

# Path Analysis

Joseph R. Duffy

Clinicians interested in neuropathologies of communication often strive to explain the behaviors they observe, diagnose, and treat. Developing these causal explanations is difficult because the *ex post facto* nature of much of our research results in a loss of experimental control, leaving us with design problems and a reliance on statistical control. In addition, many of the behaviors we wish to explain are probably a function of multiple causes. Also, the variables that may contribute to a phenomenon might not do so independently; causal variables may interact with each other as well as with their effects.

This paper will discuss path analysis, a strategy for examining causal relationships. Although it has been used for years in fields as diverse as econometrics and biology, there are few examples of its application to neurogenic communication disorders or even broader issues in speech pathology or neuropsychology. In spite of this neglect, path analysis is a potentially valuable tool for investigating neuropathologies of communication. This paper addresses the conceptual underpinnings and steps involved in using path analysis, applications of path analysis to aphasia research, and some of the major advantages and potential misuses of path analysis. Because path analysis employs correlation and regression, which have already been discussed (Tompkins, 1993), this paper focuses primarily on conceptual rather than statistical issues.

## INVESTIGATING MULTIPLE-CAUSE BEHAVIORS

Studies of neurogenic communication disorders generally recognize that abnormal behaviors are often caused by more than one factor. But such studies tend to simplify the investigation of causal variables by examining the effects of only one of them at a time. The results from studying a single factor's effect are then fit back into the framework of a more complete, but

usually vaguely stated, multivariate explanation, often with cautions against drawing causal inferences from the data. If causal inferences are drawn, the fact that the single causal variable that was studied did not interact with other causal variables in the data analysis may be ignored.

This kind of hunt-and-peck method of theory building can be useful, but offers less to our attempts to explain some behaviors than approaches that carry the admission of complex causal relationships into the research study itself. The inclusion of more than two variables (more than one causal hypothesis) captures the complexity of real interrelationships. Path analysis helps us do this.

## **BASIC CONCEPTS OF PATH ANALYSIS**

Path analysis is a type of causal modeling (or structural equation modeling) for investigating postulated relationships among variables. (For more complete conceptual and procedural introductions to path analysis and structural equation modeling, see Duffy, Watt, & Duffy, 1981; Francis, 1988; and for more in-depth information see Blalock, 1964, 1969, 1971; Duncan, 1971, 1975; Heise, 1975; and Kenny, 1979.) Strictly speaking, path analysis "is neither a statistical procedure nor an experimental design, and under no circumstances does it ever prove causality" (Duffy, Watt, & Duffy, 1981). Unlike procedures such as correlation and multiple regression, which examine only covariance between variables, path analysis assumes that several conditions for defining a causal relationship are met. Among the most important of these are the assumptions that the occurrence of one event is sufficient for the occurrence of a later event, and that the cause and effect variables covary so that a change in the level of the cause variable alters the effect variable. Finally, it is essential to specify a mechanism by which the cause produces the effect; that is, there must be a rationale for the causal relationship.

## **APPLYING PATH ANALYSIS TO APHASIA RESEARCH**

### **Example 1**

Path analysis has been used in aphasia research. This example of its application illustrates the steps and procedures associated with its use.

Many investigators have been interested in explaining why aphasic individuals often have problems with pantomime recognition and expres-

**TABLE 1. CORRELATION COEFFICIENTS\* (PEARSON R) AMONG MEASURES OF APHASIA, INTELLECTUAL IMPAIRMENT, LIMB APRAXIA, VISUAL PROCESSING DEFICITS, AND PANTOMIME RECOGNITION AND EXPRESSION DEFICITS FOR 45 APHASIC PATIENTS (DATA FROM DUFFY, WATT, & DUFFY, IN PREPARATION)**

	<i>Intellectual Impairment</i>	<i>Limb Apraxia</i>	<i>Visual Processes</i>	<i>Pantomime Recognition</i>	<i>Pantomime Expression</i>
Aphasia	.49	.55	.47	.65	.81
Intellectual Impairment		.46	.39	.44	.47
Limb Apraxia			.45	.41	.70
Visual Processes				.65	.61
Pantomime Recognition					.68

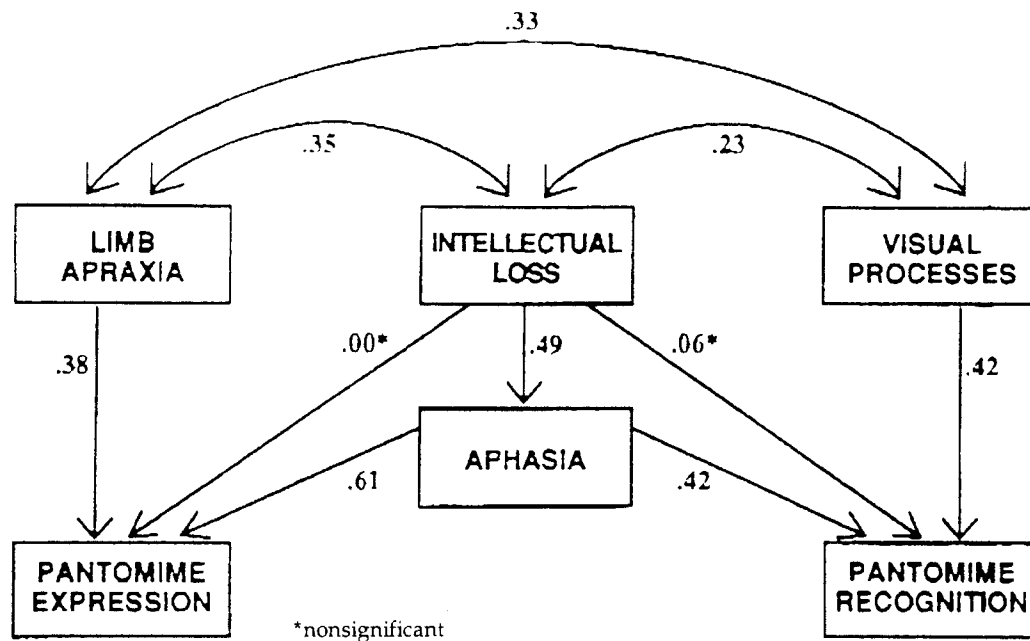
\*All correlations are significant ( $p < .001$ )

sion. A number of hypotheses exist. One is tied to input-output deficits that can occur with aphasia. For example, impairment of visual processes may lead to pantomime recognition deficits and limb apraxia may lead to pantomime expression deficits. Another theory states that pantomime deficits result from a general impairment of intelligence. A third hypothesis states that pantomime deficits result from a central symbolic deficit that simultaneously affects verbal and nonverbal symbolic abilities (for a more complete explanation of these theories, see Duffy, Watt, & Duffy, 1978; Duffy & Duffy, 1981, 1990).

Studies examining these theories often have used correlation analysis to examine the relationship between the hypothesized cause variables and the effect variables of pantomime expression and recognition. The hope was that *one* hypothesized causal factor would be strongly related to pantomime ability and that the other factors would not. These hopes were not realized. Ignoring the exact measures used by Duffy, Watt, and Duffy (in preparation) in a study of 45 aphasic patients, it turns out that all 15 of the correlations among the variables in Table 1 are significant. The magnitude of these correlations are typical of those found in similar investigations.

Path analysis is ideally suited to the untangling of these complex interrelationships.

**Step 1.** The first step in path analysis is to represent hypothesized relationships in what is known as a path diagram. The path diagram (causal



**Figure 1.** Model 1 of Pantomime Deficits (from Duffy, Watt, & Duffy, in preparation).

model) is an explicit representation of the nature of the relationships among the variables. Competing theories generate different path diagrams. Testing the plausibility of different theories makes path analysis valuable. If one model or theory is found to be plausible and others are not, the results can be considered as evidence in favor of one theory over the others (Duffy, Watt, & Duffy, 1981).

Three types of relationships can be specified in a path diagram: causal, unanalyzed, and null. A causal relationship is represented by a straight line arrow pointing from the cause to the effect; it can be direct or indirect. In Model 1 (Figure 1), depicting pantomime deficits, intellectual loss directly leads to aphasia and to pantomime expression and recognition deficits; it also indirectly causes pantomime expression and recognition deficits through its effects on aphasia, which in turn directly causes pantomime recognition and expression deficits. The model also states that pantomime expression is caused by limb apraxia and that pantomime recognition is caused by visual processing deficits.

There are also unanalyzed relationships, represented by the curved double-headed arrows among limb apraxia, intellectual loss, and visual processing deficits. Unanalyzed means that there is ambiguity about a relationship. For example, in Model 1 it is posited that visual processing deficits and limb apraxia are related to one another, but it is uncertain whether the relationship is causal or spurious.

Finally, there are null relationships, in which it is hypothesized that no theoretical linkage exists between variables. In Model 1, no relationships between aphasia and limb apraxia or between aphasia and visual processing deficits are assumed.

**Step 2.** The next step is to measure the strength of the relationships along the paths between the variables. Partial correlations generate the coefficients for the unanalyzed paths. For example, in Model 1 the path between visual processing deficits and limb apraxia is best estimated by computing their correlation while holding constant variables that might intervene or influence their relationship; thus, the path was estimated through a partial correlation between them while holding constant intellectual loss.

Regarding the causal paths, when there is more than a single causal path to a dependent variable, they are estimated by beta weights from multiple regression analyses. For example, the causal paths from limb apraxia, intellectual loss, and aphasia to pantomime expression in Model 1 are the beta weights of a multiple regression analysis of limb apraxia, intellectual loss, and aphasia regressing on pantomime expression. When there is only a single causal path, a simple zero-order correlation is the coefficient; in Model 1, the correlation between intellectual loss and aphasia was the best estimate of their causal relationship.

Each of the paths can be tested for statistical significance. In Model 1, all paths were significant except those between intellectual loss and pantomime expression and between intellectual loss and pantomime recognition.

**Step 3.** Thus far we can only compare the magnitude of the various paths. The next step in path analysis is crucial because it helps establish the plausibility of the model in terms of the observed data.

The plausibility of a model can be examined in several ways. First, the strength of the estimated coefficients can be examined. In Figure 1, the paths from intellectual loss to pantomime expression and from intellectual loss to pantomime recognition are nonsignificant. These findings suggest that the model is not plausible, because two of the hypothesized causal relationships fail to meet one of the conditions for a causal relationship—covariance.

We can also examine the model's capacity to accurately represent the relationships actually observed among the variables testing its ability to correspond to the facts by regenerating the original bivariate correlations. This process involves estimating the total relationship between two variables by adding together all of the ways in which they interact in the model. The sum of these interactions should approximate the original bivariate correlation. To illustrate, the original correlation between intellectual loss and aphasia was .49 (Table 1), which is the same as the value in Model 1, so the theory of their relationship corresponds to the fact of the

original correlation. In contrast, the total effect of intellectual loss on pantomime expression in Model 1 is the sum of the path from intellectual loss to pantomime expression, which was 0, plus the effect of intellectual loss on pantomime expression through aphasia. This latter effect is estimated by the product of the intellectual loss to aphasia path and the aphasia to pantomime expression path; in other words,  $.49 \times .61$ , which equals .30. The sum of the influences ( $0 + .3 = .3$ ) is significantly less than the original correlation of .47, so this aspect of the model must be rejected because it did not accurately regenerate the original correlation. In fact, 5 of the 15 original correlations were not accurately approximated in the analysis of Model 1's adequacy.

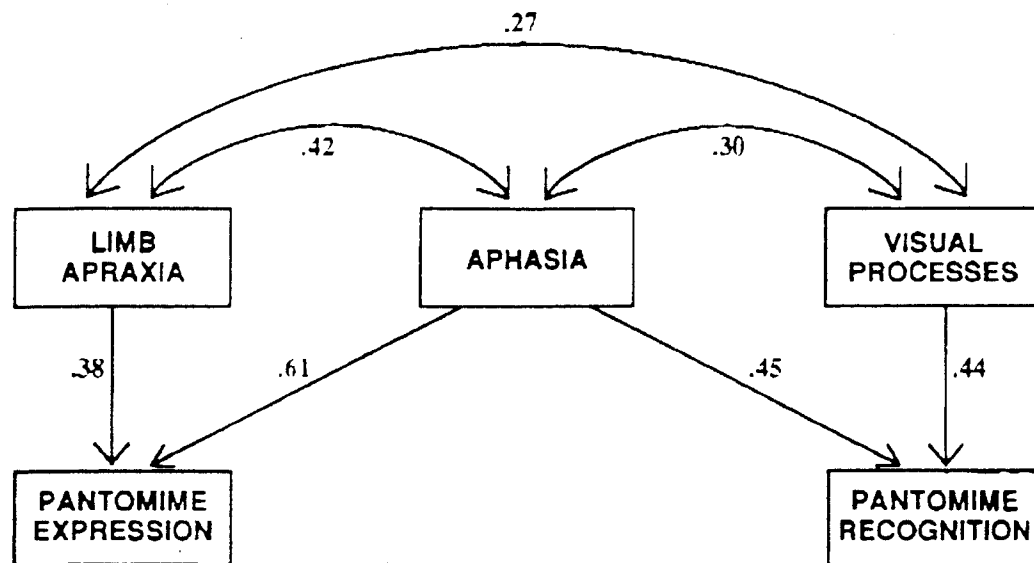
Finally, the overall goodness of fit of the model can be estimated. This is accomplished by computing the average ability of the paths in the model to regenerate the original correlations, expressed as a percentage of explained variance. For Model 1, the overall goodness of fit was 86%; not bad, but we would reject the model on the grounds already discussed.

**Step 4.** The next step in model testing is to assess alternate explanatory models. With the pantomime data, four additional models were assessed (Duffy, Watt, & Duffy, in preparation); three were found to be implausible for reasons similar to those for Model 1.

In the one plausible model (Model 2 in Figure 2), intelligence was removed because it had weak support from the rejected models and because theoretical rationales justify its elimination. Limb apraxia, aphasia, and visual processing deficits were viewed as co-occurring deficits that have unanalyzed relationships with one another. Pantomime expression deficits were viewed as directly caused by limb apraxia and aphasia, and pantomime recognition deficits were viewed as directly caused by aphasia and visual processing deficits.

The evaluation of Model 2's plausibility found that all of the path coefficients were significant; all of the relevant original correlations were regenerated by the model; the goodness of fit for the model was 93%. Finally, the model's explanatory power was fairly good, although by no means perfect—the variables in the model accounted for 76% of the variance in pantomime expression and 58% of the variance in pantomime recognition. These percentages suggest that the model might be improved by adding variables, further modifying the specified relationships, or improving the reliability or validity of the variables.

Through path analysis, four of five alternative pantomime-deficit models were rejected. The one plausible model (Model 2) suggests that pantomime deficits are the result of a central symbolic deficit as well as specific motor and visual processing dysfunctions. The path analysis process did not prove that the plausible model is correct or that it is even the best of all possible models. It did support the conclusion that the model was the only one of the five tested that proved to be plausible.



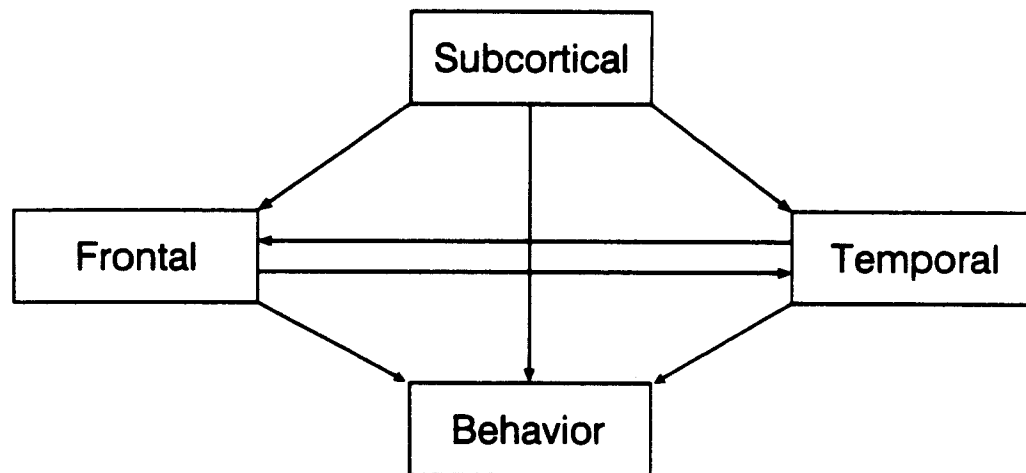
**Figure 2.** Model 2 of Pantomime Deficits (from Duffy, Watt, & Duffy, in preparation).

## Example 2

Is path analysis applicable only to the analysis of pantomime data? Metter et al. (1988) have demonstrated otherwise. They asked whether subcortical structural damage directly causes certain behavioral symptoms in aphasia or whether it does so indirectly through remote effects on cortical metabolic function. CT and PET scan data were used as indices of structural and metabolic dysfunction, and several measures on the Western Aphasia Battery were used as indices of aphasic behavior. Verbal fluency is the aphasic behavior that will be used here to illustrate Metter et al.'s use of path analysis.

A path diagram (see Figure 3) made their theory explicit. Subcortical damage was viewed as having a direct effect on aphasic behavior (fluency) as well as indirect effects through its effect on both frontal and temporal metabolism. Temporal and frontal metabolism were postulated to have a reciprocal relationship with one another.

Computation of the appropriate correlations and regressions by Metter et al. (1988) showed a significant direct path from the subcortical lesion to fluency, subcortical lesion to frontal metabolism, and frontal metabolism to fluency. The paths from subcortical lesion to temporal metabolism and from temporal metabolism to fluency were not significant. These results, among others, made plausible a conclusion that subcortical lesions have a direct effect on fluency plus an indirect effect on fluency through its remote



**Figure 3.** A model of the relationships among subcortical lesions, frontal metabolism, temporal metabolism, and aphasic behavior (fluency) (adapted from Metter et al., 1988).

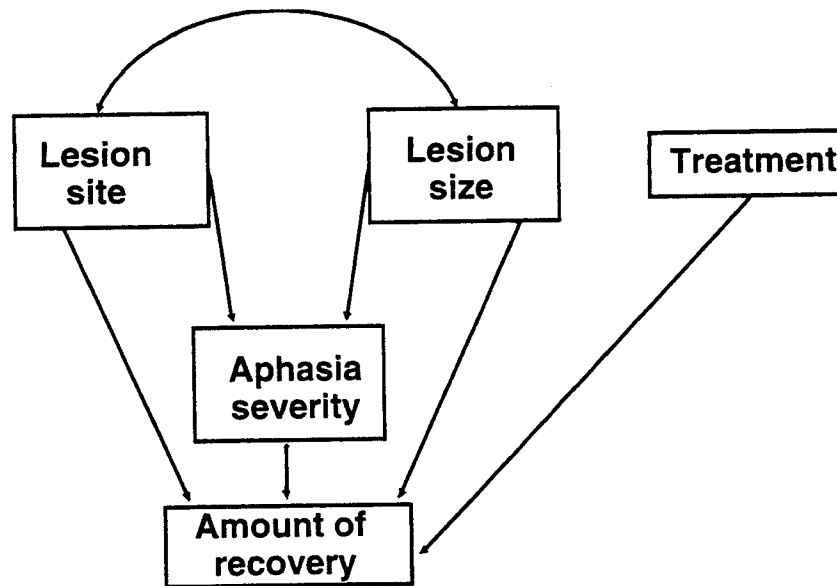
effect on frontal lobe metabolism. Path analysis was very useful to the investigation of the relationships among lesions, their remote metabolic effects, and aphasia.

### Example 3

Many clinicians are interested in the contribution of various factors to the amount of recovery from aphasia. Variables may include lesion size and site, early severity of aphasia, and treatment.

Instead of investigating these variables' independent relationship to recovery we might test a model that reflects the more likely complex interrelationships among them. For example, the path diagram in Figure 4 indicates that initial aphasia severity directly determines amount of recovery. Lesion size has a direct effect on recovery amount plus an indirect effect through its contribution to initial aphasia severity. Similarly, lesion site directly affects recovery and indirectly affects it through its contribution to initial aphasia severity. The relationship between size and site of the lesion is considered unanalyzed, and the relationship between treatment and all other variables—except amount of recovery—is considered null. Using path analysis in this instance might provide some insight into the mechanism that relates each variable to recovery amount as well as the degree to which the variables and their interactions are important to prognosis. A number of alternative models (simpler or more complex) also could be tested.





**Figure 4.** Hypothetical model of the relationships among lesion site, lesion size, initial aphasia severity, treatment, and amount of recovery from aphasia.

## CAVEATS

Certain dangers accompany the use of path analysis. First, because path analysis is designed to examine causal relationships, there is a special danger of believing that the model has described reality and has *proven* the existence of causal relationships, which is not the case. Path analysis only evaluates the plausibility of relationships specified by a particular theory. We can only hope that the model is sufficient for the purposes of the study and the practical uses to which its results will be put.

It is also important to recognize that plausible and implausible models can vary as a function of subject characteristics as well as the validity and reliability of measures generating the data. For example, the plausible model (Model 2, Figure 2) for explaining pantomime deficits was based on data from a sample of unselected aphasic patients. It is possible that the model would not have been plausible if subjects were more narrowly selected on the bases of aphasia severity or type, or lesion site, for example. Similarly, intellectual deficit was basically rejected as a plausible explanation for pantomime deficits (Model 1, Figure 1). However, our measure of general intelligence was the Raven's Progressive Matrices (Raven, 1960). The cognitive processes measured by the Raven's are very different from those measured by other tests, which might generate a more significant role for nonlinguistic cognitive functioning in our models (for further discussion of these issues, see Duffy & Duffy, 1990).

## ADVANTAGES OF PATH ANALYSIS

Duffy, Watt, and Duffy (1981) have discussed the following strengths and advantages of path analysis:

1. The path diagram requires explicit specification of the model to be tested. This facilitates communication of well-defined theory. Explicitly laying out a theoretical model forces us to be consistent and comprehensible in our theoretical statements. This enables criticism to be explicit and sharply focused.
2. The path diagram is isomorphic with the statistical properties of the postulated system of variables (Duncan, 1971)—that is, theory construction and data analysis are specified simultaneously. It is possible to look at a path diagram and know the exact statistical computations necessary to generate path coefficients and to evaluate the model's adequacy.
3. Path analysis is an operational way to distinguish between true and spurious correlation. It represents a "disciplined compromise between oversimplified explanations of obviously complex problems and the scientifically crippling effects of a mystical or everything-causes-everything-else philosophy" (Duffy, Watt, & Duffy, 1981).

Using path analysis to test alternative explanations for behavior potentially helps build a web of circumstantial evidence about the nature of complex relationships in neurogenic communication disorders. It is certainly not the answer to all of the problems we face when investigating causal explanations, but it is a useful tool for expressing some of our theories and evaluating their adequacy.

## ACKNOWLEDGMENT

The author would like to acknowledge the contributions of James H. Watt and Robert J. Duffy to his understanding of path analysis. This paper could not have been written without the benefit of their creativity, collaboration, and guidance during our research efforts together.

Note: Computational procedures for path analysis can be time-consuming. A path-analysis software program called PATHEVAL is available for academic or nonprofit use for a reproduction fee. For further information, contact Information Analysis Systems, Corp. 2, Brookside Lane, Mansfield Center, CT 06250.

## REFERENCES

- Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.
- Blalock, H. M. (1969). *Theory construction*. Englewood Cliffs, N.J.: Prentice-Hall.
- Blalock, H. M. (1971). *Causal models in the social sciences*. Chicago: Aldine.
- Duffy, J. R., Watt, J., & Duffy, R. J. (1978). Deficits of nonverbal communication in aphasic patients: A test of some current causal theories. In R. H. Brookshire (Ed.), *Clinical Aphasiology Conference Proceedings* (Vol. 8, pp. 153-163) Minneapolis, MN: BRK Publishers.
- Duffy, J. R., Watt, J., & Duffy, R. J. (1981). Path analysis: a strategy for investigating multivariate causal relationships in communication disorders. *Journal of Speech and Hearing Research*, 24, 474-490.
- Duffy, R. J. & Duffy, J. R. (1981). Three studies of deficits in pantomimic expression and pantomimic recognition in aphasia. *Journal of Speech and Hearing Research*, 24, 70-84.
- Duffy, R. J. & Duffy, J. R. (1990). The relationship between pantomime expression and recognition in aphasia: the search for causes. In G. E. Hammond (Ed.), *Cerebral Control of Speech and Limb Movements*. Amsterdam: North-Holland.
- Duffy, R. J., Watt, J. H., & Duffy, J. R. (in preparation). An investigation of the causes of pantomimic deficits in aphasia using path analysis.
- Duncan, O. D. (1971). Path analysis: sociological examples. *American Journal of Sociology*, 72, 1-16.
- Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.
- Francis, D. J. (1988). An introduction to structural equation models. *Journal of Clinical and Experimental Neuropsychology*, 10, 623-639.
- Heise, D. R. (1975). *Causal analysis*. New York: Wiley.
- Kenny, D. A. (1979). *Correlation and Causality*. New York: Wiley & Sons.
- Metter, E. J., Riege, W. H., Hanson, W. R., Jackson, C. A., Kempler, D., & van Lanken, D. (1988). Subcortical structures in aphasia: an analysis based on (F-18)-fluorodeoxyglucose, positron emission tomography and computed tomography. *Archives of Neurology*, 45, 1229-1234.
- Raven, J. C. (1960). *Guide to the standard progressive matrices*. London: H. K. Lewis and Co.
- Tompkins, C. A. (1993). Using and interpreting linear regression and correlation analysis: Some cautions and considerations. In M. Lemme (Ed.), *Clinical Aphasiology* (Vol. 21, pp. 35-46) Austin, TX: PRO-ED.