

Measuring 3-D understanding on the Web and in the laboratory

KEN SUTTON, ANDREW HEATHCOTE, AND MILES BORE
University of Newcastle, Callaghan, New South Wales, Australia

We describe the ongoing development of a psychometric test to measure understanding of three-dimensional (3-D) concepts represented in drawings. The test consists of five subtests requiring participants to complete timed choice tasks previously used in experimental investigations of 3-D understanding, and a new subtest based on the idea of true length. We examined performance in terms of accuracy and response time using both laboratory- and Web-based methods of delivery. Reliability coefficients were high for both methods (.90 and .96, respectively, for accuracy, and .87 and .95, respectively, for response time measures), but Web-based participants produced consistently lower overall scores. Issues regarding the deployment of, and recruitment for, complex Web-based experimental tasks, and factors affecting comparisons between Web and laboratory results, such as the presence of an experimenter, and interface and sample differences, are discussed.

One of the more difficult skills to learn for students studying technical drawing is the ability to understand three-dimensional (3-D) concepts. Technical drawing is a graphical communication method used by designers, architects, engineers, and tradespeople to share technical information about objects such as buildings, machinery, and engineering structures. Technical drawing skills contribute to the development of products from the conceptual stage through to the manufacturing stage. 3-D understanding is the ability to extract information about 3-D properties from two-dimensional (2-D) representations (i.e., drawings; Sutton, Heathcote, & Bore, 2005). This skill requires visual and perceptual abilities to interpret what is seen, and cognitive and spatial abilities to mentally manipulate visual representations.

Modern computer-based graphical techniques, such as computer-assisted design packages, offer increased flexibility and speed in developing 2-D representations of 3-D objects. However, they can also place greater demands on a novice user's attention, requiring him or her to focus on issues related to the software interface, and reducing effort devoted to 3-D understanding. The impetus for the present research was the need for a test that measures the various aspects of 3-D understanding relevant to skill development. We begin by briefly reviewing traditional approaches to teaching 3-D understanding and the opportunities afforded by modern software for improving that training. We then describe our test, the 3D Ability Test (3DAT). 3DAT takes advantage of computer-based delivery by measuring not only choice accuracy but also a second aspect critical to skilled performance, choice speed. Our approach is novel in this respect, in that 3DAT is more

akin to experimental rather than traditional psychometric approaches to ability measurement. We then examine the reliability of 3DAT as delivered in a laboratory setting and a Web-based setting.

Teaching 3-D Understanding

Conventional teaching of technical drawing is partly passive. Students learn from observing instructor-centered demonstrations that primarily focus on scaled models of objects encased in Plexiglas frames (Duesbury & O'Neil, 1996). Often, the views of objects are manually projected onto planes represented by the Plexiglas to illustrate what is seen from different viewing directions, and to illustrate what changes occur when an object is moved. Such learning techniques provide very little hands-on experience and encourage students to rote learn a set of rules rather than developing a deeper understanding. Rote learning can be effective for simple and familiar examples but is unreliable for complicated and novel structures. To be effective in all situations, students require a fundamental appreciation of 3-D concepts. The most active and hands-on aspect of traditional training is the production of paper models of objects. Folding and unfolding models reinforces the importance of active exploration but is time consuming, and many of the manual processes involved in this technique are not directly relevant to developing 3-D understanding.

Technical drawing is usually based on a standard set of three axes (x , y , and z) meeting at right angles at a point called the origin. The axes provide a Cartesian coordinate system for locating points, lines, and planes. Three reference planes are usually defined, one parallel to the x -, y -

K. Sutton, ken.sutton@newcastle.edu.au

axes (x, y plane), another to the x, z -axes (x, z plane), and the third to the y, z -axes (y, z plane). The details of an object are normally represented as a set of two-dimensional (2-D) drawings, where each represents a separate view of the object. 2-D views are referred to as orthographic views. To produce a set of orthographic views, an object is positioned with respect to the three axes and individual views are projected perpendicularly onto each of the reference planes. This method of projection is termed orthographic projection. The view seen through the (x, y) plane is called the top view (TV), the view seen through the (x, z) plane is called the front view (FV), and the view seen through the (y, z) plane is called the end view (EV). 2-D views given by orthographic projection are termed degenerate views, because one axis is excluded. The distance of the object from the reference planes is not critical to the shape of the projected views, because the shape is the same (isomorphic), regardless of the distance in an orthographic projection.

Isometric drawings, which provide more obvious information about the 3-D properties of an object than orthographic projections, are produced by a view that is not parallel to any axis, with the view being projected onto a plane perpendicular to the viewing direction. For an isometric drawing, the top view of a line representing the viewing direction is typically at 45° , 135° , 225° , or 315° to the x -axis. In the standard setting of most computer-assisted design software, the viewing direction will be a true angle of 35.3° to the (x, y) plane. A true angle is seen when a projection of a line or edge is parallel to a viewing plane. In isometric drawings, parallel edges running away from the viewing plane are always drawn as parallel lines. This contrasts with perspective drawings, in which parallel edges that run from the viewing plane are drawn as converging lines. However, for relatively small objects (as opposed to, say, landscapes) the difference between isometric and perspective drawings is negligible. Given this, and the dominant use of isometric rather than perspective representations in technical drawing, we focus on the former.

A common approach to developing technical drawing competency is to experience both 2-D and 3-D representations concurrently in order to develop an understanding of the relationship between the two. It is not desirable to develop 2-D or 3-D skills in isolation, and in many respects, working in 2-D may be more important than working in 3-D. James, Humphrey, and Goodale (2001) report that participants in their experiments spent more time looking at the end and front views of objects rather than three-quarter or intermediate views. They contend that these are the views where there is the greatest amount of difference in the visibility of object features. In contrast, the three-quarter views are perceptually similar. The process of working from 2-D to 3-D drawings, and working from 3-D to 2-D drawings, is the common way students build up their understanding of concepts. The ability to interpret a multiview drawing is learned by forming mental images from the 2-D views and visualizing what the object will look like in 3-D. As the complexity of objects increases, extra views are generally necessary, including sectional views (projections of planes cut through objects), ex-

ploded views (magnified projections showing individual parts separated), and assembled views (working parts in position).

Modern graphical and multimedia software offers an opportunity to improve on these traditional training techniques. Such software enables flexible implementations of analogues of all of the tasks used in traditional training. Such implementations allow the user to more easily engage in active exploration and learner-controlled manipulation of realistic 3-D models using animations that increase awareness of object properties. Evidence suggests (e.g., James, Humphrey, & Goodale, 2001; James et al., 2002) that active exploration and control of novel objects assists the learning of 3-D structures, better object recognition, and improved spatial ability. Other researchers, including Piaget (1953), Gibson (1979), Held (1965), and Neisser (1976; all as cited in James et al., 2001), have emphasized the importance of motor activity, including exploratory activity, in perceptual and cognitive development.

In order to take advantage of the training opportunity presented by modern graphical software, a valid and reliable test of 3-D understanding is required. This article reports the initial steps taken to develop such a test, the 3D Ability Test (3DAT). 3DAT addresses all of the skills emphasized in traditional training, such as understanding of different types of projections, the concept of true length, folding and unfolding, and the properties of coordinate systems. 3DAT is delivered by a computer, enabling measurement of both accuracy and speed. Speed is particularly important to the full development of expertise, because the final stage of skill acquisition is marked by a transition from mastery—the ability to perform a task in a relatively error free but slow and effortful manner—into effortless and fast performance, as exemplified by language fluency in an experienced native speaker (e.g., Fitts, 1964; Spelman & Kirsner, 2005). Studies of the development of fluency in cognitive choice tasks show that participants are able to reduce response time (RT) markedly in the transition from mastery to fluency while maintaining a high and constant, or only slightly increasing, level of accuracy (e.g., Heathcote, Brown, & Mewhort, 2000). Hence, measurement of both accuracy and RT enables 3DAT to remain sensitive to improvements throughout all stages of skill acquisition.

Computer delivery enables 3DAT to be used in both laboratory- and Web-based settings. Laboratory studies can be problematic, both because of the resources required to obtain sample size sufficient for statistical techniques used in developing a psychometric test (e.g., factor analysis), and to some extent, because it is difficult to sample a demographic representative of the general community. Web-based research provides a possible resolution to these problems, and it also provides a more valid representation of ability in the absence of the instructional support available in the laboratory and classroom. Steyvers and Malmberg (2003) and Birnbaum (2004) provide evidence that reliability and validity of data from Web studies compare favorably with data collected from parallel laboratory studies. We provide a comparison of 3DAT performance in parallel laboratory- and Web-based studies in order to compare their reliabilities and to validate the Web delivery method.

3-D ABILITY TEST

Blasko, Holiday-Darr, Mace, and Blasko-Drabik (2004) emphasize the need to use multiple spatial cognitive tasks to assess 3-D understanding. They report results from mental rotation and correct fold tasks similar to ours using a Web-based presentation (viz.bd.psu.edu/viz/). Our scale consisted of 89 items divided into six subtests. Five subtests were based on previous psychological research, including the correct fold and mental rotation tasks used by Blasko et al. (2004), whereas the sixth subtest is based on the idea of true length, an important concept in technical drawing. An edge of an object can be represented in any view of the object, but its true length is not always seen; only edges parallel to a projection plane have their true length in a projection. The items are varied in form and most are novel in design. The items are constituted of straight lines and flat planes. (3-D understanding for curved objects will be addressed in further development of 3DAT.) They were created using computer-assisted design software and saved in bitmap and GIF formats for the lab and Web studies, respectively. Image resolutions were comparable, and the different formats were required to suit the software used for the two studies. A description of each of the six subtests follows; example items may be found in the Appendix.

2-D/3-D Recognition

Objects are presented as orthographic and isometric projections. Participants select which of two alternatives of one type matched a standard of the other type (Bertoline & Miller, 1990; Cooper, 1990). Subtests use either (1) an orthographic standard or (2) an isometric standard, with eight and nine items, respectively (see Figure A1 for examples).

Correct Fold

Objects are presented as an isometric projection or as an unfolded view. Participants selected which of two alternatives of one type matched a standard of the other type (cf. Blasko et al., 2004). Subtests use either (1) an isometric standard or (2) an unfolded standard, with five items for each (see Figure A2 for examples).

True Length Recognition

Objects are presented as isometric and orthographic projections. In one subtest, participants decide which view in a set of orthographic projections shows the true length of a labeled edge in an isometric projection (True Length Recognition A). In a second subtest, participants decide which of three isometric projections shows the true length of a labeled edge in a set of orthographic projections (True Length Recognition B). There are 13 and 9 items, respectively, in the subtests (see Figure A3 for examples).

Mental Rotation

Participants decide whether a rotated isometric projection of an object matches the isometric projection of a standard or its mirror image (Metzler & Shepard, 1988). The object on the left is always in the same position and

is the referent. The object on the right can be the same or the mirror image of the referent, and its orientation in the (x, y) plane can be different. There are five matching and five mismatching items (see Figure A4 for examples).

Possible/ Impossible Structures

Participants decide whether an isometric projection can represent a 3-D object (Schacter & Cooper, 1990). The objects can be one of two types. The first (possible) is one where the projection can reasonably represent a true object. The second (impossible) displays some visual feature that could not reasonably represent an aspect of a true object. There are 6 and 13 items of each type, respectively (see Figure A5 for examples).

Dot Coordinate

Participants are shown an isometric projection of a 3-D Cartesian coordinate system and a text description of the position of a point in that system. From four orthogonal projections, participants choose the projection that corresponds to the description (Bore & Munro, 2002). There are 11 items (see Figure A6 for an example).

METHOD

Laboratory Study

Participants worked through the 89 items organized as a set of computer-controlled activities. The study was created in SuperLab 2.01, an experimental software package used for psychological research. Participants had control over the initiation of each subtest, with each subtest preceded by instructions containing an example and advice about how to respond. Practice trials for all subtests were conducted before the actual study, to allow familiarization with the subtests and response procedures. The setup was explained by the researcher, and participants could ask questions. No feedback was given during practice or testing. The study was conducted with groups of approximately 5 participants, who were taken through the practice trials to explain what was required, but no strategies to determine correct answers were discussed. Instructions emphasized an understanding of true length and the relationship between an isometric drawing and orthographic projections.

Breaks were built into the study to safeguard against fatigue; they occurred at the start of each of the nine subtests. Participants controlled the duration of the breaks by initiating the start of each subtest after reading through the instructions and studying the example provided. Excluding breaks, the study took about 60 min to complete. The subtests were presented in the same order as in the description of the 3DAT given in the last section, but the order of items within each subtest was randomized for each participant. Participants entered their responses using a six-button response pad.

Participant eligibility criteria were (1) 18 years of age or older and (2) no self-reported prior technical drawing experience. These criteria were made explicit in recruitment advertising, and no participants applied to do the experiment who did not meet them. The sample of 41 participants (32 females and 9 males) was drawn from a participant pool of psychology students in first-year university classes, who received course credit for participation.

Web-Based Study

The Web study replicated the laboratory study as closely as possible, with differences noted below. It was developed to utilize Cold-Fusion MX using Mach II methodology. As a measure to protect against poor Web experimental design, the implementation was checked against the 16 standards suggested by Reips (2002a). Because Web participants had to work independently, whereas laboratory participants' questions could be answered by the researcher,

additional explanations were considered necessary. As a result, detailed information was provided to explain the relationship between orthographic projection and isometric drawings, the experimental design, and the concept of true length. Hence, participation in the Web study was more demanding in terms of reading and understanding the test requirements than for the laboratory study, and as a consequence of this, and additional demographic information collected, it took slightly longer (75 min, on average) to complete. Participants recorded their responses by mouse-clicking a number using the same numbering scheme as for response buttons used in the laboratory study (e.g., 2, 3, or 4). The numbers were displayed on the screen but separated from the image choices. RT was measured on the client side and managed through the Web browser. From the start of each image being displayed, a JavaScript counter recorded the time until a response was received (excluding a short delay intentionally built in to accommodate image loading time). The time taken (RT) was then logged with the response of the participant.

Krantz (2001) identifies stimuli as a potential confound for comparison of our laboratory and Web results and emphasizes the need for calibration. Krantz provides reasons for calibration such as differences in monitor displays, image stability, and inconsistency of color across monitors. However, the laboratory and Web formats of 3DAT differed only minimally, because a sophisticated Web interface was used that was equivalent to the laboratory format in most aspects. The interfaces were near identical with the exception of text position, and the laboratory study required the use of a response pad, whereas participants in the Web study needed to mouse-click on numbers. The delivery of the images included a time delay before each image displayed, to allow for hardware differences (also included in the laboratory study), the average image file size was only 8 KB, and the images were simple line figures without color or rendering. The focus of our comparison was on completing the study in a quiet, controlled environment, versus completing the study over the Web using a virtually identical interface.

To identify the profile of the Web participants, a demographic section in the study asked about gender, country of residence, ethnicity, and vocation. This section also asked participants whether they were age 18 or older and whether they had previous technical drawing experience. Participants younger than 18 or those with technical drawing experience were excluded from analyses, although they were able to complete the experiment. When the test was completed, participants were provided with a score out of 89. The results from the excluded group were not recorded, and the final sample size of 30 consisted of 23 females and 7 males. Of 260 eligible participants who entered the demographics section of the study, 80 made a start on the testing phase and 48 completed it, with 18 more being excluded because of a technical problem reported later in this article. No participants were excluded on any other basis. The final sample of 30 consisted of participants from several countries, mostly from the USA (67%), and from a range of vocations, such as academic, service, professional, and military, with the majority (60%) indicating that they were students. Recruitment of Web participants was conducted through the Psychological Research on the Net site (psych.hanover.edu/Research/exponnet.html), and special psychology interest groups like those suggested by Birnbaum (2004) and Reips (2002a). Recruitment from interest groups was carried out by advertising on their Web sites. Web participants could nominate for a prize draw, with the prize being an Au\$40 gift voucher. The Web version of our test can be viewed by linking to the Web site at webapps.newcastle.edu.au/2d3dsurvey/index.cfm. Participants were not able to proceed to the actual study without first completing the demographics section and the practice trials.

RESULTS

We tested reliability by comparing Cronbach alpha coefficients and validity by comparing mean accuracy and mean RT for correct answers between our laboratory-

and Web-based samples. Reliability results are reported in Table 1. Generally, both laboratory- and Web-based subtest scores produced acceptable alpha reliability coefficients. Psychometric standards define acceptable coefficients as greater than .7, with values above .8 considered highly acceptable. Values closer to zero indicate poor consistency across items. The low alpha coefficients found for 2-D/3-D Recognition—A (1A) and Correct Fold—B (2B) in the laboratory sample were not found in the Web sample. Reliability coefficients for the combined subtests indicate high reliability in both laboratory- and Web-based studies. Noteworthy is that reliability is consistently greater for the Web study across all subtests for accuracy and for all but one subtest for RT.

Correlations between the accuracy scores for each subtest for the Web-based sample and the lab-based sample are shown in Table 2. The high internal reliability of the total scores (total of correct responses) is reflected in the high correlations generally found between subtest accuracy scores. The exceptions, as would be expected given the alpha reliability coefficients, were the 2-D/3-D Recognition—A (1A) and Correct Fold—B (2B) subtests for the lab-based sample, where correlations between these subtests and all other subtests of the tests were weak and did not reach significance. This was not found in the Web-based sample, where strong correlations between the 2-D/3-D Recognition—A (1A) and Correct Fold—B (2B) subtests and all other subtests were observed. Of additional interest were the moderate to strong correlations for the Dot Coordinate scores. This particular subtest requires considerably more reading of instructions, while also being the most difficult subtest of the six subtests presented in the instrument (see below for analysis of percentage correct by subtest). However, the correlations found suggest that the Dot Coordinate items are measuring the same ability as the other items of the test.

In order to examine the relative difficulty of each subtest and to compare the difficulties between Web- and lab-based samples, the mean percentages of correct responses given for each subtest were calculated and plotted as shown in Figure 1 (standard error of the means are also shown). Lab-based participants achieved a higher mean percent correct across all nine subtests of the 3DAT in comparison with the Web participants, although these differences

Table 1
Comparison of Cronbach Alpha Reliability Coefficients for Parallel Laboratory and Web Studies

Subtests	Web-Based Study		Laboratory Study	
	Accuracy	RT	Accuracy	RT
2-D-3-D recognition—A (1A)	.68	.76	.09	.69
2-D-3-D recognition—B (1B)	.79	.74	.48	.82
Correct fold—A (2A)	.42	.69	.38	.56
Correct fold—B (2B)	.57	.78	-.02	.51
True length recognition—A (3A)	.89	.76	.80	.68
True length recognition—B (3B)	.80	.70	.54	.50
Mental rotation (4)	.62	.83	.61	.60
Possible/impossible structures (5)	.83	.88	.74	.84
Dot coordinate (6)	.92	.94	.82	.74
All tests combined	.96	.95	.90	.87

reached significance in only four of the subtests: 2-D/3-D Recognition—B (1B), Correct Fold—A (2A), Mental Rotation (4), and Possible/Impossible Structures (5). For both Web and lab samples, the lowest mean percentage correct was for the Dot Coordinate subtest. For the Web sample, a one-way ANOVA of the percentage correct means of the nine subtests was significant [$F(8,261) = 8.5, p < .001$], with Tukey's pairwise comparisons (family error rate, $p = .05$) showing that the percentage correct mean for Dot Coordinate was significantly lower than for all other subtests with the exception of the Correct Fold—A (2A) subtest. In the lab sample, significant differences were also found [$F(8,360) = 16.9, p < .001$], with the Dot Coordinate's percent correct mean being significantly lower than all other subtests' percent correct means.

The time taken to give a correct response was recorded for both Web- and lab-based samples, and a mean time (in seconds) and the standard error of means were calculated for each subtest. As shown in Figure 2, the pattern of mean correct RTs was highly similar for both Web and lab samples, with Web participants taking longer than laboratory participants for six of the nine subtests. These differences were significant for two subtests: Mental Rotation (4) and Possible/Impossible Structures (5). The input method for the Web study (mouse movements and clicks on a Web page) may account for the slower RTs. A one-way ANOVA of the Web sample mean RT for correct responses revealed a significant effect of subtest [$F(8,248) = 18.8, p < .001$], with Dot Coordinate times being significantly longer than for True Length Recognition—A (3A), Mental Rotation (4), and Possible/Impossible Structures (5), as indicated by Tukey's pairwise comparisons. A one-way ANOVA on the laboratory-based sample mean RT for correct responses also revealed a significant effect of subtest [$F(8, 360) = 37.3, p < .001$], with Dot Coordinate correct responses taking significantly longer to produce in comparison with

Table 2
Correlations Between Subtest Accuracy Scores for Web Sample and Lab Sample

	1A	1B	2A	2B	3A	3B	4	5	6
Web-Based Study									
1B	.77**								
2A	.58**	.50**							
2B	.48**	.36*	.51**						
3A	.63**	.74**	.55**	.51**					
3B	.62**	.66**	.53**	.54**	.83**				
4	.53**	.54**	.33	.62**	.44*	.43*			
5	.59**	.56**	.60**	.70**	.64**	.71**	.66**		
6	.48**	.54**	.51**	.44**	.75**	.63**	.28	.53**	
Total	.77**	.80**	.69**	.69**	.90**	.87**	.62**	.84**	.79**
Lab-Based Study									
1B	.28*								
2A	-.13	.23							
2B	-.03	.19	.19						
3A	.15	.56**	.40**	.16					
3B	.13	.66**	.47**	.21	.65**				
4	-.08	.42**	.41**	.36*	.55**	.56**			
5	-.00	.54**	.38*	.25	.73**	.70**	.49**		
6	.21	.36*	.46**	.38*	.43**	.56**	.35*	.41**	
Total	.21	.68**	.57**	.40**	.82**	.85**	.67**	.81**	.76**

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

all other subtests. The correct RTs for Mental Rotation (4) and Possible/Impossible Structures (5) were significantly faster in comparison with all other subtests.

The pattern of results across subtests in Figure 1 is a mirror image of the pattern across subtests shown in Figure 2, suggesting that participants took longer to correctly answer harder subtests (lower percentage of correct responses) but took less time to correctly answer easier subtests (higher percentage of correct responses). This was further examined by correlating the percentage correct means with mean correct RT for each subtest, as shown in Table 3. Significant, positive, and moderate to strong correlations were found for each subtest and the total for

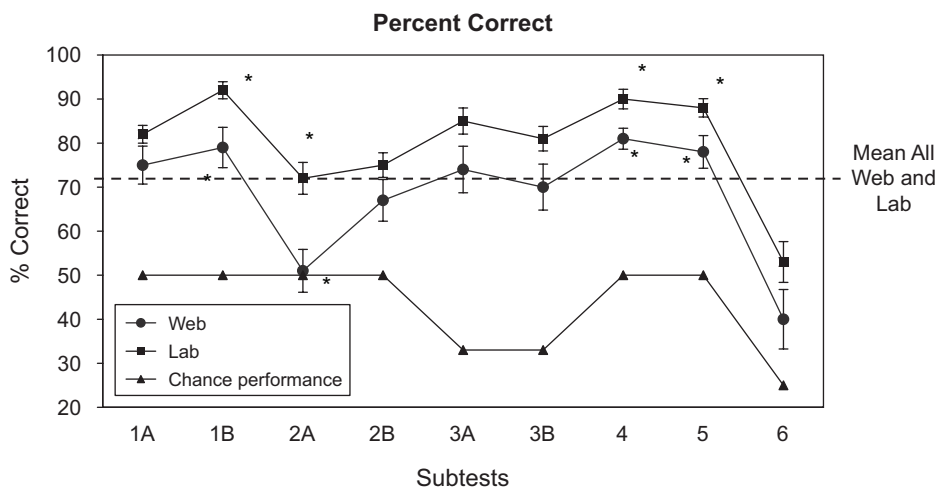


Figure 1. Plot of mean correct responses and standard error per subtest for Web and lab samples. The line labeled "chance" indicates guessing performance, given the number of test alternatives. The interpretation of laboratory results for Tests 1A and 2B should be qualified by their low reliability (see Table 1). *significant difference ($p < .05$).

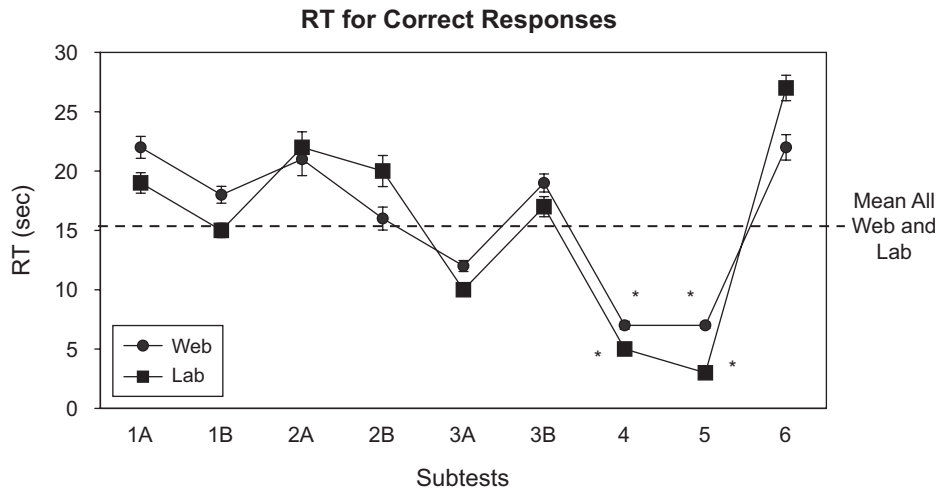


Figure 2. The comparison of mean correct response time and standard error per subtest for the lab and Web samples. In some instances, the standard error is very small and is hidden by symbols used for Web and Lab RT. *significant difference ($p < .05$).

the Web sample, indicating that the harder the items in a subtest, the longer participants took to produce a correct answer. This relationship was reflected to some extent in the lab-based sample; however, four of the correlations did not reach significance.

DISCUSSION

Results for accuracy and RT were compared across both the lab and Web studies, to demonstrate that similar patterns emerge. The comparison was not expected to show that one methodology was superior to the other, since the format of the two tests was very similar except for the input method (response pad vs. mouse-clicks).

There were no problems encountered in running the laboratory study, likely because of the controlled environment and the opportunity for the experimenter to address participants' concerns. Participants performed at a high level, considering that they did not have prior learning experiences in technical drawing, likely due to the high academic achievement required to enter the psychology program at the University of Newcastle. At this institution, students are admitted based on their University Admissions Index (UAI), and psychology students who participated in this study had a UAI of 89.1 or better. To allow some comparison, approximately 15% of school leavers who graduate after a senior high school education have a UAI of 89.1 or better.

Several issues were encountered in implementing the Web study. The need to match the Web and laboratory stud-

ies sometimes limited our ability to fully exploit the benefits of the Web interface. ColdFusion has the capability of allowing participants to click on the actual image that best represented their answer, which could have reduced the effort required to click on a number (Web version) or press a response key (laboratory version) associated with the image. Future versions of the test will take advantage of this capability. Other implementation issues included how to accommodate a variety of end-user connections and make allowances for timing differences due to network bandwidth. Since the Web study was not carried out in a laboratory setting, the participant's computer and Internet connection quality were outside the control of the study. This needed to be addressed by taking a lowest common denominator approach. Although it was possible that users on fast PCs with fast Internet connections could load the test data for each screen almost immediately, it was necessary to allow a few seconds for data to load onto slower machines before displaying any data to participants. This had no impact on the measurement of RT but may have resulted in some frustration for many participants.

Calibration issues raised by Krantz (2001) may account for some differences in results between the two studies, and the question of equivalence between sample characteristics (Buchanan, 2001) may offer further explanation. Despite the differences, results are similar, providing confidence in reliability and validity between the studies.

The main technical difficulty with Web delivery was caused by some participants discovering they could use the back button in the Web browser, which resulted in data compilation problems. Participants were not explicitly told they could not use the back button, and many may have considered this to be a reasonable practice to engage in. Of the 50 participants eliminated from the final analysis after entering the test phase, 18 were excluded because of the back button problem. Reips and Stieger (2004) point to log file analysis tools such as Scientific LogAnalyzer (psych-wextor.unizh.ch/loganalyzer/)

Table 3
Correlations Between Percentage Correct Mean Scores and Mean Response Times for Correct Answers for Web- and Lab-Based Samples

Sample	1A	1B	2A	2B	3A	3B	4	5	6	Total
Web	.46**	.49**	.79**	.69**	.52**	.55**	.36*	.41*	.94**	.74**
Lab	.24	.14	.54**	.56**	.26	.42**	.25	.38*	.83**	.37*

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

Analyzer5//) to mine data and analyze log files produced by Web servers. In many cases, these tools help detect problems associated with data collection over the Web. Although they help identify unusable data, they cannot eliminate the problems, only detect them. Hence, future Web implementations of 3DAT will disable the back button by adding JavaScript to the ColdFusion code.

The remaining 32 participants eliminated from the final analysis were excluded because their test data were incomplete. For the 89 possible responses, excluded participants provided between 2 and 82 responses. Some participants may have failed to complete because they were not informed about how many questions remained to be answered during testing. This information will be included in future versions of the test, to minimize the drop-out rate. Note that none of the remaining data sets were eliminated on the basis of patterns in responses and RTs. Screening on this basis may be necessary for Web-based data, particularly if a pattern of rapid responding indicates "clicking through," but there was no evidence of this pattern in our data.

The interface and response requirements of 3DAT are more complex than for most other Web-based tests that we are aware of. Our Web study was also more time consuming than most Web-based tests and required dedication from participants to work through detailed instructions and complete practice trials before testing began. In contrast to the laboratory study, participants worked in isolation, with no supervisory support. Collectively, these factors may have discouraged many potential participants. Of the 80 participants who made a start to the testing phase, some did not finish and others were excluded from the sample because of the data compilation issues caused by the browser back button problem.

The Web study was developed to parallel the laboratory study, so that reliability and validity could be compared. Our results support previous findings that Web-based studies can produce reliable and valid data (Birnbbaum, 2004; Steyvers & Malmberg, 2003) and extend these findings to a more complex and demanding design than has been used in many previous Web-based studies. The Web study may also have more validity in terms of generalization to the population, because it was less student based. However, drop-out rates are a concern. Reips (2002b) draws attention to the importance of examining drop-out data and using this to improve online studies. Drop-out numbers may indicate that a simpler and shorter test, and perhaps greater rewards for participation, will be required to obtain a larger sample. As well, improvements in design, such as simplifying instructions and changing the interface layout to take better advantage of the tools provided by ColdFusion, are also likely to be helpful, together with improved and more widespread promotion strategies.

Future Directions

The data collected in the studies reported here, and in the proposed extensions, will underpin the development of a set of learning tasks designed to improve 3-D understanding. Evidence suggests that spatial ability can be improved using purpose-designed learning tasks that

allow active exploration and participant interaction. For example, Harman, Humphrey, and Goodale, (1999; as cited in James et al., 2001) reported that "observers who actively rotated 3-D novel objects on a computer screen later showed faster visual recognition of these objects than did observers who had passively viewed exactly the same sequence of images of these virtual objects" (p. 111). Similarly, James et al. (2001) state from their experimental work that "active exploration of novel objects leads to better performance on later tests of object recognition" (p. 118). These results argue in favor of active exploration over passive observation to improve spatial ability. Although there is a tradition of research and educational thinking that underlines the value of *learning by doing*, James et al. (2001) point out that "this idea has not often been applied to perceptual learning in vision" (p. 118). 3DAT provides a valid and reliable way of measuring the efficacy of such developments in both laboratory- and Web-based settings.

AUTHOR NOTE

Correspondence concerning this article should be addressed to K. Sutton, School of Psychology, Faculty of Science and Information Technology, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia (e-mail: ken.sutton@newcastle.edu.au.).

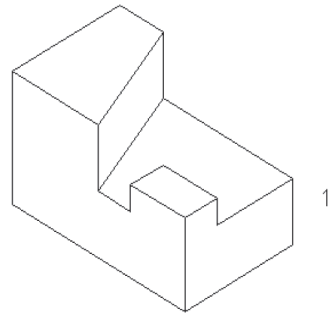
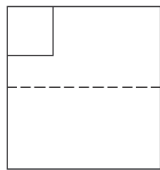
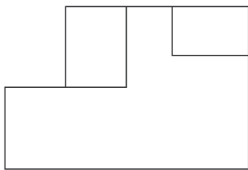
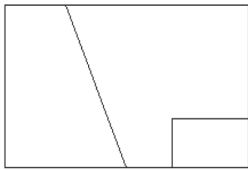
REFERENCES

- BERTOLINE, G. R., & MILLER, D. C. (1990). A visualization and orthographic drawing test using the Macintosh computer. *Engineering Design & Graphics Journal*, *54*, 1-7.
- BIRNBAUM, M. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*. (First published online as a Review in Advance, October 6, 2003.)
- BLASKO, D. G., HOLLIDAY-DARR, K., MACE, D., & BLASKO-DRABIK, H. (2004). VIZ: The visualization assessment and training Web site. *Behavior Research Methods, Instruments, & Computers*, *36*, 256-260.
- BORE, M., & MUNRO, D. (2002). Mental agility test. *Personal Qualities Assessment*. Newcastle, Australia: TUNRA.
- BUCHANAN, T. (2001). Online personality assessment: Equivalence of traditional and WWW personality measures. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 57-74). Lengerich, Germany: Pabst.
- COOPER, L. A. (1990). Mental representation of three-dimensional objects in visual problem solving and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 1097-1106.
- DUESBURY, R. T., & O'NEIL, H. F. (1996). Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections. *Journal of Applied Psychology*, *81*, 249-260.
- FITTS, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243-285). New York: Academic Press.
- GIBSON, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.
- HARMAN, K. L., HUMPHREY, G. K., & GOODALE, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*, 1315-1318.
- HEATHCOTE, A., BROWN, S., & MEWHORT, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185-207.
- HELD, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, *213*, 84-94.
- JAMES, K. H., HUMPHREY, G. K., & GOODALE, M. A. (2001). Manipulating and recognising virtual objects: Where the action is. *Canadian Journal of Experimental Psychology*, *55*, 111-120.
- JAMES, K. H., HUMPHREY, G. K., VILES, T., CORRIE, B., BADDOUR, R., & GOODALE, M. A. (2002). "Active" and "passive" learning of three-

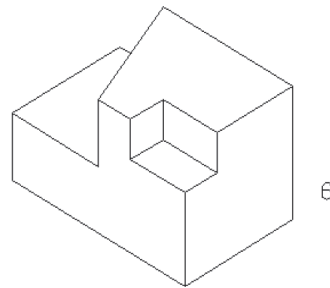
- dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, **34**, 383-390.
- KRANTZ, J. H. (2001). Stimulus delivery on the Web: What can be presented when calibration isn't possible. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 113-130). Lengerich, Germany: Pabst.
- METZLER, D., & SHEPARD, S. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 3-11.
- NEISSER, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman.
- PIAGET, J. (1953). *The origins of intelligence in the child*. London: Routledge & Kegan Paul.
- REIPS, U.-D. (2002a). Standards for Internet-based experimenting. *Experimental Psychology*, **49**, 243-256.
- REIPS, U.-D. (2002b). Internet-based psychological experimentation. *Social Science Computer Review*, **20**, 241-249.
- REIPS, U.-D., & STIEGER, S. (2004). Scientific LogAnalyzer: A Web-based tool for analyses of server log files in psychological research. *Behavior Research Methods, Instruments, & Computers*, **36**, 304-311.
- SCHACTER, D. L., & COOPER, L. A. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General*, **119**, 5-24.
- SPEELMAN, C., & KIRSNER, K. (2005). *Beyond the learning curve*. New York: Oxford University Press.
- STEYVERS, M., & MALMBERG, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 760-766.
- SUTTON, K., HEATHCOTE, A., & BORE, M. (2005). Implementing a Web-based measurement of 3-D understanding. *Proceedings of the 19th conference of the Computer-Human Interaction Special Interest Group (CHISIG) of Australia on computer-human interaction: Citizens online; considerations for today and the future* (pp. 1-4). Narabundah, Australia CHISIG.

APPENDIX

Orthographic Standard A



1

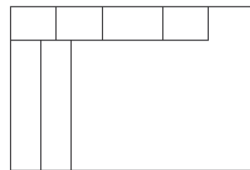
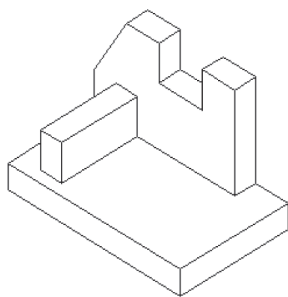


6

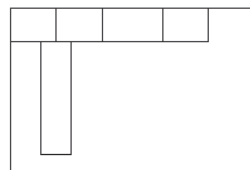
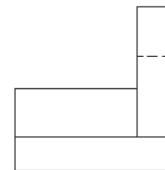
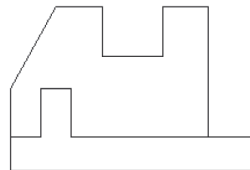
Enter the number of the 3-D object that is represented by these three views.

Isometric Standard B

Enter the number of the set of 2-D views that represent this 3-D object.



1



6

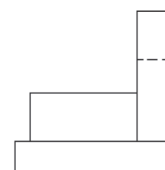
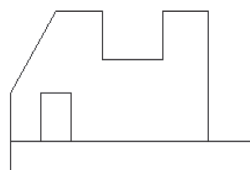
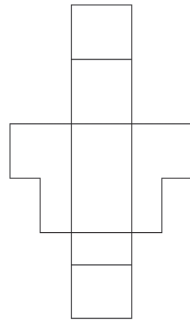
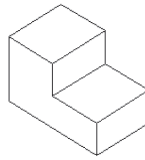


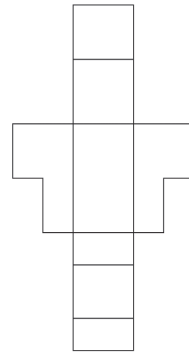
Figure A1. Examples of 2-D–3-D Recognition tasks showing Orthographic Standard A (top) and Isometric Standard B (bottom). Participants match and select from two options.

Isometric Standard A

Enter the number of the open view which you think will fold into the 3-D object shown



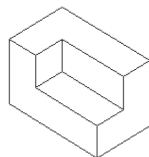
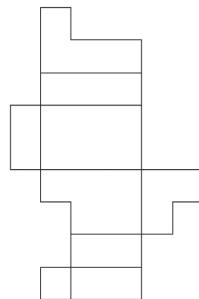
1



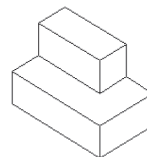
6

Unfolded Standard B

Enter the number of the 3-D view which you think matches the unfolded view shown



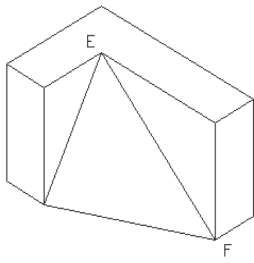
1



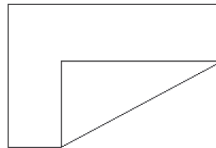
6

Figure A2. Examples of Correct Fold tasks showing Isometric Standard A (top) and Unfolded Standard B (bottom). Participants match and select from two options.

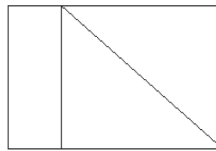
True Length Recognition A



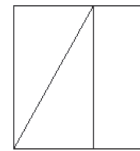
Enter the number of the 2-D view which you think shows the True Length of the edge EF.



2

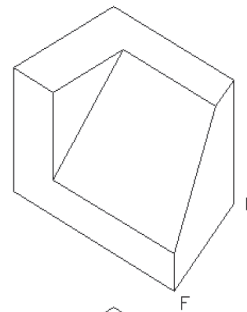
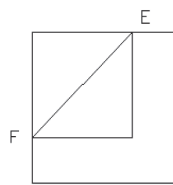
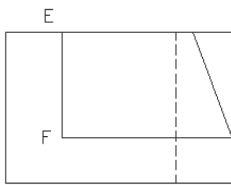
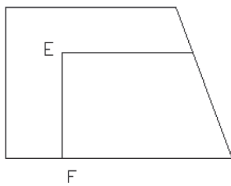


3

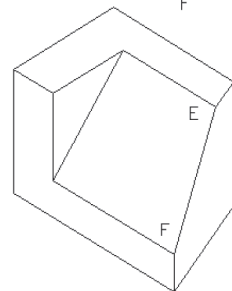


4

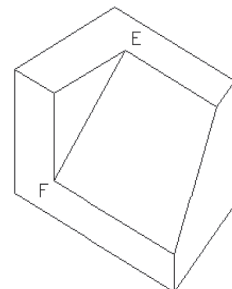
True Length Recognition B



2



3



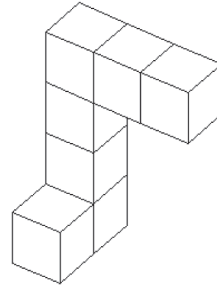
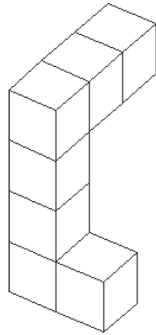
4

After studying the 2-D views above, enter the number of the 3-D view that shows the edge EF correctly labeled.

Figure A3. Examples of True Length Recognition tasks showing True Length Recognition A (top) and True Length Recognition B (bottom). Participants match and select from three options.

Standard

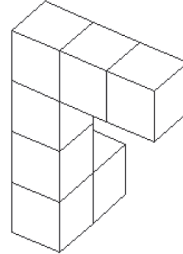
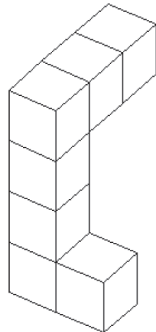
Consider the two objects and decide if you think they are the same or different.



Enter
1 = SAME
OR
6 = DIFFERENT

Mirror

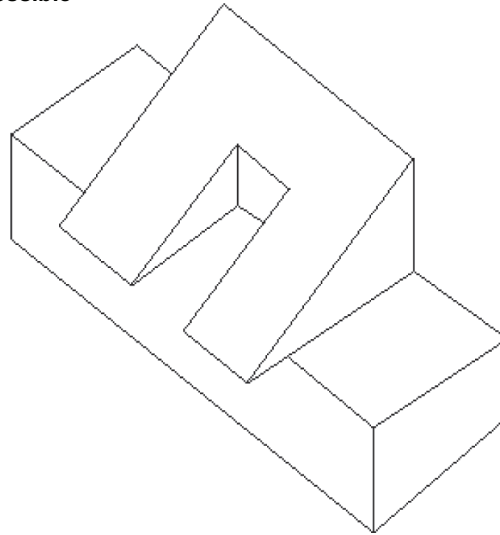
Consider the two objects and decide if you think they are the same or different.



Enter
1 = SAME
OR
6 = DIFFERENT

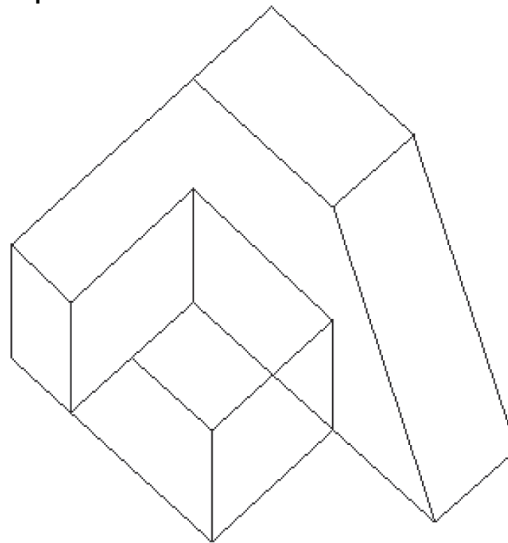
Figure A4. Examples of Mental Rotation Tasks showing Standard (top) and Mirror (bottom). Participants decide whether the two images are the same or whether one is the mirror image of the other.

Possible



ENTER
1 = Possible
OR
6 = Impossible

Impossible

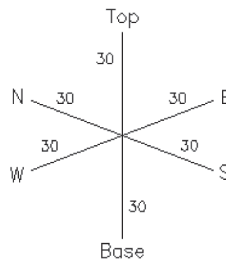


ENTER
1 = Possible
OR
6 = Impossible

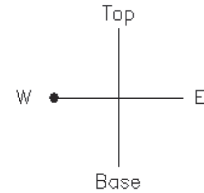
Figure A5. Examples of Possible/Impossible tasks showing Possible (top) and Impossible (bottom). Participants decide whether the projections can represent a true 3-D object.

Example of a Dot Coordinate Task

REFERENCE AXES
 Starting from where the axes intersect, a dot cannot be located more than 30 units North (N), South (S), East (E) or West (W) and not more than 60 units in Height (H) above the Base point.



EXAMPLE
 If you were looking from the South and a dot was located at:
 W = 30
 S = 0
 H = 30
 then it would look like this.



If you were looking from the North, which of the four diagrams below correctly represents the dot coordinates:

S = 15
 W = 15
 H = 45.

Enter the number of the diagram you think is correct.

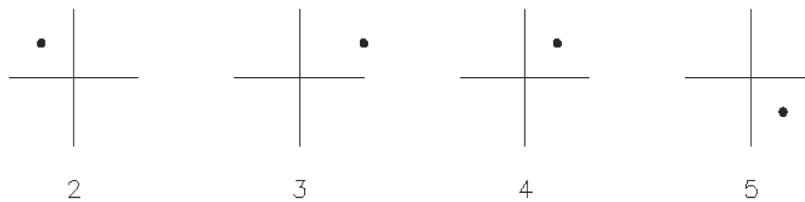


Figure A6. Participants decide which of four orthographic views corresponds to the description of a point in a Cartesian coordinate system.

(Manuscript received February 7, 2006;
 revision accepted for publication October 9, 2006.)