

Information Structure for Geometric Analogies: A Test Theory Approach

Susan E. Whitely and Lisa M. Schneider
University of Kansas

Although geometric analogies are popular items for measuring intelligence, the information processes that are involved in their solution have not been studied in a test theory context. In the current study, processing is examined by testing alternative models of information structure on geometric analogies. In contrast to the treatment of models in other studies that have appeared in the cognitive literature, the models are tested jointly as mathematical models of processing and as latent trait models of individual differences. The joint model-

ing was achieved by applying the one-parameter linear logistic latent trait model to predict response accuracy from information structure. The results supported the model that distinguished between spatial distortion and spatial displacement transformations, which have opposite effects on item difficulty. Further, no significant sex difference in overall accuracy or processing were observed. Implications of the results for processing mechanisms and test design are discussed.

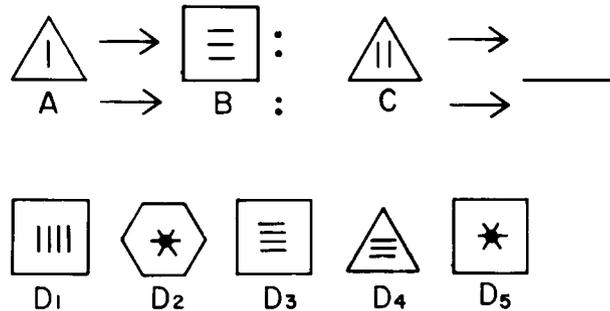
Geometric analogies are popular items for measuring nonverbal intelligence. They are included on tests such as the Cognitive Abilities Test (CAT), the Lorge-Thorndike Intelligence Test, and the American Council on Education Examination (ACE). An example of a geometric analogy that is similar to CAT items is given in Figure 1, along with structural notation. The task is to select from the alternatives (D_1 to D_5) to find a figure that is related to C in the same way that B is related to A.

Similar to most types of intelligence test items, validation studies on geometric analogies have been mainly concerned with trait organization rather than the nature of processing involved in item performance. That is, person scores are correlated with other tests or achievement and learning, but studies of processing are infrequently reported. Thus, the psychometric validation research has not clarified the nature of the reasoning process measured by geometric analogies.

A new type of validity research, cognitive component analysis of aptitude (e.g., Pellegrino & Glaser, 1979; Sternberg, 1977a; Whitely, in press), seeks to understand test validity by identifying the information processes that contribute to performance. Cognitive component analysis has been applied to several types of aptitude test items, including verbal analogies (Pellegrino & Glaser, 1979; Sternberg, 1977a, 1977b; Whitely, 1977; Whitely & Barnes, 1979), spatial rotation and orientation (Egan, 1979), reading comprehension (Frederiksen, 1980), and deductive reasoning (Sternberg, 1979).

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 5, No. 3, Summer 1981, pp. 383-397
© Copyright 1981 Applied Psychological Measurement Inc.
0146-6216/81/030383-15\$1.75

Figure 1
A Geometric Analogy That is Similar
to CAT Test Items



Three major studies have examined information processing on geometric analogies. Evans (1968) provided a detailed specification of both representation and processing by constructing a computer program that could solve the ACE geometric analogies. Evans' program includes the following processes: (1) pattern feature encoding, so that the figures are represented in a form that is recognized by the program; (2) rule generation, where the relationships between stimulus pairs (A:B, C:D_i) are inferred; and (3) rule comparison, where the A:B relationship is compared for equivalence to the C:D_i relationships. Although Evans' program is a very detailed description of processing, it should be noted that the program represents *artificial* or machine intelligence, which may not correspond to the processing mechanisms that are employed by human problem solvers.

Sternberg (1977b), in contrast, employed mathematical modeling of subject performance data to test his componential model of analogical reasoning. Again, ACE analogies were used, and Sternberg found that his processing model gave better accounts of the data than those postulated by Evans (1968). The important difference is that Sternberg's (1977b) model contains an additional component in the stem (mapping), followed by an evaluation of the response options with respect to a specific hypothesis (this is called application). In contrast, Evans (1968) postulates unconstrained inference attempts for C:D_i, which are then evaluated for comparability to A:B.

Recently, Mulholland, Pellegrino, and Glaser (1980) studied item information structure in mathematical models of both item difficulty and response time on geometric analogies. Using a set of geometric analogies that were constructed to represent specified information features, they found that a model with two major information variables—number of elements and number of transformations—yielded high prediction of the item performance data.

Although these three studies provide a good beginning for clarifying what is measured by geometric analogies, they may have little impact on the psychometric literature for two reasons. First, and most importantly, person parameters are not explicitly included in the process models. Although the mathematical models are tested for individual subjects (c.f. Sternberg, 1977b; Mulholland et al., 1980), no parameters to account for person differences in overall accuracy or response time are postulated. This is an especially important limitation, because the goal of psychometrics is the measurement of individual differences. Furthermore, very precise accounts of test performance from joint models of person and item characteristics are becoming increasingly commonplace in psychometrics due to the application of latent trait models. However, the popular latent trait models are trait models, not process models.

Second, and also important, the cognitive component studies do not diagnose the contributions of the multiple processes to specific test items. That is, the mathematical models examine overall data

accountability but do not directly include component item parameters. Such parameters are needed if the results from cognitive component analysis are to be used for item selection and test development.

The current study examines the nature of processing on geometric analogies by modeling item difficulty from information structure. Unlike previous studies, the mathematical model of processing is a special latent trait model that estimates person ability as well as process contributions to item difficulty.

Models Relating Information Structure and Item Difficulty

Three alternative mathematical models are postulated to relate information structure to item difficulty. All three models consist of two basic types of information: the number of elements in the A stimulus and the number of transformations involved in converting A to B in the analogy. The models vary, however, in the types of transformations that are specified.

Model I is the most parsimonious of the three models, as it does not distinguish among types of transformations. This model is similar to Mulholland et al.'s (1980) model. Model II distinguishes among two major types of transformations: spatial displacements (e.g., rotation) and distortions (e.g., size or shape changes). It is hypothesized that these two types of transformations have different relationships to item difficulty. Consistent with previous literature on spatial rotation (Shepard & Metzler, 1971), items with more displacement transformations should be more difficult. However, items with more distortion transformations should be easier. Research on pattern discrimination has indicated that objects that have more discrepant features are more easily identified (Cooper, 1976; Pomerantz, Sager, & Stoeber, 1977).

Model III distinguishes among specific types of transformations in each item, such as size or shape changes. Previous research on geometric analogies may have involved a rather limited set of transformations, which could differ in their relationship to item difficulty. Model III is the least parsimonious model.

It should be noted that sex differences on geometric analogies could be expected because of their dependence on spatial displacement transformations. That is, males are frequently found to perform somewhat better than females on tests that involve visualization skills, as in rotation transformations (McGee, 1979). However, sex differences in geometric analogies have been inconsistent. Mulholland et al. (1980) report substantial sex differences in accuracy, while normative data on the CAT geometric analogies finds that females are slightly more accurate (Thorndike & Hagen, 1974, p. 21). However, these studies have not examined possible sex differences in processing models for item performance. Possibly the processing algorithms that are applied to solve geometric analogies are different between men and women. Processing algorithm differences could moderate the relationship of information structure to item difficulty.

The Linear Logistic Latent Trait Model

Although the linear logistic latent trait model (Fischer, 1973) has been available in European literature for several years, it is virtually unknown in the American psychometric literature. The linear logistic latent trait model, along with the multicomponent latent trait model (Whitely, 1980), estimate processing contributions to item difficulty.

The linear logistic latent trait model is similar to the Rasch (1961) model, in that the only item characteristic is difficulty. That is, no discrimination or guessing parameters are postulated. However, items vary in the processing operations that are involved in performance. If processing complex-

ity or information structure can be specified by an a priori theory or mathematical model, item difficulty can be factored into contributions from the separate processing operations.

To show the relationship of the linear logistic model to the Rasch model, consider the following equation, which is the Rasch model for the probability that person j passes item i :

$$P(x_{ij} = 1) = \frac{\exp(\xi_{ij} - \sigma_i)}{1 + \exp(\xi_{ij} - \sigma_i)} \tag{1}$$

where

- ξ_j = ability level for person j , and
- σ_i = difficulty for item i .

Fischer postulates that item difficulty can be explained by a subset of m complexity components, according to the following linear model:

$$\sigma'_i = \sum_m f_{im} \eta_m + a \tag{2}$$

where

- f_{im} = the number of operations of type m that are involved in solving item i
- η_m = the difficulty of processing operation m , and
- a = a normalization constant.

Thus, if an $i \times m$ matrix of information structure for the items containing the f_{im} values can be scored, the processing difficulty parameters, η_m , can be estimated.

The complete linear logistic model is given as follows:

$$P(x_{ij} = 1) = \frac{\exp(\xi_{ij} - (\sum_m f_{im} \eta_m + a))}{1 + \exp(\xi_{ij} - (\sum_m f_{im} \eta_m + a))} \tag{3}$$

The Rasch model is a special case of the linear logistic model, where information structure is an identity matrix, with each item defining a separate η_m , and a equal to zero. Table 1 presents the information structure for the first 10 geometric analogies in Model II, scoring three complexity factors: num-

Table 1
Information Structure for Ten
Geometric Analogies

Item	Information Structure (Model II)		
	Elements	Distortions	Displacements
1	1	2	0
2	1	1	0
3	1	2	0
4	1	0	1
5	1	1	1
6	1	0	1
7	2	0	1
8	4	0	2
9	1	1	1
10	2	2	1

ber of elements, number of spatial distortions, and number of spatial displacements. The latter two factors refer to the number of transformations required to convert A into B in the analogy.

It should be noted that the linear logistic model does *not* require that the complexity factors are uncorrelated. Furthermore, a given complexity factor may have a value of zero for a given item, meaning that the factor is not involved in processing the item. However, the model does require unidimensionality among the items, since only one parameter is estimated for each person.

Fischer (1973) presents conditional maximum likelihood estimates for the parameters, which can be obtained from Fischer's program for the model (Fischer & Formann, 1972). Thus, an estimate for the difficulty of each complexity factor is obtained, along with standard errors and error correlations. Furthermore, likelihood ratio chi-square tests can be used to compare alternative models. Under conditional estimation procedures for latent trait models, the difference in log likelihoods between alternative models multiplied by -2 are asymptotically distributed as chi-square. Similarly, complexity models can be compared between groups, such as male versus females, by the chi-square test. That is, the log likelihood for the total group, when treated as a single sample, is compared to the sum of the log likelihoods when estimates are obtained for each separate group.

Method

Subjects

The subjects were 211 undergraduates who were participating in the experiment to earn credit in an introductory psychology course. The sample included 110 females and 101 males.

Materials and Design

The 30 geometric analogies were from Levels G and H of the Cognitive Abilities Test (Thorndike & Hagen, 1974), which is appropriate for 12th grade and beyond. The items were presented on an overhead projector for a fixed interval of 25 seconds. This interval corresponds to the average time available under standard test instructions. However, since accuracy, rather than speed, was the dependent variable of interest, the fixed presentation rate was implemented so that item differences would depend on relative complexity but not test position. Standard test instructions were given preceding the items to assure that the stated speed/accuracy requirements were the same as the standard CAT. Items were presented in the same order in which they appear on the CAT.

Scoring Information Structure

Number of elements was scored in two ways. Elements (1) counted the number of features in the A term. This is similar to the Mulholland et al. (1980) scoring of number of elements, which ranged from one to three. However, psychometric geometric analogies, such as the CAT items, contain substantially more complex figures than the experimental analogies in Mulholland et al. (1980); and, as noted in that study, the number of separate elements that subjects actually represent may depend on the transformation to be made to the B term. Elements (2) then, counted the number of features of the A term that were varied independently in the analogy. Elements (1) and Elements (2) were each scored by two raters with reliabilities of .81 and .83, respectively.

Seven types of transformations were also scored: number (i.e., adding, removing, or dividing elements), shading, size, shape, rotation, reflection, and spatial exchanges. It should be noted that Evans' (1968) program included only the number, rotation, reflection, and spatial exchanges transfor-

mations; whereas the Mulholland et al. (1980) study included only size, rotation, reflection, and number. Thus, the CAT items represent a larger set of transformations than has been studied previously.

Three types of transformations involve spatial *displacement*; that is, one or more elements of A are disoriented in B. These types include rotation, reflection, and exchanges, which are usually regarded as involving visualization or mental imagery. The other four types involve spatial *distortions*, since the A stimulus must be contorted in some way to become B. These types include size, shading, shape, and number, none of which would require mental imagery.

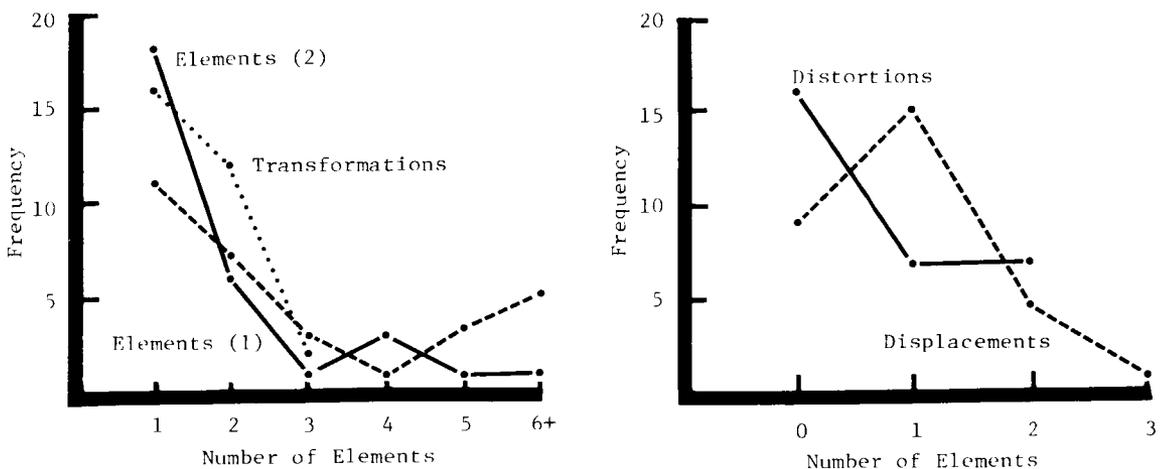
Data Analysis

Three mathematical models of item difficulty were tested by comparing data fit to the three information structure matrices: (1) Model I, scoring Mulholland et al.'s (1980) variables of number of elements and number of transformations; (2) Model II, scoring number of elements, as above, but subdividing number of transformations into two types—spatial distortions and spatial displacements; (3) Model III, scoring number of elements and number of transformations separately within the seven categories described above. Then, the best model is examined for overall utility by comparing likelihood of the data using the processing model to the likelihood obtained for the Rasch model. Notice that the linear logistic model estimates difficulty parameters for the complexity factors, which are less than the number of items. Thus, better fit is expected for the Rasch model, which contains one parameter for each item. However, item difficulties are reproduced from the linear logistic model through the prediction model described in Equation 2. Thus, a regression of the Rasch model parameters on the item difficulty obtained from the linear logistic model is another method for examining the adequacy of the information structure model to characterize performance.

Results

Figure 2 presents frequency polygons for item information structure. These data show the “experimental design” of information requirements that is implicit in the items. It can be seen that Ele-

Figure 2
Frequency Polygons for Information Structure



ments (1) led to higher frequencies for large values than Elements (2). However, Elements (1) and Elements (2) were highly correlated ($r = .75, p < .01$) and the correlation of Elements (2) with item difficulty ($r = .52, p < .01$) was only slightly higher than the correlation of Elements (1) with item difficulty ($r = .48, p < .01$). Thus, only Elements (2) was examined in the models. The correlation between Elements (2) and number of transformations was not significant ($r = .15, p < .05$).

The bottom of Figure 2 shows the frequency of number of distortion and displacement transformations. It can be seen that 16 items did not involve any distortion transformations, whereas 9 items did not involve any displacements. Thus, 25 of the 30 items contained only one type of transformation. The occurrence of the two types of transformations in different items is further shown by the significant negative correlation between number of distortions and displacements ($r = -.65, p < .01$). Figure 2 also shows that most items contained only one transformation and that only two items in the set required more than two transformations.

The information structure for the specific operations involved in each item (Model III) is shown in Table 2. Also shown in Table 2 are conditional maximum likelihood estimates of the Rasch item difficulties, which were obtained as a special case of the linear logistic model (Fischer & Formann, 1972). The Rasch item difficulties were obtained (1) to evaluate the fit of the information structure models and (2) to compare the linear logistic estimates to the Rasch estimates by an ordinary least squares regression of the Rasch item difficulties on the item difficulties that are reproduced from the linear logistic model.

Models of Information Structure

Conditional maximum likelihood estimates for the information complexity factors were obtained separately from Fischer and Formann's (1972) program for the linear logistic latent trait model. Table 3 shows the complexity factor estimates and associated standard errors for Model I, based on Mulholland et al.'s (1980) variables. Number of transformations was significant, but number of elements was not significant. These parameters are roughly comparable to the predictor weights that would be obtained by ordinary least squares estimation of the Rasch difficulty parameters (obtained from a standard program) regressed on the information matrix variables (i.e., number of elements and number of transformations). For comparison, the regression weights were estimated with the 30 Rasch item difficulties as the dependent variable. The regression weights were .55 and .00, respectively, for number of elements and number of transformations, which are quite similar to the parameters in the linear logistic model that are presented in Table 3. Also presented in Table 3 is a , the normalization constant, for Model I, which is comparable to the regression intercept estimate of -1.03 . Notice also that the correlations of the parameter estimation errors are presented in Table 3. This correlation results from the correlation among number of elements and number of transformations in the test information design.

Table 4 presents the complexity factor estimates for Model II. The log likelihood goodness-of-fit test showed that Model II was significantly better than Model I ($\chi^2_1 = 454.14, p < .01$). This value is -2 times the log likelihood difference between the models. It has one degree of freedom because one additional factor was estimated in Model II.

Unlike Model I, number of elements was not significant in Model II. Both types of transformations, however, were significant; but the direction of the relationship was reversed. For displacements, increased number of transformations predicted high item difficulties; whereas for distortions, increased number of transformations predicted lower item difficulties. Furthermore, the absolute values for weights of the two types of transformations differed significantly ($t = 10.00, p < .05$) so that displacements had a larger influence on difficulty than distortions.

Table 2
Rasch Item Difficulties and Information Structure for 30 Geometric Analogies

Item	Difficulty θ	Elmt	Numb	Information Structure-Model III						Rotate
				Shade	Size	Shape	Reflt	Exchg	Rotat	
1	.356	1	2	0	0	0	0	0	0	0
2	-1.193	1	1	0	0	0	0	0	0	0
3	-1.121	1	0	2	0	0	0	0	0	0
4	-1.269	1	0	0	0	0	0	0	0	1
5	-1.532	1	0	0	1	0	0	0	0	1
6	-1.053	1	0	0	0	0	0	0	0	1
7	-1.193	2	0	0	0	0	1	0	0	0
8	.705	4	0	0	0	0	1	1	1	0
9	-2.010	1	1	0	0	0	0	0	0	1
10	.129	2	1	0	0	1	0	0	0	1
11	-2.589	1	0	1	0	1	0	0	0	0
12	-2.010	1	1	0	1	0	0	0	0	0
13	.775	5	0	0	0	0	1	1	1	0
14	-.163	1	1	0	0	0	0	0	0	0
15	2.032	4	0	0	0	0	0	2	0	0
16	-2.010	1	0	0	0	1	0	0	0	0
17	.129	1	0	1	1	0	0	0	0	1
18	.246	2	2	0	0	0	0	0	0	0
19	1.374	1	0	0	0	0	0	0	0	1
20	2.118	4	0	0	0	0	1	0	0	2
21	.198	1	0	0	0	0	1	1	0	0
22	.988	1	1	0	0	0	0	0	0	1
23	1.864	1	0	0	0	0	0	0	0	1
24	2.459	6	0	0	0	0	1	1	1	0
25	1.271	1	0	0	0	0	0	0	0	1
26	-.988	2	0	1	0	0	0	0	0	0
27	.435	3	0	0	0	0	0	1	0	0
28	.099	2	0	0	0	0	0	0	0	1
29	1.230	2	0	0	0	0	1	0	0	0
30	3.094	1	0	0	0	0	1	1	0	1

Table 3
Linear Logistic Estimates for Model I

Information Complexity	Complexity Parameter	Standard Error	t-value
Elements	.39**	.02	19.50
Transformations	.07	.05	1.40

**p ≤ .01 (a = -.84)

To clarify further the nature of the complexity factor effects on difficulty, Rasch item difficulties were plotted by number and type of transformation for the 25 items that involved only distortion or displacement transformations. It can be seen in Figure 3 that only a slight difference in difficulty is observed between items with one versus two distortions, whereas the difference between one and two displacements is quite substantial. Also evident in Figure 3 is that displacement items are substantially more difficult than distortion items.

The distortion and displacement items were then divided into two subsets to test for homogeneity. Gustafsson's (1977) likelihood ratio chi-square test for item subsets was not significant ($p \geq .05$), indicating that the distortion and displacement items measure the same ability.

Table 5 shows the complexity factor estimates from Model III, which examined the effects of specific transformations on difficulty. A significant increase in goodness of fit was obtained over Model II, as indicated by the log likelihood chi-square test ($\chi^2 = 65.79, p \leq .01$). It can be seen in Table 3 that the seven types of transformations were not related to item difficulty in the same way. Transformations involving number and shading had no contribution to difficulty in the CAT items. However, transformations involving reflection, exchange, and rotation had highly significant positive relationships to difficulty; and those involving shape and size had a highly significant *negative* relationship to difficulty.

Figures 4 and 5 present the regression of the Rasch difficulty parameters and the predictions from Model II and Model III, respectively. A predicted item difficulty from the model is given by multiplying the information structure data by the complexity parameter estimates, using Equation 2. For Model III, the parameter estimates given for the eight complexity factors can be multiplied by the information structure data, as follows, for a given item, such as Item 8 in Table 2:

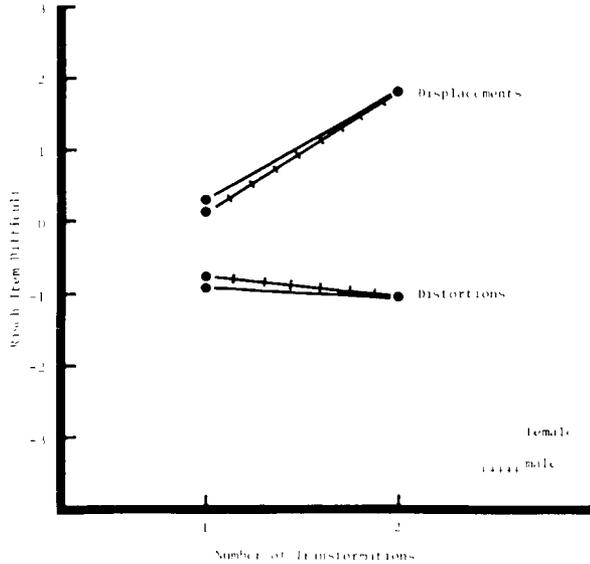
Table 4
Linear Logistic Latent Trait Estimates for Model II

Information Complexity	Complexity Parameter	t-value	Standard Errors and Error Correlations ^a
Elements	.01	.33	<u>.03</u>
Transformations			
Distortion	-.27	-5.40**	-.05 <u>.05</u>
Displacement	.96	13.71**	-.61 <u>.49</u> <u>.07</u>

**p ≤ .01 (a = -.73)

^aunderlined values are the standard errors.

Figure 3
 Rasch Model Difficulties Plotted
 by Type and Number of Transformations



$$\sigma'_s = 4(-.07) + 0(.11) + 0(-.08) + 0(-.76) + 0(-.95) + 1(1.09) + 1(1.28) + 0(1.09) = 2.09$$

It can be seen that only those transformations or elements that are involved in the item, as indicated by the information matrix, enter into the difficulty prediction.

The regression presented in Figures 4 and 5 show fairly good prediction of the Rasch item difficulties from the complexity factor model. These regressions yield Pearson correlations of .68 and .73, respectively, for Model II and Model III, so that Model III accounted for 7% more of the variance in item difficulty than Model II.

Table 5
 Linear Logistic Latent Trait Estimates for Model III

Information Complexity	Complexity Parameter	t-value	Standard Errors and Error Correlations							
Elements	-.07	-1.75	<u>.04</u>							
Transformations										
Number	.11	1.57	-.24	<u>.07</u>						
Shading	-.08	-.72	-.17	.35	<u>.11</u>					
Shape	-.76	-5.84**	-.14	.00	-.03	<u>.13</u>				
Size	-.95	-6.33**	.04	.04	.18	.09	<u>.15</u>			
Reflection	1.09	9.90**	-.67	.48	.32	.19	.07	<u>.11</u>		
Exchanges	1.28	10.67**	-.81	.42	.31	.16	.00	.61	<u>.12</u>	
Rotation	1.09	15.57**	-.32	.43	.37	.07	.09	.43	<u>.16</u>	<u>.07</u>

**p ≤ .01 (a = -.78)

Figure 4
 Regression of Rasch Item Difficulties
 on Model II Item Difficulties

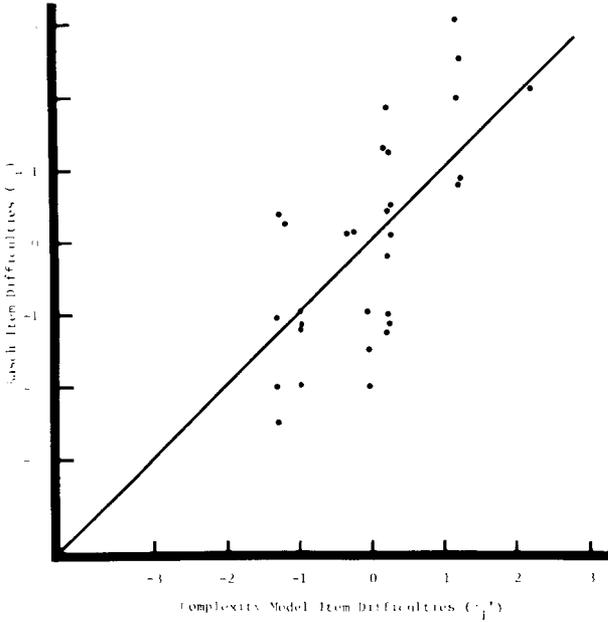
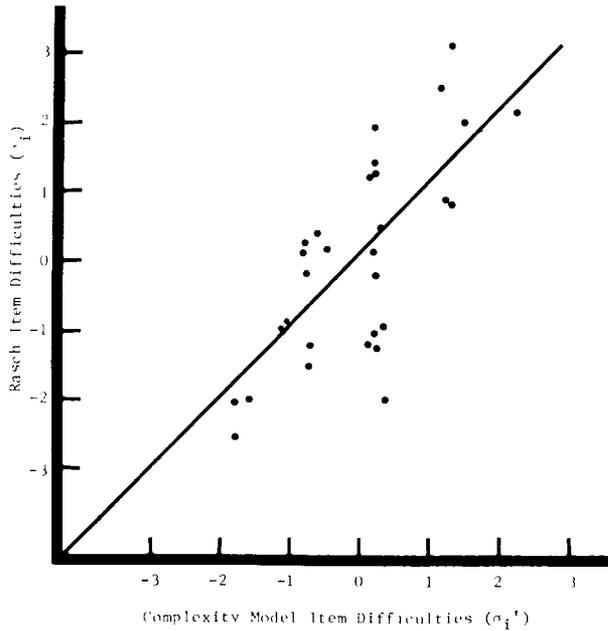


Figure 5
 Regression of Rasch Item Difficulties
 on Model III Item Difficulties



A last test compared the likelihoods of the data between Model III and the standard Rasch Model (i.e., the information structure is an identity matrix). A highly significant difference was found ($\chi^2_{30} = 484.95, p \leq .01$), which indicates that more complex information structures are needed to fully account for item difficulty.

Sex Differences

The mean raw scores were 22.29 for females and 22.18 for males; the difference was clearly non-significant ($t = .16, p > .05$). Table 6 presents the complexity factor estimates, by sex, for each model of information structure. Reported in the two last columns are the significance tests. The chi-square values are based on the log likelihood differences of the data between estimating the model on the full group and estimating the parameters separately by group. For Model I, in the total sample, two parameters were estimated, whereas separating the sample by sex gave two parameters for each group, as shown in Table 6. The chi-square has two degrees of freedom, since two additional estimates are obtained when the data are divided by sex. The chi-square results indicated no significant sex differences for all three models, when the parameters were considered as a set. Table 6 also shows t values for group comparisons of each separate parameter of the three models. Only the parameter for spatial distortion transformations in Model II differed significantly. Distortions were more (negatively) related to item difficulty for females than for males.

Table 6
Complexity Parameters by Sex

Model	Females	Complexity Parameters		
		Males	t-value	χ^2_{model}
Model I				4.26
Elements	.39	.40		(df=2)
Transformations	.02	.16		
Model II				5.06
Elements	.01	.02	.05	(df=3)
Transformations				
Spatial Distortion	-.38	-.16	2.11**	
Spatial Displacement	.87	1.05	1.35	
Model III				9.86
Elements	-.04	-.10	-.48	(df=8)
Transformations				
Number	.04	.19	-1.06	
Shading	-.20	.04	-.75	
Shape	-.99	-.54		
Size	-1.20	-.73	1.58	
Reflection	1.02	1.17	-.75	
Exchanges	1.06	1.52	-1.85	
Rotation	1.00	1.19	-1.27	

** $p \leq .05$

Discussion

The current study examined the adequacy of some relatively simple latent trait models of information structure to account for performance on geometric analogies. All three models that were examined considered just two sources of processing complexities, the A stimulus elements and the A to B transformation. Obviously, however, geometric analogy items have several additional sources of complexity that were not considered in the information structure models, and these sources have had importance in other research on geometric analogies. Sternberg (1977b), for example, includes the A to C transformation, as well as the complexity of the response alternatives. The significant difference between the processing models and the Rasch model in the current study no doubt reflects the role of these other sources of processing complexity. Nevertheless, however, the results indicate that the relatively simple models of information structure give a good account of item difficulty. The item difficulties that were predicted from the best processing models, Model II and Model III, correlated .68 and .73, respectively, with Rasch-scaled item difficulty. Thus, the models in the current study accounted for a major portion of variance in item difficulty.

The comparisons between the three models of information structure indicated that the various types of transformations between A and B terms of geometric analogies have different relationships to item difficulty. Model I, which did not distinguish between types of transformations, indicated that number of transformations was not significantly related to item difficulty. However, distinguishing between two types of transformations in Model II not only yielded substantially better fit, but both types of transformations had significant contributions to difficulty. As predicted, number of spatial displacement transformations was positively related to item difficulty, while number of spatial distortions was negatively related. Apparently the nonsignificant findings for number of transformations in Model I resulted from summing over types of transformations with opposing influences on item difficulty. Model III, which separated transformations into several categories, yielded significantly better fit than Model II. However, it accounted for only 7% more variance. Since five additional parameters were estimated in Model III, the increase in prediction does not seem large enough to support this more complex model.

In any case, the data clearly indicate that different types of transformations do not have similar influences on item difficulty. These results are inconsistent with Mulholland et al.'s (1980) study of geometric analogies but are consistent with other cognitive research on spatial displacements and spatial distortions. That is, increased spatial displacement leads to greater difficulty in figure comparisons (Shepard & Metzler, 1971), while increased distortion has the opposite effect (Pomerantz et al., 1977). These reversed effects of transformations on task difficulty suggest qualitatively different processing mechanisms for items that require displacement versus distortion comparisons. Many theoretical models of comparisons between distorted figures assume that the major processing component is self-terminating encoding of the figure features (Nickerson, 1972). The process terminates when a difference between two figures is detected, so that as the number of transformations increases, a difference is identified earlier. On the other hand, research on spatial displacements has supported a three-component model (Just & Carpenter, 1976), which includes search, transformation/comparison, and confirmation, but not self-terminating encoding. Thus, conceptually distinct processing mechanisms have been identified in displacement and distortion comparisons.

The opposing effects of different types of transformations on item difficulty has some obvious implications for test construction. The test developer can explicitly increase distortion transformations to obtain easy items, but can decrease displacement transformations for the same effect. Furthermore, since the current results indicated that distortion items are easier than displacement items, the relative proportions of these two item types should vary systematically with the intended level of the

test. Thus, in a multilevel test, such as the CAT, the younger levels should have relatively more distortions, and the higher levels should have more displacements.

The opposing effects of different types of transformations also have some possible implications for test validity. Since the linear logistic test model contains both person and item process parameters, it is possible to examine the unidimensionality of ability over items with different processing characteristics. Since most items in the current study involved either distortion or displacement transformations, but not both, two sets of items with different processing mechanisms may be readily formed. The homogeneity of the two item sets, with respect to ability, was supported, however, by a likelihood ratio chi-square test. Thus, the two item sets do not measure different abilities.

At first glance, the item homogeneity results seem inconsistent with the information structure results. If distortion and displacement items involve somewhat different processing mechanisms, as indicated by the information structure data, then it would seem that they should measure different abilities. Such support has been obtained for processing components on verbal aptitude test items (Whitely, in press). However, there are several reasons why items that depend on somewhat different processing mechanisms may not measure different abilities. First, individuals may not vary in their ability to perform some processing components. Some processing events may be essential in performance, such as marking the chosen response option on the answer sheet, but executed nearly error-free by all subjects. Second, abilities on some processing components may not be reliably measured. This occurs most frequently when the item set is extremely easy on a particular component, so that even a low ability person is highly likely to execute the component correctly. Third, abilities on theoretically distinct processing components may be so highly correlated that they may not define meaningful dimensions of individual differences.

In the current study, either of the last two sources could have led to item homogeneity. Distortion "abilities," if they do meaningfully exist, may not have been reliably measured because the distortion items were quite easy. Further, high positive correlations probably exist between the distortion and displacement processes if they could both be reliably measured, due to probable communalities in their cognitive development.

The failure to obtain sex differences in either test total scores or the information structure parameters indicates equivalent performance and processing mechanisms for males and females. Although sex differences on geometric analogies have been inconsistent in previous research, it is interesting to compare these results to the more well-supported sex differences in spatial visualization and rotation. Many geometric analogies in the current study involve spatial displacements, and sex differences could be expected. Since sex differences were not observed, however, it would seem that the role of possible individual differences in visualization and rotation is relatively trivial in the complex reasoning problem that is defined by geometric analogies.

In conclusion, solving geometric analogy items involves two different processing transformations, which have opposing effects on item difficulty. Although the data did not indicate that different abilities were involved in the two transformations, they do suggest how a test developer can design items explicitly for difficulty by controlling information structure.

References

- Cooper, L. A. Individual differences in visual comparison processes. *Perception and Psychophysics*, 1976, 19, 433-444.
- Egan, D. E. Testing based on understanding: Implications from studies of spatial ability. *Intelligence*, 1979, 3, 1-15.

- Evans, T. G. A program for the solution of geometric analogy intelligence test questions. In M. Minsky (Ed.), *Semantic information processing*. Cambridge MA: MIT Press, 1968.
- Fischer, G. H. The linear logistic model as an instrument in educational research. *Acta Psychologica*, 1973, 37, 359-374.
- Fischer, G. H., & Formann, A. K. *An algorithm and a FORTRAN program for estimating the item parameters of the linear logistic test model* (Research Bulletin No. 11). Vienna, Austria: Psychologisches Institut der Universität Wien, October 1972.
- Frederiksen, J. R. Components skills in reading: Measurement of individual differences through chronometric analysis. In R. E. Snow, P. Federico, W. E. Montague (Eds). *Aptitude, learning, and instruction* (Vol. 1). Hillsdale NJ: Erlbaum, 1980.
- Gustafsson, J.-E. *The Rasch model for dichotomous items: Theory, applications, and a computer program* (Research Report No. 63). Mölndal, Sweden: University of Göteborg, Institute of Education, December 1977.
- Just, M. A., & Carpenter, P. A. Eye fixations and cognitive processes. *Cognitive Psychology*, 1976, 8, 441-480.
- McGee, M. G. Human spatial abilities: Psychometric studies and environmental genetic, hormonal, and neurological influences. *Psychological Bulletin*, 1979, 6, 889-901.
- Mulholland, T., Pellegrino, J. W., & Glaser, R. Components of geometric analogy solution. *Cognitive Psychology*, in press.
- Nickerson, R. S. Binary-classification reaction time: A review of some studies of human information processing capabilities. *Psychonomic Monograph Supplements*, 1972, 4 (Whole No. 65), 275-318.
- Pellegrino, J. W., & Glaser, R. Cognitive correlates and components in the analysis of individual differences. *Intelligence*, 1979, 3, 187-214.
- Pomerantz, J. R., Sager, L. C., & Stoeber, R. J. Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Performance and Perception*, 1977, 3, 422-435.
- Rasch, G. On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics*. Berkeley: University of California Press, 1961.
- Shepard, R. N., & Metzler, J. Mental rotation of three-dimensional objects. *Science*, 1971, 171, 701-703.
- Sternberg, R. J. Component processes in analogical reasoning. *Psychological Review*, 1977, 31, 356-378. (a)
- Sternberg, R. J. *Intelligence, information-processing and analogical reasoning: The componential analysis of human abilities*. Hillsdale NJ: Lawrence Erlbaum, 1977. (b)
- Sternberg, R. J. The nature of mental abilities. *American Psychologist*, 1979, 34, 214-230.
- Thorndike, R. L., & Hagen, E. *Cognitive abilities test*. New York: Houghton-Mifflin, 1974.
- Whitely, S. E. Information-processing on intelligence test items: Some response components. *Applied Psychological Measurement*, 1977, 1, 465-476.
- Whitely, S. E. Multicomponent latent trait models for ability tests. *Psychometrika*, 1980, 45, 479-494.
- Whitely, S. E. Modeling aptitude test validity from cognitive components. *Journal of Educational Psychology*, in press.
- Whitely, S. E., & Barnes, G. M. The implications of processing event sequences for theories of analogical reasoning. *Memory and Cognition*, 1979, 1, 323-331.

Acknowledgments

This research was partially supported by the National Institute of Education Grant Number NIE-6-7-0156 to Susan E. Whitely, principal investigator. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Author's Address

Send requests for reprints or further information to Susan E. Whitely, Psychology Department, University of Kansas, Lawrence KS 66045.