Cognitive Cubes: A Tangible User Interface for Cognitive Assessment

Ehud Sharlin, Yuichi Itoh[*]**, Benjamin Watson**[+]**, Yoshifumi Kitamura**[*]**, Steve Sutphen, Lili Liu**[o]

Dept. Comp. Science
University Alberta
221 Athabasca Hall
Edmonton, Alberta
Canada T6G 2E8
+1 780 492 4768
{ehud,steve}
@cs.ualberta.ca

[*] Human Interface
Engineering Lab
Osaka University
2-1 Yamadaoka, Suita
Osaka 565-0871, Japan
+81-6-6879-7752
{yuichi,kitamura}
@eie.eng.osaka-u.ac.jp

[+]Dept. Comp. Science
Northwestern University
1890 Maple Ave
Evanston, Illinois
USA 60201
+1 847 491 3710
watsonb@cs.nwu.edu

[o]Dept. Occupat. Therapy
University Alberta
3-14 Corbett Hall
Edmonton, Alberta
Canada T6G 2G4
+1 780 492 5108
lili.liu@ualberta.ca

ABSTRACT

Assessments of spatial, constructional ability are used widely in cognitive research and in clinical diagnosis of disease or injury. Some believe that three-dimensional (3D) forms of these assessments would be particularly sensitive, but difficulties with consistency in administration and scoring have limited their use. We describe Cognitive Cubes, a novel computerized tool for 3D constructional assessment that increases consistency and promises improvements in flexibility, reliability, sensitivity and control. Cognitive Cubes makes use of ActiveCube, a novel tangible user interface for describing 3D shape. In testing, Cognitive Cubes was sensitive to differences in cognitive ability and task, and correlated well to a standard paper-and-pencil 3D spatial assessment.

KEYWORDS: Cognitive assessment, neuropsychological assessment, tangible user interfaces, three-dimensional user interfaces, constructional ability, spatial ability

INTRODUCTION

The assessment of cognitive spatial and constructional ability is an important clinical tool in the diagnosis and monitoring of brain disease or injury [6,9]. It is also indispensable in scientific study of cognitive brain function. Techniques for assessment include asking patients or participants to perform purely cognitive tasks such as mental rotation, as well as constructional tasks involving arrangement of blocks and puzzle pieces into a target configuration. These constructional tasks have the advantage of probing not only pure spatial ability, but also the ability to perceive, plan, and act in the world. Studies suggest that assessment with three-

dimensional (3D) forms of these tasks may be most demanding and sensitive [9]. However, use of 3D tasks in assessment has been limited by their inherent complexity, which requires considerable examiner training, effort and time if scoring is to be consistent and reliable.

To overcome the limitations of existing 3D cognitive assessments, we have designed Cognitive Cubes, an automated tool for examination of 3D spatial cognitive ability. Cognitive Cubes makes use of ActiveCube [7], a Lego™-like tangible user interface for description of 3D shape. With Cognitive Cubes, users attempt to construct a target 3D shape, while each change of shape they make is automatically recorded and scored for assessment. Cognitive Cubes is the first computerized tool for constructional assessment, combining the increased sensitivity of 3D constructional tasks with the efficiency, consistency, flexibility and detailed data collection of automation.

In the following, we review the need for and methods of cognitive and constructional assessment. We then describe the Cognitive Cubes system in detail, comparing it to related systems using tangible user interfaces (TUIs). We conclude with a rigorous experimental examination of the sensitivity and utility of Cognitive Cubes.

COGNITIVE AND CONSTRUCTIONAL ASSESSMENT

The use of physical objects as tools for cognitive assessment is well established, especially for the assessment of constructional ability, which integrates perception and motor response in a spatial context [9]. Constructional ability is related to a number of everyday skills, and constructional deficits and disorders can be associated with brain lesion [6,9], Alzheimer's disease (AD) [6] and other impairments.

Constructional ability is most often assessed using assembly, building or drawing tasks. These assessments are non-verbal, relatively culture-free and can be very sensitive to and selective for constructional ability alone. Research shows that two dimensional (2D) tests may be less sensitive to constructional deficits than 3D tests [9].

Figure 1: A virtual prototype display in `Cognitive Cubes`.

In a typical constructional assessment, the participant is presented with a spatial pattern and is asked to mimic it by manipulating or assembling physical objects [9]. The test administrator scores participant performance using measures such as time to completion and accuracy, or more demanding observations such as order of assembly and strategy analysis.

2D constructional assessment is already used widely. For example, the well-known Wechsler intelligence test includes two 2D constructional subtests, Block Design and Object Assembly. In the former, the participant arranges red and white blocks to copy a presented pattern. In Object Assembly, the participant solves a 2D puzzle. Measures for both tests are based on time and accuracy [6,9].

3D constructional assessments are far less common. Two examples are Block Model from Hecaen *et al.* and Three Dimensional Constructional Praxis from Benton *et al.* [9]. In both of these tests the participant tries to match a 3D prototype using wooden blocks, and is scored on time and accuracy. The use of Lego blocks was suggested for 3D tests, but to our knowledge was never implemented. Given the complexity of the target shapes in these 3D assessments, manual scoring of even simple measures such as accuracy can be difficult. Manual scoring of denser measures such as order and strategy would certainly require a very skilled, trained and alert assessor.

Computer-assisted cognitive assessment is becoming increasingly common and most major paper-based tests will soon be or have been automated [6]. The advantages of automating an assessment include reductions in professional assessor time, elimination of assessor bias, and improvements in test reliability. Computerized assessments can implement adaptive testing (tailoring questions to preceding performance) and enable collection of measures that otherwise would require the close attention of a professional. On the other hand, automated assessments are easily administered by unqualified personnel, and automated assessments are not the same as their manual parallels,

Figure 2: A participant using `Cognitive Cubes`.

requiring extensive and often overlooked recalibration. To our knowledge, no automatic constructional assessment exists, either 2D or 3D.

COGNITIVE CUBES

`Cognitive Cubes` is an automated system for constructional cognitive assessment. In each trial, the participant attempts to construct with `ActiveCube` the prototype shape displayed in front of her with a projector (figs. 1-2). The prototype rotates slowly at 2.7 rpm, providing 3D depth information. Connecting a cube causes one chime to sound, disconnecting a cube another. The system records the time and 3D location of each (dis)connect event. When the participant is satisfied with the match between her construction and the prototype, she informs the assessor, who advances the system to the next trial.

Hardware and Software

`Cognitive Cubes` is based on a novel TUI for describing shape called `ActiveCube` [7]. The TUI consists of a set of plastic cubes (all 5 cm/edge) that can be connected to one another using simple male-female connectors on each face, forming both a physical shape and a network topology. Each cube and cube face has a unique ID. A host PC is connected to a special *base cube* and communicates with the small CPUs in each cube through a broadcast mechanism to sense the (dis)connection of any cube. Since all cubes have the same size, the current topology represents a unique collective shape.

All the cubes in our version of `ActiveCube` were the same color. One blue stripe on each face indicated how to orient the male-female connectors for proper assembly.

The test administrator can choose the next task type and prototype shape through the host PC. Although the assessment process is largely automatic, the administrator can choose to stop the assessment at any point if, for example, the participant is not making any progress.

After assessment, `Cognitive Cubes` analyzes its collected data offline to calculate the 3D similarity between the
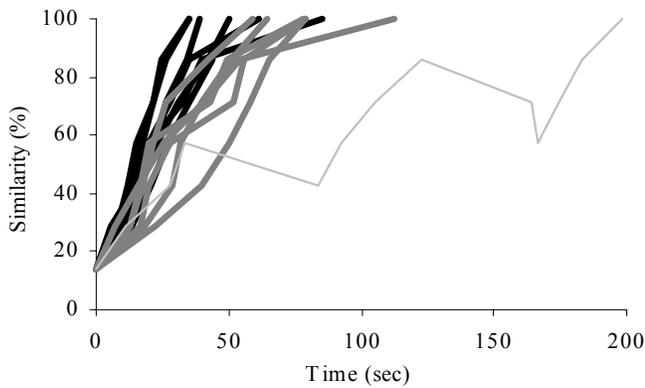
Figure 3: Similarity vs. time graphs ("similarity functions") for all participants in one task. Black shows young participants, gray elderly, and the thin line is one AD participant.

participant's structure *s* and the prototype *p*. Similarity is calculated at each connect or disconnect event, meaning for example that a five step participant assembly will result in five different similarity calculations. The equation for similarity is

$$S = 100 \cdot \left(\frac{|i|}{|p|} - \frac{|s|-|i|}{|p|} \right)$$

where *i* is an intersection of *s* and *p*, and |*i*|, |*s*|, and |*p*| are the number of cubes in *i*, *s* and *p*. *S* is maximized over all possible intersections *i* produced by rotating or translating *s*. Intuitively speaking, similarity is the number of intersecting cubes minus the number of remaining "extra" cubes in the participant's structure, normalized by the number of cubes in the prototype. Figure 3 graphs similarity vs. time ("similarity functions") for all participants in one Cognitive Cubes task.

We make *similarity* at task completion, calculated as described above, one of our four assessment measures. The remaining three are: *last connect*, the time elapsed from the start of the task to the last cube connect or disconnect; *derivative*, the differences between two successively measured similarities in a task divided by the time elapsed between those measurements (local "slope" of the similarity function), averaged for all such pairs in a task; and *zero crossings*, the number of times the local slope crossed zero. We sometimes use the terms "completion time", "rate of progress", and "steadiness of progress" as substitutes for *last connect, derivative*, and *zero crossings*.

Related Systems

Ishii and Ullmer define TUIs as devices that *give physical form to digital information, employing physical artifacts as representations and controls of the computational data* [14]. We focus here on spatial TUIs, or *interface devices that use physical objects as means of inputting shape, space and structure into the virtual domain*. Our work on Cognitive Cubes is motivated by the belief that TUIs are uniquely suited to spatial applications.

The first working TUI prototypes were part of the pioneering work of Frazer and Aish during the early 80s [1,4]. These spatial systems enabled a user to input 3D shapes into a computer by physically assembling blocks for CAD prototyping. Later, the Geometry Defining Processors project supported 3D spatial input for engineering system definition and analysis [2].

More recently, AlgoBlocks [13] and Triangles [5] enabled 2D spatial input using physical objects. AlgoBlocks was used in introductory programming education, while Triangles applied a spatial metaphor to storytelling for children. MERL's blocks [3] enable 3D design using Lego-like blocks, useful in children's play and game level prototyping. Unlike ActiveCube, which can provide real time updates of each connect or disconnect event, MERL's blocks are assembled and constructed off-line, and then digitally sampled in a relatively slow process.

System Strengths

Cognitive Cubes offers many advantages over existing tools for constructional assessment. These include:

Consistency. Any assessment is of little use if comparisons between different sets of its results are not trustworthy. Existing tools for 2D assessment have been in use for some time and are quite consistent. However, while 3D constructional assessments have been proposed previously, the complexity of the shapes and tools involved have made them difficult to administer consistently. As the first automated constructional assessment, Cognitive Cubes is extremely consistent in administration and scoring.

Sensitivity. Assessments must also respond as sensitively as possible to the strength or weakness of cognitive abilities. Because they incorporate a demanding level of complexity, 3D constructional assessments have shown particular promise in this regard. Cognitive Cubes incorporates this sensitivity. Moreover, automation allows Cognitive Cubes to monitor elements of performance that are ignored in other assessments, including actions during the assessment itself (not just the final result).

Cost efficiency. Although the hardware used by Cognitive Cubes will be more costly than that used in many other assessment tools, we expect that overall, it will reduce costs. Because Cognitive Cubes is automated, the level of training and expertise required by personnel employing it will be relatively inexpensive. In addition, automation should ultimately allow adaptive testing, isolating the level of cognitive ability much more quickly and thus reducing testing time.

ASSESSING COGNITIVE CUBES

Is Cognitive Cubes as sensitive and consistent as we would like to believe? In this section we describe our experiments, designed to find answers to this question. We begin with a discussion of our general experimental approach and the testing and resulting adjustment of it in a pilot study. Next we describe the cognitive sensitivity study, which examined the response of Cognitive Cubes to known participant and task cognitive factors. Finally, we present the test comparison study, which compared

Cognitive Cubes' results to those obtained with a standard paper-and-pencil 3D assessment.

General Methodology

Cognitive Cubes tasks were designed with two principles in mind. First, tasks should be as diverse and as interesting as possible, ensuring that participant interest remains high and that the assessment is sensitive to a range of cognitive ability levels. Second, participants should move gradually from easy to difficult tasks. This eases participants into a familiarity with the novel interface and allows quick identification of thresholds in participant cognitive ability, permitting participants to drop out without frustration as soon as they show their capabilities.

We designed four task types. *Intro* tasks were simple practice trials, designed to introduce the participant to the Cognitive Cubes. A cube appears on the display after each new connection, indicating the next cube to attach. The *follow* task type also provided step-by-step guidance, but the tasks were much more difficult. *Match* tasks provided no cube by cube guidance, but rather displayed a complete virtual prototype for the participant to construct using their own approach. In all three of these task types, the starting point for the participant's construction was the base cube. On the other hand, with *reshape* tasks the participant started from a more complex initial condition (always the same 7-cube 3D construct) – in all other respects *reshape* was exactly like *match*.

Being guided, *intro* and *follow* required less cognitive planning than *match* and *reshape*. Since it started from a complex, somewhat arbitrary shape, we expected *reshape* tasks to require more planning effort than *match* tasks. For this reason task types were placed in an *intro*, *follow*, *match* and *reshape* order. Within each task type, tasks were organized roughly according to their relative difficulty, taking into account the number of cubes, symmetry, and 2D or 3D shape.

Intro shape complexity was minimal, while *follow* tasks used moderately complex shapes. In *match* tasks shapes reached their greatest complexity, while shape complexity was moderated for *reshape* in light of the heightened demands on cognitive planning.

Design

During each Cognitive Cubes assessment, a participant performed 39 tasks in this order: 6 *Intro*, 8 *follow*, 15 *match* and 10 *reshape* tasks. 14 tasks were 2D, 25 were 3D.

Apparatus and Procedure

The participant sat at a table with only Cognitive Cubes placed in front of her. A front projector behind and above the participant displayed a 125 cm diagonal image at a viewing distance of 185 cm, all in a brightly lit room. The assessment administrator sat in front of an adjacent table with the host PC.

The experiment was conducted with a strict written protocol read out loud to the participant. The participant was introduced to the system, the experiment, and its purpose,

and then read an information letter. She was told that she might stop the experiment at any time, and asked to sign a consent form. The participant was interviewed quickly, answering questions concerning age, education, occupation, experience in 3D design, construction sets, computer games, general health and handedness.

The participant was asked to be "as fast and as precise as possible", and was told that "the system is recording your actions". She was told that there was no time limit, and that she may decide when she had finished each task, but that she should do "the best she could" in building each shape. She was asked to connect one cube at a time to the cube structure (avoiding offline interaction), unless reconnecting a chunk of the structure that had fallen off. On the other hand, removing several cubes at once was perfectly fine.

The system never provided feedback about construction correctness. In the first few *intro* tasks the participant was guided closely by the administrator, both verbally and physically. Guidance was gradually reduced and after the *intro* tasks, ceased. Between the *follow* and *match* tasks the participant was asked to take a short break. During the *reshape* tasks, planning was encouraged by reminding the participant that "the system is counting your steps". Finally, the participant was interviewed for her impressions of the system and the experiment. Performing the complete assessment took roughly 90 minutes on average.

If the system hardware failed, the participant was asked to repeat the interrupted trial. When the participant was not showing progress in a task after 5 minutes, the administrator suggested ending the task. Skipped tasks tended to be more difficult tasks.

Pilot Study

Our extensive piloting included 14 young, healthy participants who each performed the entire Cognitive Cubes assessment. They ranged in age from 22 to 43, with 3 females and 11 males. From this study emerged a strict written protocol which was not changed afterward. A few task parameters (e.g. prototype rotation rate) were altered following participant feedback.

We quickly absorbed two lessons in this study concerning the complexity and structure of task shape. First, since the cubes' ability to support their own weight was limited, certain prototypes were modified. Second, since matching even ten cube prototypes was quite challenging, and since fewer cubes resulted in constructions that were structurally more sound, all shapes were restricted to at most ten cubes.

We also learned that Cognitive Cubes suffers infrequently (table 1) from three types of system errors. In the most severe connection error the system reported cube (dis)connections that did not in fact occur. These events are always excluded from our analyses. The less severe crash errors occurred when the system simply stopped responding. We decided to allow the participant to repeat tasks with crash errors. Last, the cube construction sometimes collapsed, usually when the participant applied too much

Table 1: Error trials and filtered or repeated trials in the cognitive sensitivity study.

Trial Description	Frequency
Total trials:	621 (100%)
Error trials:	
Connect errors	26 (4.2%)
Sys crashes w/o connect error	13 (2.1%)
Trials without error:	582 (93.7%)
Remaining affected trials:	
Repetitions	21 (3.6%)
Filtered collapses	81 (13.9%)

Table 2: Completed trials in the cognitive sensitivity study by participant group, shape and task type.

Group	Overall	Shape Type		Task Type			
		2D	3D	intro	follow	match	reshape
young	270	97	173	41	56	103	70
elderly	246	97	149	42	56	88	60
AD	63	27	36	12	15	24	12

Table 3: Results of three way ANOVA in cognitive sensitivity study. Intro trials, AD participants excluded.

Indep Meas	Depend Meas	ANOVA
age	last connect	$F(1,13)=23.82, p<.00005$
age	derivative	$F(1,13)=71.21, p<.00001$
task type	last connect	$F(2,26)=4.7, p<.01$
task type	similarity	$F(2,26)=4.96, p<.01$
task type	zero crossings	$F(2,26)= 7.58, p<.001$
task type	derivative	$F(2,26)=34.32, p<.00001$
shape type	last connect	$F(1,13)=37.24, p<.00001$
shape type	similarity	$F(1,13)=3.9, p<.05$
shape type	zero crossings	$F(1,13)=13.07, p<.0005$
shape type	derivative	$F(1,13)=137.15, p<.00001$
ttype x stype	last connect	$F(2,26)=3.22, p<.05$
ttype x stype	similarity	$F(2,26)=6.02, p<.005$
ttype x stype	zero crossings	$F(2,26)=3.93, p<.05$
ttype x stype	derivative	$F(2,26)=4.16, p<.05$

force. The participant typically reconnected the collapsed cubes immediately without administrator intervention. We filter for these collapse errors by locating multiple cubes simultaneously disconnected and within 10 seconds, simultaneously reconnected. We then "snip" the error from the similarity function.

Cognitive Sensitivity Study

To confirm and improve the sensitivity of Cognitive Cubes, we studied its response to three factors known to correspond to differences in cognitive ability: participant *age* (34-, 54+), *task type*, and *shape type* (2D, 3D).

Since cognitive ability declines gradually with increasing *age*, we expected younger participants in this study to perform better with Cognitive Cubes than older participants. As cognitive load of a task increases, cognitive abilities are stressed, leading us to expect better performance with *task types* requiring less planning. Similarly, we have already noted the heavier cognitive demands involved in working with 3D *shape*. We anticipated better performance with 2D shapes than with 3D shapes.

Because cognitive ability decreases with the progression of AD, we made a preliminary study of the sensitivity of Cognitive Cubes to that form of dementia. No cure for AD exists, but its early detection can have an enormous impact on palliative care and quality of life [10]. Although the numbers of AD patients in this study were small, we include some limited results here.

Participants and Methods

The cognitive sensitivity study had 16 participants, recruited on and off campus, ranging in age from 24 to 86, with 4 females and 12 males. 7 of the participants were young, 7 elderly and 2 had mild AD. Methods were as described above, but participants also performed a mini mental state evaluation.

Results

In these results, we exclude connection errors and system crashes, and filter collapse errors as described in the pilot study. During experimentation, we repeated trials with system crashes and included them in analyses. Frequencies of these errors and repetitions are listed in table 1.

Because there were so few AD participants, we exclude them from any analyses of variance. As noted above, elderly participants often were not able to complete all tasks, unbalancing the *age* factor in our ANOVAs (recall that *task type* and *shape type* are already unbalanced in the design). We list the number of trials completed in table 2.
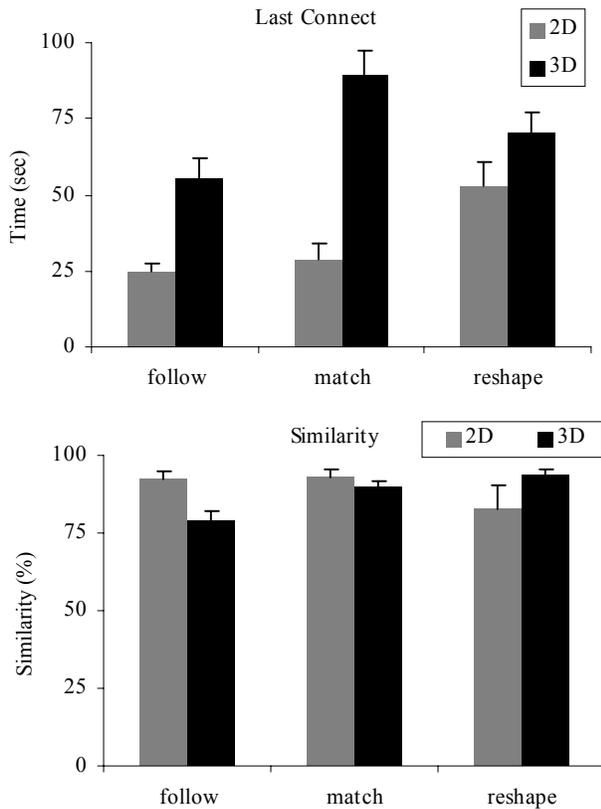
Results are presented in tables 3-4 and figures 4-7. We analyzed results using one 3-factor (2 *age* x 3 *task type* x 2 *shape type*) unbalanced ANOVA for each of the *last connect, similarity, zero crossings*, and *derivative* measures. We exclude *intro* task type and AD participant results from the analyses.

All three factors produce main effects in line with our expectations. Participant performance varies significantly by *age* (table 4), with elderly participants needing more time to complete each task, and showing a low rate of progress. By all four dependent measures, participant performance is also significantly affected by *shape type* (figs. 4-7). 2D shape construction is completed more quickly, more accurately, and with a higher and steadier rate of progress. Finally, *task type* also has significant effects on all four measures. *Follow* is the easiest of the task types, enabling quick completion and a high, steady rate of progress toward the target shape. However, shape *similarity* is lowest with the *follow* task. Participants perform the *match* and *reshape* tasks with roughly equal completion times and *similarities*, but the rate of progress in the *match* task type is higher and steadier.

The only significant interaction, by all four measures, is between *shape type* and *task type* (figs. 4-7). In general, for 2D tasks, *follow* and *match* are roughly equal in difficulty by all four measures. *Reshape* is more difficult. For 3D tasks, *follow* is simplest, followed by *match*, then *reshape* − with *last connect* time a lone exception among the measures: *reshape* tasks are completed more quickly than *match* tasks.

Table 4: The main effect of *age* on *last connect* and *derivative*. Last connect is measured in seconds, derivative in similarity per second. Means are presented with standard error in parentheses.

Dependent Measures	Age		
	young	elderly	AD
Last connect (sec)	48.45 *(3.5)*	76.33 *(5.3)*	91.24 *(8.8)*
Derivative (sim/sec)	3.02 *(.13)*	1.97 *(.1)*	1.43 *(.16)*

Table 5: Pre-CC MRT/Cognitive Cubes correlations. Significance: (p<.1) in bold, (p<.05) in italics.

Dependent Measures	Overall	Shape Type		Task Type		
		2D	3D	follow	match	reshape
Last connect	-0.38	**-0.49**	-0.35	***-0.63***	-0.35	-0.24
Similarity	0.03	-0.36	0.17	0.16	-0.09	0.08
Zero crossings	-0.23	0.07	-0.25	-0.14	-0.45	0.11
Derivative	**0.51**	0.38	*0.57*	0.43	**0.50**	0.34

Figures 4-5: The effect of the *shape x task type* interaction on *last connect* and *similarity*.

Figures 6-7: The effect of the *shape x task type* interaction on *zero crossings* and *derivative*.

Test Comparison Study

Having studied the sensitivity of Cognitive Cubes to factors related to cognitive performance, we turn to a comparison of Cognitive Cubes to a known tool for 3D spatial assessment: the Mental Rotation Test (MRT) [11,12,15]. Like Cognitive Cubes, the MRT is 3D and spatial, leading us to expect a strong relationship between the two assessments, particularly with 3D tasks. However, since the MRT does not include any of Cognitive Cubes' constructional, planning or motor task components, we might anticipate the relationship to be limited to simpler tasks such as *follow*.
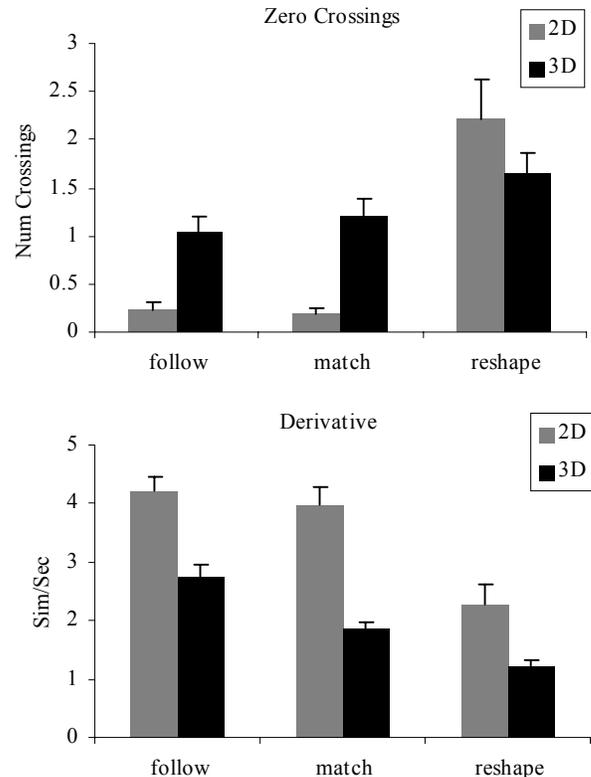
Participants and Methods

The test comparison study's 12 participants had ages ranging from 18-36, with 4 females and 8 males. Participants were all volunteers recruited on and off campus. Procedure followed the general methodology except that participants took the MRT test before and after the Cognitive Cubes assessment.

Results

Unlike the previous study, this experiment used a homogenous set of participants without any elderly or AD participants. Completion rates were therefore uniformly high. System error rates were similar to those reported in the cognitive sensitivity study.

Cognitive Cubes results in all four measures are compared to MRT results obtained both before (pre-CC) and after (post-CC) the Cognitive Cubes assessment. Interestingly, post-CC MRTs are markedly improved (almost all in the 90th percentile). While it is well-known that repeating the MRT brings improved performance, improvements are in this case well above the normally reported repetition improvement rate of roughly 5%.

We perform our MRT/Cognitive Cubes comparisons using correlations. Because they reached ceiling and lost sensitivity, correlating post-CC MRTs to Cognitive Cubes measures would be a meaningless exercise; we do not present them here.

The correlations of pre-CC MRT and Cognitive Cubes are presented in table 5. The measure with the most significant overall correlation (and the only reaching

marginal significance) is the *derivative* measure. Correlations to *zero crossings* are low. Correlations to *similarity* are also low, perhaps because *similarities* are uniformly high. Correlations to *last connect* are also high. Correlations are only slightly stronger for 3D than 2D shapes, while correlations are strongest with *follow* tasks, slightly weaker with *match* tasks, and completely untrustworthy with *reshape* tasks.

DISCUSSION

In this section we review the results of our experimental examination of Cognitive Cubes, including outcomes both expected and surprising, as well as implications for the use and improvement of Cognitive Cubes.

Limitations

We begin, however, by noting several reasons for caution when drawing inferences from our experimental work. First, our experiments were motivated by the desire to improve and learn about the Cognitive Cubes tool, and not by basic scientific questions about cognitive function. This led us to choose an experimental design that was not balanced in *shape type* or *task type*, so that we could emphasize those tasks that seemed to us most promising in assessment. ANOVA results should therefore be interpreted with care.

We did not randomize or counterbalance ordering of task type and difficulty, instead we used a rough order of increasing difficulty. This enabled participants beginning to struggle with the current task to skip following tasks with which they most likely also would struggle, which proved crucial in retaining elderly and AD participation. At the same time, this decision not to randomize or counterbalance introduced a practice effect that must be reckoned with in the effect of the *task type* factor.

Since the elderly and AD participants struggled with tasks most often, only the stronger of these participants completed the more difficult tasks (table 2). This unbalanced the *age* factor and made task performance by these participant groups appear better than it would otherwise. Even so, mean task performance by elderly and AD groups was still worse than performance by the young.

Finally, we introduced filtering and repetition into our trials to handle the remaining hardware shortcomings of our prototype. The frequency of repetition was relatively low, and as we already have noted, analyses excluding them were very similar to those shown here. Filtering was more frequent, and since structural collapses were more likely with elderly participants and during 3D tasks, it may have distorted the results of the *age* and *shape type* analyses. However, since our filter only affected Cognitive Cubes events which removed and immediately replaced multiple blocks in less than ten seconds (otherwise an infrequent and discouraged participant action), we are confident that we have controlled this potential confound well.

Confirmations

First, we note that the ActiveCube hardware component of Cognitive Cubes performs quite well, given it is a prototype. While spot repairs were sometimes required, the hardware continued to function through well over 50 hours of demanding use. Participants had few complaints, were engaged and interested, and were usually having fun.

Spatial cognitive performance is known to decline with increasing age, cognitive load, and shape complexity, so it is reassuring and gratifying to see these trends in Cognitive Cubes' measures. The only exception to this trend is in the effect of *task type* shown by *similarity*, which is lower for the *follow* task, and thus less similar to the target despite less cognitive load. Since *follow* was always the first task type, this may be a side effect of ordering: participants had not yet reached peak performance when they were performing *follow* tasks.

Preliminary results indicate that Cognitive Cubes is sensitive to mild AD. Though we examine Cognitive Cubes with only two mild AD participants (fig. 3 and table 4), outcomes show strong differences between unaffected elderly and AD participants. Further work should examine whether Cognitive Cubes can discriminate between AD and other explanations of constructional weakness.

Finally, not only does Cognitive Cubes respond well to known cognitive factors, but certain of its component measures also have sensitivities similar to a 3D assessment already in wide use: the MRT. Other Cognitive Cubes components promise additional sensitivities.

Surprises

Since our primary goal was system evaluation and not cognitive research, we did not form any hypotheses about interactions among the cognitive experimental factors of *age*, *task type*, and *shape type*. However, we are pleasantly surprised that Cognitive Cubes is sensitive to an interaction between the *task* and *shape type*. One possible explanation of the interaction is that with 2D target shapes, the additional cognitive planning load of *match* (vs. *follow*) is minimized. At the same time, since the starting point in *reshape* is 3D, 2D targets still require significant cognitive planning. With 3D target shapes (by all measures except *last connect*), the cognitive load increases steadily from *follow* to *match* to *reshape*. The *last connect* exception may indicate the added time it takes to move from a 3D to a 2D target shape. Alternatively, it may also result from the combined effect of participant dropout and practice.

Contrary to our expectations, both 2D and 3D task types produce good correlations to the 3D MRT. We believe this may well be attributable to task difficulty. While the MRT asks the user to perform a small set of relatively simple 3D mental rotations, Cognitive Cubes challenges participants to construct a single shape, which may be small or large, 2D or 3D. Which is more like the MRT, building from scratch a complex 3D shape, or a simple 2D shape? The answer is unclear, and thus the lack of clarity in the *shape type* correlations.

The improvement from pre- to post-CC MRT is unexpected, but very intriguing. Could Cognitive Cubes be used as a

form of cognitive therapy or training, for example in rehabilitation?

Implications

Is Cognitive Cubes *a useful tool for assessment?* Our experience certainly indicates great promise. Despite being a prototype, the ActiveCube hardware component stood up well to intense use and proved to be quite intuitive for our participants. In experimental evaluation, the system as a whole was sensitive to well-known cognitive factors and compared favorably to an existing assessment. Automation introduced a previously unachievable level of reliability and resolution in 3D measurement and scoring. Despite all this, Cognitive Cubes is not yet ready for regular use.

How might Cognitive Cubes *be prepared for use in the field?* The gap between a good prototype and a reliable tool is a large one. Use in clinical or research settings would require significant improvements in cost, reduction of connection and system errors, and improvements in structural strength. These are fairly typical requirements for the development of any technology. In addition, extensive testing would be required to identify the distribution of scores typically achieved with Cognitive Cubes. In this way, assessors can reliably decide whether or not a score is exceptional.

How might Cognitive Cubes *be improved?* The system could be greatly improved with a more polished notion of task difficulty, which then might be used to weigh assessment results over multiple tasks into a composite score. In this study we use *shape type* as a rough approximation of difficulty (without weighting composite scores), but certainly the number of blocks needed to build a shape should also be a factor. Researchers and thinkers in a variety of fields have proposed numeric measures of shape complexity [8]; Cognitive Cubes could be a good mechanism for testing their relevance to humans.

One the most unique strengths of Cognitive Cubes is its ability to capture each step of task progress – closely mirroring the cognitive processing of the participant. With the same data used to build similarity graphs, it is also possible to build decision trees reflecting the participant's chosen path through the space of possible cube-by-cube construction sequences, probing this dynamic process even more deeply. Participant trees might then be categorized for assessment with Bayesian network analysis.

What does Cognitive Cubes *imply for 3D and tangible user interfaces?* We are by no means alone in advocating that the interface suit the application, a maxim repeated so often now that it can seem trite. We believe that Cognitive Cubes is a convincing example of the truth in this maxim, and as such an example it makes a compelling case for the utility and the future of 3D and tangible UIs.

CONCLUSION

In this paper, we have described the design and evaluation of Cognitive Cubes, to our knowledge the first system for the automated assessment of 3D spatial and constructional ability. Cognitive Cubes makes use of ActiveCube, a 3D building-block TUI for describing 3D shape. Cognitive Cubes offers improved sensitivity and reliability in assessment of cognitive ability and ultimately, reduced cost. Experimental evaluation with 43 participants confirms the sensitivity and reliability of the system.

REFERENCES

1. Aish R., "3D Input for CAAD Systems", Computer-Aided Design, 11 (2):66-70, Mar. 1979.
2. Anagnostou G., Dewey D. & Patera A. T., "Geometry Defining Processors for Engineering Design and Analysis", The Visual Computer, 5:304-315, 1989.
3. Anderson A., Frankel J.L., Marks J., Agarwala A., Beardsley P., Hodgins J., Leigh D., Ryall K., Sullivan E. & Yedidia J.S., "Tangible Interaction + Graphical Interpretation: A New Approach to 3D Modeling", SIGGRAPH 2000, July 23-28. New Orleans, LA, 2000.
4. Frazer J. H., *An Evolutionary Architecture*, Architectural Assoc 1995.
5. Gorbet, G. M., Orth, M. and Ishii, H., "Triangles: Tangible Interface for Manipulation and Exploration of Digital Information Topography", Proc. CHI '98, 49-56, April, 1998.
6. Groth-Marnat G., *Handbook of Psychological Assessment*, 3rd ed., New York: John Wiley & Sons, 1997.
7. Kitamura Y., Itoh Y. and Kishino F., "Real Time Interaction with ActiveCube", CHI 2001 Extended abstracts, 2001.
8. Koenderink, J.J., *Solid shape*, MIT Press, Cambridge, MA 1990.
9. Lezak M.D., *Neuropsychological Assessment*, NY: Oxford UP 1983.
10. Liu L., Gauthier L. & Gauthier S., "Spatial Disorientation in Persons with Early Senile Dementia of the Alzheimer Type", The American Journal of Occupational Therapy, 45 (1): 67-74, Jan. 1991.
11. Peters, M., Laeng, B., Latham, K., Johnson, M., Zaiyouna, R. & Richardson, C., "A Redrawn Vandenberg & Kuse Mental Rotation Test: Different Versions and Factors that Affect Performance", Brain and Cognition, 28, 39-58, 1995.
12. Rizzo, A.A., Buckwalter, J.G., Neumann, U., Kesselman, C., Thiebaux, M., Larson, P., & Van Rooyan, A., "The Virtual Reality Mental Rotation/Spatial Skills Project: Preliminary Findings". CyberPsychology and Behavior, 1(2), 107-113, 1998.
13. Suzuki, H. & Kato, H., "Interaction-Level Support for Collaborative Learning: AlgoBlock - An Open Programming Language", Online: http://www-cscl95.indiana.edu/cscl95/suzuki.html
14. Ullmer B. & Ishii H., "Emerging Frameworks for Tangible User Interfaces", in *Human-Computer Interaction in the New Millennium*, by John M. Carroll (Ed), Addison-Wesley, 2001.
15. Vandenberg, S.G. & Kuse, A.R., "Mental Rotations: a Group Test of Three-Dimensional Spatial Visualization", Perceptual and Motor Skills, 47, 599-604, 1978.