2 On the Mismatching of Levels of Abstraction in Mathematical-Statistical Model Fitting

John R. Nesselroade
John J. McArdle
University of Virginia

In “Are Theories of Learning Necessary?“, Skinner (1950) used the term ***theory*** to refer to “any explanation of an observed fact which appeals to events taking place somewhere else, at some other level of observation, described in different terms, and measured, if at all, in different dimensions” (p. 193). Skinner thus identified several levels of abstraction between observable events and the explanations for those events, and he questioned the value of such multilevel explanatory systems. **Hebb (1949),** Reese and **Overton (1970),** and others elucidated the formal representations and empirical requirements for a “model.” **Cattell (1966b, 1966c)** discussed the form of a model, especially one based on formal mathematical and statistical features (see also Learner, 1978; Salmon, 1971). Following these latter notions we presume that there is merit in fitting abstract models to empirical data, but we take a critical look at some issues that arise because of the involvement of model fitting with different levels of abstraction of concepts and relationships.

The Growth of Model Fitting

Mathematical-statistical model fitting of data was given considerable attention in the 1940s and **1950s.** Although modeling seemed to be a major “growth industry” in physics, chemistry, economics, and biology it did not become so in psychology and sociology until the later 1960s. The delay is somewhat surprising because the need was certainly there-the problems facing psychology and sociology were no **simpler** than those in the other disciplines. Doubtless, there were many reasons for this longer lag time, one of which was the lack of sufficiently powerful computational machinery available to the interested psychologist and sociologist. Another was the deeply entrenched research and graduate training focus on using data to reject or accept null hypotheses rather than lo assess the plausibility and goodness of fit of explanatory models (Cohen, 1994; Morrison & Henkel, 1970; Mosteller & Tukey, 1977). Since the 1960s, we have witnessed a surge of interest in mathematical-statistical modeling of behavioral and social science phenomena. Increased interest in mathematical-statistical modeling is as evident in the study of development and change as it is in any area of psychology. Indeed, students of development are participating actively in a stunning increase in both the amount of mathematical-statistical modeling and in the level of comprehensiveness and sophistication of its use (see, e.g., Bollen & Long, 1993; Collins & Horn, 1991; Connell & Tanaka, 1987; Hertzog & Schaie, 1988; Horn & McArdle, 1980).

Instead of going lo the "calculator room" as was common in the 1940s and 1950s or the "comp center" as was the case in the 1960s and 1970s, we now sit in our own offices before our private video display terminal and "estimate models." Attracted as if by electronic pheromones, modelers fly with abandon to their consoles and attach themselves, separating when they are convinced that another important line of work has been perpetrated. Perhaps this development is not too surprising. After all, the personal computer that now appears on any secretary's desk is more than 100 times more powerful than those of the largest computer centers of the late 1960s, and the software is far more "user friendly." Perhaps even more important, granting agencies are not reluctant lo fund such activities.

The quite noticeable changes in our observable mathematical-statistical modeling behavior have included important alterations in the way data analyses are conceptualized and discussed. From some quarters, there has been a strong push away from loosely defined analytical models in favor of those that are more explicitly and precisely specified. For example, models that in the past were called analyses of variance and exploratory factor analyses, even at times with some reverence, are now often denigrated by reviewers. Some of the the more respectful critics point out that these models are instances of saturated linear comparisons and unrestricted factor analyses (McArdle & Nesselroade, 1994). Rotationally indeterminant factor models now often are supplanted by highly constrained comparisons and what are loosely labeled "confirmatory factor analysis" models. Instead of testing hypothesized "no differences," more and more investigators are now testing the goodness of fit between a precisely specified model and empirical data, examining how satisfactorily that model accounts for variafion in variables of theoretical interest and providing confidence boundaries around these statements.

Although the newer modeling and estimation procedures continue to be

confined largely to relatively simple linear, additive models, their capabilities for representing more complex relationships (e.g., incorporating nonlinear constraints in the specification of the models being fitted) are growing. Indeed, the current formulations and accompanying computational machinery are quite impressive, even by 1990 standards. Despite the elegance of fhe methods and techniques available for representing reality in mathematical and statistical models, however, it is important to remember that mathematical and statistical modeling procedures are the tools and not the craftsmen; they are the instruments and not the musicians. As with quality tools and fine musical instruments, full realization of the promise of these analytic devices requires a high level of familiarity and knowledge on the part of users. The more skillfully these implements are used, the more impressive and valuable are the outcomes that fhey help to produce, and the greater will be the gains from applying them to substantive problems and issues. Thus, the old adage, "technology does not produce science," continues to accrue validity.

Focus of the Chapter

A benefit of the currently available modeling innovations has been that users are forced to provide an explicit rendering of hypothesized interrelationships among implicated concepts and variables in order to invoke the computational machinery t hat yields estimates of f he model's parameters. A drawback, however, is that even "half-baked" notions can be lent an air of considerable respectability when put in the guise of an explicit model, just as a mediocre composition can be made to sound acceptable (or al least better) when performed on fine musical instruments by a competent, formally attired orchestra.

The primary objective of this chapter is to identify and discuss selected matters that we believe are germane lo the productive application of mathematical-statistical model fitting procedures. The ideas that we emphasize have substantial histories in the psychological literature (e.g., attifude-behavior and treatment-outcome relationships) from which important lessons can be learned, but in the contemporary modeling context, these discussions have not yet received the attention that we believe they merit.

Our principal contention is that one of the activities necessary lo realize the best that current mathematical-statistical modeling procedures have lo offer is to attend explicitly to the matter of tailoring concepts to relational statements in regard to their levels of abstraction. The tailoring must take into account one's theory-driven objectives. In current parlance, the tailoring is an aspect of model specification. It is on fhis mafter that we try to center the present discussion most directly. Admittedly, there is likely lo be an iterative, successive approximation character to any fruitful application of modeling procedures in a given substantive domain as theory and data are used to exploit each other (Catteil, 1966c; Learner, 1978). But persistence in the efforts to forge stronger and stronger links both between observable and latent variables and among latent variables is a key element of scientific research and a constant reminder that validation of models is a never-ending process.

Consider some common examples of what we mean by levels of abstraction of concepts and relationships and the matter of mismatch between them. One does not attempt to explain the workings of a radio or television set to a 4-year-old child using concepts such as differences in potential, micromhos of conductance, or phase shifts. Such concepts are not a good match to relationship statements of the form "the signal goes in there and the sound comes out here." As central as the abstract concepts are to the design of electronic circuitry, their level of abstraction is not suited to the conceptual machinery of the 4-year-old. Similarly, one does not recite principles of lift and drag to a 6-year-old child who is experiencing difficulty keeping a kite aloft. Depending on the child's age, this kind of explanation may come later when the kite is lost and the child is trying hard to hold back the flow of tears. In some sense, the child's behavior determines the appropriate level of explanation, but how do we know just what is appropriate in a given circumstance?

To take yet another example, when one is trying to construct an explanatory account of dividend size for stockholders, it seems far more appropriate to focus on the relationship between magnitude of a company's annual stock dividend and its amount of annual sales than between magnitude of the annual stock dividend and annual level of "'extraversion" of members of the sales force. Even if extraversion is a valuable component of successful selling, the relational coefficient between it and annual sales is certainly not the most appropriate parameter estimate to report to disappointed stockholders (unless they all happen to be differential psychologists). Here, we usually count on the stockholders to tell us if we are at an appropriate level of explanation.

The value of a mathematical-statistical model rests to a considerable extent on the match between the structural model (e.g., a structural-causal set of relationships) and the latent variables to which it is being applied: Are the substantive concepts at the optimal level of abstraction for the nature of the relationships embodied in the structural model? Such a question underscores the important role that latent variables play in using abstract relationships to account for observable manifestations. But such relationships are what theory is about. How do we know when we have achieved a "good" match? Usually we only have the consensus of a local group of scholars. However, we do have at least one superordinate concept—factorial invariance (McArdle & Cattell, 1994; Meredith, 1964, 1993; Thurstone, 1957)—that can be used both as a hammer (to impose conditions that data may or may not meet) and as a chisel (to separate quantitative and qualitative differences and changes) in shaping models of reality from the standpoint of matching of levels of abstraction. Other promising tools and concepts are mentioned later in the chapter.

In model building and theory testing there should be deliberate, explicit concern for possible discrepancies among levels of abstraction and for the implications of those discrepancies. Admittedly, what are discrepancies for one purpose might not be for another. We offer some examples that illustrate the importance of the match in levels of abstraction and examine the concepts in relation to three areas that are germane to the interests of developmentalists where mathematical-statistical modeling is prevalent—factor analysis, personality measurement, and behavior genetics.

On purely technical grounds, we acknowledge that the match of levels of abstraction of concepts and relationships might not be critical per se to the estimation process. As long as the model is identified, generally its parameters can be estimated, regardless of how ill-specified the model might be (e.g., McArdle & Prescott, 1992). Interpreting the level of model-data congruence in relation to the further elaboration of explanatory frameworks is another story. The match between levels of abstraction is critical to the interpretation and evaluation of the outcome of the model fitting exercise, the further evolution and elaboration of an explanatory scheme, and the advancement of an area. In common parlance, "Sure, it can be estimated, but can sense be made out of it?" What we aim to do is to point the way to more deliberate exploration of the matter of match between levels of abstraction in concepts and relationships to help further the use of model fitting in applications of particular interest to developmentalists.

COMPLETENESS OF DATA USED IN MODELING SCHEMES

Before proceeding to a more systematic examination of examples pertinent to the notion of match between the levels of abstraction of concepts and relationships found in the psychological literature, we briefly consider some more general aspects of the empirical substrate of model fitting. These concerns are addressed for the purpose of laying useful groundwork for subsequent consideration of the levels of abstraction issues that follow.

The data used in modeling schemes arise in the context of a research design, however well or poorly conceived it may be. Elsewhere, the concepts of selection and selection effects (Aitken. 1934; Lawley, 1943-1944; Meredith, 1964, 1993; Pearson, 1903) have been used to emphasize the fact that empirical research always involves limited (selected) data and that the resulting selection effects impinge on the inferences one draws from those data (Horn & McArdle, 1992; McArdle, 1994, 1995; McArdle & Cattell, 1994; McArdle & Nesselroade, 1994; Nesselroade, 1983, 1991a; Nesselroade & Jones, 1991). Here we briefly review the key arguments in relation to the concerns of the present chapter.

The principal idea is that research design is a multimodal selection operation, the modalities of which are persons, variables, and occasions of measurement and other bounding dimensions of the data manifold described by Cattell (1952b, 1966a). The specific array of data obtained in a given empirical study is the product of a set of selection operations defined on one or more of these modalities. The general idea is portrayed in Fig. 2.1. When empirical data are collected, they are invariably selected from a vast universe of potential observations. As a consequence of the selection operations that define a specific collection of data, the obtained data

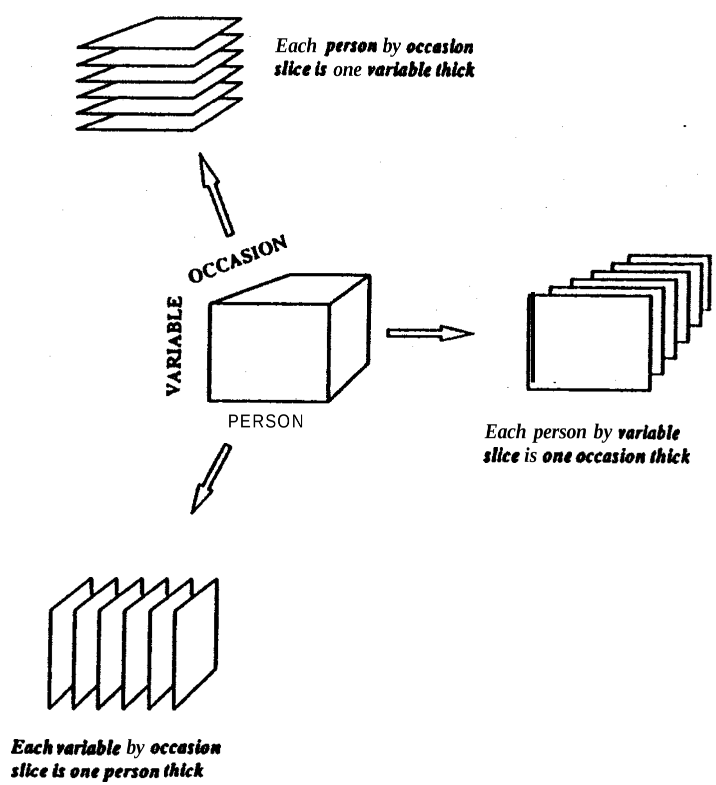

FIG. 2.1. Research design and data collection: "Slicing" the data box.

inevitably harbor selection effects with respect to those dimensions along which selection occurred. On the one hand, these selection effects jeopardize the external validity of the conclusions one reaches about relationships from an examination of those data. On the other hand, a better understanding of their nature may enable investigators to design research with deliberately placed "holes" for more economical but still valid data collection (for a review, see McArdle, 1994).

Suppose one's data are derived from a design that involves a large, exquisitely representative sample of persons. Such representative sampling with respect to persons is very expensive, so typically one must compromise with regard to other data modalities, for example, by using a minimum set of variables to identify key concepts or measuring the persons only one time. A large scale survey, for instance, might involve measuring broad-gauge concepts such as anxiety or extraversion with two or three self-report items. The resulting data are very narrowly selected with respect to the variables dimension of the data box, and this variable selection jeopardizes estimates of relationships of anxiety or extraversion thus measured to other concepts and variables. Hence, the matter of levels of abstraction of concepts is inextricably tied to the concerns of selection of data with respect to choice of variables. Just as obviously, data reflecting only one time of measurement (cross-sectional data) are very narrowly selected with respect to the occasions dimension of the data box, and the selection effects that inhere in the data because of the severe selection with respect to time virtually eliminate meaningful discussion of stability and change and curtail the discovery of time-based relationships among thk variables.

We elaborate a little more concerning each of the three selection modalities: persons, variables, and occasions of measurement. It is selection with respect to the variables dimension of the data box that leads most directly into the issues of concern for this chapter, although from the perspective of a developmentalist, it is difficult to exaggerate the threat of selection effects due to selection with respect to the occasions dimension.

Person Selection

Whether one uses convenience samples or broadly constituted ones, unless some representative sampling scheme is employed, the data that are used in model fitting will reflect the effects of selection with respect to persons. These effects, in turn, can affect the model-fitting process in a variety of ways. We need not say much more about selection with respect to the persons dimension of the data box here. It is with respect to this modality of selection that researchers seem to be the most sensitive and for which the most notable progress has been made in specifying how adjustments can be made in current modeling practice (e.g., Berk, 1983; Heckman, 1979; Heckman & Robb, 1986; Muthèn, 1984; Rubin, 1987).

Variable Selection

It is with respect to the variables dimension of the data box that selection effects may well take their greatest toil with respect to model-fitting exercises, in part because the tools for assessing and correcting for such selection effects are not nearly as well articulated and rigorous as they are for the effects due lo selection with respect to the persons dimension of the data box. In the case of latent variable modeling, the appropriateness of the choices of observed variables or indicators to mark the latent variables has long been understood to dictate the limits of usefulness and generality of the model-fitting outcomes (e.g., Cattell, 1952a). Moreover, the omission of key variables from model specification and observation is known to bias the remaining parameter estimates in deleterious ways (e.g., Goldberger, 1972). The mathematical and statistical tools for minimizing the effects of these problems may still be somewhat deficient, but the general issue of how to build strong, convincing links between observed and unobserved variables-and thereby among unobserved variables- has been a concern for many decades (e.g., Cattell, 1952b; Humphreys, 1962; Thurstone, 1938).

Occasion Selection

Many important concepts and relationships are defined in relation to time, a fact that pertains to (heir levels of abstraction. Elsewhere (Nesselroade, 1983; Nesselroade & Jones, 1991), it has been argued that, for developmentalists, selection with respect lo occasions of measurement is one of the most influential aspects of research design. A single occasion of measurement, for example, as in cross-sectional research designs, represents an extreme case of selection with respect to the occasions modality of the data box. Indeed, from this perspective, one of the primary differences between cross-sectional and longitudinal research designs is the degree of severity of selection with respect to occasions. Regarding selection with respect to occasions, cross-sectional and longitudinal designs can be argued to be quantitatively, but not qualitatively, different.

A simple longitudinal design of two occasions of measurement (e.g., pre-post designs) represents a slightly more generous selection of data with respect to the occasions modality- but how much more generous is it? According to Rogosa (1988), it may not be much more generous in the framework of conventional panel designs. But if the two times of measurement are selected with differing time lags, the limitations can be diminished (see McArdle & Woodcock, 1996). The nature of selection with respect to

the occasions dimension of the data box and the accompanying selection effects has not yet been formulated with the level of mathematical and statistical rigor as is the case for selection with respect to the persons modality, but the general selection formulations implicate it as a serious design concern.

MATCHES AND MISMATCHES IN LEVELS OF ABSTRACTION

Factor Analytic Models

A distinction found in the modeling literature that bears directly on the matter of levels of abstraction has been referred to as the measurement versus the structural model (e.g., Jöreskog & Sörbom, 1979) and as the outer model versus the inner model (e.g., Lohmöller, 1989; Wold, 1975). The measurement or outer model, which links manifest and latent variables, by definition involves concepts at different levels of abstraction. The structural or inner model consists of the latent variables and the relational statements among them. The distinction between measurement and structural or inner and outer models is not required for purely mathematical or statistical reasons per se (for proof, see McArdle & McDonald, 1984), but there is no doubt that the distinction is a fundamental aspect of contemporary modeling philosophy. Without this particular difference in level of abstraction, modeling by means of latent variables would have no future because there would be no way to devise empirical tests of deductions stemming from the model.

Examples to be presented later involve both measurement and structural aspects of modeling. In ‘the measdrement model case, the presence of different levels of abstraction of concepts and variables is understood; indeed, it is a deliberate feature of the modeling. Nevertheless, the possible range of matches between levels of abstraction of concepts, variables, and the relational statements connecting them is wide and deserves explicit discussion. The structural or inner model is very much at the center of the issues on which we are focusing.

The relationships between latent and manifest variables are reflected in regession-like weights (e.g., factor loadings) in some cases and in correlation coefficients (e.g., factor structure values) in others (see, e.g., McArdle, 1994). The factor analytic literature since the 1940s contains some of the most enlightening discussions of the measurement model. The discussions are found in that literature because one of the primary uses of factor analysis is to examine relationships between observed and latent variables (e.g., the information contained in the factor pattern and factor structure matrices). This venerable goal remains salient today despite great advances in the methods and techniques of factor analysis (McArdle & Cattell, 1994).

To get a fix on the key issues, it is useful to remind ourselves of what was being attempted by factor analysts in the middle decades of the 20th century. In brief, many of them were trying to specify certain content areas and to delineate a basic set of dimensions—a reference frame—that spanned those content domains, for example, as general traits of interindividual differences in human abilities. Some content areas (e.g., human abilities) were broad in conception; others were relatively narrow (e.g., honesty). In personality research, for example, investigators such as Cattell (1950), Eysenck (1952), and Guilford (1959) sought to identify major personality dimensions that "spanned the space" of personality. Indeed, this general line of inquiry continues even today as witnessed by contemporary efforts to clarify further the nature and generality of the "Big Five" personality traits (e.g., Goldberg, 1993; McCrae, 1989; cf. Block, 1995; Mershon & Gorsuch, 1988).

Illustrative is the work of Raymond B. Cattell (1950, 1957), who over 6 decades has been intent on unearthing a set of basic dimensions or factors from a domain of observable variables to represent human personality (see, e.g., Angleitner, 1991). Cattell had to address two major problems before factor analysis could be used to approach this task efficiently. The first was the circumscription of the domain of content, and the second was the inclusion of that content in a representative manner in empirical research. Cattell argued that "everyday" language was the place to start in defining the domain of personality because it would mirror all the important aspects of behavior (e.g., all those aspects noticeable enough to have earned a verbal label). This argument is now referred to as *the lexical hypothesis* (Goldberg, 1990). Cattell took the extant list of more than 4,000 trait adjectives that had been compiled by Allport and Odbert (1936) as the first approximation to the domain—the *personality sphere*. These adjectives were the basis for developing the personality dimensions found, for example, in self-report measures such as the 16 Personality Factor (16PF) test (Cattell, Eber, & Tatsuoka, 1970) and were also used as the basis for a series of studies based on behavior ratings. Block (1995) summarized and critiqued Cattell's and others' efforts to determine personality structure by this approach.

Other researchers also have argued for the validity of content domain specification and representation in the establishment of psychological concepts (e.g., Cronbach & Meehl, 1955; Humphreys, 1962; Nunnally, 1967). Humphreys (1962), for example, argued that domain identification and representative sampling schemes such as Cattell developed were steps in the right direction but did not satisfactorily resolve the density and sampling of variables questions. He further elaborated the organization of domains

and how the construction of measures should be based on principles that matched the level of abstraction of the variables to one's purposes. Humphreys illustrated the basic ideas using tools (saws, hammers, wrenches, etc.) as a domain and developed his arguments around familiar aspects of these items. His proposals gave further substance to the importance of distinguishing among levels of abstraction of concepts and their roles in relational statements. For instance, hand tools is a more inclusive class than either saws or wrenches. The manifest variable *number of teeth* is more sensibly related to the concept *saws* directly and to the concept *hand tools* indirectly by means of the saws concept, rather than directly related to hand tools. Although one could describe other handtools (e.g., wrenches) as having zero saw teeth, zero saw teeth in a wrench does not mean the same thing as zero saw teeth in a saw.

As we noted in the introductory section, the validity of a mathematical–statistical model rests somewhat on the match among the levels of abstraction of the structural model and the variables. We now turn to some other examples that further illustrate the importance of this match in levels of abstraction. To conclude, we examine these ideas specifically in relation to behavior genetics modeling.

There are many examples of matching (and mismatching) of levels of abstraction of concepts and relationships in the behavioral and social sciences. Because of its explicit mathematical–statistical properties, the factor analysis model was used previously to illustrate some key concepts pertinent to the theme of this chapter. We again draw from that literature to render more concrete the notion of disparity among levels of abstraction of concepts and relationships in the realm of mathematical–statistical modeling.

Personality Measurement

Through the application of the factor model, a number of important structural discoveries have been made, especially in the domains of temperament and human abilities. In the course of this work, a number of issues arose concerning the level of abstraction of concepts and relationships as they pertain to the modeling of observed variables. The theory of fluid and crystallized intelligence (Horn & Cattell, 1966), for example, resulted from dissatisfaction with the level of abstraction represented by Spearman's concept of general intelligence vis-à-vis that of actual test performance.

Different Levels of Abstraction. In the personality domain, Cattell (1957) distinguished among life-record (L) data, questionnaire (Q) data, and objective performance test (T) data as the products of three distinct and important media of observation employed by students of behavior. He argued that part of the proof of the fundamental nature of putative basic personality factors was that they would be detectable across the different observational media. Thus, he argued that one should be able empirically to align or "match" factors as the same whether they were obtained from L, Q, or T data. Cattell's initial attempts to provide this evidence of generality were challenged and the success of attempts to match personality factors across observational media was enthusiastically debated in the early 1960s (e.g., Becker, 1960; Cattell, 1961; Peterson, 1965).

In the aftermath of the discussions, there emerged a conception that illustrates the levels of abstraction notion very directly and that has helped considerably in the evaluation of subsequent personality and ability trait research. Cattell (1965) articulated a distinction between *order* and *stratum* in describing the nature of factors. Order had to do with the mechanics of factor analyzing data. The initial factors were first-order factors. If the intercorrelations of those first-order factors were themselves factored, the factors obtained were second-order factors, and so on. Stratum characterized the level of the factors in the theoretical scheme in which they were embedded. First-stratum factors, for instance, occupied the first level of abstraction in a theoretical scheme that linked multiple levels of factors to observable measures. Thus, depending on the level of abstraction represented in the variables being factor analyzed (e.g., single items vs. scale scores), one could obtain second-stratum factors in a first-order analysis or in a second-order analysis.

More concretely, what in Cattell's scheme were the second-stratum concepts of extraversion and anxiety could be obtained as first-order factors if one factored the scales (sums of item subsets) of the 16PF or as second-order factors if one factored the items of the 16PF. For example, the counterparts of the broad second-stratum factors of extraversion and anxiety that were found at the first order when T data were factored appeared at the second order when questionnaire items were factored. Thus, Cattell argued that until the proper level of abstraction of concepts (factor strata) was resolved, the theoretically specified relationships (factor isomorphism across media of observation) could not be evaluated against empirical data.

Consistency and Generality of Personality Traits. Another example of the salience of matching levels of abstraction of concepts and relationships comes from the literature on trait consistency and generality. Characterizing the personality, attitudinal, and evaluation literature since the 1960s have been the debates about the small to middling correlations usually found between measures of personality traits and measures of behaviors purported to be influenced by those traits (e.g., Epstein, 1983; Epstein & O'Brien, 1985; Mischel, 1968; Mischel & Peake, 1982; Steyer, 1987; Steyer & Schmitt, 1990), between expressed attitudes and behaviors (Fishbein & Ajzen, 1974), and between treatments and outcomes in evaluation research (Wittmann, 1988a, 1988b, 1991). Ranging through these discussions is a series of proposals and counter proposals concerning the role of data aggregation in empirical tests of the hypothesized linkages between broad dispositions and more specific behaviors in personality and attitudinal research and between generalized treatments and relatively specific outcomes in evaluation studies. Researchers have investigated the pros and cons of a variety of ways to aggregate information to explore the limits of these relationships. The issues have centered around aggregation of information in relation to different concepts of reliability and validity, and the exchanges among protagonists have helped to clarify a number of issues in important ways.

Various investigators have tried to formalize the nature of relationships across domains under appropriate aggregation of variables in ways that might lead to the expected relationships among different classes of variables. Wittmann's (1988a, 1988b, 1991) synthesis, which involves multiple data box representations and a multivariate generalization of the concept of reliability, emphasized the levels of abstraction issues under discussion. He used Brunswik's "lens model" to illustrate how *symmetry* and *asymmetry* apply to the matching of concepts in levels of abstraction when the objective is to optimize predictability.

Wittmann's argument is represented in Fig. 2.2. Empirical work supports Wittmannk (1988a, 1988b) contention that more symmetric choices of variables in regard to generality or level of abstraction on the antecedent and outcome sides of equations will lead to considerably stronger relationships. For example, from the domain of evaluation research, Wittmann illustrated how outcome criteria can be tailored to input variables to match them in level of abstraction and, thereby, yield correlations in the .70–.80 range instead of the usually found .20–.30 range. Wittmann's examples of mismatches include linking broad social programs on the treatment side with specific indicators on the outcome side or, equally unpromising, (mis)matching highly specific treatments with global measures of outcome.

More concretely. in relation to Fig. 2.2, health interventions might range from very national ones such as healthy behavior suggestions delivered via the mass media (e.g., as Level 3 predictors) through state-wide, school-based physical exercise programs (as Level 2 predictors) to local applications of fluoride in community water supplies (as Level 1 predictors). Examined outcomes might range from very general assessments of overall health via epidemiological statistics (e.g., as Level 3 criteria) to performance on physical fitness tests for teenagers (as Level 2 criteria) to incidence of tooth decay in children (as Level 1 criteria). Although the Level 3 teatment (mass media health suggestions) might exert some positive influence on the

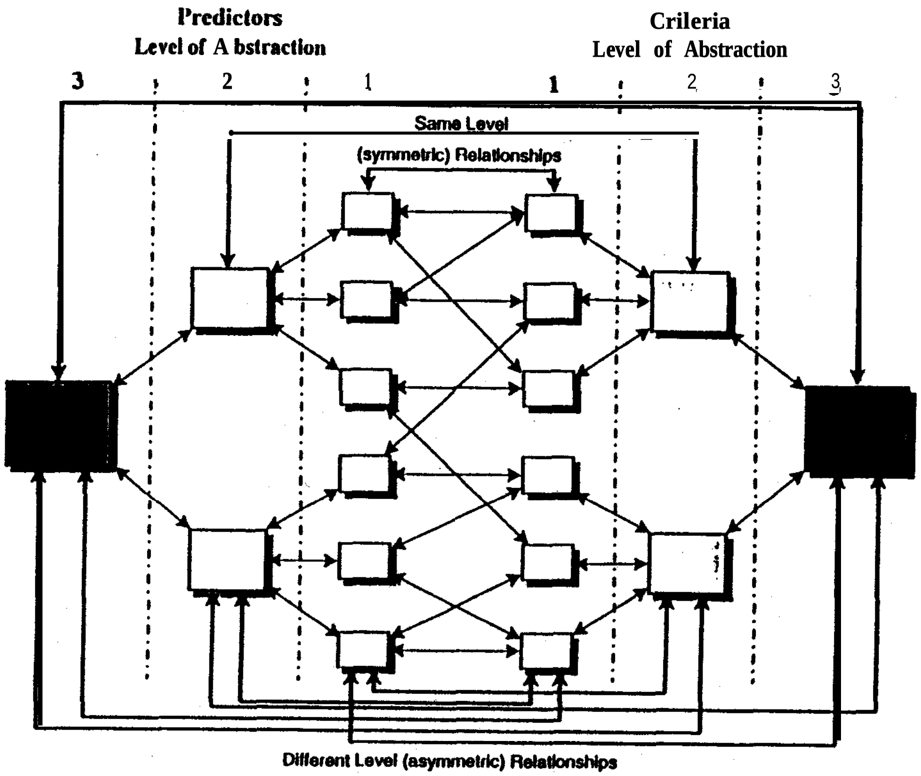

FIG. 2.2. Symmetric and asymmetric relationships among levels of abstraction (after Willmann, 1991).

Level 1 outcome (incidence of tooth decay in children), the relationship is bound to be less dramatic than that commonly reported between the introduction of fluoride and the incidence of tooth decay in children. By the same token, the local introduction of fluoride in water supplies cannot be expected to exert a major effect on general assessments of overall health.

More recently, Brody (1994) made arguments similar to Wittmann's regarding achieving increased estimates of heritability coefficients by means of more aggregation of measures. However, aggregation purely for the sake of increasing the magnitude of stability and other kinds of validity coefficients without concern for relationships that the very act of aggregation may be obscuring is naive. The literature on the trait-state distinction in personality research, for example, contains many examples of how extreme enthusiasm for high stability coefficients, sometimes obtainable by gross aggregation of scores over time, has tended to detract from the systematic examination of personality functioning (Nesselroade, 1988; Nesselroade & Boker, 1994).

Behavior Genetics Modeling

In the remainder of this chapter, we examine some issues of match and mismatch in a salient arena of application of mathematical-statistical modeling- behavior genetics. We draw attention explicitly to the levels of abstraction issue and examine the role it plays in estimating the parameters of pertinent models.

Since the 1970s, mathematical-statistical modeling has been increasingly used in behavior genetics research (e.g., Boomsma & Molenaar, 1986; Dolan, 1990; Martin & Eaves, 1977; McArdle & Goldsmith, 1990; Neale & Cardon, 1992). Some part of this increase is due to the ease with which both simple and complex biometric genetic models can be fitted to data using, for example, LISREL and other computer programs. Also, some part of the growth is due to a more general awareness of the potential value of genetic information as illustrated, for example, by the level of federal support received by the Human Genome Project.

Models of the impingement of genetic and environmental influences on phenotypic variables, some simple, some complicated and elaborate, have been fitted to both cross-sectional and longitudinal data. Not surprisingly, there have been disagreements concerning the use of behavior-genetic models for understanding behavioral and psychological phenomena, both statistically and substantively and in both stability and change contexts. It is our contention that the bases for some of the disagreements can be attributed to the levels of abstraction issues identified previously. What one researcher regards as an optimal combination of levels of abstraction is regarded by another to mix them up, thereby yielding peculiar information. We explore this notion, in part to see if some light can be shed on the future conduct of behavior genetics analysis of psychological/behavioral variables. A couple of examples clarify the importance of the matching of levels of abstraction ideas in the context of behavior genetics models.

***Attitudes* and *Heritability*.** Tesser (1993) analyzed responses to a series of attitude items in terms of relationships between the heritabilities of the responses to the items and other item characteristics including response latencies and resistance to change. He concluded that the data were consistent with an interpretation that attitudes manifesting higher heritabilities were "stronger" than attitudes of lower heritability. "Stronger" meant that pertinent items were responded to more quickly, were more difficult to change, and were more consequential in attraction among individuals.

Tesser (1993) asked rhetorically whether these data suggested that there is a gene for specific attitudes analogous to a gene for eye color. His answer was: "I doubt it. However, one can imagine a number of mechanisms by which more directly heritable physical differences might play themselves out in specific attitudes in a particular cultural milieu" (p. 139). Tesser identified sensory structures, body chemistry, intelligence, temperament and activity level, and conditionability as examples of more directly heritable differences. Thus, in relation to Fig. 2.2, Tesser identified several different levels of abstraction of concepts on the outcome side-for example, eye color versus conditionability versus attitudes-and identified those al one level (eye color) as more directly related to genetic influences than those found at other levels.

Televison Viewing and Heritability. Plomin, Corley, Defries, and Fulker (1990, 1992) and Prescott, Johnson, and McArdle (1991) had an exchange concerning the former's conclusions regarding behavior genetics modeling of the influences on television viewing in early childhood. Several aspects were highlighted, but the question of Prescott et al. (1991), "Does viewing time reflect passivity, need for stimulation, or some other characteristic of adaptive significance?" (p. 431), illustrates once more the issue of level of abstraction of concepts and the matter of matching them lo the set of relationships being modeled. The response of Plomin et al. (1992), "Behavioral genetic research is a reasonable first step in answering questions about the origins of individual differences, and such research is heuristic, raising other interesting questions of the sort listed" (p. 76), reinforces the previous emphasis on the successive approximation and gradual model refinement that appear necessary to scientific advance.

The examples involving attitudes and heritability and television viewing and heritability highlight the following fact. The phenotypic variance of any variable that one chooses to study can be decomposed biometrically into different sources of variance and the results summarized with statements such as "*x* is largely genetic" or "*y* is largely environmental." Such decomposition of phenotypic variance into different sources is always possible regardless of the level of abstraction of the particular variable relative to the levels of abstraction of biological and other behavioral/psychological variables that might be implicated. As was mentioned earlier, parameters of identified models can be estimated, but interpretation of the estimates can require additional information. When levels of abstraction are mixed, it is possible that some relevant information will be confounded in misleading ways with the levels of measurement.

Just what outcomes, such as those described previously, mean in terms of mechanisms, levels of abstraction, and complexity of relationships is the kind of information that needs ultimately to be sought if behavior genetics analysis is to contribute in a major way to our understanding of ontogeny across the life span. Developing valid models of the interrelationships among biological and behavioral phenomena rests heavily on explicit regard for the matching of levels of abstraction of the relationships and their referent concepts on both the biological and behavioral sides.

Matching Levels of Abstraction. What are the essential dimensions of concepts in relation lo the questions being examined here? We have mentioned some existing taxonomies of abstractions that bear on the problem. These include the hierarchical and the-reticular frameworks found in the factor analytic literature (Cattell, 1965) and the notions of symmetry and asymmetry discussed by Wittmann (1988a, 1988b, 1991). The general issue concerns how the matching of levels of abstraction of concepts bears on the fitting of behavior genetics models to data. Stated another way, what is the nature of the relationship between levels of abstraction of concepts and relationships and the interpretation of outcomes when behavior genetics models are fitted to behavioral and psychological variables? For example, given the current state of behavior genetics models with regard to levels of abstractions of the concepts on the antecedent side-genetic and environmental informalioq-are there more or less appropriate levels of abstraction for variables on the outcomes (phenotypic) side? Which criteria can be developed for ascertaining the answer in a given circumstance?

These issues can be illustrated al several levels using the basic concepts of structural factor analysis illustrated in Fig. 2.3 and 2.4. At a first level, consider the standard univariate biometric model. As Fig. 2.3 illustrates, this model assumes that a measured variable (*Y*) is a function of three or more uncorrelated latent components: (a) additive genetic influences (*A*), (b) unspecified influences common (*C*) to all members of the same family, and (c) largely unspecified nongenetic or environmental influences including errors of measurement (*E*). Information from members of different family configurations (MZ and DZ twins, parents and children, adopted and biological family members, etc.) is used to estimate the parameters of

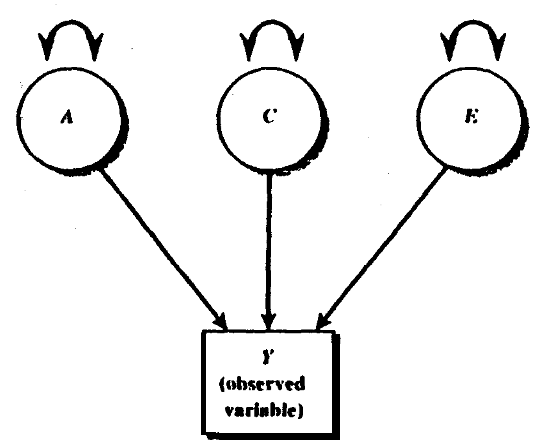

FIG. 2.3. Univariate biometric model (*A* = additive genetic influences, C = unspecified influences common to members of the same family, *E* = largely unspecified nongenetic or environmental influences including errors of measurement).

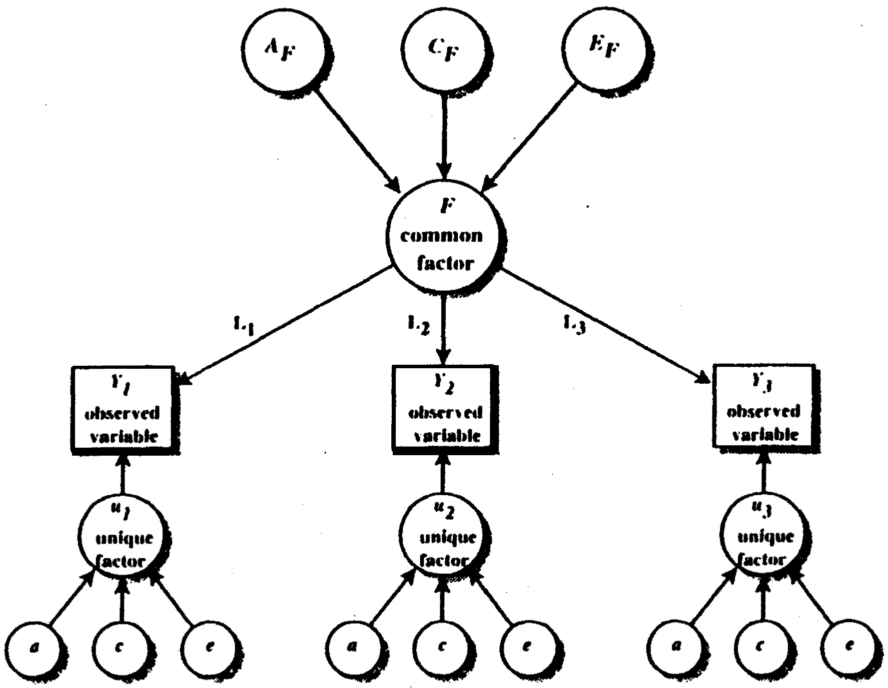

FIG. 2.4. **Multivariate** psychometric common factor model (A_F = additive genetic influences on the common factor, C_F = unspecified influences that are common to members of the same family acting on the common factor, E_F = largely unspecified nongenetic or environmental influences including error of measurement acting on the common factor).

the model. The resulting parameters of the model are often reported directly or as a complex, explained variance ratio $\sigma^2_A/(\sigma^2_A + \sigma^2_C + \sigma^2_E)$, termed the ***heritability coefficient***. The denominator of the heritability coefficient is directly assessable only in toto, as variance in the phenotype, with the component variances being estimable only indirectly.

As with the use of most complex mathematical-statistical models, the validity of statements about behavior genetic influences based on heritability estimates is subject to several concerns. The usual sampling limitations of any explained variance ratio apply to heritability, and confidence boundaries for the estimates are critical to their interpretation. Also, heritability estimates are obtained under multiple latent variable assumptions about orthogonality, linearity, homoscedasticity, covariances, and interactions. These assumptions are often untested and, in some cases, untestable, a limitation that is often ignored in subsequent discussions. Yet, it is on the basis of the heritability estimate that claims such as "x is largely genetic" or "y is largely environmental" that were mentioned previously are rested..

Most important, however, from the perspective of this chapter, is the fact that heritability estimates may be biased by the levels of abstraction involved in the modeling. To illustrate, consider the model that is presented in Fig. 2.4. This model renders explicit an additional level of abstraction by including the latent variable F(a common factor) between the two levels of abstraction that are represented in Fig. 2.3. For example, measures of primary mental abilities are commonly a focus of behavior genetics investigations. Considerable evidence exists (Horn & Cattell, 1966) that there is at least one additional level of abstraction- for example, fluid and crystallized intelligence factors-intervening between the ability test scores and the A, C, and E of Fig. 2.3 and 2.4. Fin Fig. 2.4 is a function of the orthogonal A, C, and E, and the observed variables, Y_k, are functions of the common factor F with factor loadings, L_k, and their respective unique factors, u_k. If this model represents the "true" relationship among different levels of abstraction, then the parameter estimates of the model in Fig. 2.3 are statistically biased, and the heritability estimate would be biased as well.' The heritability of the common factor, which now contains no random measurement error (see McArdle & Goldsmith, 1990), depends on three influences A, C_f, and E_f. However, if this model is correctly specified, then the variance of any observed variable, Y_k, is also a function of the biometric components of the unique factor (which does contain random measurement error). For any Y_k, the degree to which the modeling approach represented in Fig. 2.3 approximates the values of those variances of interest shown in Fig. 2.4, namely A,, C_f, and E_f, is a direct function of the size of the loading L_k. The common factor model represented in Fig. 2.4 can be falsified by a multivariate data set, but the univariate representation of Fig. 2.3 does not afford that scientifically desirable possibility. In general, only minor aspects of the univariate biometric model can be rejected at this level of measurement.

The factor-based model of Fig. 2.4 reflects one of the most fundamental concepts in behavioral analysis-all measured scores are composed, in part, of one or more common factors, and the common factors are involved in the chain of structural influences and effects. In this model, the common factors reflect the geneotypic and environmental influences, and these influences are measured in the observed variables in a way consistent with the strength of the relationship between the variables and the factors (for more details, see McArdle & Goldsmith, 1990). In the history of behavior genetic analyses, however, a variety of alternative multivariate models have been postulated. One classical alternative model (first presented by Loehlin

'The MATH EMATICA computer program for the symbolic evaluation of the biis in these models can be obtained as an ASCII file under the subdirectory \PUBLIC_ ACCESS\JRN\LEVELS1994 from the Anonymous FTP server at the University of Virginia (VIRGINIA.EDU).

& Vandenberg, 1968; Martin & Eaves, 1977) suggests that all variables are influenced by the same additive genetic factor (A) and some other (orthogonal) nongenetic factor (*E*). This biometric-factor concept might be an appropriate interpretation of the genotypic action at some level of abstraction. One interesting feature of these two models is that the psychometric genetic model (Fig. 2.4) requires parameters that are a subset of the biometric factors concept. Thus, these competing alternatives may be adequately tested at any well-defined level of measurement (for details, see McArdle & Goldsmith, 1990). and the test provides an important clue to the appropriate level of abstraction.

At any level of abstraction, however, the potential relationship between these kinds of latent variables and genetic information needs to be considered carefully (McArdle, 1994). Many examples of striking genotypic behavioral influences are reported in the literature (see any issue of the journal Behavior Genetics). Another often overlooked point is that the genetic information in a standard biometric model is obtained only as a latent "genotype" that is estimated under restrictive assumptions (orthogonality, linearity, homoscedasticity, no interactions, etc.). This level of evidence is not the same as, nor does it give the necessary empirical basis for, the existence of a "gene" in the sense defined and used in molecular biology. Behavioral geneticists have recognized this level of abstraction problem and made a first move toward a more fundamental level of analysis (e.g., DNA analysis using quantitative trait Loci; Plomin et al., 1994) However, direct relationships between genetic information at low levels of abstraction (e.g., chromosomal information) and important behavioral tendencies, especially within the normal response range, at a higher level of abstraction remain to be demonstrated. This last statement is issued more as a challenge to, than an indictment of, those attempting to explore the biological/behavioral interface.

CONCLUSION

In concluding this chapter, there are several points that we wish to emphasize. Our emphases are intended to further the use of mathematical--statistical modeling in the social and behavioral sciences. Seen from the somewhat historical perspective we have tried to develop, there are reasons to be concerned with the matter of level of abstraction of concepts in relation to mathematical-statistical models and the implications that derive from them. These modeling procedures have moved us away from a focus on the null hypothesis to looking more at the production and evaluation of viable alternative hypotheses. This shift in emphasis has had several positive effects, including attracting more thought to the way even simple statistical analyses are conducted. For example, with proper design considerations, error variance can now be isolated so that estimates of interrelationships among variables are "automatically" corrected for attenuation due to measurement errors.

The examples we have mentioned from the personality domain illustrate several key points bearing on the present discussion. First, the history of psychology reflects concern with the question of matching level of abstraction between concepts and relationships in the development of format mathematical-statistical models. Second, different configurations of abstraction levels in concepts and relational statements are possible and the choice can be guided by notions such as symmetry elaborated by Wittmann (1988a, 1988b, 1991) and others. Third, there is evidence to suggest that when, in light of one's purposes, an optimal match between levels of abstraction of concepts and relationships in modeling is sought, there can be striking increases in the strength of empirically estimated relationship. Unfortunately, no approach to determining an optimal match of levels of abstraction of concepts and relationships is generally accepted, although Wittmann's use of the concept of symmetry is promising in this regard. Moreover, some of the analytical tools discussed by Wittmann (1988b), such as set correlations, appear suited to evaluating the appropriateness of putative matches in levels of abstraction between predictors and criteria. These particular waters are being further "muddied" by indications that behavioral consistency may be found at the level of interindividual differences in patterns of intraindividual variability (Nesselroade, 1991b; Shoda, Mischel, & Wright, 1994)

We reiterate that aggregating information merely to maximize predictive validity is not the solution to every problem regarding aggregation. Deliberately specified differences in level of abstraction can be a key feature of modeling, as is the case, for example, of measurement models as referenced previously and in multitrait-multimethod analyses (e.g., Widaman, 1985).

From our perspective, one of the most valuable contributions that the structural modeling advances since the 1970s have given is the clarification of measurement issues and procedures by means of the measurement model and its distinction from the structural model in the framework of structural modeling programs. This has helped to clarify important issues of two of the pillars of modern psychological measurement -reliability and validity in their several guises (Cronbach & Meehl, 1955; McArdle & Prescott, 1992). Further, the formalization of the measurement model has made it possible to distinguish rigorously and explicitly among levels of abstraction in the concepts presumed to lie behind the observable measures.

There is no doubt that ascertainment of optimal levels of abstraction of concepts and relationships as discussed here is a part of the evolutionary stance we have to take with theory development in science. One has to find a "toe hold" and work from there. What we have tried to do is sensitize investigators to the idea that level of abstraction should be part of their conscious concerns when they are engaged in testing and improving their models. Having a testable hypothesis is an important concern whether the researcher is engaged in confirmatory or exploratory analysis. We believe, moreover, that testing hypotheses about the level of factorial invariance (abstraction) is among the most important experiments of all. Failing to be aware of this matter can seriously impede the development of better fitting models.

The evolutionary-developmental process of model elaboration has to include increased attention to the various modes of data selection discussed earlier. A given set of observable variables will mark some levels of abstraction much more validly than others. For example, medical doctors listen to self-reports of patients but tend not to act precipitously on those reports until they have supplemented them with other kinds of diagnostic information. Similarly, short shrift can no longer be given to the occasions dimension of the data box. Extending definitions and models over the repeated measurements dimension constitutes an important step toward the implementation of dynamic representations of interesting phenomena.

What do the next couple of decades hold for psychology with regard to mathematical-statistical modeling? Whether or not one accepts the progression of scientific maturation presented by West (1985), it is difficult for developmentalists to deny the appeal of being able to represent their substantive phenomena in rigorous dynamic formulations. And in fact, there are indications in the literature that mathematical dynamic representations of social and behavioral phenomena are just around the corner (e.g., Arminger, 1986; Baltes, 1987). Building and testing dynamic formulations of any meaning and significance involve, of necessity, mathematical-statistical modeling. Just as surely, these activities will force us to confront explicitly the matter of levels of abstraction in teaming concepts with relational statements.

ACKNOWLEDGMENTS

The generous support of the Max Planck Institute for Human Development and Education and the MacArthur Foundation Research Network on Successful Aging in the preparation of this manuscript is gratefully acknowledged by the first author. Research by the second author was supported by a grant from the National Institute on Aging (AG- 07137). We also thank Werner W. Wittmann for his valuable comments on an earlier version of the chapter.

REFERENCES

Aitken, A. C. (1934). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society, 4,* **106-110.**

Allport, 0. W.. **& Odbert,** H. S. **(1936).** Trait names: A psycho-lexical study. Psychological *Monographs,* 47. **(211),** l-171.

Angleitner, A. **(1991).** Personally psychology: Trends and developments. ***European Journal*** *of Personality. 5.* **185-197.**

Arminger, 0. (1986). Linear stochastic differential quation models for panel data with unobserved variables. In N. B. **Tuma** (Ed.), ***Sociological*** *methodology* (Vol. **16,** pp. **187-213).** Washington, DC: American Sociological Association.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental* Psychology. 23(J). **61** l-626.

Becker, W. C. **(1960).** The matching of behavior rating and questionnaire **personality** factors. *Psychologkol* ***Bulletin, 57,* 201-2** *12.*

Berk, R. A. **(1983).** An **introduction** lo sample selection bias in sociological data. ***American Sociological*** *Review, 48, 386-398.*

Block. J. (1995). A contrarian view of the five-factor approach to personality description. *Amerkan Psychologist.* ***117,*187-215.**

Bollen, K. A., **&** Long, J. S. (1993). ***Testing structural equation*** *models.* **Newbury** Park, CA: sage.

Boomsma, D. I., & Molenaar. P. C. M. (1986). Using **LISREL** lo analyze **genetic** and environmental covariance structure. *Behovior Genetics.* ***16,*237-250.**

Brody, N. (1994). Heritability of traits. ***Psychological*** *Inquiry. 5,* **117-1** 19.

Cattell, R. **B. (1950).** *Personality. New* York: McGraw-Hill.

Cattell, R. B. **(1952a).** *Factor* ***analysis:*** *An introduction and* ***manual*** *for the psychologist and social* ***scientist.*** *New* York: Harpr.

Cattell, R. B. **(1952b).** The **three** basic factor analytic research designs- their interrelations and derivatives. ***Psychological Bulletin,*** *49.* **499-520.**

Cattell, R. **B.** (1957). *Personality ond* ***motivation structure and measurement.*** Yonkers, NY: World Book.

Cattell, R. B. **(1961).** Theory of situational. instrument, second order. and **refraction** factors in **personality structure** research. *Psychological* ***Bulletin, 58,*160-174.**

Cattell, R. **B. (1965).** Higher order factor structures and reticular-vs.-hierarchical formulae for their interpretation. In C. Banks **&** C. L. Broadhurst (Eds.), *Studies in psychology in honor of Sir Cyril Burt (pp. 223-266).* London: University of London Press.

Cattell, R. B. **(1966a).** The data box: Its ordering of total resources in terms of possible relational systems. In R. B. **Cattell** (Ed.), *Handbook of* ***multivariate experimental psychology*** (pp. **67-128).** Chicago: Rand McNally.

Cattell, R. B. **(1966b).** The meaning and strategic use of factor analysis. In R. B. **Cattell** (Ed.), *Hondbook of* ***multivariate experimental*** *psychology* (pp. 174-243). Chicago: Rand McNally.

Cattell, R. B. **(1966c).** Psychological theory and scienlifii method. In R. B. **Cattell** (Ed.), *Handbook of* ***multivariate experimental psychology*** (pp. l-18). Chicago: Rand McNally.

Cattell, R. B.. **Eber,** H. W., **& Tatsuoka,** M. M. (1970). ***The 16 personality factor test*** *hundbook.* Champaign, IL: **Institute** for Ability and Personality Testing.

Cohen, J. (1994). The **earth** is round **($p < .05$).** *American* ***Psychologist,*** *49. 997-1003.*

Connell, J. P., **&** Tanaka, J. S. (Eds.). (1987). Special section on structural equation modeling. ***Child*** *Development 58(l).* l-75.

Cronbach, L. J.. **& Meehl,** P. **(1955). Construct** validity in psychological tests. *Psycho/o&u/ Bulletin, 52. 281-302.*

Dolan, C. V. (1990). *Biometric decomposition of phenotypic means in human samples.* Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51, 360-392.*

Epstein, S.. & O'Brien, E. J. (19115). The person-situation debate in historical and current perspective. *Psychological Bulletin, 98. 513-537.*

Eysenck, H. J. (1952). *The scientific study of personality.* London: Routledge & Kegan Paul.

Fishbein, M. L., & Ajzen. 1. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psyrhological Review, 81. 59-74.*

Goldberg, L. R. (1990). An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology, 59.* 1216-1229.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48. 26-34.*

Goldberger, A. S. (1972). Structural equation models: An overview. In A. S. Goldberger & 0. D. Duncan (Eds.). *Structural equation models in the social sciences* (pp. I-18). New York: Academic Press.

Guilford, J. P. (1959). *Personality. New* York: McGraw-Hill.

Hebb, D. 0. (1949). *The organization of behavior. New York:* Wiley.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrika, 45,* 153-161.

Heckman, J.. & Robb, R. (1986). Alternativemethods for solving theprobkm of selection bias in evaluating the impact of treatments on outcomes. In H. Wainer (Ed.). *Drawing inferences from self-selected samples* (pp. 63-107). New York: Springer.

Hertzog, C., & Schaie, K. W. (1988). Stahility and change in adult intelligence: 2. Simultaneous analysis of longitudinal means and covariance structures. *Psychology and* Aging, 3. 122-130.

Horn, J. L.. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57, 253-270.*

Horn, J. L., & McArdle, J. J. (1980). Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), *Aging in the 1980s: Psyrhological issues* (pp. 503-541). Washington, DC: American Psychological Association.

Horn, J. L., & McArdk, J. J. (1992). A practical guide to measurement invariance in research on aging. *Experimental Aging Research, 18*(3), 117-144.

Humphreys, L. G. (1962). The organization of human abilities. *American Psycho/o&/, 17,* 475-483.

Jöreskog, K. G,. & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt.

Lawky, D. N. (1943-1944). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh* (Section A), 62, 28-30.

Learner. E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data. New York:* Wiley.

Loehlin, J. C., & Vandenberg, S. G. (1968). Genetic and environmental components in the covariation of cognitive abilities: An additive model. In S. G. Vandenberg (Ed.), *Progess in human behavior genetics (pp.* 261-285). Baltimore: Johns Hopkins University Press.

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least* squares. Heidelberg, Germany: Physica-Verlag.

Martin, N. G., & Eaves. L. J. (1977). The genetical analysis of covariance structures. *Heredity, 38.79-95.*

McArdk. J. J. (1994). Factor analysis. In R. J. Sternberg (Ed.), *The encyclopedia of intelligence* (pp. 422-430). New York: Macmillan.

McArdk, J. J. (1995). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*(4), 409-454.

McArdk, J. J.. & Cattell, R. B. (1994). Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research. 29*(1), *63-l 13.*

McArdk, J. J.. & Goldsmith, H. H. (1990). Alternative common factor models for multivariatc biometric analyses. *Behavior Genetics, 20,* 569-608.

McArdle, J. J., & McDonald. R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *The British Journal of Mathematical and Statistical Psychology, 37, 234-251.*

McArdk. J. J., & Ncssclroade, J. R. (1994). Structuring data to study development and change. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. *223-267).* Hillsdale. NJ: Lawrence Erlbaum Associates.

McArdle, J. J., & Prescott, C. A. (1992). Age-based construct validation using structural quation models. *Experimental Aging Research, 18*(3), *87-l 15.*

McArdle, J. J.. & Woodcock. R. W. (1996). *Extending lest-retest data to examine developmental time-lag components.* Unpublished manuscript. Department of Psychology, University of Virginia.

McCrae, R. R. (1989). Why I advocate the five-factor model: Joint factor analyses of the NEO-PI with other instruments. In D. M. Buss & N. Cantor (Eds.). *Personality psychology: Recent trends and emerging directions* (pp. 237-245). New York: Springer.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29.* 177-185.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58, 525-543.*

Mershon. B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability in real-life criteria? *Journal of Personality and Social Psychology. 55*(4), 675-680.

Mischel, W. (1968). *Personality and assessment. New* York: Wiley.

Mischel, W., & Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review, 89,* 730-755.

Morrison, D. F., & Henkel, R. E. (Eds.). (1970). *The significance /es/ controversy-A reader.* Chicago: Aldinc.

Mosteller, F., & Tukcy, J. W. (1977). *Data analysis and regression: A second course in statistics.* Reading. MA: Addison-Wesley.

Muthèn, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49.* 115-132.

Neale, M. C.. & Cordon. L. R. (1992). *Methodology for genetic studies of twins and families.* Boston: Kluwer Academic.

Nesselroade, J. R. (1983). Temporal selection and factorial invariance in the study of development and change. In P. B. Baltcs & 0. G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol. 5. pp. 60-87). New York: Academic Press.

Nesselroade, J. R. (1988). Some implications of the trait-stale distinction for the study of development over the life-span: The case of personality. In P. B. Baltes, D. L. Featherman. & R. M. Lerner (Eds.), *L&-span development and behavior* (Vol. 8, pp. 163-189). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nesselroade, J. R. (1991a). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 92-105). Washington, DC: American Psychological Association.

Nesselroade, J. R. (1991b). The warp and the woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Pakrmo (Eds.). *Visions of aesthetics, the environment. & development: The legacy of Joachim F. Wohlwill* (pp. 213-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nessclroade, J. R., & Bokcr. S. M. (1994). Assessing constancy and change. In T. P.

Heatherton & J. L. Weinberger (Eds.). *Can personality change? (pp. 121–147).* Washington, DC: American Psychological Association.
Nesselroade, J. R., & Jones. C. J. (1991). Multi-modal selection effects in the study of adult development: A perspective on multivariate, replicated, single-subject, repeated measures. *Experimental Aging Research, I I,* 2 I -27.
Nunnally, J. C. (1967). *Psychometric theory. New* York: McGraw-Hill.
Pearson, K. (1903). On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London* (Section A), *200,* I-66.
Peterson, D. R. (1965). *The scope* and generality of verbally defined personality factors. *Psychological Review, 72, 48-59.*
Plomin, R., Corley, R., DeFries, J. C.. & Fulker, D. W. (1990). Individual differences in television viewing in early childhood: Nature as well as nurture. *Psychological Science, I,* 371–377.
Plomin. R., Corley. R., DeFries, J. C., & Fulker, D. W. (1992). Children's television viewing: Response to Prescott et al. *Psychological Science, 3. 75-76.*
Plomin. R., McClearn, G. E., Smith. D. *L.,* Vignitti, S.. Chorney. M. J., Chorncy, K., Vendetti. C. P., Karsada, S., Thompson, L. A., Detterman, D. K.. Daniles, J.. Owen, M.. & McGuffin, P. (1994). DNA markers associated with high versus low IQ: The IQ quantitative trait loci (QTL) project. *Behuvior Genetics, 24*(2), 107-I 18.
Prescott, C. C.. Johnson, R. C., & McArdle, J. J. (1991). Genetic contributions tô television viewing. *Psychological Science,* 2.430-43 I.
Reese. Ii. W.. & Ovcrton, W. F. (1970). Models of development and theories of development. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 115–145). New York: Academic Press.
Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith. & S. C. Rawlings (Eds.). *Methodological &sues in aging research* (pp. 171–209). New York: Springer.
Rubin, D. B. (1987). *Multiple imputation for nonresponse-in surveys. New York:* Wiley.
Salmon, W. C. (1971). *Statistical explanation ond statistical relevance.* Pittsburgh, PA: University of Pittsburgh Press.
Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into idiographic analysis of personality. *Journal of Personality and* Social *Psychology.* 67, 674-687.
Skinner, B. F. (1950). Arc theories of learning necessary? *Psychological Review, 57,* 193–216.
Steyer, R. (1987). Konsistcnz und Spezifitaet: Definition zweier zentraler Begriffe der Differentiellen Psychologie und ein einfacha Modell zu ihrer Identifikation [Consistency and specificity: Definition of two crucial concepts of differential psychology and a simple model for their identification]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 8, 245-258.*
Strycr. R., & Schmitt, M. J. (1990). The effects of aggregation across and within occasions on consistency, specificity, and reliability. *Methodika,* 4, 58-94.
Tesser, A. (1993). The importance of heritability in psychological research: The case of attitudes. *Psychological Review, 100,* 129–142.
Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.
Thurstone. L. L. (1957). *Mu/riple factor analysis.* Chiigo: University of Chicago Press.
West, B. J. (1985). An *essay on the importance of being nonlinear.* Berlin: Springer-Verlag.
Widaman. K. (1985). Hierarchically nested covariancc structure models for multitrait-multimethod data. *Applied Psychological Measurement. 9.* I-26.
Wittmann, W. W. (1988a, August). *Brunswik symmetry and successfully predicting human behevior.* Paper presented at the 24th International Congress of Psychology. Sydney, Australia.
Wiltmann. W. W. (1988b). hlultivariatc reliability theory: Principles of symmetry and

successful validation strategies. In J. R. Naxlroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 505–560). *New York:* Plenum.
Wiltmann. W. W. (1991). Mets-analysis and Brunswik symmetry. In G. Albrecht & H.-U. Otto (Eds.), *Social prevention and the social sciences: Theoretical controversies, research problems, ond ● voluorion strategies* (pp. 381-393). Berlin: deGruyter.
Wold, H. 0. A. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.). *Quantitative sociology: International perspectives on mathematical and statistical model building* (pp. 307–357). New York: Academic Press.

The West Virginia University Conferences on Life-Span Developmental Psychology

Datan/Greene/Reese: Life-Span Developmental Psychology: Intergenerational Relations
Cummings/Greene/Karraker: Life-Span Developmental Psychology: Perspectives on Stress and Coping
Puckett/Reese: Mechanisms of Everyday Cognition
Cohen/Reese: Life-Span Developmental Psychology: Methodological Contributions
Reese/Franzen: Biological and Neuropsychological Mechanisms: Life-Span Developmental Psychology

BIOLOGICAL AND NEUROPSYCHOLOGICAL MECHANISMS

Life-Span Developmental Psychology

Edited by

Hayne W. Reese
West Virginia University

Michael D. Franzen
Department of Psychiatry
and Allegheny *Neuropsychiatric Institute*
Allegheny General Hospital, Pittsburgh

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1997 Mahwah, New Jersey