Journal of Educational Measurement
Spring 1996, Vol. 33, No. 1, pp. 29–39

Cognitive Design Principles and the Successful Performer: A Study on Spatial Ability

Susan E. Embretson
University of Kansas

An important trend in educational measurement is the use of principles of cognitive psychology to design achievement and ability test items. Many studies show that manipulating the stimulus features of items influences the processes, strategies, and knowledge structures that are involved in solution. However, little is known about how cognitive design influences individual differences. That is, does applying cognitive design principles change the background skills and abilities that are associated with successful performance? This study compared the correlates of two spatial ability tests that used the same item type but different test design principles (cognitive design versus psychometric design). The results indicated differences in factorial complexity in the two tests; specifically, the impact of verbal abilities was substantially reduced by applying the cognitive design principles.

An important trend in educational measurement is the use of principles of cognitive psychology to design achievement and ability test items. Mislevy (1993) notes that contemporary educational measurement strives to measure individual skills and abilities that are more diagnostic of learning and problem-solving processes. Recently, Wittrock and Baker's (1991) book *Testing and Cognition* examines the potential of cognitive design principles in diverse areas of testing.

Applying cognitive principles to design test items requires analyzing how item stimulus content influences the cognitive processes, strategies, and knowledge structures that are involved in item solving. Stimulus features which are hypothesized to influence irrelevant processes may be controlled, and features which influence relevant processes may be selected (see Embretson, 1985; Whitely, 1981). Mathematical modeling of item difficulty from item stimulus features has supported the importance and independence of multiple processes for many item types that appear on ability and achievement tests. A recent example is a study on mathematical problem solving (Embretson, 1995) in which item difficulty is well predicted by stimulus features that operationalize the difficulty of various processes in Mayer, Larkin, and Kadane's (1984) theory.

However, it is not clear that applying cognitive design principles during test development impacts findings about the nature of individual differences. If tests are designed to operationalize different aspects of cognition which depend on different background skills and abilities, then the tests should be expected to differ in their external correlates. However, comparing the correlates of whole tests that are based on different design principles has received little attention in the literature. Given the positive manifold of abilities and skills, it is possible that little influence is exerted.

A recently developed test of spatial ability, the Spatial Learning Ability Test (SLAT; Embretson, 1994), illustrates the application of cognitive design principles to item construction. It is well known that spatial tasks are not necessarily solved by spatial processing strategies. Early studies found evidence of nonspatial strategies on spatial items from self-reports (e.g., Barratt, 1953). More recently, Just and Carpenter's (1985) studies found that some participants used orientation-free verbal descriptions to solve a cube comparison task. Furthermore, Just and Carpenter's results suggest that verbal strategies are routinely employed on other spatial tasks, including both three-dimensional rotation tasks (e.g., Shepard & Metzler, 1971) and spatial orientation tasks (Guilford & Zimmerman, 1947). Similarly, Kyllonen, Lohman, and Woltz (1984) found evidence for a feature-matching strategy on a complex spatial integration task.

SLAT items were designed to maximize spatial processes and to minimize verbal processes by applying findings from several experimental cognitive studies. Like items on the Space Relations Test of the Differential Aptitude Battery (DAT-S), SLAT items are spatial folding tasks in which folded three-dimensional objects (e.g., a cube) must be compared to an unfolded view presented in a stem. However, the DAT-S and the SLAT differ in design principles. Whereas DAT-S item development reflects traditional psychometric principles, SLAT item features were manipulated for difficulty on two major spatial processes: rotation and folding. Furthermore, verbal strategies were minimized by constructing distractors that could not be falsified by a feature matching strategy.

A comparison of item stimulus content (Embretson, 1993) revealed that, unlike the SLAT, many DAT-S distractors can be falsified by feature matching rather than spatial analogue processing because they have perceptual mismatches with the stem figure. Furthermore, also unlike the SLAT, many DAT-S item stems require minimal spatial processing for the stem to be folded into the keyed alternative. Mathematical models of item difficulty and response times on the SLAT and the DAT-S strongly support the idea that the SLAT involves less verbal analytic processing than the DAT-S (see Embretson, 1994).

The studies of the SLAT described above examine the construct representation aspect of construct validity, which concerns the processes, strategies, and knowledge structures involved in item solving (see Embretson, 1983, for further definition). However, they do not concern the nomothetic span aspect of construct validity, which concerns the relationship of test scores to other measures of individual differences. Because nomothetic span depends on construct representation, it is expected to differ between two tests if construct representation differs between those tests. More specifically, the question addressed by the current study is whether SLAT performance is correlated with different abilities and skills than is DAT-S performance.

The current study examines the impact of cognitive design on findings about individual differences by comparing the nomothetic span of the SLAT to that of the DAT-S. Reference tests were selected to measure various spatial and verbal abilities. Confirmatory factor analysis for multiple groups is applied to test specific hypotheses about the factor loadings of the SLAT and the DAT-S on the reference tests. The postulated factorial complexity of the reference tests will be discussed before the method and results.

The Factorial Complexity of Spatial Tests

The results on construct representation, as presented above, have implications for the intercorrelations of SLAT and DAT-S scores with other tests. Both the SLAT and the DAT-S items can involve spatial processing and hence should correlate highly with other spatial tests. However, the traditionally designed DAT-S items may involve substantial verbal analytic processing and thus should be more highly correlated with some aspect of verbal ability than are the SLAT items. The factorial structure of spatial ability tests is complex and needs to be considered more fully before some specific hypotheses about SLAT and DAT-S factor loadings are considered.

Several different sets of spatial ability factors have been proposed in the psychometric literature. Lohman, Pellegrino, Alderton, and Regian's (1987) review concluded that two factors involving the mental manipulation of objects have been supported consistently. Spatial Visualization involves complex spatial manipulations, such as rotation, folding, or reflection of complex objects. For example, the paper folding task, paper form board, and the object folding task load on this factor. Spatial Relations (often labeled Spatial Orientation in the literature) involves more simple spatial manipulations, namely, rotating objects. Two-dimensional figure rotation and cube comparisons load on this factor. Recently, Carroll's (1993) comprehensive review of human cognitive abilities also supported these two factors.

According to Just and Carpenter's (1985) analysis, many spatial test items may involve verbal analytic processing. That is, spatial test items may be solved by converting the visual stimuli to a linguistic code and then applying verbal reasoning. Thus, for at least some individuals, relative success on spatial items is due to verbal ability rather than spatial ability. Consequently, the correlations between spatial ability tests represent the confounded influence of verbal ability factors and spatial ability factors. One special verbal ability, verbal coding, is probably especially important in the verbal processing of spatial stimuli. If the spatial stimuli are complex and irregular, often no standard verbal label adequately characterizes the relationships among elements. Thus, the ability to verbally code the spatial stimuli is probably essential to successful verbal processing of spatial test items.

If many spatial test items are solved by verbal analytic processing, then spatial reference factors do not clearly represent spatial ability. Residualizing spatial ability factors for linguistic coding ability could control at least partially for verbal analytic processing. Thus, a factor structure in which the spatial factors are controlled for linguistic coding ability may provide a more plausible account of the relationships among spatial and verbal tests than a simple reference factor structure with correlated factors.

Hypotheses

Several hypotheses are tested in the current study using confirmatory factor analysis, as follows.

Hypothesis 1. A residualized factor structure is hypothesized to explain the relationship between verbal and spatial tests more adequately than does a reference factor structure. A reference factor structure consists of correlated factors in which most tests load on only one factor. However, it is hypothesized that the correlations of a linguistic coding test with the spatial tests will not be well fit by the reference factor structure. Specifically, because linguistic coding may be involved in processing spatial stimuli, it may be too highly correlated with the spatial tests to serve as an indicator of only verbal ability in the reference structure. If linguistic coding is specified as a control factor for the spatial tests, the remaining correlations between spatial tests will more clearly represent spatial processing. Thus, a residualized factor structure, in which the spatial tests are controlled for linguistic coding, will provide better fit to the data than a reference factor structure.

Hypothesis 2. Both the SLAT and the DAT-S are hypothesized to load on the complex spatial ability factor, Spatial Visualization, rather than on Spatial Orientation.

According to the literature on spatial ability (see Carroll, 1993), folding tasks of the sort contained in the SLAT and the DAT-S involve complex spatial processing. Consequently, a factor model with loadings for the SLAT and the DAT-S on Spatial Orientation, which involves simple rotations, will not fit as well as a model with loadings on Spatial Visualization, which involves more complex spatial processing.

Hypothesis 3. Verbal reasoning abilities are hypothesized to be irrelevant to spatial task solutions. Linguistic coding is hypothesized to be the essential aspect of verbal reasoning in spatial tasks. If linguistic coding is controlled in the spatial tests, as in the residualized factor structure described above, adding loadings of the DAT-S and the SLAT on the verbal reasoning factor will not lead to better fit.

Hypothesis 4. The SLAT is hypothesized to be less related to linguistic coding ability than is the DAT-S. The cognitive design principles behind SLAT items, as compared to DAT-S items, specifically minimized the role of linguistic processing strategies in item solving. Thus, the SLAT loadings on linguistic coding are smaller than the DAT-S loadings.

Method

Tests

The test battery consisted of the SLAT and the DAT-S plus eight cognitive reference tests (see Table 1 for listing). Form 1P of the SLAT and Form V of the DAT-S were used in the comparisons. The cognitive reference tests were selected to represent the spatial ability and verbal reasoning factors. Spatial ability was measured by four tests from the *Kit of Factor-Referenced Cognitive Tests* (Ekstrom, French, & Harman, 1987): Cube comparisons and card rotations were selected to measure the Spatial Orientation factor, which is the same as Lohman et al.'s (1987) Spatial Relations factor, and the paper folding test and the form board test were selected to measure the Spatial Visualization factor.

Table 1

Descriptive Statistics and Comparisons on the Eight Cognitive Reference Tests in Two Groups

Variable	Factor	Group 1 Mean	Group 1 S.D.	Group 2 Mean	Group 2 S.D.	ANOVA F	ANOVA p
Inference Test	Verbal Reasoning	13.75	3.59	13.32	3.81	.67	.414
Deciphering Language	Verbal Reasoning	14.96	5.05	14.75	4.75	.09	.762
Paper Folding	Spatial Visualization	12.50	4.05	11.67	3.61	2.35	.127
Cube Comparison	Spatial Orientation	27.13	7.25	25.56	7.05	2.44	.119
Verbal Classification	Verbal Reasoning	18.02	3.84	17.63	3.16	.64	.424
Verbal Analogy	Verbal Reasoning	17.87	3.02	17.26	3.46	1.75	.187
Form Board	Spatial Visualization	22.56	8.64	20.97	9.34	1.57	.212
Card Rotation	Spatial Orientation	12.14	4.57	11.75	4.57	.37	.542

Verbal Reasoning was measured by four tests: The verbal analogy test and the verbal classification test were selected from the Cognitive Abilities Test, and the inference test and the deciphering languages test were selected from the *Kit of Factor-Referenced Cognitive Tests.* The deciphering languages test was selected also because it involves linguistic coding; that is, the test involves translating symbols into language to evaluate simple sentences. Thus, the deciphering languages test has a dual role in the factor models; it measures Verbal Reasoning, and it serves as a control factor for verbal analytic processing.

Participants

The participants were 209 undergraduates from a large Midwestern university. They took part in the study to earn credits toward grades in an introductory psychology course.

Design

Because the SLAT and the DAT-S contain the same item type, carry-over effects may be expected if both tests are administered to the same participants. Counterbalancing test order does not fully control for this effect. Examinees may develop strategies on the first test that are then applied to the second test, regardless of their effectiveness or appropriateness. Thus, responses to the second test are affected by exposure to the first test. Consequently, in the current study two independent groups were used to examine the factorial complexity of the two tests. In Group 1, participants received the SLAT as the first test; in Group 2, participants received the DAT-S as the first test. Confirmatory factor analysis for multiple groups was used to estimate the loadings of the SLAT and the DAT-S on the reference test factors. That is, by constraining the loadings of the reference tests on the factors across groups, the loadings of both the SLAT and the DAT-S on the common factors can be estimated from the two sets of independent data.

Procedure

Participants were randomly assigned to two groups, which varied according to whether the SLAT or the DAT-S was the first test. The test order was as follows: (1) the SLAT or the DAT-S, (2) verbal classification, (3) verbal analogy, (4) card rotations, (5) form board, (6) cube comparisons, (7) paper folding, (8) deciphering languages, and (9) the inference test.

Results

Preliminary Analyses

Frequency distributions were prepared to identify outliers. Observations more than 3 standard deviation units from the mean were trimmed for each variable. The final sample size was 97 in Group 1 and 90 in Group 2.

Table 1 presents the means and standard deviations of each test within each group. A multivariate analysis of variance indicated that the groups did not differ significantly on the cognitive reference tests, $F(8, 166) = 1.35, p = .22$. Consistent with the overall test, none of the univariate analyses of variance was significant for the reference tests (shown on Table 1). The uniformity of the covariances for the reference tests was also compared across groups. The Box *M* test was not significant, $\chi^2(36) = 37.17, p = .42$, as expected. Thus, the covariances for the reference tests are statistically equal, which indicates that the reference tests can be fit by the same factor structure across groups. Specifically, all factor loadings, error variances, and factor covariances for the reference tests in the various postulated structures will be constrained across groups.

Table 2 presents the intercorrelations among the tests. Some exploratory analyses were performed on the reference tests to determine if the correlations were explained by three ability factors, Spatial Visualization, Spatial Orientation, and Verbal Reasoning, as expected. Factor structures that were unspecified for content and unconstrained across groups were fit to the pooled sample to determine the number of reference factors, as recommended by Jöreskog (1971). A one-factor model did not fit the data, $\chi^2(40) = 147.05, p < .001$. A two-factor model was significantly better than the one-factor model, $\Delta\chi^2(14) = 104.21, p < .001$, but it did not fit, $\chi^2(26) = 42.84, p = .02$. A three-factor model, however, led to significant improvement, $\Delta\chi^2(12) = 30.40, p = .01$, and the data did not deviate significantly from the model, $\chi^2(14) = 12.44, p = .42$. A four-factor model was also attempted, but the additional factor did not improve fit significantly, $\chi^2(10) = 11.12, p = .62$.

Confirmatory Factor Analysis

Figure 1 shows the postulated residualized factor model and the estimated loadings. In EQS, a program for structural equation modeling (Bentler, 1989), models that are fit to multiple groups can contain somewhat different subsets of variables in each group, as is true in the current study. For those variables that are common across groups, the estimates may be constrained to the same value. For those variables that are unique in each group, the estimates are based only on the data from that particular group. Thus, in Figure 1, the estimates for

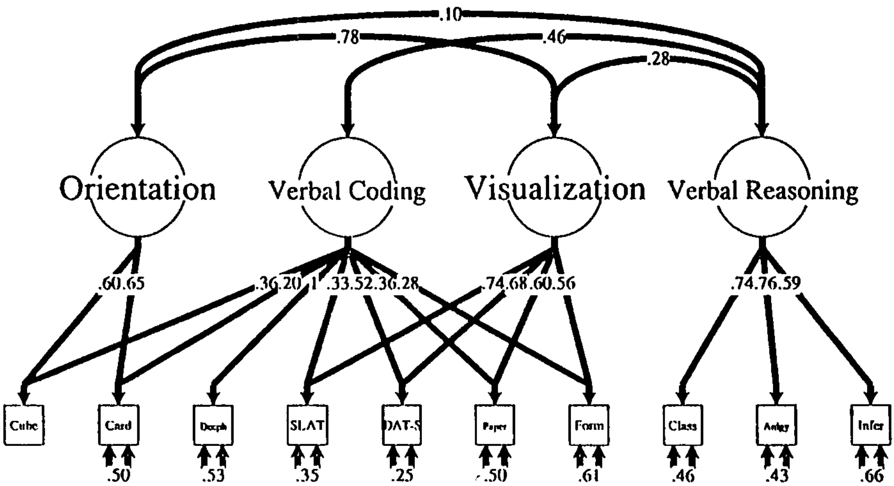

FIGURE 1. *The residualized factor structure for the relationship of the SLAT and the DAT-S with the reference factors*

the factor loadings, error variances, and factor correlations for the reference tests are constrained across the two groups. The factor loadings and error variances for the SLAT and the DAT-S are estimated only from the group in which they were administered.

Figure 1 shows a residualized factor structure with three ability factors, Spatial Orientation, Spatial Visualization, and Verbal Reasoning. It is a residualized factor structure because linguistic coding ability is a control factor for the spatial tests. Linguistic coding is identified by specifying the deciphering languages test to load fully on the factor and by specifying all spatial tests to have split loadings on the factor. Similar to a methods factor in a multimethod-multitrait analysis, linguistic coding is uncorrelated with the spatial factors for which it serves as a control. Deciphering languages is specified also to measure verbal reasoning (through its correlation with Verbal Reasoning).

Table 3 presents the likelihood ratio χ^2 tests to evaluate fit for several models. Also shown is the comparative fit index (see Bentler, 1990), which quantifies overall model quality (it ranges from 0 to 1). Table 3 shows the goodness of fit for the factor model in Figure 1 as Residualized Factor I. Table 3 shows that the data departed significantly from the model at the .05 significance level but not at the .01 significance level. Furthermore, the comparative fit index was very high.

The first hypothesis, as noted in the introduction, was tested by comparing the overall fit of the model shown in Figure 1 to a simple reference factor structure. In the reference structure, the linguistic coding loading was deleted from all spatial tests, and thus deciphering languages was permitted to load only on Verbal Reasoning. Table 3 shows that this model yielded a relatively higher χ^2 and a much lower comparative fit index. The decrease in goodness of fit, as compared to Residualized Factor I, was significant, $\Delta\chi^2(7) = 28.10$, $p < .01$.

Table 2

Correlations Among Reference Tests, SLAT and DAT-S*

	1	2	3	4	5	6	7	8	9	10
1) Inference Test	1.00									
2) Deciphering Language	.24	1.00								
3) Paper Folding	.25	.36	1.00							
4) Cube Comparison	.11	.36	.43	1.00						
5) Verbal Classification	.46	.34	.15	.13	1.00					
6) Verbal Analogy	.43	.36	.19	.13	.56	1.00				
7) Form Board	.28	.28	.46	.40	.19	.28	1.00			
8) Card Rotation	.11	.20	.37	.46	.14	.20	.33	1.00		
9) DAT-S	.32	.41	.60	.45	.23	.35	.53	.46	1.00	
10) SLAT	.21	.44	.56	.48	.21	.36	.45	.43	.69	1.00

* Correlations among reference tests are pooled across groups, while correlations for SLAT and DAT-S are from separate groups.

Therefore, the spatial tests are more related to linguistic coding than is explained by the factor correlations with Verbal Reasoning.

To test the second hypothesis, the SLAT and the DAT-S were specified to load on Spatial Orientation rather than Spatial Visualization. All other loadings were specified as in the basic model. This model, shown as Residualized Factor II in Table 3, did not fit as well as Residualized Factor I. Furthermore, the correlation between the two spatial factors was theoretically implausible ($r = .97$).

The third hypothesis was tested by allowing the SLAT and the DAT-S to load also on the Verbal Reasoning factor, which is shown in Table 3 as Residualized Factor III. This model did not fit significantly better than Residualized Factor I, $\Delta\chi^2(2) = 1.34, p > .05$. Thus, the impact of Verbal Reasoning on the SLAT and the DAT-S is adequately represented by the factor intercorrelations and the role of linguistic coding as a control factor.

Thus, the three hypotheses concerning the overall model were confirmed by the results. Some additional analyses were undertaken to attempt to improve the basic model in Figure 1. In EQS, model modification is indicated by the Lagrange multiplier (LM) test, which can be interpreted as the approximate improvement in fit upon releasing previously constrained parameters. The LM test indicated that model fit could be improved by releasing some constraints of the reference factor loadings across groups. The three most significant constraints with the highest LM test values were released accordingly. Table 3 shows that this model did fit the data. Furthermore, the improvement in fit was significant, $\Delta\chi^2(3) = 9.89, p < .01$. However, the comparative fit index of .97 differed little from the already high comparative fit index for Residualized Factor I. Because model fit

Table 3
Confirmatory Factor Analysis on Alternative Factor Models

Model	df	χ^2	p	Fit Index
1. Residualized Factor I	61	81.19	.04	.96
2. Reference Factor Structure	68	109.29	.01	.91
3. Residualized Factor II (DAT-S & SLAT on Spatial Orientation)	61	90.98	.01	.94
4. Residualized Structure III (I + DAT-S & SLAT on Verbal Reasoning)	59	79.84	.03	.96
5. Residualized Structure IV (I + Released Group Constraints)	58	71.30	.12	.97

was little improved by releasing these constraints across groups, and because the prior analysis of the reference test covariance matrices had supported uniformity, this model was deemed not worth the likelihood of capitalizing on sample-specific error. Furthermore, the relative pattern of loadings, including the loadings for the SLAT and the DAT-S, was identical to the pattern in Residualized Factor I.

Lastly, the fourth hypothesis was tested by investigating the pattern of the standardized factor loadings and standardized error variances in Figure 1 for the best model, Residualized Factor I. All loadings shown were highly significant ($p < .01$). As predicted by the fourth hypothesis, the SLAT had a much lower relationship to linguistic coding than the DAT-S in the basic model (loadings of .33 and .52, respectively, both of which are rounded to .7 in Figure 1). Further, the SLAT had a slightly higher loading on Spatial Visualization than the DAT-S (loadings of .74 and .68, respectively).

Discussion

The current results suggest that spatial ability factors confound verbal and spatial processing if not corrected for verbal analytic processing. As noted in the introduction, spatial reference tests themselves do not necessarily define purely spatial factors. Not only is their factor structure inconsistent across studies (Carroll, 1993), but experimental studies (e.g., Just & Carpenter, 1985) suggest that performance on spatial tests can represent either spatial or verbal processing, if spatial stimuli can be converted to a linguistic code.

In the current study, a residualized factor structure, in which the spatial test correlations were controlled for linguistic coding, better accounted for the correlations between verbal and spatial tests than did a simple reference factor structure. In the residualized factor structure, linguistic coding ability was partialed out of the spatial ability tests. The remaining correlation of the spatial tests with Verbal Reasoning was small, and adding split loadings for the spatial tests on Verbal Reasoning did not improve model fit. Thus, linguistic coding ability was supported as the primary aspect of verbal reasoning in spatial tests. Lastly, the spatial tests loading on the Spatial Visualization factor were more related to linguistic coding than were the tests loading on the Spatial Orientation factor. These results support Just and Carpenter's (1985) hypothesis that the role of verbal analytic processing is especially large in more complex spatial tasks.

Most importantly for evaluating the role of cognitive design principles, the results from the factor structures suggest that verbal ability is more important in the nomothetic span of the DAT-S than in that of the SLAT. In the residualized factor structure, both the SLAT and the DAT-S loaded on the complex spatial ability factor, Spatial Visualization. However, not only was the SLAT the highest loading test on Spatial Visualization, but the SLAT had a much smaller relationship to linguistic coding ability than the DAT-S. Because the test for linguistic coding had a moderate correlation with the Verbal Reasoning factor, higher indirect correlations with the other verbal tests were produced for the DAT-S than for the SLAT.

In conclusion, this study found that cognitive design principles can influence nomothetic span. It was shown that verbal analytic processing strategies were reduced by applying cognitive design principles to spatial items, with consequent impact on nomothetic span. Thus, cognitive design principles can be effectively employed to control the nature of what is measured by a test. Because inappropriate strategies can be reduced by cognitive design principles, one would also expect better differential validity for such tests. Cognitive design may be especially important for aptitude tests because differential validity has not been strongly supported for many aptitude tests (e.g., Anastasi, 1985).

References

Anastasi, A. (1985). *Psychological testing.* New York: Macmillan.

Barratt, E. S. (1953). An analysis of verbal reports of solving spatial problems as aid in defining spatial factors. *The Journal of Psychology, 36,* 17–25.

Bentler, P. M. (1989). *EQS: Structural equations program manual.* Los Angeles: BMDP Statistical Software, Inc.

Bentler, P. M. (1990). Comparative fit indices in structural equation modeling. *Psychological Bulletin, 107,* 238–246.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1987). *Manual for the Kit of Factor-Referenced Cognitive Tests* (Research Rep. NR-150-329). Princeton, NJ: Educational Testing Service.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Embretson, S. E. (1985). Multicomponent latent trait models for test design. In S. E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics.* New York: Academic Press.

Embretson, S. E. (1994). Applications of cognitive design systems to test development. In C. Reynolds (Ed.), *Advances in cognitive assessment: An interdisciplinary perspective* (pp. 107–135). New York: Plenum.

Embretson, S. E. (1995). A measurement model for linking individual learning to pro-

cesses and knowledge: Application to mathematical reasoning. *Journal of Educational Measurement, 32,* 277–294.
Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426.
Just, M., & Carpenter, P. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137–172.
Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology, 76,* 135–145.
Lohman, D. F., Pellegrino, J. W., Alderton, D. L., & Regian, J. W. (1987). Dimensions and components of individual differences in spatial abilities. In S. H. Irvine & S. E. Newstead (Eds.), *Intelligence and cognition: Contemporary frames of reference.* Boston: Martinus Nijhoff Publishers.
Mayer, R., Larkin, J., & Kadane, P. (1984). A cognitive analysis of mathematical problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2). Hillsdale, NJ: Erlbaum.
Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Erlbaum.
Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science, 171,* 701–703.
Whitely, S. E. (1981). Measuring aptitude processes with multicomponent latent trait models. *Journal of Educational Measurement, 18,* 67–84.
Wittrock, M., & Baker, E. (1991). *Testing and cognition.* Englewood Cliffs, NJ: Prentice Hall.

Author

SUSAN E. EMBRETSON is Professor, Department of Psychology, University of Kansas, Lawrence, KS 66045; sembret@stat1.cc.ukans.edu. *Degree:* PhD, University of Minnesota. *Specializations:* measurement, cognition, and quantitative methods.