

A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities

JAMES W. PELLEGRINO

University of California, Santa Barbara, California

EARL B. HUNT

University of Washington, Seattle, Washington

RONALD ABATE

University of California, Santa Barbara, California

and

SIMON FARR

University of Washington, Seattle, Washington

A battery of 10 computerized tests of spatial ability is described. It includes 5 tests that require reasoning about static spatial displays and 5 tests that require reasoning about dynamically displayed spatial information. An integrated software package for task presentation and data analysis is described, along with a summary of results from a validation study comparing performance on the computer-based tests with performance on standardized paper-and-pencil tests of spatial abilities. Finally, research applications of the current battery are discussed.

Over the last decade, we and others have been involved in research that focuses on information-processing component analyses of individual differences in cognitive abilities (see, e.g., Sternberg, 1985). As a result, there are now sophisticated theories and models of performance on a wide range of cognitive tasks, including performance on typical tests of spatial ability (see, e.g., Lohman, Pellegrino, Alderton, & Regian, 1986; Pellegrino, Alderton, & Shute, 1984; Pellegrino & Kail, 1982).

Spatial ability is better thought of as a domain of abilities than as a single ability or skill. Multivariate studies of the domain have identified several factors (Lohman, 1979; McGee, 1979). One of the most important and clearly defined is *spatial relations* ability, which refers to the capacity to rapidly transform objects "in the mind's eye," as is required when one "mentally rotates" an object about its center (Shepard & Cooper, 1982). Conventional psychometric tests of spatial relations ability include the Primary Mental Abilities Space test (Thurstone & Thurstone, 1949). A second major factor is *spatial visuali-*

zation, which is best thought of as the ability to deal with complex visual problems that require imagining the relative movements of internal parts of a visual image. Psychometric tests that tap spatial visualization include the paper-folding (surface development) task in the Differential Aptitude Battery (DAT; Bennett, Seashore, & Wesman, 1974) and the Minnesota Paper Form Board Test (Likert & Quasha, 1971). Another well-defined visual-spatial factor is *perceptual speed*, which involves rapid encoding and comparison of visual forms.

Virtually all current spatial tests depend on the conventional paper-and-pencil test format. This restricts the form of the test items severely, since the visual scenes that the examinee must reason about cannot contain moving elements. Also, although it is possible to determine how many items an examinee can pass in a fixed time, it is not possible to examine the time that a person spends on an individual item, or the time spent on various identifiable subparts of the spatial problem posed by a given test item. This is an issue because speed and accuracy in solving parts of a problem may reflect different psychological skills (Pellegrino & Kail, 1982). More generally, different people may trade off between speed and accuracy of performance in different ways and measures of both speed and accuracy may be needed to adequately assess skill (Pachella, 1974).

It has been pointed out that both of these problems may be remedied by computer-administered testing (Hunt & Pellegrino, 1985). Visual displays with moving elements

This project was supported by Grant N660001-85-C-0017 from the Naval Personnel Research and Development Center to Earl B. Hunt and James W. Pellegrino. The software package and user's guide may be obtained for a service fee of \$30 from James W. Pellegrino, Graduate School of Education, University of California, Santa Barbara, CA 93106 or from Earl B. Hunt, Department of Psychology, N 1-25, University of Washington, Seattle, WA 98195. Please make checks payable to the University of California or the University of Washington. Requests for reprints should be addressed to James W. Pellegrino.

(dynamic displays) can be presented in computer-controlled testing. In addition, computer-controlled testing makes it possible to record accuracy and latency measures each time an examinee attacks a problem. The latter advantage applies both to tasks using dynamic displays and to those using displays without moving elements (static displays). Thus, one can embed componential-analysis procedures in computer-based testing procedures, thereby providing for more refined measures of processing ability (Hunt & Pellegrino, 1985; Lohman et al., 1986).

Elsewhere, we have argued that the advent of computer-controlled testing offers the possibility for substantial improvements in spatial ability testing (Hunt & Pellegrino, 1985). Whether or not that possibility should be realized depends on the answers to three questions. Does computerized testing involving static displays evaluate the same abilities as conventional paper-and-pencil tests? Can tests using dynamic displays be designed to test a dimension of ability that is different from the abilities evaluated using static displays? Finally, do the finer measures available through computerized testing make possible more precise measurement of ability, and thus better prediction of on-the-job performance? The latter is the question of most interest in applied psychology. An attempt to answer it directly, however, could be both fruitless and extremely expensive if the first two questions were not examined beforehand.

We have completed a project designed to address the aforementioned issues (Hunt et al., 1987). As part of the project, we developed a computer-administered test battery containing 10 spatial reasoning tasks. Five of the tasks involve static visual displays and five involve dynamic visual displays. These computer-based tasks were administered to a large sample of young adults ($N=170$) who also took a battery of conventional paper-and-pencil tests. The purpose of this paper is to describe the task battery and software package for task presentation and data analysis. In the next two sections we briefly describe the tasks included in the computer-administered battery of spatial tests, the logic behind the selection and implementation of each task, and the types of performance measures provided by each task. More complete details about individual task design can be found in the report by Hunt et al. (1987). In subsequent sections we briefly discuss characteristics of the available software package, results of the validation study, and research applications of the current test battery.

INFORMATION-PROCESSING TASKS INVOLVING STATIC DISPLAYS

The five static spatial reasoning tasks that we selected for computer implementation and administration had to meet a set of criteria. First, we wanted tasks that had well-defined information-processing characteristics, tasks that had been previously analyzed and validated with respect to underlying component processes. Second, whenever possible the task and its derived performance/processing

measures should have some history of use in the study of individual differences in spatial ability or imagery ability. Third, the full set of static tasks should include tasks measuring all of the major visual-spatial factors (e.g., perceptual speed, spatial relations, and spatial visualization).

Perceptual Comparison

This task was selected to provide measures of perceptual speed, that is, the ability to rapidly encode figures and make visual comparisons. The computer task is based on visual comparison research conducted by Cooper (1976). On each trial a pair of random shapes is presented. The examinee is to indicate whether or not the shapes are exactly identical. The shapes are generated by connecting 6, 8, 10, 12, or 14 randomly chosen points on a plane. Mismatching comparison figures are generated from a standard figure by slightly moving one or more of the original points. Two aspects of this task are of interest. First, on trials where there is a mismatch, the degree of mismatch is varied and decision time should be a monotonically decreasing function of this variable. The slope of the difference detection function provides an index of the speed (efficiency) of detecting feature differences. Second, stimulus complexity is manipulated by varying the number of points (from 6 to 14) used to generate the original shape. The slope of the stimulus complexity reaction time function provides an index of encoding and comparison efficiency. Measures derived from this task include average response time and accuracy (measures similar to those obtained from standard perceptual speed tests), as well as slope and intercept measures reflecting efficiency of difference detection and efficiency of encoding and comparison processes.

Mental Rotation

This task was selected to provide a measure of spatial relations ability. The computer-controlled mental rotation task closely resembles tasks previously used to assess individual differences in spatial relations ability (Mumaw, Pellegrino, Kail, & Carter, 1984). On each trial a pair of polygons is presented on the computer display. The task is to judge their identity and to respond by pressing one of two keys. The individual trials represent the combination of two variables: angular disparity of the stimuli (ranging from 0° to 180° in 20° increments) and match type (positive or negative). Nonmatching pairs are created by a mirror-image reversal of the stimulus. Measures of separate processes are estimated from the slopes and intercepts of the linear functions relating response time to angular disparity. General and specific latency and accuracy parameters are estimated separately for positive and negative match conditions, because other research has shown that performance in the positive and negative conditions often differs (see Pellegrino & Kail, 1982).

Surface Development

This task was selected as one of two spatial visualization tasks. Spatial visualization ability is often assessed by surface development tasks, which contain representa-

tions of flat, unfolded objects and completed three-dimensional shapes. The task that we used is based on previous work on individual differences in solving surface development problems (Alderton & Pellegrino, 1984; Pellegrino, 1984). On each trial, the individual is presented with two figures. The left-hand figure shows a cube "unfolded" along its edges so that it lies flat. The base of the cube is labeled, and two or three sides are marked by dots. The right-hand figure presents a two-dimensional projection of a cube "refolded" and slightly rotated so that the top, front, and right lateral surfaces are visible. The task is to determine whether the flat figure on the left could be refolded and rotated (if appropriate) to form the cube shown on the right of the screen. Different folding patterns are used, in which the number of mental folds that have to be made and the number of surfaces that have to be mentally "carried along" during such folds are systematically varied (Shepard & Feng, 1972). Performance measures include mean latency and accuracy, as well as slopes and intercepts of the linear functions relating response time to problem complexity as defined by the number of mental folding operations.

Integration of Detail in an Image

This task was selected as a second prototypical spatial visualization task. Spatial visualization ability is often assessed by form board tasks requiring the integration of elements to form a composite image: for example, a set of shapes must be concatenated in a certain way to form a completed shape or puzzle. Mumaw and Pellegrino (1984) and Poltrock and Brown (1984) developed variants of this type of task that permit assessment of various image-integration processes as well as of the capacity of the visual buffer. The task that we used is similar to that developed by Poltrock and Brown. On each trial, the individual is presented with an array of regular shapes with various edges marked by specific letters. The number of pieces in the array varies from three to six. The individual studies the pieces and tries to determine the composite image that would be created by appropriately aligning the pieces with corresponding edge markings. The time to perform the integration should vary with the number of shapes to be integrated. Following the integration phase, the individual is shown a complete shape and told to decide whether it represents the correct integrated image. The slope of the latency function for image integration provides an index of the efficiency of several imagery processes, and the accuracy score provides an index of buffer capacity.

Adding Detail to an Image

One of the components of Kosslyn's (1980) imagery theory is the addition or deletion of detail in a mental representation (image). To do this, several subprocesses are required. Examples are the "Put" operation, in which a component part is placed at a point in an image, and

the "Find" operation, in which an image is examined to determine whether or not it contains a feature. Poltrock and Brown (1984) developed a task that provides indices of the efficiency of these and other processes. Individuals are asked to image a base form and then add details (dots) to it at specified locations. We implemented a variant of this task in which the base form is a six-pointed star. Trials vary with respect to the total number of dots (four to seven) to be sequentially added to the base form prior to presentation of a single composite image containing the appropriate number of dots. The subject's task is to then decide whether the composite image is a correct or incorrect final product. From this task we derive several latency measures, including a slope and intercept of the latency function for adding successive details. We also analyze changes in accuracy as a function of problem complexity; this analysis provides estimates of a person's visual-memory buffer capacity.

INFORMATION-PROCESSING TASKS INVOLVING DYNAMIC DISPLAYS

In contrast to the availability of information on human performance on static spatial tasks, there is little in the experimental literature on individual differences in motion extrapolation. Therefore, the tasks we implemented to measure dynamic spatial reasoning had less history of development than the tasks described above. We attempted to design tasks that coincided with a rational analysis of what appears to be required in this domain of performance. The basic dynamic visual-spatial problem a person has to solve is to predict where a moving object is going and when it will arrive at its predicted destination. This skill can be divided into three components. First, the observer may need to remember the path the object has just traveled in order to extrapolate its future path. Second, given the path the object has traveled thus far, the observer must be able to predict or extrapolate the future path. Third, to predict when an object will arrive at a destination, the observer must extrapolate its speed. We developed three tasks to assess these three components of dynamic spatial reasoning.

It appears that many dynamic spatial problems require—in addition to judgments about the path and speed of a single moving object—judgments of relative, rather than absolute, speed. For instance, if two objects are moving toward destinations, the observer may wish to know whether they will arrive at the same time or, if not, which will arrive first. We developed a task that assessed the ability to make such comparative speed judgments.

Although it is reasonable to break up judgments about absolute and relative visual motion into component tasks, it is possible that judgments about motion are made in a holistic fashion. Therefore, at least one dynamic spatial reasoning task should require coordination of time, direction, and motor movement judgments. The act of mak-

ing a coordinated judgment might itself be a significant source of individual differences, and we designed a task that required such performance.

Path Memory

This task was designed to assess a person's memory for the path of moving objects. On each trial, a small square moves across the computer screen three times. The square follows a parabolic path, starting at the lower left of the computer screen. Either the first and second paths are the same and the third path is different, or the second and third paths are the same and the first path is different. The observer indicates which path differs from the second. The computer then reports whether or not the response was correct.

Three parameters determine the paths: the starting height of the parabola, the height of the apex of the parabola, and the horizontal distance from the start of the parabola to the apex. Within each of these dimensions, eight levels of difficulty are established. The easier the level of difficulty, the larger the difference between the unique path and the two identical paths. An adaptive staircasing method is used to measure performance. The level of difficulty is increased when the subject answers two consecutive trials correctly, and decreased when an error is made. A separate staircase is computed for each of the three different methods of changing the parabola. The dependent measure is the average difficulty level of the last two thirds of the trials in each staircase series.

Extrapolation

This task was intended as a measure of the ability to extrapolate from an observed to an expected path. Three types of curves are presented: a straight line, a sine wave, and a parabola. A portion of the curve is shown on each trial, starting from the left side of the computer screen and ending 41%, 52%, or 63% of the distance across the computer screen. The observer uses a joystick to move an arrow up or down along a vertical line on the right side of the computer screen to indicate where the curve will intersect the line. The computer displays the remaining portion of the curve as soon as the response is made. The dependent measure is the difference between the correct answer and the subject's answer. Separate scores are computed for each of the three types of curves.

Arrival Time for One Object

In this task, the observer has to make an absolute judgment of velocity. On each trial, a square moves horizontally from the left side of the computer screen toward a vertical line on the right. One quarter to one half of the way across the computer screen, the object disappears. The observer presses a key when he or she thinks the object would have crossed the line, had it continued moving on the same course at the same speed. The dependent measure is the difference in time between the correct answer and the subject's answer. Both absolute differences (accuracy) and signed differences (bias) are computed.

Arrival Time Comparison for Two Objects

This task requires judgments of relative speed of motion. On each trial, two different objects are presented, each moving toward its destination at a constant speed. The destinations are displayed as horizontal or vertical lines. A fifth of the way to the destination, the objects disappear. The subject then reports which object would have arrived at its destination first. The computer informs the subject whether or not the response is correct.

There are five variations on this task. In the first variation, the objects move in perpendicular paths toward different destinations. In the second variation, the objects move in perpendicular paths toward the same destination. In the remaining three variations, the objects move in parallel horizontal paths. In the third variation, the two paths are near each other and the destinations are vertically aligned. In the fourth variation, the two paths are near each other, and the starting locations are vertically aligned. The fifth variation is like the third except that the paths are not near each other; one is at the top of the computer screen and the other is at the bottom.

The staircasing method of measuring performance in the path memory task is also used here. Eight levels of difficulty are established by varying the size of the difference in arrival time for the two objects. The staircasing is run separately on the five different variations of the task. The dependent measure is the average level of difficulty for the last two thirds of the trials.

Intercept

This task was designed to measure the ability to combine the extrapolation of both speed and path. The task is like a video game in which a player attempts to shoot down a moving object. A small rectangular target moves from left to right at a constant horizontal speed. The target moves along a horizontal, sine-wave, or parabolic trajectory. When the subject presses a key on the keyboard, a triangularly shaped object (called a "missile") begins to move straight upward at a constant velocity. The subject attempts to time the missile's "launch" so that the missile will collide with the target. The dependent measure is the vertical distance between the missile and the target when the target crosses the missile's path.

SOFTWARE PACKAGE

A fully integrated software package representing all programs for task presentation, data analysis, and printing of data tables is contained on five disks. The primary disk, STARTUP, contains the master menu for selection of a task for presentation, data analysis, or printing of results. Menus are used throughout to simplify use of the software; instructions about disk changing are presented where appropriate. The 10 task-presentation programs are contained on two disks and any of the 10 tasks may be individually presented in any order desired. Each task program includes a set of instruction screens explaining the task and providing practice problems before the start of

the actual problem set. Table 1 provides information about the number of problems in each task and the approximate time for completion of the full problem set.

Two of the disks contain programs for analyzing the trial-by-trial data recorded for each task. The analysis programs provide detailed descriptive statistics, such as the mean, median, standard deviation, and minimum and maximum value, organized by problem type. Also provided are derived statistics based on fitting models of task performance, for example, slopes and intercepts of the function relating reaction time to angular disparity in the mental rotation task. Data are also provided for odd and even trial halves for various performance measures, thereby permitting derivation of reliability estimates. A single disk contains programs for the annotated printing of the data files produced by the data analysis programs.

All programs are written in UCSD Pascal and run on an Apple II+ or IIe computer. Hardware requirements include 64K of memory, two disk drives, a Mountain Hardware Clock Card, and a monochrome monitor. One of the programs also requires use of a joystick.

SUMMARY OF TEST VALIDATION RESULTS

As noted earlier, the computer-based tasks were administered to 170 young adults along with a set of standardized paper-and-pencil tests. In the present context we can provide only a general summary of the validation results. For more detailed results and discussion, see Hunt et al. (1987). One important issue is whether the computer-based tasks of static and dynamic spatial reasoning behave in a principled manner and produce measures that have substantial reliability. The group and individual-subject data revealed that performance in all tasks was consistent with general expectations about within-task processing. The overall latency and accuracy measures for the static reasoning tasks had substantial reliability, with the majority of reliability coefficients above .90. Within each static task we also derived various component-process latency measures. Examples include the slope and intercept of the mental rotation function and the slope of the function relating integration latency to the number of elements to be integrated. These component-process latency measures also had substantial

reliability: the average reliability coefficients of such within-task measures ranged from .67 to .91. For the dynamic tasks, the reliability coefficients of the performance measures ranged from .63 to .99, with the exception of the path memory task, which had a .50 reliability coefficient.

The results from a variety of multivariate analyses, including exploratory and confirmatory factor analyses and canonical correlation analyses, supported two main conclusions. The first is that our theory-based, computer-administered static spatial reasoning tasks can be used to replace and augment current paper-and-pencil procedures for the assessment of spatial ability. The computer-based static tasks and measures encompass the variance contained in traditional paper-and-pencil tests and also provide measures of unique variance associated with speed versus accuracy of processing. The second major conclusion is that the ability to deal with moving elements and dynamic spatial relations is separate from the abilities associated with reasoning about static spatial information.

RESEARCH APPLICATIONS

The tasks contained in the present battery can be used in several research applications. One primary area of application is the further study of developmental and individual differences in spatial processing. Of the tasks contained in this battery, only mental rotation has any history of use in developmental information-processing research with children (e.g., Kail, 1986; Kail, Pellegrino, & Carter, 1980). In addition, most developmental and aging research on spatial abilities has been restricted to use of gross accuracy measures from paper-and-pencil tests. The present tests provide more refined latency, accuracy, and process measures, thereby permitting more detailed analyses of the nature of developmental change in spatial processing abilities. The same possibilities exist for research focusing on sex differences in spatial abilities, as well as research on the heritability of specific spatial abilities.

A second primary area of application is research on predictive validity. One issue is whether the predictive validity of static spatial tests is significantly increased by the use of more refined measures that separate latency and accuracy of processing. A second issue is whether the ability to represent and reason about dynamically changing spatial relations better predicts real-world performances that intuitively seem to require this ability. Examples include machinery operation and piloting. There is reason to believe that this may be the case, since this ability is separate from those assessed by conventional spatial ability tests. To date, there have been no studies examining the utility of these computer-controlled tests of static or dynamic spatial reasoning as predictors of performance in situations outside the laboratory. As noted at the beginning of this paper, this is an area of considerable interest and remains as a topic for future research.

Table 1
Task Duration and Item Set Characteristics

| Task | Number of Items | Task Duration (min) |
|--------------------------|-----------------|---------------------|
| Perceptual Comparison | 300 | 25 |
| Mental Rotation | 280 | 30 |
| Surface Development | 192 | 50 |
| Integrating Detail | 48 | 30 |
| Adding Detail | 60 | 30 |
| Path Memory | 72 | 20 |
| Extrapolation | 108 | 20 |
| Arrival Time (1 object) | 80 | 10 |
| Arrival Time (2 objects) | 250 | 25 |
| Intercept | 72 | 15 |

REFERENCES

- ALDERTON, D., & PELLEGRINO, J. W. (1984). *Analysis of mental paper folding*. Unpublished manuscript, University of California, Santa Barbara, CA.
- BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. (1974). *Differential Aptitude Test*. New York: The Psychological Corporation.
- COOPER, L. A. (1976). Individual differences in visual comparison processes. *Perception & Psychophysics*, *19*, 433-444.
- HUNT, E., & PELLEGRINO, J. (1985). Using interactive computing to expand intelligence testing. *Intelligence*, *9*, 207-236.
- HUNT, E., PELLEGRINO, J. W., ABATE, R., ALDERTON, D., FARR, S., FRICK, R., & McDONALD, T. (1987). *Computer controlled testing of spatial-visual ability* (Tech. Rep.). San Diego, CA: Naval Personnel Research and Development Center.
- KAIL, R. (1986). Sources of age differences in speed of processing. *Child Development*, *57*, 969-987.
- KAIL, R., PELLEGRINO, J. W., & CARTER, P. (1980). Developmental changes in mental rotation. *Journal of Experimental Child Psychology*, *29*, 102-116.
- KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- LIKERT, R., & QUASHA, W. H. (1971). *Revised Minnesota paper form board test (Series AA)*. New York: The Psychological Corporation.
- LOHMAN, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Tech. Rep. No. 8). Palo Alto, CA: Stanford University, Aptitude Research Project.
- LOHMAN, D., PELLEGRINO, J. W., ALDERTON, D., & REGIAN, J. W. (1986). Dimensions and components of individual differences in spatial abilities. In S. Irvine & S. Newstead (Eds.), *Intelligence and cognition: Contemporary frames of reference* (pp. 253-312). Dordrecht, The Netherlands: Martinus Nijhoff.
- MCGEE, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*, 889-918.
- MUMAW, R. J., & PELLEGRINO, J. W. (1984). Individual differences in complex spatial processing. *Journal of Educational Psychology*, *76*, 920-939.
- MUMAW, R. J., PELLEGRINO, J. W., KAIL, R., & CARTER, P. (1984). Different slopes for different folks: Process analyses of spatial aptitude. *Memory & Cognition*, *12*, 515-521.
- PACHELLA, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Erlbaum.
- PELLEGRINO, J. W. (1984, April). *Information processing and intellectual ability*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- PELLEGRINO, J. W., ALDERTON, D., & SHUTE, V. J. (1984). Understanding spatial ability. *Educational Psychologist*, *19*, 239-253.
- PELLEGRINO, J. W., & KAIL, R. (1982). Process analyses of spatial aptitude. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 311-366). Hillsdale, NJ: Erlbaum.
- POLTROCK, S. E., & BROWN, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence*, *8*, 93-138.
- SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- SHEPARD, R. N., & FENG, C. (1972). A chronometric study of mental paper folding. *Cognitive Psychology*, *3*, 228-243.
- STERNBERG, R. J. (ED.) (1985). *Human abilities: An information processing approach*. New York: W. H. Freeman.
- THURSTONE, L. L., & THURSTONE, T. G. (1949). *SRA primary abilities*. Chicago: Science Research Associates.