Structural equation modeling (SEM) is a relatively new research method for analyzing multivariate data. This article provides an introduction and overview of what SEM is and identifies some potential uses of this technique for occupational therapy researchers. The application of SEM in the areas of test construction and treatment effectiveness is highlighted, some cautionary measures related to this technique are presented, and resources that provide more detailed and advanced reading on the subject are included.

Rehabilitation is a complex process involving the interplay of a multitude of humanistic and environmental variables that are intricately intertwined. Concepts of function and dysfunction are likewise multidimensional. SEM, therefore, offers promise as a method particularly useful in theory-based research and in the development of test instruments.

The most popular computer program for performing SEM, LISREL (linear structural relations), was developed in the early 1970s (Joreskog & van Thillo, 1973) and has since been updated (Joreskog & Sorbom, 1993). Bender (1985) developed the EQS program, among many others, which is now available.

What Is SEM?
SEM is firmly in the tradition of empirical science. Differences in human behavior are often described by unobserved variables, such as motivation, quality of life, disability, intelligence, and independence. In SEM, these constructs are called latent variables. An example of a latent variable related to occupational therapy is the concept of purposeful activity. In SEM, this theoretical construct would be defined by several observed variables known as indicators or manifest variables. Purposeful activity might be defined, for example, by assigning values related to the importance of the activity for a given person in the areas of productivity, leisure, and health maintenance.
Although the correlation between two variables does not imply a causal relationship between them, a causal relationship between two variables does imply a correlation between them. This is the underlying premise of SEM. Although SEM does not prove a causal hypothesis, it gives us a way to judge how consistent the causal model is with observed data (Moore, 1995).

For clarity, it may be helpful to differentiate path analysis and CFA, which both may be subsumed under SEM. Path analysis examines the causal links among observed variables. Crowley and Fan (1997) described path analysis as a technique using multistep regression analysis. Correlations among variables are examined by breaking them into four components: (a) direct effects, (b) indirect effects, (c) unanalyzed (due to correlated cause), and (d) spurious (due to common causes). Unfortunately, the major weaknesses of this approach are that all variables are assumed to be measured without error and that the technique lacks the statistical mechanisms to test the fit between the hypothesized path model and empirical data (Crowley & Fan, 1997).

CFA examines causal relationships between observed variables and the factors (or latent variables) with which they are associated. Therefore, CFA concerns itself with only the measurement model portion of the general SEM. CFA was the technique I used in an article published in this issue of the American Journal of Occupational Therapy (AJOT) (Mulligan, 1998), and it is useful for examining the construct validity of measurement instruments (see Crowley & Fan, 1997). Finally, the full structural model examines the causal links both between observed variables and latent variables and among the latent variables themselves.

A graphic representation is one of the easiest ways to explain SEM. Figure 1 presents a hypothetical full SEM model. Conventionally, an oval represents a latent variable, and a rectangle represents an observed variable. A straight line with an arrow represents the hypothesized effect that one variable has on the other, whereas curved lines with arrows on both ends represent hypothesized correlations between two variables without implying any causal relationship. As indicated in Figure 1, the SEM model can be viewed as having two

measurement models and a structural model.

However, like other statistical techniques, appropriate application involves meeting some basic assumptions. First, there are a few statistical assumptions inherent in the estimation method used. The most widely used estimation procedure in SEM is maximum likelihood estimation (MLE) (Stevens, 1992). The value of MLE is that this procedure accounts for the measurement error in the SEM analysis. MLE assumes that no single variable or group of variables perfectly explains another in the data set and that the observed variables have a multivariate normal distribution (Stevens, 1992). Therefore, models with observed variables that correlate at or above .90 cannot be estimated. MLE is fairly robust to the multinormality assumption, which is one of the reasons it is the most popular estimation procedure used. However, when sample sizes are small, dichotomous or categorical data are used, or the model is complex, MLE is particularly sensitive to violations of multinormality (Norris, 1997). Byrne (1994) described some methods for handling nonnormal data. It is assumed that the covariance matrix (not the correlation matrix) is being analyzed because the statistical theories for the MLE procedure (and others) were developed on the basis of the covariance matrix (Crowley & Fan, 1997).

Second, sample size must be considered because it is assumed in SEM that samples are so large that they approach infinity (Bollen, 1989). Generally, the larger the sample, the more stable the parameter estimates. Although researchers in the field do not necessarily agree on what is a sufficiently large sample, Boomsma (1987) suggested that 200 cases is a bottom-line number. It has been suggested that a minimum of 5 to 10 cases per estimated parameter is necessary (Floyd & Widaman, 1995). This requirement is a limiting factor in the use of this technique in occupational therapy research because it is often difficult to obtain large clinical samples. Sample size is related to power because the larger the sample, the more power the test will have. Consequently, when large samples are used, minor discrepancies between sample data and theoretical models may be found to be significant and, therefore, cause a researcher to reject a model with reasonably good fit. More thorough discussions related to issues of power are provided by Kaplan (1995); MacCallum, Browne, and Sugawara (1996); and Norris (1997).

Test Construction

My study in this issue of AJOT (Mulligan, 1998), an example of CFA, examined the construct validity of the Sensory Integration and Praxis Tests (SIPT) (Ayres, 1989). In the past, the term factor analysis has been used to represent exploratory factor analysis (EFA); however, with the increase use of CFA, it is important to make a theoretical distinction between EFA and CFA (Crowley & Fan, 1997).

In EFA, one wants to explore the empirical data to discover and detect characteristic features and interesting relationships without imposing any specific model on the data. EFA is a data-reducing technique that often strives to detect hypothetical constructs that can be measured by some observed variables. For example, during the development of SIPT, Ayres (1989) found through EFA that scores on the space visualization, figure-ground perception, design copy, and motor accuracy subtests loaded together on a factor that she identified as visuopraxis. Exploratory procedures are typically conducted in the early phases of test development or experimentation and are useful for generating theoretical models. One of the pitfalls of using exploratory procedures is that the models derived from them are data driven. Therefore, they are greatly influenced by the particular sample used, and replication is essential to validate them.

In CFA, the model is specified a priori (before analysis begins) and is formulated on the basis of theory or previous research. In other words, in the previous example, the researcher would have predicted that these four test items were related to one another and specified those relationships in the model. The model is then subjected to an analysis with empirical data to determine whether the model "fits" or is consistent with the data.

I tested a model of sensory integration dysfunction that was derived from many previous EFAs and that had been identified by experts in the field (Mulligan, 1998). Certain SIPT subtests were grouped together to form higher-order constructs, and then it was hypothesized that the proposed model would fit the empirical data. Examination of the parameter estimates, the chi-square statistic, goodness-of-fit measures, and standardized residuals assisted in determining whether the hypothesized model fit the data. A detailed review of these measures is beyond the scope of this article, but reviews of them can be found elsewhere (Bentler, 1990; Bentler & Bonnett, 1980; Bollen & Long, 1992, 1995; Croweley & Fan, 1997; Hu & Bentler, 1995; Joreskog & Sorbom, 1993; Marsh, Balla, & MacDonald, 1988; Muliak et al., 1989; Norris, 1997).

Because the model in CFA is theory driven, the model construction is not affected by the particular sample used in the study. Another advantage of using CFA is that the computer generates modification indices, which suggest changes to the model that would improve the fit of the data. Examples of CFA used for similar purposes are increasing and are evident in many disciplines, including psychology (Chi & Duda, 1995; Gleaves & Eberenz, 1995; Harris, 1995) and special education (McGrew, Bruininks, Thurlow, & Lewis, 1992).

The study by McGrew et al. (1992), for example, examined models of community adjustment for young adults with mental retardation. A total of 239 subjects with mild to severe mental retardation participated in the study. Data were collected with two assessments: a 142-question
The maturity of a scientific discipline is undoubtedly reflected in the specificity and complexity of the research questions posed and tackled by scholars in the field. To keep up with and be able to address complex research questions, an increase in the number of options, flexibility, and sophistication of research design and statistical models is necessary.

In addition to using CFA to evaluate construct validity, SEM may be used to measure test–retest reliability (Hoyle & Smith, 1994). For illustrative purposes, Figure 2 depicts a path diagram that may be used for evaluating the stability of a measurement instrument (X represents the latent variable) at Weeks 1 and 2. The curved line represents the correlation between the latent variable at the two time periods. One advantage of using this approach versus traditional dependent t tests is that the measurement error has been removed. In addition, note that a curved line runs from the measurement error associated with each indicator. Those paths account for relationships between the same indicators that are not attributable to the relationship between the latent variable at Week 1 and Week 2. Therefore, correlated error, a source of bias, is removed before estimating the test–retest correlation.

The previous examples demonstrate two ways that SEM can assist researchers in the development of new measurement tools. Occupational therapy practitioners are continually challenged to find good clinical assessments to measure therapy progress and outcomes as well as to function as dependent measures in traditional types of experimental research. SEM provides flexible options for researchers attempting to develop psychometrically sound assessments and can be used to increase the rigor involved in the evaluation of both validity and reliability.

**Treatment Effectiveness**

The maturity of a scientific discipline is undoubtedly reflected in the specificity and complexity of the research questions posed and tackled by scholars in the field. To keep up with and be able to address complex research questions, structural means models (Bentler, 1990) for the purposes of estimating the mean values of the latent constructs studied so that differences between means (the effects of the treatment modalities) could be studied. Although the use of SEM in this context is complex, it is available. SEM for comparing group differences does not ensure more statistical power than traditional techniques, such as multiple analysis of covariance. However, what SEM does offer is more accurate estimates of treatment effects because of error-free dependent variables, more accurate evaluation of construct stability over time, and corrections for selection bias (Aiken et al., 1994).

Hoyle and Smith (1994) reported that the type of research questions that are best suited for addressing hypotheses about group differences are of the form, “For whom was Treatment X effective?” or “Under what conditions was Treatment X effective?” Because of the tremendous complexity of the research questions posed and tackled by scholars in the field, to keep up with and be able to address complex research questions, an increase in the number of options, flexibility, and sophistication of research design and statistical models is necessary.

Hoyle (1994) described the emergence of second-generation research and suggested that the use of SEM is one way to address research questions posed by subject matter with an accumulation of prior research and complexity.

Aiken et al. (1994) illustrated the use of SEM in outcome research through their report of a study that compared the effects of two different treatments (methadone maintenance and outpatient counseling) on the daily lives of persons with drug addiction 2 months after treatment began. They reported that SEM (particularly with the latent variables—domestic activities, leisure, health, quality of life, social interaction, drug and criminal involvement—used in their study) is the preferred method for understanding the underlying structure of the measuring instruments. In addition, subject characteristics between the two groups differed (the methadone treatment was given to persons severely addicted to heroin, and counseling was given to persons with less severe addictions), and SEM was helpful in minimizing the effects of bias resulting from subject selection and attrition.

Aiken et al. (1994) implemented a technique called structural means models (Bentler, 1990) for the purposes of estimating the mean values of the latent constructs studied so that differences between means (the effects of the treatment modalities) could be studied. Although the use of SEM in this context is complex, it is available. SEM for comparing group differences does not ensure more statistical power than traditional techniques, such as multiple analysis of covariance. However, what SEM does offer is more accurate estimates of treatment effects because of error-free dependent variables, more accurate evaluation of construct stability over time, and corrections for selection bias (Aiken et al., 1994).

Hoyle and Smith (1994) reported that the type of research questions that are best suited for addressing hypotheses about group differences are of the form, “For whom was Treatment X effective?” or “Under what conditions was Treatment X effective?” Because of the tremendous complexity of the research questions posed and tackled by scholars in the field, to keep up with and be able to address complex research questions, an increase in the number of options, flexibility, and sophistication of research design and statistical models is necessary.

Hoyle (1994) described the emergence of second-generation research and suggested that the use of SEM is one way to address research questions posed by subject matter with an accumulation of prior research and complexity.

Aiken et al. (1994) illustrated the use of SEM in outcome research through their report of a study that compared the effects of two different treatments (methadone maintenance and outpatient counseling) on the daily lives of persons with drug addiction 2 months after treatment began. They reported that SEM (particularly with the latent variables—domestic activities, leisure, health, quality of life, social interaction, drug and criminal involvement—used in their study) is the preferred method for understanding the underlying structure of the measuring instruments. In addition, subject characteristics between the two groups differed (the methadone treatment was given to persons severely addicted to heroin, and counseling was given to persons with less severe addictions), and SEM was helpful in minimizing the effects of bias resulting from subject selection and attrition.

Aiken et al. (1994) implemented a technique called structural means models (Bentler, 1990) for the purposes of estimating the mean values of the latent constructs studied so that differences between means (the effects of the treatment modalities) could be studied. Although the use of SEM in this context is complex, it is available. SEM for comparing group differences does not ensure more statistical power than traditional techniques, such as multiple analysis of covariance. However, what SEM does offer is more accurate estimates of treatment effects because of error-free dependent variables, more accurate evaluation of construct stability over time, and corrections for selection bias (Aiken et al., 1994).

Hoyle and Smith (1994) reported that the type of research questions that are best suited for addressing hypotheses about group differences are of the form, “For whom was Treatment X effective?” or “Under what conditions was Treatment X effective?” Because of the tremendous complexity of the research questions posed and tackled by scholars in the field, to keep up with and be able to address complex research questions, an increase in the number of options, flexibility, and sophistication of research design and statistical models is necessary.
number of factors that are influential in rehabilitation efforts and ultimate outcomes, SEM is one technique that may help us account for and study these factors when examining the effects of occupational therapy treatments.

SEM has been used in longitudinal studies. Hays, Marshall, Wang, and Sherbourne (1994) examined the physical and mental health of a large sample of persons with chronic illness at baseline, after 2 years, and again after 4 years. They used various self-report measures of physical functioning, including role limitations, general health perceptions, social functioning, emotional well-being, and energy or fatigue. Although their research design was complex, they outlined their methods in detail. They concluded that SEM is well suited to evaluate longitudinal models involving a combination of direct and indirect effects among theoretical constructs and that important information can be gained about the nature of the relationships between constructs over time.

Limitations and Cautions

Some of the potential problems with SEM include model misspecification, difficulty meeting the assumptions of the method, and obtaining sufficiently large sample sizes. It is likewise technically challenging to become familiar with the software and the mechanics of model specification and analysis. In addition, universally accepted methods for evaluating goodness-of-fit are lacking, and despite the quantitative output, much interpretation ultimately relies on the researcher’s subjective experience. Crowley and Fan (1997) reported that researcher judgment is necessary in both specifying and testing a model. Critiques of SEM that highlight these concerns in more detail are provided by Bryk and Raudenbush (1987) and Burchinal and Appelbaum (1991).

Conclusion

The goal of this article was to acquaint occupational therapy researchers with the potential of SEM for designing and analyzing clinical research. SEM applications related to test construction and treatment effectiveness were highlighted, although there are many other applications. Hoyle and Smith (1994) encouraged the use of SEM by stating that SEM “is an attractive complement and on occasion, alternative to traditional approaches to design and analysis such as ANOVA, factor analysis, and multiple regression” (p. 439). For readers who want a more extensive review of the topic, good introductions to SEM have been written by Bollen (1989), Byrne (1994), Hayduk (1987), Loehlin (1992), Mueller (1996), and Schumaker and Lomax (1996).

SEM is an exciting move toward increasing the sophistication and complexity of behavioral and social research. An examination of occupational therapy literature reveals an increase in the amount of qualitative research over the past 10 years. This increase might be due to advancements in qualitative methodology and rigor or the need for methods of scientific inquiry that allow for the examination of a multitude of variables or factors that influence function, dysfunction, and recovery from illness and the relationships among such variables. One way to move forward is to use SEM to evaluate some of the conceptual frameworks and complex models developed in occupational therapy. By forcing explicit statements of relationships reflecting our theories and by encouraging good measurement, both our way of thinking and our analysis will be enhanced.

References


Hoyle, R. H., & Smith, G. (1994). Formulating clinical research


