

A Generative Analysis of a Three-Dimensional Spatial Task

Issac I. Bejar
Educational Testing Service

The feasibility of incorporating research results from cognitive science into the modeling of performance on psychometric tests and the construction of test items is considered, particularly the feasibility of modeling performance on a three-dimensional rotation task within the context of item response theory (IRT). Three-dimensional items were selected because of the rich literature on the mental models that are used in their solution. An 80-item, three-dimensional rotation test was constructed. An inexpensive computer system was also developed to administer the test and record performance, including response-time data. Data were collected on high school juniors and seniors. As expected, angular disparity was a potent determinant of item difficulty. The applicability of IRT to these data was investigated by dichotomizing response time at increasing elapsed times, and applying standard item parameter estimation procedures. It is concluded that this approach to psychometric modeling, which explicitly incorporates information on the mental models examinees use in solving an item, is workable and important for future developments in psychometrics. *Index terms: cognitive psychology, continuous response, item response theory, mental rotation, response latency.*

The majority of psychometric theorizing and testing practice has been unconcerned with very detailed descriptions of the mental processes that underlie performance on test items. Instead, the focus has been on broader constructs, such as aptitude and abilities. As a result, the test item—the building block for constructing a test—does not play as large a role as it perhaps should in either test construction or psychometric modeling, although that may be changing (e.g., Sum-

mer 1989 issue of *Journal of Educational Measurement*).

The alternative perspective is that test items, far from being easily replaceable entities, are important in their own right. Accounting for differences among items, with respect to their psychometric characteristics, will reveal information about what a test measures (e.g., Bejar, 1985; Carroll, 1976; Egan, 1979; Embretson, 1983), just as accounting for differences in total score variation provides such information.

Whereas the preferred methodology for modeling test score variation has been factor analysis, the methodologies of cognitive science now appear to be suited to undertake the validation of tests from the above perspective. In fact, Pellegrino (1988) claimed that “cognitive scientists have been engaged wittingly and unwittingly in a process of construct validation. . .” (p. 52). The focus of this research was on mental rotation, and on the broad objective of exploring the feasibility of integrating cognitive research, psychometric modeling, and item construction into a generative framework (Bejar, in press). Specifically, the present research attempted to capitalize on the significant amount of research produced in the last twenty years in the area of spatial cognition. Much of this research has focused on the type of representation used by examinees to solve spatial problems. Myers (1958) conducted research on spatial ability for the College Board in the 1950s, and said that “we use the term ‘spatial ability’ to represent a complex family of abilities with unknown interrelationships. We do not yet know of a terminology that permits a more precise and efficient language” (p. 24).

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 14, No. 3, September 1990, pp. 237-245
© Copyright 1990 Applied Psychological Measurement Inc.
0146-6216/90/030237-09\$1.70

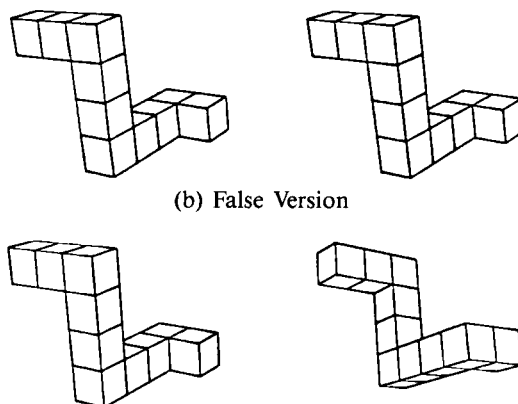
By contrast, Cooper and Shepard (1984) recently concluded, after reviewing their work on mental rotation, that

In spite of some unresolved issues, the close match we have found between mental rotation and their counterparts in the physical world leads inevitably to speculations about the functions and origin of human spatial imagination. It may not be premature to propose that spatial imagination has evolved as a reflection of the physics and geometry of the external world. The rules that govern structures and motions in the physical world may, over evolutionary history, have been incorporated into human perceptual machinery, giving rise to demonstrable correspondences between mental imagery and its physical analogues. (p. 114)

In the intervening period significant research and theorizing have occurred from the factor-analytic and the experimental perspective. Much of that work has been reviewed (Corballis, 1982; Lohman, 1979), and suggests the presence of three mental factors: spatial relations, spatial visualization, and spatial orientation. These factors have been investigated by cognitive psychologists (e.g., Pellegrino & Kail, 1982), but a specific task under the spatial relations factor—the three-dimensional mental rotation task—has received a majority of the attention. A typical stimulus used by Shepard & Metzler (1971) for this kind of research appears in Figure 1. The task for the examinee with this type of item is to rotate the right-hand figure to the same orientation as the left-hand figure; elapsed time and correctness are recorded.

The most significant finding from this line of research has been the seemingly universal finding that one feature of these stimuli—angular disparity between the two figures—controls the response time (e.g., Cooper, 1980; Shepard & Metzler, 1971). By contrast, it is usually difficult to obtain a priori predictions of the psychometric difficulty of an item or its response time (e.g., Bejar, 1983). Moreover, estimating item difficulty is not the same as explaining item difficulty.

Figure 1
 Sample Three-Dimensional Item
 (a) True Version



The absence of a valid performance model for solving the item may justify a focus on estimating rather than explaining difficulties. To the extent that the performance model is concerned with the effort required to solve an item, it may be feasible to accurately predict an item's psychometric characteristics, especially its difficulty. If this can be done with enough precision, it is in principle an alternative procedure—or at minimum an additional source of information—for estimating the difficulty of items. For example, it may be feasible to obtain valid estimates of difficulty by combining information about the psychological demands of the items using a small sample of examinees, instead of administering the test to a large sample of potential examinees. The implementation of this approach would require procedures for estimating the parameters in a psychometric model that are capable of incorporating "prior" information into the estimation process.

The foundation is being laid for incorporating prior information into the estimation of psychometric parameters (e.g., Bock & Atkin, 1981; Mislevy, 1988; Swaminathan & Gifford, 1981; Tsutakawa & Lin, 1984). Moreover, commercially available computer programs exist that are capable of handling some forms of prior information (Assessment Systems Corporation, 1984;

Mislevy & Bock, 1982). Because the production of collateral information on an item would be based on an understanding of how examinees solve the item, the possibility also exists that at some point it might be possible to build (on the basis of that knowledge) systems whereby an item writer could receive feedback on the likely psychometric characteristics of a prospective item before it is ever administered to an examinee (Bejar, in press). The capability to accurately predict an item parameter may become urgently needed if proposals succeed at prohibiting test publishers from pretesting as part of a standard test administration.

Psychometric Modeling of Spatial Rotation Data

The determination of psychometric difficulty in terms of item attributes can be used, therefore, as a criterion of success for the psychometric modeling of spatial ability. In the present research, psychometric difficulty was linked to angular disparity. The mildest criterion was that difficulty should increase as angular disparity increases. A stronger criterion was that difficulty should increase in a linear fashion with angular disparity, or some transformation thereof. If such a relationship could be established, the data could be interpreted more descriptively. In the absence of guessing, as well as in the absence of such a linkage, difficulty in an item response theory (IRT) model was defined as the point on the ability scale at which there was a 50-50 chance of responding correctly. When difficulty is linked to an item attribute, such as angular disparity, performance can be referenced to that attribute. Thus ability could be referred to as the level required to achieve a 50-50 chance of success on a task involving a certain degree of angular disparity. In short, relating psychometric parameters to a mental model of the item solution process is likely to improve the interpretation of psychometric results.

Because the three-dimensional mental rotation item does not require problem solving, the time required to obtain a correct response is directly interpretable as the efficiency with which the

mental rotation takes place. Therefore, both accuracy and response time could be taken into consideration when fitting a psychometric model to these data. A consideration of both responses suggested an expansion of the criterion described above. In addition to undergoing an increase in difficulty as a function of angular disparity, the relationship between angular disparity and difficulty should remain the same as the elapsed time to perform the task is increased. For example, the regressions of difficulty on angular disparity at 2, 3, and 4 seconds are linear, parallel to each other, and equidistant.

Method

Modeling Response Latency

A strategy for incorporating response latency into psychometric modeling has been proposed by Bloxom (1985); therefore, the relevant literature is not reviewed. Instead, the focus is on a discussion of modeling response latency as an extension of models for dichotomous data. A dichotomous item response model was fitted to response times from a set of 80 three-dimensional rotation items. The objective was to determine whether a more refined psychometric model should be attempted, rather than to provide the definite calibration of these data.

A common model for the probability of dichotomous response when time is not a factor is the two-parameter logistic model

$$P(u_i = 1|\theta) = \frac{1}{1 + \exp[1.7a_i(\theta - b_i)]}, \quad (1)$$

where a_i is the discrimination, b_i is the difficulty parameter, θ is ability, and $D = 1.7$ is a constant. In a situation where interest is on the probability of a correct response after a certain period of time has elapsed, it would be expected (1) that the longer an item is considered, the higher the probability of a correct response, at least with certain item types; and (2) that the function which models that probability at any elapsed time would have the same discrimination value. This means that the discrimination parameter is constant across time, but the difficulty parameter de-

creases as a function of time and the probability of a correct response after increasing elapsed times is solely a function of time. Micko (1969) applied this idea by specifying the dichotomous item response model to be the Rasch model.

Although it is possible, in principle, to model response time with a dichotomous item response model, it is also possible to generalize a dichotomous model as Samejima (1983) did. The continuous response is converted to a 0-1 interval in her generalization. For response latency, this means that the response is expressed as a proportion of the total time allowed for responding to an item. If the time limit is 15 seconds, for example, a response latency of 5 seconds would be .33. Samejima referred to the response expressed in this manner as z , but there was nothing in the model concerning whether the response time is for a correct or an incorrect response. Such a distinction must be made for scoring purposes; but the focus in this research was on characterization of the items, not the respondents.

One possible approach to estimating θ would be to treat responses as correct if a correct response is produced after s seconds, and as not-correct-yet if an incorrect response is given. Then incorrect responses would be treated as incomplete responses, and these would indicate that after s seconds, the individual had not yet produced a correct response. Also, in statistical terminology, incorrect responses would be treated as censored observations, as they are in survival analysis (e.g., Miller, 1981).

The probability that an examinee of ability θ would take longer than z to respond is given by

$$\begin{aligned} P_{z_i}^*(\theta) &= \int_{-\infty}^{a_i(\theta - b_{z_i})} D \exp(-Dt) [1 + \exp(-Dt)]^{-2} dt \\ &= \{1 + \exp[-Da_i(\theta - b_{z_i})]\}^{-1} \\ &= \frac{1}{1 + \exp[-Da_i(\theta - b_{z_i})]} \end{aligned} \quad (2)$$

which is similar to Equation 1, except that now the difficulty parameter is a function of response time z . The difficulty function b_{z_i} is not constrained to any particular shape other than that

it be monotonic. The fit of a linear function was investigated, particularly b_{z_i} s of the form

$$b_{z_i} = \phi_i + \delta_i z \quad (3)$$

ϕ_i could be further decomposed into components associated with figure attributes, but was not here because the focus was on the adequacy of a linear difficulty function when angular disparity and time were taken into consideration. An interesting implication of this model is its possible compatibility with the "slope and intercept" methodology commonly used by cognitive researchers interested in individual differences (e.g., Bejar, 1986; Lansman, Donaldson, Hunt, & Yantis, 1982).

Participants and Procedure

To secure data for the study, 160 high school students were recruited from a local high school. The items used consisted of 80 pairs of three-dimensional Shepard-Metzler figures. The following eight figures were used: A1, A2, A3, C1, C2, C3, E1, and E2. For each figure, true and false pairs were constructed by rotating at angular disparities of 20, 60, 100, 140, and 180. The true pairs were constructed by rotating the same figure along the picture axis. The false pairs were constructed by rotating the mirror image instead. Altogether, there were 16 items at each angular disparity. The resulting 80 items were videotaped and placed on a videodisc using the 3M mastering process.

The items were presented in two different orders. In one, the examinees saw items at 100, 60, 180, 140, and 20 degrees. At each angular disparity there were 16 possible items, and one of those was chosen at random. With the second ordering, examinees were presented with items of 20, 140, 180, 60, and 100 degrees. Approximately one-half of the examinees took the items in each order.

The instrumentation for each data collection station consisted of the following components:

1. A 64K microcomputer with 1 disc drive (Radio Shack 26-3127),
2. A videodisc player (Pioneer PR 8210),

3. An Amdek Color I Monitor,
4. A joystick (Radio Shack, 26-3012),
5. A computer-to-videodisc interface (specially constructed for the project).

The microcomputer was programmed to control the videodisc, as well as to record the responses. Response time was recorded in "ticks" of 1/60th of a second. Examinees responded by means of the joystick; a "yes" response was signaled by moving the joystick forward, and a "no" response was indicated by moving the joystick backward.

Because of the potential for unfamiliarity of the equipment on the part of the examinees (at least as a psychological testing device), careful attention was given to the instructions. Instructions were tested with several students who were not part of the study to insure that they were fully understandable. Students were told that they were to respond by moving the joystick and that they were to respond as quickly as possible without sacrificing accuracy (for these instructions, see Bejar, 1986).

The examinee's first task was to respond to a simple task—that of indicating whether an arrow was pointing up or down. This was done in order to familiarize each examinee with the response device, as well as to time their reaction time to a task with almost no cognitive load. From their responses it was then possible to obtain an estimate of their motor speed; these data, however, were not analyzed as part of this study.

After the arrow task, examinees were given instructions on the rotation test. This included manipulation of an animation sequence containing a true and a false item, which allowed them to become familiar with true and false items at all possible angles. Also, seeing the rotation in real time may have encouraged the examinees to use a rotation strategy to solve the items. The examinees were allowed to do six practice items, which were followed by 80 real items. There was a 15-second time limit for each item, and after 15 seconds a "time-out" message was given if the examinee had not responded. However, examinees could pace themselves in the sense that they con-

trolled when the next item was administered. After each item they were told whether they had responded correctly or incorrectly.

Parameter Estimation

With the growing interest in the psychometric modeling of response time (Bloxom, 1985; Scheiblechner, 1985; Thissen, 1983), it is likely that estimation procedures tailored to response time will be forthcoming. In the meantime, it is possible to obtain estimates of item parameters through estimation procedures designed for the dichotomous case. That was the approach taken here. Parameters were estimated by successively dichotomizing response time after elapsed times of 3, 4, 5, 6, and 7 seconds, and fitting a one-parameter logistic model at each dichotomization point. Imposing a one-parameter model across angular disparities implemented the constraint that discrimination be constant across time.

The item response data matrix for each item consisted of five sets of columns corresponding to elapsed times of 3, 4, 5, 6, and 7 seconds. Each set, in turn, consisted of ten columns corresponding to angular disparities of 20, 60, 100, 140, and 180 degrees in true and false versions. Each examinee was represented as a set of five rows corresponding to the five elapsed times. This allowed the response status of an examinee to be coded at each interval. For example, if an examinee responded after 3.5 seconds, a 0 or 1 would be entered in the first row of the first column set, and a 2 for not presented would be entered everywhere else.

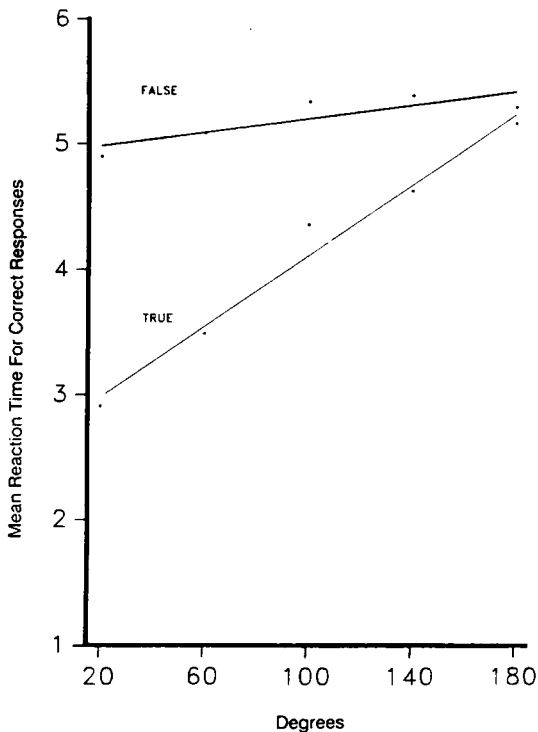
Each of the eight data matrices was analyzed separately with BILOG (Mislevy & Bock, 1982), specifying a one-parameter logistic model. The resulting estimates were rescaled with respect to the distribution of θ estimates derived from the EAP algorithm.

Results

Unlike the typical mental rotation experiment in which examinees receive a great deal of practice time, the examinees in this study spent altogether no more than forty minutes, including

instruction and practice items, on the mental rotation task. Therefore, it is important to verify that the usual finding concerning the linearity of response time on angular disparity was replicated in this case. Figure 2 shows the relationship between angular disparity and response time for correct responses for true and false items averaged across the eight basic figures.¹ It suggests that there was, for the most part, a good linear fit to the data. The largest residual was at 100 degrees. Apart from this, there was relatively little scatter around the best-fitted line. The figure

Figure 2
 Relationship Between Response Time and Angular Disparity for True and False Items Averaged Over Eight Items

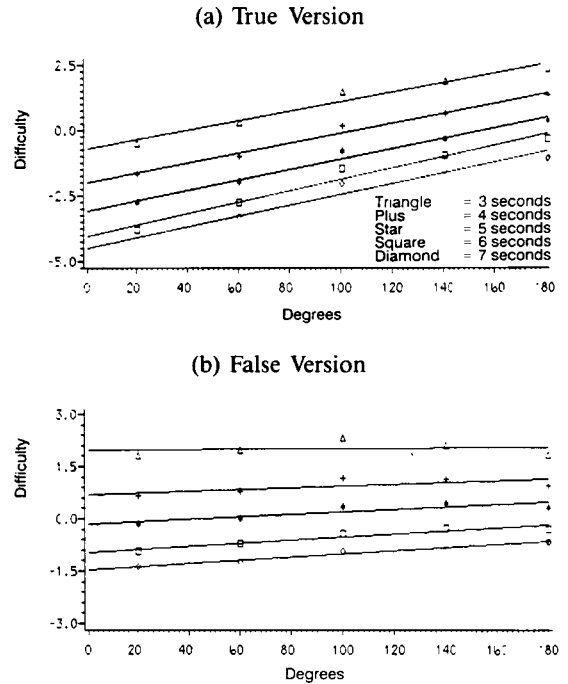


¹Because the model does not specifically distinguish between correct and incorrect responses and focuses on the modeling of response time, it is more appropriate to plot response time rather than response time of correct responses in the context of model fitting. Nevertheless, the plot with mean correct response time was also done, but similar results were obtained.

also shows that the impact of angular disparity on response time was less potent for false items.

The results for each item can be found in Bejar (1986). The expected results of equal slope of regression of difficulty and angular disparity at increasing elapsed times were observed, as well as the equal separation of the regressions. Figures 3a and 3b present the averaged results across the eight items for true and false versions. As can be seen, there is a slight departure from linearity at 100 degrees, which was seen in each of the items.

Figure 3
 Relationship Between Difficulty and Angular Disparity Averaged Across Eight Items



Discussion

The motivation behind this research was both substantive and methodological. The substantive interest was the assessment of the convergence, or lack thereof, of two approaches to individual differences: a psychometric approach, and an approach based on information processing psychology. Hunt and MacLeod (1978) expressed concerns that the differences between these two

approaches may be irreconcilable (see Pellegrino, 1988, for a different perspective). They cite the tendency for psychometricians to focus on global scores, such as proportion correct, versus the tendency of the cognitive psychologist to focus on more reductionist parameters, such as slope and intercepts of performance on task attributes. This study demonstrated that the dichotomy need not exist, at least not when IRT is used as a psychometric framework. Similarly, person characteristic curve methodology (Carroll, Meade, & Johnson, in press; Trabin & Weiss, 1983) is well-suited to characterize examinee performance in a psychologically-meaningful fashion.

This study also demonstrated that the IRT framework can be expanded when the response of interest is response time; in so doing, the results also demonstrated the possibility of encompassing both global and reductionist views of individual differences within the same measurement framework.

The methodological interest in this research was purely psychometric, and was focused on the development of a generative approach to psychometric modeling (Bejar, in press). The term "generative approach" refers to a methodology in which generation of items is controlled by an algorithm that encodes sufficient knowledge about the mental processes underlying performance on an item in order to be capable of anticipating the psychometric characteristics of the item before it is administered to examinees. The goals of this methodology are a natural extension of adaptive testing (Weiss, 1983). In an adaptive test, a computer retrieves the item that is most informative for a given individual from an existing database containing previously calibrated items. With a generative approach, however, an item can be created specifically for an individual instead of being retrieved, and can be done in such a way that the anticipated psychometric characteristics of the items are maximally informative for each individual.

One of the goals of adaptive testing has been the achievement of high measurement precision

throughout the ability range. This is a special concern with dichotomous items, when a balance must be struck between the overall level of precision and the distribution of precision at different levels of ability. Such a balance takes care of itself in the move from dichotomous to continuous responses, in that for continuous response models information may be high throughout the ability range, and equally high at all ability levels in some cases (Samejima, 1973). Therefore, that goal of adaptive testing seems to be automatically satisfied through the use of continuous responses. Nevertheless, there may be advantages to adapting the angular disparity to individuals of different abilities as a means of insuring a common response strategy by all examinees (Lohman, in press, p. 129).

Two essential ingredients in implementing a generative approach are (1) a psychometric framework of sufficient flexibility, and (2) a knowledge base about performance on the item. The three-dimensional cube item was chosen for this study because of the potent effect of angular disparity on performance in mental rotation items. It was therefore possible to establish the relationship between angular disparity and difficulty based on a few points. By manipulating this feature of the item, it became possible to generate items of any arbitrary difficulty. In a practical implementation of this approach that could be presented at rotations ranging from 20° to 180°, there would be 30 basic items. Although everyone would be presented all 30 items, the rotation at which they actually would be presented would be different for different individuals.

Although results of the study showed that the idea of generating items of arbitrary difficulty is indeed feasible, there are some problems that must be borne in mind with even relatively simple stimuli, such as the three-dimensional rotation items. Specifically, performance on false items is not a function of angular disparity; on the surface, this finding suggests that performance on the false items is controlled by a different combination of mental processes (see Carter, Pazak, & Kail, 1983).

References

- Assessment Systems Corporation. (1984). *User's manual for the MicroCAT testing system*. St. Paul MN: Author.
- Bejar, I. I. (in press). A generative approach to psychological and educational measurement. In N. Frederiksen, R. J. Mislevy, and I. I. Bejar (eds.) *Test theory for a new generation of Tests*. Hillsdale NJ: Erlbaum.
- Bejar, I. I. (1983). Subject matter experts' assessment of item statistics. *Applied Psychological Measurement*, 7, 303-310.
- Bejar, I. I. (1985). Speculations on the future of test design. In E. Embretson (Ed.), *Test design: Contribution from education and psychometrics* (pp. 279-294). New York: Academic Press.
- Bejar, I. I. (1986). *A psychometric analysis of a three-dimensional spatial task* (Rep. RR-86-19-ONR). Princeton NJ: Educational Testing Service.
- Bloxom, B. (1985). Consideration in psychometric modeling of response time. *Psychometrika*, 50, 383-397.
- Bock, R. D., & Atkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, 46, 443-459.
- Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new structure of intellect. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale NJ: Erlbaum.
- Carroll, J. B., Meade, A., & Johnson, E. S. (in press). Test analysis with the person characteristic function: Implications for defining abilities. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in education, psychology and social science: A book in honor of Lee J. Cronbach*. San Francisco CA: Jossey-Bass.
- Carter, P., Pazak, B., & Kail, R. (1983). Algorithms for processing spatial information. *Journal of Experimental Child Psychology*, 36, 284-304.
- Cooper, L. A. (1980). Spatial information processing: Strategies for research. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), *Aptitude, learning and instruction. Vol. 1: Cognitive processes analyses of aptitudes* (pp. 149-176). Hillsdale NJ: Erlbaum.
- Cooper, L. A., & Shepard, R. N. (1984). Turning something over in the mind. *Scientific American*, 251(6), 106-114.
- Corballis, M. C. (1982). Mental rotation: Analysis of a paradigm. In M. Potegal (Ed.), *Spatial abilities: Developmental and psychological foundations*. New York: Academic Press.
- Egan, D. E. (1979). Testing based on understanding: Implications from studies of spatial ability. *Intelligence*, 3, 1-15.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 175-197.
- Hunt, E., & MacLeod, C. M. (1978). The sentence verification paradigm: A case study of two conflicting approaches to individual differences. *Intelligence*, 2, 129-144.
- Lansman, M., Donaldson, G., Hunt, E., & Yantis, S. (1982). Ability factors and cognitive processes. *Intelligence*, 6, 347-386.
- 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. F. (in press). Estimating individual differences in information processing using speed-accuracy models. In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation and methodology: The Minnesota Symposium on Learning and Individual Differences*. Hillsdale NJ: Erlbaum.
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences. *Psychological Bulletin*, 86, 889-918.
- Micko, H. C. (1969). A psychological scale for reaction time measurement. *Acta Psychologica*, 30, 324-335.
- Miller, R. G. (1981). *Survival analysis*. New York: Wiley.
- Mislevy, R. J. (1988). *Exploiting collateral information in the estimation of item parameters: Final report* (Rep. RR-88-53-ONR). Princeton NJ: Educational Testing Service.
- Mislevy, R. J., & Bock, R. D. (1982). *BILOG: Item analysis and test scoring with binary logistic models*. Mooresville IN: Scientific Software, Inc.
- Myers, C. (1958). *Some observations of problem solving in spatial relations tasks* (Rep. RB-58-16). Princeton NJ: Educational Testing Service.
- Pellegrino, J. W. (1988). Mental models and mental tests. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 49-59). Hillsdale NJ: Erlbaum.
- Pellegrino, J. W., & Kail, R. (1982). Process analysis of spatial aptitude. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale NJ: Erlbaum.
- Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika*, 38, 203-219.
- Samejima, F. (1983). *A general model for the homogeneous case of the continuous response* (Res. Rep. No. ONR/RR-83-3). Knoxville: University of Tennessee, Department of Psychology.
- Scheiblechner, H. (1985). Psychometric models for speed-test construction: The linear experiential model. In S. E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics*. New

- York: Academic Press.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Swaminathan, H., & Gifford, J. A. (1981). *Bayesian estimation in the three-parameter logistic model* (Rep. LR-119). Boston: University of Massachusetts, School of Education.
- Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Ed.), *New horizons in testing* (pp. 179-203). New York: Academic Press.
- Trabin, T. E., & Weiss, D. J. (1983). The person response curve: Fit of individuals to item response theory models. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 83-108). New York: Academic Press.
- Tsutakawa, R. K., & Lin, H. Y. (1984). *Bayesian analysis of item response curves* (Res. Rep. 84-1). Columbia, University of Missouri, Department of Statistics.

Weiss, D. J. (Ed.). (1983). *New horizons in testing*. New York: Academic Press.

Acknowledgments

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-83-C-0761, Contract Authority Identification No. NR 150-531. The author is indebted to Professor Roger Shepard of Stanford University for providing the three-dimensional figures used in this study. The author thanks Min Hwei Wang for performing the data analysis, and Robert Mislevy for assisting in the estimation of item parameters.

Author's Address

Send requests for reprints and further information to Isaac I. Bejar, Law School Admissions Service, P.O. Box 40, Newtown PA 18940, U.S.A.