Captions

Figure 1. Tarski's World: world module, sentence module, inspector module and keyboard module.

Figure 2. A typical example of a problem situation in which the students are guided in their problem solving process.

Figure 3. A typical example of the items given in the first part of the transfer tests measuring knowledge about the conditional within the domain of Tarski's World. The answer are 'yes, this cube is large' and 'no, there is no tetrahedron'.

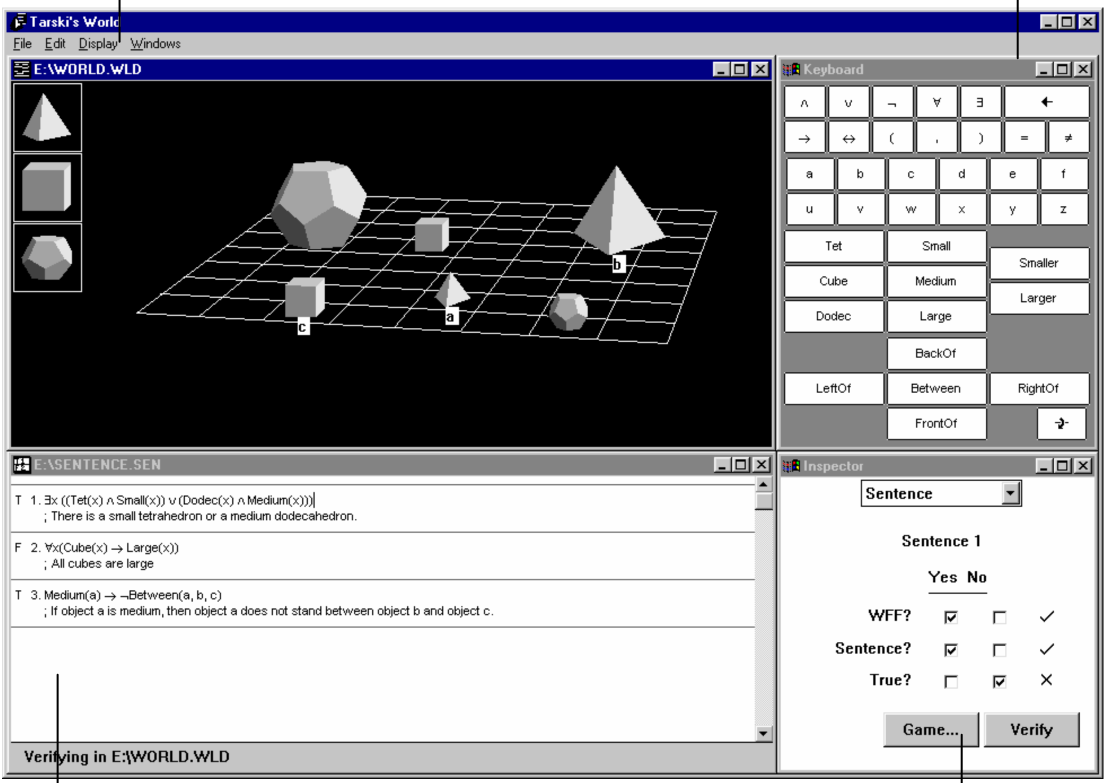
Figure 4. A typical example of the items given in the second part of the transfer tests measuring knowledge about the conditional in an everyday life setting. The answer is 'yes, the statement is true'.

Figure 5. A typical example of the items given in the third part of the transfer tests measuring knowledge about the conditional with a Wason Selection Task (Wason, 1966). The answer is to turn over the cards with the E and the 5 written on them.

Figure 6. The scores of the students on the pre-, post and retention test of the Tarski's World items, the everyday-life items and the Wason items as a function of the conditions of the students.

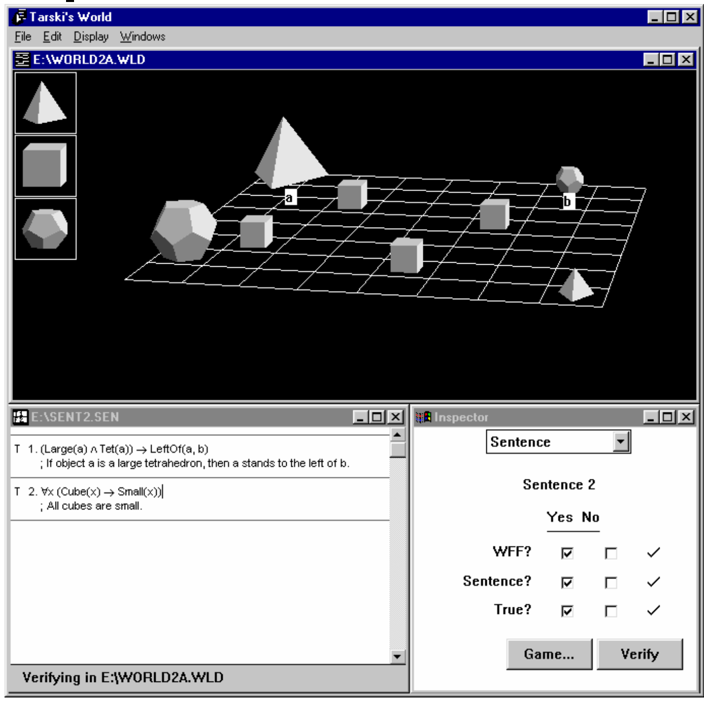
World module: the student can construct a world of geometric objects by mouse clicking on the objects on the left of the screen. The objects can vary in shape (Tet, Cube, Dodec), size (Small, Medium, Large) and position (LeftOf, RightOf, BackOf, FrontOf, Between).

Keyboard module: the student can construct logical formulae by mouse clicking on the keys.



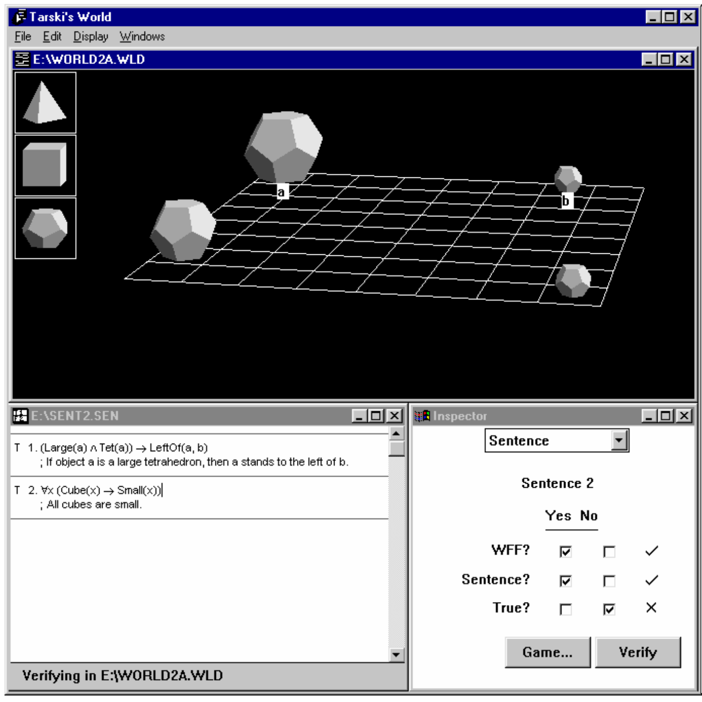
Sentence module: the sentences the student is constructing by using the keyboard module appear in this screen. The computer ignores the semicolons followed by sentences in English. By giving T(rue)s and F(alse)s, the computer gives feedback about the truth value of the sentences in the world given in the world module.

Inspector module: the student can check whether a logical sentence is syntactically correct (WFF?), whether all variables are bound to a quantifier (Sentence?) and whether the formula is true (True?) in the world given in the world module, by selecting a box. The computer gives immediate feedback (✔, ✖).



Assignment:

1. Translate the two given sentences into correct logical expressions.
2. Predict whether the sentences are true or false in the given world and verify this
3. Change the world in such a way that:
   - no cubes are left
   - all tetrahedrons are changed into dodecahedrons
4. Predict again whether the sentences are true or false in the changed world.



Assignment:

1. Translate the two given sentences into correct logical expressions.
2. Predict whether the sentences are true or false in the given world and verify this
3. Change the world in such a way that:
   - no cubes are left
   - all tetrahedrons are changed into dodecahedrons
4. Predict again whether the sentences are true or false in the changed world.

Given are two **true** statements:

1. If there is a cube, then this cube is large.
2. If there is a tetrahedron, then this tetrahedron is small.

A. Given is that there is a cube.

- Is this cube large?

  ○ yes, this cube is large

  ○ no, this cube is not large

B. Given is that there are no small objects.

- Is there a tetrahedron?

  ○ yes, there is a tetrahedron

  ○ no, there is no tetrahedron

Given is statement S1.

S2: **If I go on holiday to Italy, then I go by plane.**

This year, I'm going to Spain and not to Italy. Does this make statement S1 true, false or can't you tell?

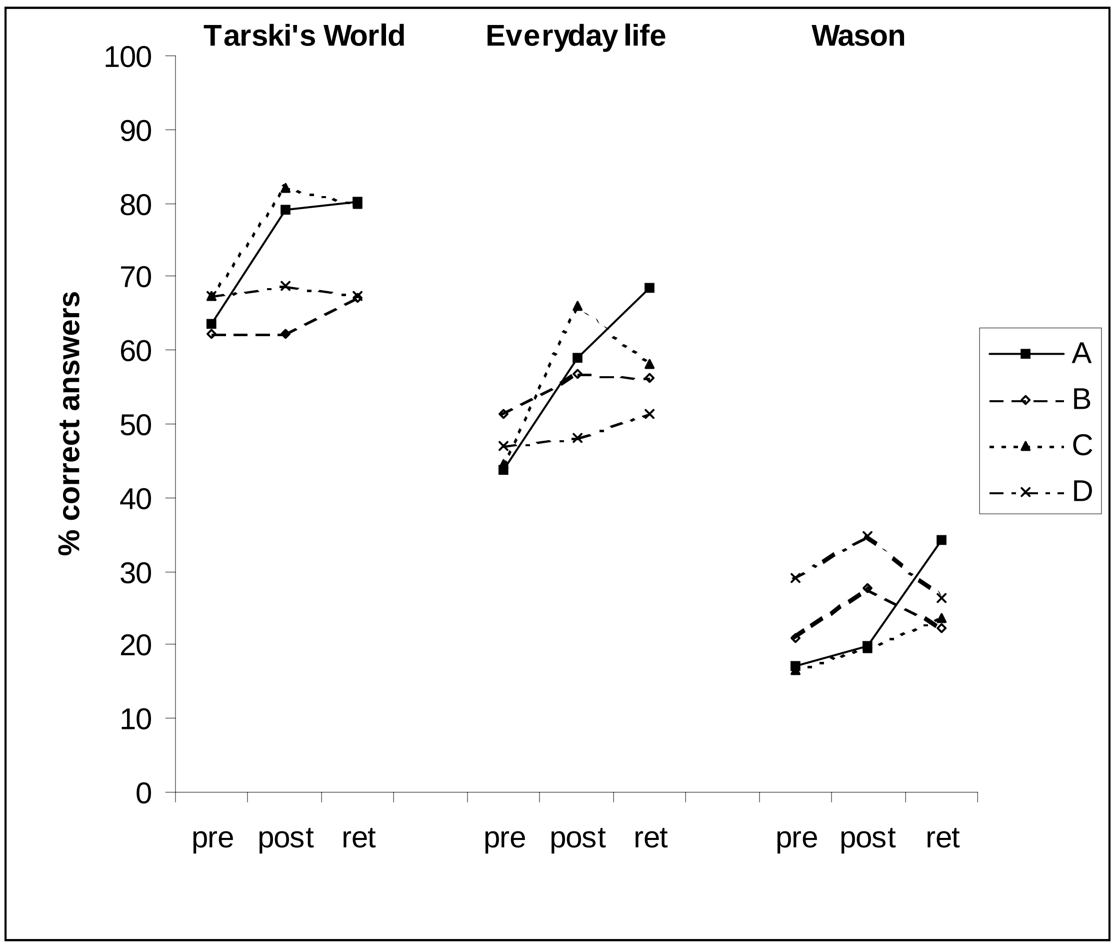
- Statement S1 is true
- Statement S1 is false
- You can't tell whether statement S1 is true or false

Below are four cards. On each card there is always a *letter* on one side and a *number* on the other side. A card never contains two numbers or two letters. There is a rule that says:

If there is an E on one side, then there is a 4 on the other side.

Which cards do you have to turn over in order to decide whether the rule is true or false?

E	K	4	5



Tarski's World
Everyday life
Wason
% correct answers
100
90
80
70
60
50
40
30
20
10
0
pre post ret
pre post ret
pre post ret
A
B
C
D