M. Fjeld, S. Guttormsen Schär, D. Signorello, H. Krueger (2002): Alternative Tools for Tangible Interaction: A Usability Evaluation.
In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, pp. 157-166.

Alternative Tools for Tangible Interaction: A Usability Evaluation

Morten Fjeld, Sissel Guttormsen Schär, Domenico Signorello, Helmut Krueger
IHA, Swiss Federal Institute of Technology
Clausiusstr. 25, CH-8092 Zurich, Switzerland
morten@fjeld.ch, guttormsen@iha.bepr.ethz.ch, domenico.signorello@wcom.ch, krueger@iha.bepr.ethz.ch

Abstract

In this work we compare an in-house designed Tangible User Interface (TUI) with three alternative single-user tools through an empirical investigation. These three alternative tools are a 3D physical, a 2D cardboard, and a mathematical tool. We expected the 3D physical to perform best, followed by the TUI, the 2D cardboard, and the mathematical tool. A pilot study was first carried out, the results of which were used to design a major experiment. Participants solved the same positioning problem, each using one of the four tools. The mathematical tool was not used in the experiment. In the experiment, trial time, number of user operations, learning effect in both preceding variables, and user satisfaction were measured. The 3D physical tool significantly outperformed the 2D cardboard tool. It also outperformed the TUI, but only in user satisfaction. This justifies the value of researching TUI systems and carrying out usability studies with such systems.

Keywords

Augmented Reality, Tangible Interfaces, Input Devices, Direct Manipulation, Three-Dimensional Interfaces, Cognitive Support, User Evaluations.

INTRODUCTION

This paper reports on an empirical evaluation of a Tangible User Interface (TUI) by comparing the TUI with three alternative tools for single-user problem solving. The TUI employed is called BUILD-IT [3, 4] and was designed to support multi- and single-user spatial planning activities. Following the tradition of Augmented Reality, projected light replaces the use of screens as the output medium. Grouped around a table and employing tangible physical bricks, users can select and manipulate virtual models within the scene that they are planning. Employing computer vision technology, the system enables its users to cooperate in a virtual environment for planning a real-world project, such as a room, a school, or a factory (Fig. 1). The usability of the BUILD-IT system may be reflected in its support for 'intuitive' and 'direct' interaction. In this work, quantifying such usability, the effect of *cognitive support* underlying the practice of single user spatial planning was investigated. This effect should be valid across different forms of systems and apply to single and to multiple users. In a pilot study and a major evaluation, we compared BUILD-IT against alternative tools, being a 3D physical, a 2D cardboard, and a mathematical tool. The mathematical tool was not used in the experiment. The specific task employed was a spatial positioning problem requiring an analytical work-style. Specific actions required for task solution were concurrent positioning and rotation. In the experiment, we found significant differences between the tools in trial time, number of user operations, and in user satisfaction.

Figure 1: The TUI evaluated is called BUILD-IT. It offers a plan and a side view (left). The system is operated in the plan view using bricks (right).

COGNITIVE SUPPORT

The aids offered by a tool have been conceived of by Gibson [5] in terms of *affordances*. For example, a pencil is held in such a way that it fits the hand, ignoring less appropriate ways that it might be grasped. "The pencil *affords* being held in this way as a result of its length, width, weight, and texture, all with respect to the size, configuration, and musculature of our hand" [9]. Most of these properties and relationships are visible. The possible interaction with an object or an environmental feature can be determined simply by looking at it. Since the use of alternative tools to solve an analytical problem is investigated, the affordances of those tools may be seen as cognitive support [2]. One way to understand the facets of cognitive support is found in the decision support techniques of Zachary [11]. Zachary suggested six decision support techniques, but only four of them are considered as relevant to guide the tool design reported. Zachary defines these four techniques as follows[1]:

---

[1] The remaining two techniques, defined by Zachary [2] as follows: *process models*, assisting in projecting the future course of complex processes, and *judgement amplification/refinement techniques*, helping in quantification and debiasing of heuristic judgements, are less relevant since we employ a positioning problem in a static environment operating on one scale at a time.

1: *Representation aids*, assisting in expression and manipulation of a specific representation of a decision problem.

2: *Information control techniques*, helping in storage, retrieval, organisation, and integration of data and knowledge.

3: *Analysis and reasoning techniques*, supporting application of problem-specific expert reasoning procedures.

4: *Choice models*, supporting integration of decision criteria across aspects and/or alternatives.

While the design of the tools was guided by the four decision support techniques [11], the comparison of the same tools was based on the cognitive support they provided [2].

Tool Design

The first decision support technique **(1)** guides us to focus on problem representation, which was realized through the design and comparison of alternative planning tools. The second technique **(2)** guides us to design tools that can be easily learned by their users. The third technique **(3)** guides us to design tools for different strategies and levels of expertise. The fourth technique **(4)** guides us to design tools facilitating rational decision-making.

Quantification of Cognitive Support

Although the cognitive support of a tool is related to a user's mental model of that tool [2], an action-based approach was followed in this work. That is, the effect of cognitive support was indirectly measured in terms of time and user actions employed to accomplish a pre-defined task. First, cognitive support of a tool was measured in terms of how it contributes to lower task solving time. Second – knowing that exploratory user behaviour in positioning tasks may reduce cognitive load [7] – cognitive support was measured in terms of user actions. Third – assuming that a good fit between users' mental model of a tool and how it actually works is beneficial – cognitive support was measured in terms of how it contributes to user satisfaction.

Measuring the Cognitive Support of the Tools

The evaluation presented here consists of a pilot study, followed by an experiment. In the pilot study, four alternative tools were examined. For the experiment, three of these four tools were selected and the cognitive support of each of these tools was examined. A comparative evaluation of the TUI against alternative planning tools was carried out.

PROBLEM REPRESENTATIONS THROUGH ALTERNATIVE TOOLS

Whereas numerous prototypes for tangible interaction have been proposed [6, 8, 10], less effort has been expended to measure the usability of such systems and how they may influence the nature of collaborative work. Motivated more by the former and partly by the latter of these deficiencies, the purpose of this evaluation was to assess the quality of BUILD-IT as a planning tool and thereby to justify the value of usability research of such systems. The BUILD-IT system is typically used for collaborative planning tasks in interior architecture, city planning, or production plant layout. Investigating different levels of abstraction, we found that three-dimensional (3D) physical replicas and two-dimensional (2D) cardboard models are two characteristic representations. A more abstract representation is a combination of tools such as pen, compass, stencil, and calculator. These three representations led us to construct three alternative planning tools. The tools were synthetically constructed as alternative means to solve a spatial laser-positioning problem and had the sole purpose to evaluate BUILD-IT in a comparative context of planning tools.

An additional, screen-based test case was initially considered, using a Computer-Aided Design (CAD) system. However, such systems do not offer concurrent positioning and rotation as a standard. This would have limited the task design, or required the implementation of non-standard functions. A more serious challenge with CAD is that the instruction time would go far beyond what is required by the other tools, unless CAD experts were employed. Therefore, at this stage of research, it was decided not to compare BUILD-IT with a screen-based tool.

Problem definition

Since the aim was to compare alternative tools for planning and layout activities, it was necessary to define a typical problem that could be represented by all tools. Typical task employed in the study of human computer interaction have varying difficulty, from search (easiest), to positioning, path pursuit, and pursuit tracking (most difficult) [1].

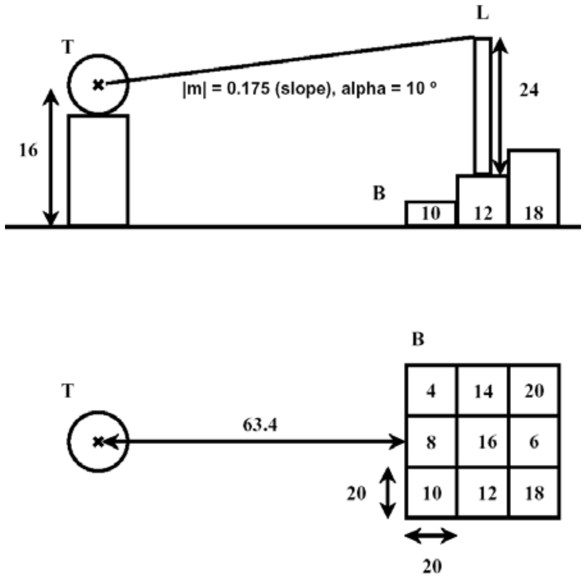

Figure 2: Problem definition: planimetric side view (top) and plan view (bottom), measurements in mm. The block heights in the top part (10, 12, 18) were not given to participants. They are shown here for clarity.

Given that focus of use is production plant layout, a typical task is to bring 3D models, e.g. welders and assembly lines, into a certain spatial relation. In practice, such tasks are carried out under demanding spatial and time constraints. In order to emulate real-life work situations, while still controlling major usability factors, it was decided to define a positioning task in terms of spatial constraints, while measuring trial time, number of user operations, and user satisfaction as the dependent variables. A laser-positioning task, given by a schematic description (Fig. 2) and the instructions below was the result.

Instructions

- The task is to find the block where the light beam of the laser source **L** hits the target **T**. The light beam should be as close as possible to the centre of the target and there is only one solution.
- The scene is shown from the side (called planimetric side view containing **T**, **B**, and **L**) and from above (called planimetric plan view containing **T** and **B**).
- Nine square blocks **B** with different heights form a three-by-three matrix.
- The target **T** with height 16 mm is situated at a distance 63.4 mm from the matrix.
- The laser source **L** with height 24 mm and slope 0.175 can be placed on any of the nine blocks and within each individual block.

TOOL DESCRIPTION

BUILD-IT is first described, followed by PhysicalBlocks, Cardboard, and Mathematics. Excluding BUILD-IT, the tools are order by expected cognitive support (high to low):

Tool 1 – BUILD-IT

For the BUILD-IT tool (Fig. 3), medium cognitive support was expected. This tool employs a virtual modelling of the blocks, target, and laser source. No navigation of the views is permitted and the views remain fixed, as shown in Fig. 3. The participants use one brick with which they manipulate the virtual laser source.

Tool-dependent instructions

The participants were shown how to select and handle the virtual model representing the laser source. It was shown how the plan view and the side view provide complementary, task-relevant information.

Tool 2 – PhysicalBlocks

For the PhysicalBlocks tool (Fig. 4), high cognitive support was expected. This is a physical task realization, scaled up by a factor of four, in comparison to the measurements given in Fig. 2. It consists of nine metal blocks mounted on a cassette, a standard laser source (maximum diameter: 15 mm, length: 50 mm), and a target made of a metal pin with a 5 mm by 15 mm wide metal flag. The investigator can adjust the height of the metal flag, target position, and block positions.

Tool-dependent instructions

The participants were shown how to position and rotate the laser source between different blocks and within individual blocks. They are not allowed to manipulate the cassette, the blocks, or the target flag. Before the solution has been found, the laser beam hits a wall, indicating the direction of the light beam.

Tool 3 – Cardboard

For the Cardboard tool (Fig. 5), medium cognitive support was expected. This tool is a planimetric tool scaled up by a factor of two in comparison to the measurements given in Fig. 2. Most elements in this tool are produced in cardboard and are (Fig. 5a, bottom to top): a cardboard ruler emulating the floor; a combined planimetric plan and side view showing the nine block positions in a plan view aspect and the floor with the target in a side view perspective; a cardboard laser source; a metal ruler emulating the laser beam; the nine cardboard blocks configured by the task definition. Solving a task with this tool may typically follow these steps (a to k, Fig. 5):

a) start-up,
b) selecting a first block,
c) fixing distance of the block with the floor ruler,
d) rotating that distance into the planimetric side view,
e) putting the block at rotated distance,
f) putting the laser source at the block,
g) putting the beam ruler at the laser; testing where beam goes (above the target),
h) testing a second block (below the target),
i) testing the closer edge of the same block (still below the target),
j) testing the farther edge of the same block (still below the target), and
k) testing a third block (hits the target).

Tool-dependent instructions

The principles of planimetric work were explained, i.e. working in an x-y projection (plan view) and a y-z projection (side view). It was explained that these two projections were integrated into one tool. Then, an example given shown how one single block is tested, and in particular how a distance is fixed in the planimetric plan view and rotated into the planimetric side view. The blocks are already at their position, giving the task definition. However, in case a participant looses the position of a block, the link is given by the number on the reverse side of the block and the corresponding height on the position. As seen in Fig. 5, blocks already tested are laid back with a slightly different position, indicating that they were already tested. Height figures noted on the blocks are not used to solve the problem.

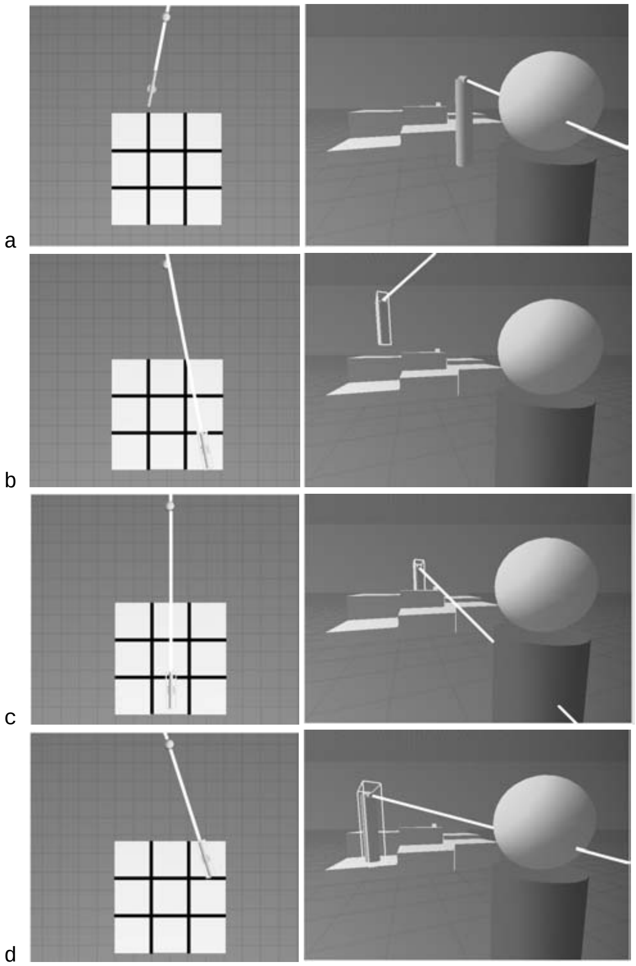

Figure 3: The BUILD-IT tool; plan view is shown in the left column and side view in the right. Typical steps: a) start-up; b) testing a first block (above the target); c) testing a second block (below); d) testing a third block (hits the target).

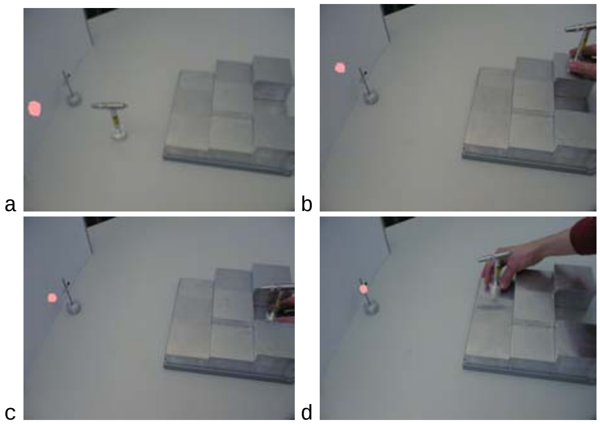

Figure 4: The PhysicalBlocks; typical steps are: a) start-up; b) testing a first block (above the target, hits the wall); c) testing a second block (below the target, hits the wall); d) testing a third block (hits the target). The laser beam spot is redrawn for print reasons.

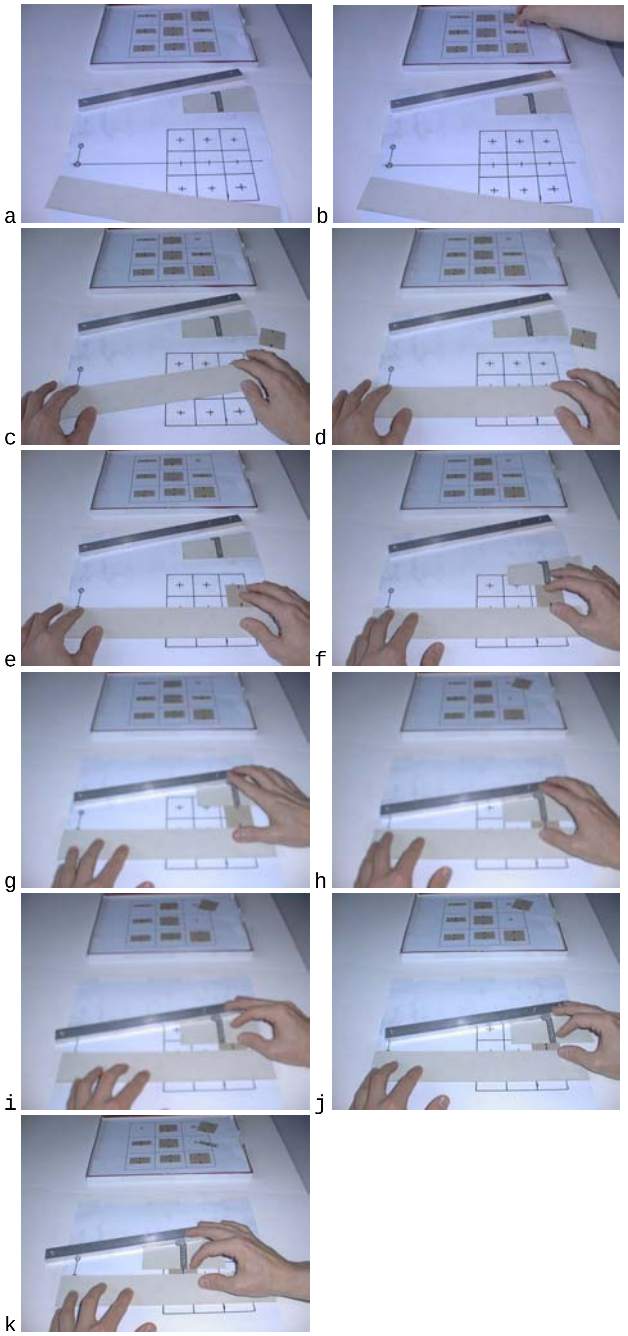

Figure 5: The Cardboard tool: some typical situations are: a) start-up, b-g) testing a first block (above the target); h-j) testing a second block (below the target); k) testing a third block (hits the target).

Tool 4 – Mathematics

For the Mathematics tool (Fig. 6), low cognitive support was expected. This tool consists of the problem definition in Fig. 2, a sheet of paper, a pen, a compass, a calculator, and a stencil.

The participants were familiar with mathematical skills from Grammar school. With the given tools, they could therefore choose to employ geometry, vector or linear function calculus to solve the task.

Tool-dependent instructions

A planimetric representation with an x-y (plan view) projection and a y-z (side view) projection of the task domain was used (Fig. 2). The target is marked **T** and a laser source marked **L** has an inclination of alpha = 10° (slope m = 0.175). The heights of **L**, **T**, and the blocks **B** are indicated as in Fig. 2.

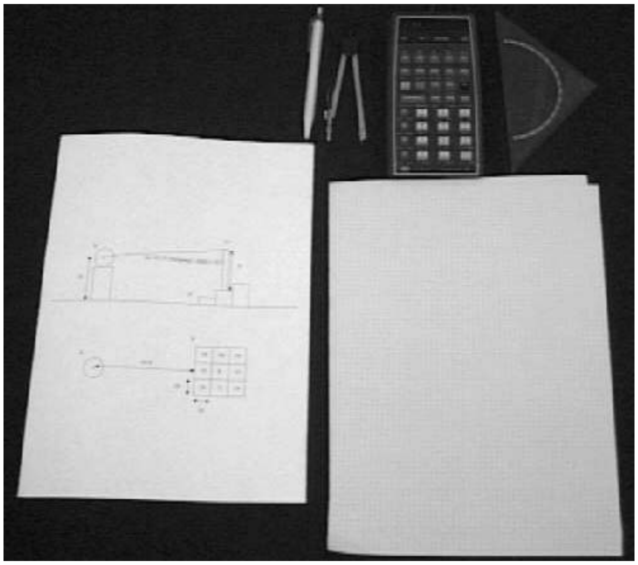

Figure 6: Mathematics: planimetric representation (left), a blank sheet (right), a pen, a compass, a calculator, and a stencil.

Relations among the tools

It was expected that the three alternative tools, having different commonalties with BUILD-IT, would give different cognitive support, as outlined in Table I.

Table I: The four tools, their representational characteristics, and expected cognitive support.

Tool	Problem representation	Commonality with BUILD-IT	Expected Cognitive Support
BUILD-IT	Augmented Reality	-	Medium
Physical Blocks	Physical, one-to-one	3D	High
Cardboard	Planimetric	Planimetric; 2D	Medium
Mathematics	Written	Computer resources	Low

Questionnaire

A questionnaire captured pre-experimental data, experimental feedback, and raised several open-ended questions. Besides gender and age, pre-experimental data was frequency, purpose, and literacy in computer use.

The experimental feedback, that is user satisfaction, was measured by three questions. All choices were ordered from low to high satisfaction:

- How was the **clarity** of the task formulation? (The exclusive choices were: not at all; not; neutral; good; very good).
- How did you perceive the **difficulty** of the task? (The exclusive choices were: very difficult, difficult; neutral; easy; very easy).
- How was the **suitability** of the tool you used to solve the task? (The exclusive choices were: of no help at all; of no help; well suited; of help; of high help).

PILOT STUDY

The pilot study was not intended to deliver statistical evidence, but simply to give approximate values guiding the set-up of the major experiment. The major experiment, described later, should last a maximum of one hour for each participant. The pilot study examined all four tools and aimed to optimise the experimental set-up with respect to:

- Design
- Procedure
- Instructions
- Time per trial (max. 10 minutes)
- Difficulty of the task-tool combinations
- User satisfaction with the given tool alternatives
- Relevance of tool comparison

Design

To avoid between-tool learning, a between-subject design examining four alternative tools was chosen. Each tool was randomly assigned to two participants, requiring a total of eight participants. For each trial, trial time was measured.

Participants

The participants were eight undergraduate or graduate students. They all had a Grammar school or equivalent degree, this assured a certain level of skills in mathematics and geometry. There was one female and there were seven male participants aged between 25 and 30 years. No importance was given to whether participants were left- or right-handed. The participants were not paid.

Task and Apparatus

All participants solved the same task. Each tool was assigned to two participants.

Procedure

The pilot study was performed with single participants during the daytime where each participant solved one and the same task only. The study was carried out in an office with the investigator sitting next to the participant. The structure of the study is as follows:

a) Description of tool, task, and task completion (approx. 10 min.)
b) Task-dependent instructions as given for each tool with a demo trial (1 - 8 min.)
c) One unaided trial with registration of trial time.
d) The participant filled out the questionnaire.

Results of Pilot Study

The data collected from the experimental set-up indicates a relation between tool (independent variable) and mean trial time (dependent variable) and is presented in Table II. The first three tools appear to give sufficiently small and comparable values for trial completion time. The fourth tool, Mathematics, was very different from the other tools, hardly representing a relevant comparison. The questionnaire was slightly altered for clarity and consistency. Instructions were considerably improved.

Table II: Mean trial time.

Tool	Trial time
BUILD-IT	2 minutes
PhysicalBlocks	18 seconds
Cardboard	3 minutes 30 seconds
Mathematics	26 minutes

EXPERIMENT

The experiment presented next, examined the BUILD-IT, PhyscialBlocks, and Cardboard tools, and was carried out with the aim to learn more about the cognitive support of these tools. For two reasons the Mathematics tool was not examined any further. Firstly, Mathematics required much more time than the other tools. Secondly, whereas alternative tools for 'intuitive' and 'direct' spatial design processes are examined, Mathematics requires much more reflection.

With the three tools selected, we measured and compared cognitive support of each one by the effect they had on trial time, on blocks tested, and on user satisfaction.

Hypotheses

It was conjectured that PhysicalBlocks would give the highest cognitive support, since it allows users to grasp the problem and solve it efficiently and intuitively (Fig. 7).

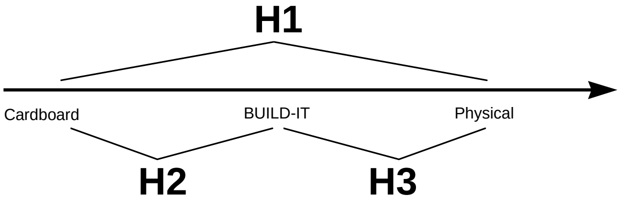

Figure 7: Expected cognitive support (x-axis) of each tool and the pair-wise relations between each tool is tested by hypotheses H1-H3, given in Table III.

It was expected that BUILD-IT would give the second best cognitive support, since it offers a representation close to the physical situation by using 3D virtual models. Then, it was conjectured that the Cardboard tool would offer the third best cognitive support, since it applies 2D representation and requires more abstract thinking from the participants. The hypotheses were stated in the conventional form, as appropriate for statistical testing and given in Table III.

Table III: Hypotheses.

H1: Cardboard gives less cognitive support than PhysicalBlocks.
H2: Cardboard gives less cognitive support than BUILD-IT.
H3: BUILD-IT gives less cognitive support than PhysicalBlocks.

Design

As in the pilot study, a between-subject scheme was employed as means of eliminating between-tool learning. Each tool was assigned to ten participants. There was a total of 12 task variations, one for demonstration trial (always the same), one for an aided trial (always the same), and a maximum of 10 for unaided experimental trials (see procedure section for details). This was meant to be sufficient for all participants to solve a minimum of five tasks correctly without solving any task more than once. The ten task variations for unaided trials were permuted by latin squares so that trial-variation combinations were equally distributed.

Operationalization of Cognitive Support

To examine the hypotheses as outlined in Table III, the effect of cognitive support was broken into a set of five measurable variables. These were: i) trial time, ii) number of user operations, or blocks tested in a trial, iii) learning effect in trial time, or difference in trial time between last and first correctly solved task, iv) learning effect in blocks tested, or difference in blocks tested between last and first correctly solved task, and v) user satisfaction with the given task-tool combination. Based on these variables, cognitive support was measured using the following five criteria:

C1: Lower trial time.
C2: More blocks are tested per trial. That is, exploratory action is likely to have taken place, hence reducing cognitive load [7].
C3: Lower trial time in last than first trial solved, which is learning effect in trial time.
C4: Fewer blocks tested in last than first trial solved, which is learning effect in blocks tested.
C5: Higher user satisfaction with task-tool combination used, given by the mean of perceived clarity of task formulation, perceived task difficulty, and perceived tool suitability.

Participants

The participants were thirty undergraduate or graduate students. There were 13 female and 17 male participant aged between 20 and 36 years. No importance was given to whether participants were right- or left-handed. The participants were paid CHF 10 each.

Task and Apparatus

The same tasks and apparatus, with exception of the Mathematics tool, as in the pilot study. Each tool was assigned to ten participants.

Procedure

The user study was performed with single participants during the daytime. Each participant solved one aided and five unaided task varieties. It was carried out in an office with the investigator sitting next to the participant. The experiment consisted of the same general procedure as in the pilot study:

a) Description of tool, task, and task completion
b) One demo trial (three false block, then the correct block, were tested)
c) One aided trial (questions were answered and help was offered at need)
d) Each participant had to perform five unaided trials with correct answer; the three last correct ones in a closed sequence. After having solved a task, the participants rang a bell. For all trials, the blocks tested, trial time, and whether the indicated block was the correct one were registered. The next task was initialised without the participant watching.
e) The participant filled out the questionnaire.

Logging

For each trial, the number of blocks tested was registered, with one being the minimum logged value. Blocks tested more than once were counted each time. Trial time was registered when the participant rang a bell.

Learning effects were normally observed during demo and aided trials, additional effects could be observed throughout the unaided trials. Learning effects were only registered in the unaided trials.

Results of Experiment

The mean values of the trial time and blocks tested are shown in Table IV. Box plots of the same data are shown in Fig. 8. To test the criteria **C1** – **C4**, an analysis was carried out to investigate differences between BUILD-IT, PhysicalBlocks, and Cardboard. Consequently, similar to Balakrishnan et al. [1], a multiway ANOVA, here with a General Linear Model (GLM), was used. There were four independent variables: tool (PhysicalBlocks, BUILD-IT, Cardboard), participant (1-30), trial (T1-T8), and task. Participant was hierarchically nested with tool. The interaction between tool and task, called tool x task, was examined.

Table IV: Means of trial time and blocks tested.

Tool	Trial time	Blocks tested
BUILD-IT	37.3 seconds	7.2 blocks
PhysicalBlocks	25.4 seconds	5.5 blocks
Cardboard	52.8 seconds	2.7 blocks

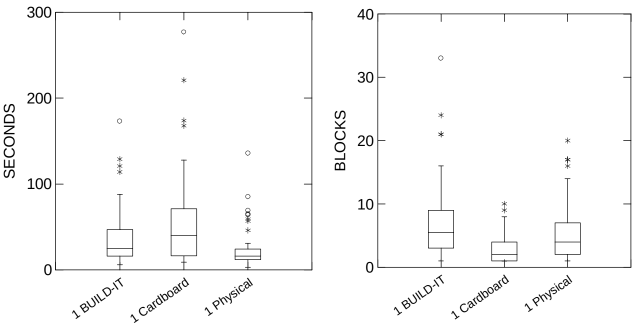

Figure 8: Box plots of trial time (left) and blocks tested (right).

There were four dependent variables, examining **C1** – **C4**: trial time (**C1**), blocks tested (**C2**), improvement in trial time from first to last correct trial (**C3**), and improvement in blocks tested from first to last correct trial (**C4)**. Table V shows significant effects for: tool, task, participant, and tool x task interaction. Table VI shows a post-hoc comparison of tool effects, alpha = 0.05.

Table V: GLM results, significant effects only.

Dependent	Independent	df	F-ratio	p-value
trial time	tool	2	12.652	< 0.001
— " —	task	9	4.156	< 0.001
— " —	participant (tool)	27	3.387	< 0.001
blocks tested	tool	2	23.435	< 0.001
— " —	task	9	6.344	< 0.001
— " —	tool x task	18	2.410	0.003

Table VI: Post-hoc comparison for tool effects.

Dependent; independent	Pair	Pairwise diff.	p-value
trial time [s]; tool	BUILD-IT & Cardboard	15.5.	0.030
trial time [s]; tool	Cardboard & PhysicalBlocks	-27.4	< 0.001
blocks tested; tool	BUILD-IT & Cardboard	-4.493	< 0.001
blocks tested; tool	Cardboard & PhysicalBlocks	2.811	< 0.001

Learning Effects

There were no significant learning effects. Hence, **C3** and **C4** were not fulfilled. For each tool as a function of trial number (T1–T8), trial time (Fig. 9, top row) and blocks tested (Fig. 9, bottom row) were plotted. These plots will be discussed later.

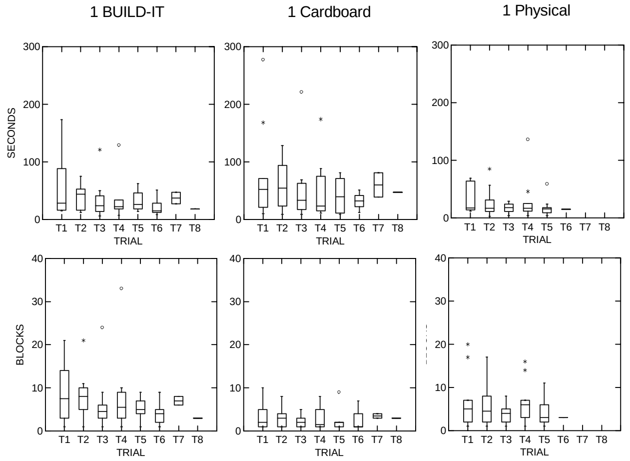

Figure 9: Improvement in trial time (top) and blocks tested (bottom); BUILD-IT (left), Cardboard (middle), PhysicalBlocks (right).

Subjective Preferences

Table VII gives participants' ratings for each tool and classification. The classifications were enumerated using a balanced scale [-2,-1,1,2], ranging from low to high satisfaction. Only the mean rating is given. This enables us to examine **C5** as follows: The user satisfaction with PhysicalBlocks is considered to be clearly higher than with BUILD-IT and Cardboard. The difference between BUILD-IT and Cardboard is considered not to be of importance. This is reflected in the **C5** column of Table VIII.

Table VII: Mean users satisfaction per criterion and total mean.

Tool	mean clarity	mean difficulty	mean suitability	Total mean
Physical-Blocks	1.7	1.2	1.1	1.3
BUILD-IT	1.4	0.5	0.9	0.9
Cardboard	1.3	0.3	1.4	1.0

Control for Third Variables

There was no significant effect for gender or age. Also, there was no significant effect for frequency, purpose, and literacy in computer use.

Summarizing the Results

The results over all criteria (**C1-C5**) can now be summarized and decisions in terms of the original hypotheses (**H1-H3**) can be taken, as indicated in Table VIII. Three criteria are fulfilled for **H1**, one criterion for **H2**, and one criterion for **H3**. Hence, **H1** was mainly confirmed. The hypotheses **H2** and **H3** were not sufficiently confirmed.

Table VIII: Criteria fulfilled per hypothesis.

Hypothesis	C1	C2	C3	C4	C5	# criteria fulfilled
H1	Yes	Yes	No	No	Yes	3 (of 5)
H2	Yes	Yes	No	No	No	2 (of 5)
H3	No	No	No	No	Yes	1 (of 5)

The major findings can be summarized as follows:

a) Less time used, more blocks tested, and higher user satisfaction with PhysicalBlocks than with Cardboard.
b) Less time used, more blocks tested with BUILD-IT than with Cardboard.
c) Higher user satisfaction with PhysicalBlocks than with BUILD-IT.

Further, the minor findings can be summarized as follows:

a) Considerable, but not significant, learning of BUILD-IT in trial time and blocks tested.

b) Considerable, but not significant, learning of Cardboard in blocks tested, having little impact on trial time.
c) Considerable, but not significant, learning of PhysicalBlocks in trial time.

Using open-ended questions, participants were also asked to justify their choice and to comment on overall system usability. Their statements were classified as positive [+], negative [–], or neutral [±], giving in brackets the number of participants making the same statement if more than one:

- **PhysicalBlocks**
  - + : interesting, intuitive tool easy to use, haptic feedback made task solving easy
  - – : context of task was not known
  - ± : task and tool could not be separated, distance had an impact on height
- **BUILD-IT**
  - + : interesting (3)
  - – : low accuracy in rotation difficult (5), difficult to learn coordination, side view of little use, slow graphical update, did not perceive that laser beam had a slope
  - ± : unusual tool, usability of side view only became clear later, learning required, higher trial time while using visual cues than reflection
- **Cardboard**
  - + : time passes quickly (2)
  - – : accuracy and speed were disparate aims, tool is inaccurate and results must be controlled repeatedly, the highest blocks were clearly not candidates for a solution
  - ± : some training improved trial time considerably (3), unusual tool, gradually changed strategy from trying to reflecting, discovered that small blocks gave solution when situated left and high blocks gave solution when situated right.

DISCUSSION

We have seen that the 3D physical tool and the TUI both give significantly more cognitive support than the 2D cardboard-based tool. However, , the TUI gives less cognitive support than the 3D physical tool in terms of user satisfaction. We conclude that the cognitive support offered by a Tangible User Interface (TUI) – like the BUILD-IT system – comes close to the physical world in terms of trial time and number of user operations, but not in terms of user satisfaction. From a logistical point of view, the preparation of 3D models for a TUI is quicker and cheaper than constructing 3D physical models like the PhysicalBlocks presented. Hence, to take out the full benefit offered by tangible interaction, it is necessary to focus on the user satisfaction of TUIs.

The interpretation of number of user operations for a given tool still remains an issue. It may be related to the ease of use of each tool, to how each tool lends itself to exploratory actions, to repeated interactions due to low accuracy, or to users preferring reflection to interaction.

The significant effect of task on trial time and blocks tested, as shown in Table V, means that different tasks had different difficulty. Such variation in difficulty is inherent to the nature of the positioning task employed. Given the design of the experiment, task effect was taken sufficiently account of not to bias the other results.

The significant tool x task interaction for blocks tested, as shown in Table V, means that specific tasks required more or less testing of blocks depending on the tool used. This may be explained by variation in precision among the tools. Precision may be seen as part of a tool's cognitive support. Given the design of the experiment, tool x task interaction was taken sufficiently account of not to bias the other results.

Although there was no significant learning effect, some trends could be observed in Fig. 9. When using BUILD-IT (Fig. 9, left column), there seems to be a learning trend in terms of trial time and blocks tested. In using Cardboard (Fig. 9, middle column), there seems to be a learning trend in terms of blocks tested but not in terms of time used. In using PhysicalBlocks (Fig. 9, right column), there seems to be a learning trend in terms of time used but not in terms of blocks tested. We interpret that with BUILD-IT, trial time improved because participants learned how to use the tool. With PhysicalBlocks, trial time improved without any observable change in use, indicating that users simply became more confident. With Cardboard, trial time did not improve in spite of an observable change in use.

The numbers describing user satisfaction – as they are given in Table II – are worthy of comment. First, it makes sense that the task could be most clearly *explained* using the PhysicalBlocks tool. Second, it also makes sense that the *difficulty* of the task was perceived as lower using the PhysicalBlocks tool than the two other ones. Third, the Cardboard tool was seen as the most *suited* one to solve the task. This preference may appear surprising and calls for explanation. Some explanation can be found in examining how users perceived the help offered by the two other tools. On the one hand, five participants stated that low accuracy in rotation made it difficult to use BUILD-IT for the task presented. On the other hand, there was hardly any separation between the task and the PhysicalBlocks tool, making the question about tool suitability somehow vague. These two factors might have helped the Cardboard tool to come out of the ranking as the most suited one to the task.

Finally, we try to understand the results in terms of the decision support techniques. In a first approach, the representation aids guided us to focus on problem representation, which was realized through the design of alternative planning tools. The information control techniques guided us to design for learning, a process that may take place during instruction, during problem solving, or both. We observed no learning after the instruction. The analysis and reasoning techniques guided us to design tools for different strategies and levels of expertise. Different strategies were best observable in Cardboard), as trial time varied considerable among participants without being reflected in a similar variation in number of user operations. The choice models guided us to design tools facilitating rational decision-making. Since participants tested fewer blocks with Cardboard, it can be assumed that this tool had less capacity to facilitate rational decision-making.

The intention behind the TUI system evaluated and the way the experiment was carried out may appear contrary. While the intention behind the in-house designed TUI is to support 'intuitive' and 'direct' interaction, the experiment reported was based on an analytical task. Hence, the task studied might have been too analytical for the original purpose. Future experiments may employ tasks demanding less reflection, for instance a search task [1]. Tasks demanding more reflection, such as path pursuit and pursuit tracking [1], are poorly suited to this application.

CONCLUSION AND FUTURE RESEARCH

We have compared an in-house designed Tangible User Interface (TUI) with three alternative single-user tools through an empirical investigation. These three alternative tools were a 3D physical, a 2D cardboard, and a mathematical tool. We expected the 3D physical to perform best, followed by the TUI, the 2D cardboard, and the mathematical tool. A pilot study was first carried out, the results of which were used to design a major experiment. Participants solved the same positioning problem, each using one of the four tools. Mainly due to weak performance in trial time, the mathematical tool was not used in the experiment. In the experiment, trial time, number of user operations, learning effect in both trial time and user operations, and user satisfaction were measured. The 3D physical tool significantly outperformed the 2D cardboard tool. It also outperformed the TUI, but only in terms of user satisfaction. This justifies the value of researching TUI systems. In particular, the results justify carrying out user studies with such systems, aiming for higher user satisfaction.

The tool design was inspired by four decision support techniques. For future research, it may be worth to examine how each of these techniques can predict the cognitive support of a tool. We believe that each of the decision support techniques can be recast into a few questions, or heuristics. These heuristics may provide a valuable compliment to usability evaluation.

To anchor this work in a real-world context, we are actually developing a screen-based test case. This will require non-standard functions to be programmed in a Computer-Aided Design (CAD) system, at considerable expense. It remains an issue whether participants without CAD experience can solve the problem studied after a reasonably short period of instruction.

The results came out of an experiment with single users only. Their validity for other TUI systems and for collaborative planning calls for further investigation.

ACKNOWLEDGMENTS

We thank Marc Arial, Nicholas Ironmonger, Martina Merz, Elke Reuss, Kent Riopelle, Tilman Schildhauer, Martin Sundin, and Wendy Alexandra Franks for commenting on this paper. Valuable help also came from Martin Bichsel (development), Georg Kralidis (statistical analysis), and Matthias Rauterberg (interaction design and Ph.D.-mentor). PhysicalBlocks was constructed by Hanspeter Honegger.

REFERENCES

[1] Balakrishnan, R., Kurtenbach, G.: Exploring bimanual camera control and object manipulation in 3D graphics interfaces, *Proceedings of the CHI'99*, ACM, pp. 56-62 (1999).

[2] Bødker, S.: *Through the Interface: A Human Activity Approach to User Interface Design*, Hillsdale, NJ: Lawrence Erlbaum (1991).

[3] Fjeld, M., Voorhorst, F., Bichsel, M., Krueger H., Rauterberg, M.: Navigation Methods for an Augmented Reality System, *Video program CHI 2000*, ACM, pp. 8-9 (2000).

[4] Fjeld M., Lauche K., Bichsel M., Voorhorst F., Krueger H., Rauterberg, M.: Physical and Virtual Tools: Activity Theory Applied to the Design of Groupware, *CSCW*, Kluwer, Volume 11 (1-2), pp. 153-180 (2002).

[5] Gibson, J.J.: *The ecological approach to visual perception*, London, L. Erlenbaum (1986).

[6] Ishii, H., Ulllmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms, *Proceedings of CHI '97*, ACM Press, pp. 234 - 241, (1997).

[7] Kirsh, D., Maglio, P.: On Distinguishing Epistemic from Pragmatic Action, Cognitive Science, 18, pp. 513-549 (1994).

[8] Starner, T.: *Wearable Computing and Contextual Awareness*. Ph.D. Thesis, MIT (1999).

[9] St. Amant, R.: User interface affordances in a planning representation, *Human-Computer Interaction*, 14(3), pp. 317-354 (1999).

[10] Ullmer, B., Ishii, H.: Emerging frameworks for tangible user interfaces, *IBM Systems Journal*, 39(3 & 4), pp. 915-931, IBM (2000).

[11] Zachary, W.: Cognitively Based Functional Taxonomy of Decision Support Techniques, *Human-Computer Interaction*, v.2 n.1, pp. 25-63 (1986).