The Evaluation of Cognitive Constructs Using Structural Equation Modeling

Morton Ann Gernsbacher & H. H. Goldsmith
University of Texas at Austin

Though, in anthropomorphic terms, the field of cognitive psychology has barely grown out of young adulthood, its development thus far has been markedly influenced by several of its sister sciences (e.g., theoretical linguistics, artificial intelligence, engineering science, philosophy). These influences are most noticeable within the structural models cognitive psychologists propose to describe various aspects of cognitive processes. This paper presents an approach to structural modeling that arose from work within the field of econometrics and has recently been successfully applied to describe behavioral processes within the fields of sociology and biometrical genetics. We believe this approach to be particularly well suited for modeling cognitive behavior because it helps to resolve several of the paradoxes involved with other methods of modeling.

Rather ironically, cognitive psychologists demonstrate less catholic tastes in the methodology they employ to verify their models than they do in outlining their models. Most cognitive psychologists collect behavioral data from classical 2 by 2 designs and then analyze these data with standard analyses of variance. Some potential problems arise when we attempt to extrapolate from two or four cells of means (or a series of such sets of data) to a more elaborate description of an underlying process. First, although we manipulate our experimental variables in a binary fashion, very often these variables occur naturally in a continuous distribution and it is this latter distribution we often imply in our models. Second, in order to obtain laboratory control, we systematically investigate the effects of different variables by pitting them in a series of one-to-one contrasts. Thus, the entire model is seldom tested simultaneously and we never gain an appreciation for the extent to which the model can account for all the phenomena it purports to explain. Third, the <u>degree</u> of the relationship between the theoretical constructs represented by the model and the variables used to measure these constructs is often left unquantified. For example, how well is the construct we call "long-term memory" indexed by the percentage of words correctly recalled after a two-week interval? In actuality, any time we test a theoretical model we are simultaneously testing the adequacy of a "measurement model." With the approach we shall describe today, this fact is made explicit. Moreover, the processes of model fitting and model testing are integrated into the same procedure.

Overview of Structural Equation Modeling

Structural equation modeling (SEM) is a comprehensive system for testing systems of linear hypotheses involving both observed, or "manifest," variables and theoretical constructs, or "latent" variables. Best known among the family of SEM approaches is the Linear Structural Relationships (LISREL) approach (Joreskog & Sorbom, 1978) which we used in the present demonstration. It is important to note that the SEM approach requires that the investigator have an explicit theory guiding his/her research and that all latent constructs in the theory be assessed, preferably with multiple convergent measures. The implementation of SEM involves three major steps:

(1) <u>Specification of the model</u>. The SEM approach requires the investigator to specify, <u>a priori</u>, <u>a model in</u> which <u>theoretical constructs</u> are hypothesized to be functionally related to observed variables. In cognitive psychology, these variables might be behavioral measures such as reaction time or performance accuracy on a laboratory task or they could be stimulus characteristics such as orthographic regularity or semantic meaningfulness as indexed by a normative scale. The theoretical constructs are such entities as lexical familiarity, memory span, or the structure of semantic categories. Furthermore, "causal" (or functional) relationships among the theoretical constructs must be specified. Thus, SEM can be thought of as a procedure that combines elements of traditional multiple regression, factor analysis, and path analysis.

(2) Model estimation. Once the model is specified, the values for the parameters in the model are simultaneously estimated using an iterative procedure, which in the case of LISREL is a maximum likelihood algorithm. The input for this analysis is a variance-covariance matrix for all of the observed variables measured in the study. The magnitude of three types of parameters can be estimated. Values for the hypothesized causal relationships can be interpreted as partial regression coefficients. Values for relationships that are not specified as being causal can be thought of as correlations. Finally, the residual, or unexplained variation, in the latent variables and in the manifest variables is estimated. The statistical significance of each of these parameters (with exceptions noted later) can also be computed.

(3) Evaluation of goodness-of-fit of the model. The statistical evaluation of the overall fit of the model to the data is a crucial element of the SEM approach, an advantage that distinguishes SEM from most other data analytic procedures. A chi-square statistic, with its associated degrees of freedom, is the most common indicator of the likelihood that the observed variance/covariance matrix could have emerged if the specified model were "true." The number of degrees of freedom in a model is the remainder when the number of parameters being estimated is subtracted from the unique number of observed variances and covariances. If this difference is negative, the model is, of course, not identified. Larger values of chi square, relative to the degrees of freedom, indicate a poorer fit of the model to the data. It is this goodness-of-fit evaluation that places SEM within the group of confirmatory, rather than exploratory, statistical procedures.

The Constructs of Category Structure and Category Verification

Over a decade of research in the field of cognitive psychology has been aimed toward investigating human semantic memory. The general consensus emerging from this body of work is that semantic memory is organized in a highly systematic and orderly fashion. One of the most commonly described units of organization within this store is the semantic category. Several principles have been proposed to underlie the organizational structure of such categories. The most popular of these are the principles of association frequency, semantic distance, featural overlap and typicality.

According to the principle of association frequency, membership in a semantic category is a function of the frequency with which a category member, such as ROBIN has been previously associated with a particular category concept, such as BIRD, and vice-versa. Many of you will recognize this principle as underlying many specimens of the very familiar breed of network models of conceptual knowledge. When speaking about semantic categories, we have simply substituted the term "category concept" for the term "superordinate" and the term "category member" for "subordinate." The principle of association frequency is usually assessed by collecting normative data upon the frequency with which subjects will mention a category member in response to a category name, and vice-versa.

A second principle proposed to underlie the structure of semantic categories is based purely upon degree of intra-category similarity. The general procedure used to assess this principle is to ask subjects to rate the similarity of pairs of members from a particular category. These data are then submitted to a multidimensional scaling procedure that places the category members in Euclidean space such that the metric distances between category members is inversely and monotonically related to their semantic similarity. And, according to this principle, category membership is a function of a member's scaled position within the multidimensional configuration revealed for that category.

A third principle that has been proposed to underlie the structure of semantic categories is that of typicality. Typicality simply refers to the degree to which each member of a category is believed to be a good exemplar of its category. For example, most subjects will rate a ROBIN as being a very typical member of the category BIRDS while a CHICKEN is usually rated to be much less so.

A fourth principle that has been proposed to underlie the structure of semantic categories involves the notion of features. Features are attributes or properties of a semantic concept. Though the possession of a feature by a concept can only be present or not present, semantic features themselves are believed to vary in a more continuous fashion in regard to their importance in defining category membership. For example, a feature of the category BIRDS such as "has feathers," might be more important in defining category membership, while a feature such as "perches in trees" might be less criterial. Given that each category member can also be described by a set of characteristic features, according to the principle of featural overlap, membership in a semantic category is a function of the number and type (more or less criterial) of categorically descriptive features shared between the category member and the category concept.

Clearly, each of these principle that have been proposed to underlie the structure of semantic categories implies the existence of a theoretical construct, namely Category Structure. Thus, in the nomenclature of SEM, Category Structure is a latent variable. Also common to each of these structural principles is an empirical prediction. Each principle predicts that Category Structure affects behavioral processes. The laboratory measure commonly used to evaluate this prediction is performance on a timed category verification task. In this task, subjects are presented with the name of a category member and asked to verify that it belongs to its appropriate category. Both the speed with which a subject responds (i.e, reaction time) and his/her accuracy of response (i.e., error rate) are provided. Thus, it has been proposed that measures of these four principles are related to the latent variable Category Structure. Category Structure, in turn, has been proposed to influence another latent variable, what we will call Category Verification. The latter variable is measured by reaction time (RT) and error rate. Here is a prime example of the proposal of implicit constructs and their causal relationships that requires evaluation in a simultaneous fashion.

In this study, we collected several independent sets of data upon one rather large sample of items. These items were eight semantic categories (viz., FRUITS, VEHICLES, FURNITURE, VEGETABLES, BIRDS, SPORTS, CLOTHING) and twenty each of their respective members. The data collection was arranged into two stages. In the first stage, measures of each of the four structural principles were obtained for the entire set of items in procedures identical to those employed by previous researchers. In the second stage, measures of performance (i.e., speed and accuracy) on the speeded category verification task were obtained, using the previously measured items as experimental stimuli. A different group of 50 undergraduates at the University of Texas participated in each aspect of the data collection. These subjects provided us with a data base composed of 24,000 measures of association frequency, 4826 measures of feature criteriality, 80,000 measures of feature possession, 76,000 measures of semantic distance, 800 measures of typicality, 12,750 RTs (with the effects of word length removed), and 136 erroneous responses. These data were reduced to a more manageable 6 X 6 (four structural principles and two performance measures) correlation matrix that we proceeded to analyze using the SEM approach.

Application of Structural Equation Modeling

From the theoretical guidance outlined above, the model in Figure 2 has been specified, estimated, and evaluated. Before discussing the model itself, we should clarify the notation used in Figure 1. Manifest variables are represented by rectangles, latent variables are depicted as circles, the direction of causal or functional relationships is specified by arrows, and unexplained relationships are denoted by curved lines. Each figure that we could draw using the symbols in Figure 1 and the conventions of path analysis specifies a series of linear equations that are simultaneously solved by the LISREL procedure. For the analysis presented here, we used the RAM parameterization (McArdle & McDonald, 1981) of version IV of the LISREL program (Joreskog & Sorbom, 1978).

Let us now consider each of the portions of the model. On the left side are the manifest variables hypothesized to represent different, but correlated, aspects of the latent variable Category Structure. In one sense, we can think of Category Structure as a factor and the standardized partial regression coefficients .69, -.57, .82, and .70 as factor loadings for the 4 observed variables. Note that these coefficients are high and that the residual variance in each manifest variable is low. The negative value for semantic distance is simply due to the fact that the Multidimensional Scaling program (ALSCAL) used to derive the measure scales dissimilarities rather than similarities.

The two undirected relationships at the extreme left of Figure 1 represent unexplained associations between residual variation between Semantic Distance and Typicality and between Typicality and Featural Overlap. The existence of these weak but statistically significant relations means that there is systematic covariance within these pairs of manifest variables that is not common to the other two variables.

Switching our attention to the right side of Figure 1, we find the measurement model for Category Verification. Gernsbacher (1982) has empirically demonstrated that a combination of Reaction Time and Error Rate is a much more comprehensive measure for evaluating performance in many speeded cognitive paradigms than either of the two measures alone. Even the low error rate observed in these data (viz., mean percentage of error = 5%) contributes substantially to the Category Verification latent variable. We do notice, however, that RT is the stronger contributor.

The key prediction of the model is that Category Structure bears a functional relationship to (or "causes") performance on the Category Verification task. Thus, we examine the directed path in the center of Figure 1. The standardized partial regression coefficient of .95 shows that the influence of Category Structure on verification performance is strong indeed! This rather startling degree of predictability from a collection of paper-and-pencil measures completed in a classroom to choice RT performance obtained under highly standardized laboratory conditions alerts us to the potential power of the SEM approach. Compare this regression coefficient of .95 to the zero order Pearson product-moment correlations of .56, -.49, .69, and .49 between, respectively, Association Frequency, Semantic Distance, Typicality, and Featural Overlap, and RT in these same data.

There are some other noteworthy points concerning the estimation procedure. The sets of directed relations on both sides of Figure 1 are simultaneously derived so as to maximize the predictability of Category Verification from Category Structure. Thus, we can conclude from the fact that Typicality is the principle most strongly related to Category Structure that Typicality also is the best single predictor of Category Verification. From the rules of path analysis, we know that the magnitude of this prediction is .82 X .95 = .78. Still employing the rules of path analysis, we can account for all of the standardized variance in, say RT, by computing (.85 X .85) + .28 = 1.00.

Our next task is to evaluate the adequacy of the model as a whole. As specified in Figure 1, the model does fit the data rather well. In addition, the first-order derivatives (supplied by LISREL-IV) for each of the potential parameters of the model are uniformly low, thus indicating no local areas of lack of fit in the model.

In many cases, the chi square statistic may lead to rejection of the model. The LISREL-IV program provides information on the loci of lack-of-fit that permits one to change the model to improve fit. Ideally, the investigator would next collect fresh data and attempt to confirm the revised model. In actual practice, the model is often "fixed" to improve fit based on preliminary attempts to fit the model. In fact, the three undirected relations in Figure 1 were added to the model in this fashion.

Within the constraints of our data, we can pit rival structural equation models against one another. The result is a test of the relative ability of the two alternate theories to account for the observed covariation. One model we were interested in specified that the four categorization principles should contribute equally to the Category Structure latent variable, i.e., that the four partial regression coefficients be constrained to be equal during the maximum likelihood estimation process. When we evaluated this model, we found a chi square of 51.88 with 10 df. In an opposing model, all conditions were equivalent, except that the four principles of categorization were allowed to vary freely in their estimated contribution to the latent variable. This latter model yielded a chi square of 42.56 with 8 df. Neither model fits very well, but achieving optimal fit is not the purpose of this comparison. A statistical comparison of the two models shows that X diff = 9.32 with 2 df, a significant difference (p = .01). This comprehensive test of models offers evidence that the four principles are *not* equivalent measures of Category Structure.

Concluding Remarks

One goal of this paper has been to convey something of the potential of the SEM approach for model building and testing within the realm of cognitive psychology research. In attempting to do this, we have underemphasized the difficulties of the approach. Perhaps it is appropriate to conclude with some cautions, that are discussed more fully by Horn and McArdle (1980). Given that SEM, like any multivariate maximization procedure, capitalizes on chance relationships in the data, replication of complex findings is mandatory. Issues of identification can be quite intractable, so much so that studies not initially conceptualized with SEM in mind are often unsuitable for SEM analysis. On the other hand, SEM has potentials we have not explored; for example, interactive terms can be entered into the systems of equations and multiple groups of subjects or items can be analyzed simultaneously. On balance, the approach is worth the attention of a discipline that utilizes complex models and is in need of methods for testing them comprehensively.

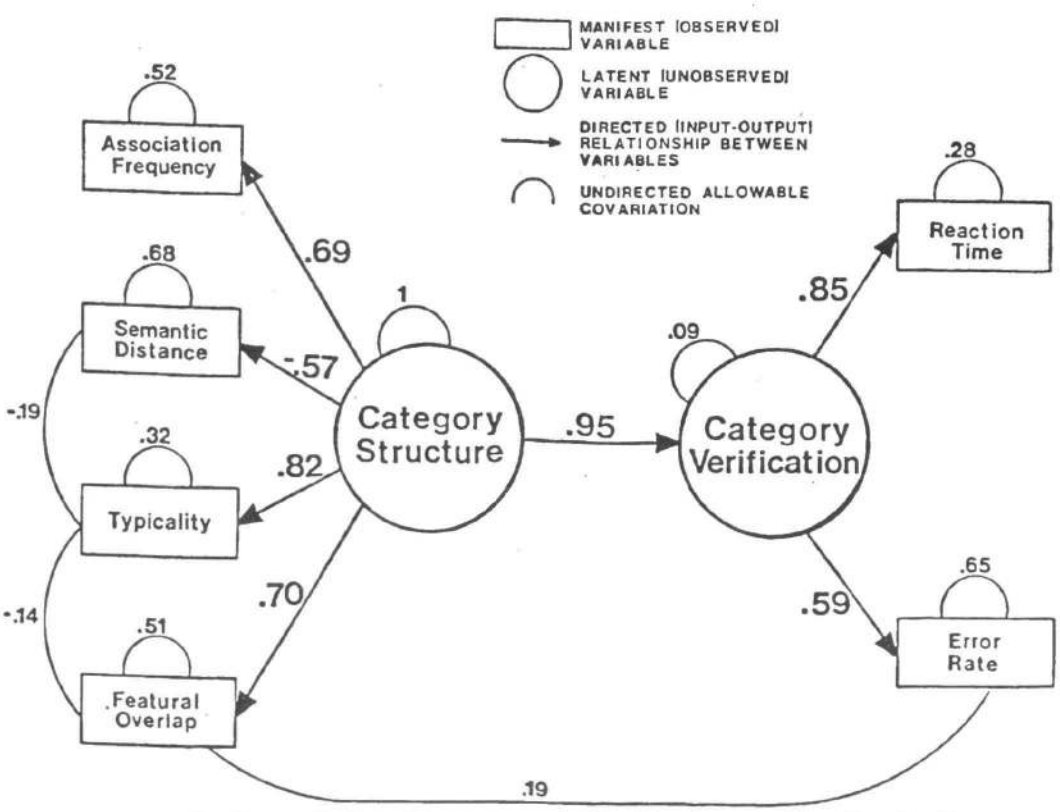
Gernsbacher, M. A. *On the use of a "new" performance variable to measure cognitive processing.* Manuscript submitted for publication, 1982.

Horn, J. L., & McArdle, J. J. Perspectives on mathematical / statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), *Aging in the 1980's: Psychological issues.* Washington, D. C.: American Psychological Association, 1980.

Joreskog, K. G., & Sorbom, D. *LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood.* Chicago: National Educational Resources, 1978.

McArdle, J. J., & McDonald, R. P. *A simple algebraic representation for structural equation models.* Unpublished manuscript, University of Denver, 1981.

FIGURE 1. A Structural Equation Model of the Relationship Between Category Structure and Category Verification



Goodness-of-fit: chi square = 3.89, df = 5, p = .57. Standardized estimates are shown. All values are at least twice the size of their standard errors.