Organizational Research Methods

A General Approach for Representing Constructs in Organizational Research
Richard P. Bagozzi and Jeffrey R. Edwards

The online version of this article can be found at:
http://orm.sagepub.com/cgi/content/abstract/1/1/45

Additional services and information for *Organizational Research Methods* can be found at:

Email Alerts: http://orm.sagepub.com/cgi/alerts

Subscriptions: http://orm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 42 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://orm.sagepub.com/cgi/content/refs/1/1/45

A General Approach for Representing Constructs in Organizational Research

RICHARD P. BAGOZZI
University of Michigan Business School

JEFFREY R. EDWARDS
University of North Carolina

A key assumption underlying methods of construct validation is that constructs and their indicators are represented at the appropriate depth (i.e., the specificity versus generality of constructs and their indicators). This article presents a framework that depicts constructs and indicators at various depths and provides guidelines for choosing from among these depths. The framework is then integrated with methods of construct validation based on the confirmatory factor analysis of multitrait-multimethod (MTMM) matrices. The authors apply these methods to the measurement of work values, using the Work Aspect Preference Scale (WAPS). Results show that the WAPS performs better when used to represent relatively specific work values as opposed to more global, general values. Further analyses supported the generalizability of the WAPS factor structure for men and women, although gender differences were found on structured means for several latent value dimensions.

Rating scales are used throughout organizational research to operationalize theoretical constructs and make policy decisions. Among other areas, rating scales form the bases for investigating task performance (Landy & Farr, 1980), job satisfaction (Ironson, Smith, Brannick, Gibson, & Paul, 1989; Robinson, Shaver, & Wrightsman, 1991), organizational citizenship (Organ & Konovsky, 1989), and such antecedents of performance and satisfaction as job characteristics (Hackman & Oldham, 1975; Pierce, McTavish, & Knudsen, 1986), commitment (Allen & Meyer, 1990; Mowday, Steers, & Porter, 1979; Vandenberg & Self, 1993), motivation (Warr, Cook, & Wall, 1979), identification (Mael & Ashforth, 1992), and work values (Pryor, 1987; Ravlin & Meglino, 1987; Super, 1973).

A central concern with any rating scale is its construct validity, or the extent to which the measures constituting a scale capture the underlying construct of interest (Campbell & Fiske, 1959; Cook & Campbell, 1979; Schwab, 1980). A wide variety

of procedures are used in organizational research to establish construct validity, such as reliability estimation and the examination of relationships of the focal measure with measures of other theoretically relevant constructs. These procedures are integrated in the multitrait-multimethod (MTMM) matrix approach, which entails the joint examination of internal consistency, convergence of multiple measures of the same construct, and distinctiveness of measures of different constructs (e.g., Becker & Vance, 1993; Vance, MacCallum, Coovert, & Hedge, 1988; Williams, Cote, & Buckley, 1989).

A premise built into any approach to construct validation, including the MTMM approach, is that constructs are represented at the appropriate depth. By construct depth, we refer to the specificity versus generality of a construct and its associated operationalizations. Specific constructs may refer to narrowly defined phenomena (e.g., satisfaction with pay) or fine-grained aspects of some broader construct (e.g., normative commitment), whereas general constructs typically correspond to global, holistic phenomena (e.g., overall job satisfaction) or combinations of multiple specific constructs (e.g., organizational commitment as reflected in affective, continuance, and normative components). Likewise, specific operationalizations are often individual items phrased in narrow terms (e.g., "It would be very hard for me to leave my organization right now, even if I wanted to"), whereas general operationalizations might represent items worded in broad terms (e.g., "Overall, I am committed to this organization") or aggregates of more specific items (e.g., the sum of responses to items measuring satisfaction with supervisors, subordinates, and coworkers).

In some MTMM applications, analyses are performed at the individual item level, such that items corresponding to a trait-method unit (i.e., a construct measured with one method) are assigned to a single factor, and relationships among factors are examined through the use of either first-order or higher order trait factors and method factors (e.g., Marsh & Hocevar, 1988). In other applications, analyses are performed at the scale level (i.e., averages or sums of items), in which scales representing trait-method units are assigned to both trait and method factors (e.g., Bagozzi & Phillips, 1982). Although these two approaches represent fundamentally distinct construct depths, the rationale for choosing one over another is often unclear.

Moreover, decisions regarding construct depth must often be made in conjunction with decisions as to the *dimensionality* of the construct under examination. For example, MTMM analyses may be used to examine both unidimensional and multidimensional traits. The former focus on discriminant validity between unidimensional traits and other unidimensional constructs, with additional emphasis on the convergence of multiple indicators of a single construct of interest across methods (i.e., the validity diagonals) and perhaps within method (i.e., reliability estimates). The latter address validity for several dimensions of a multidimensional construct, examining not only the convergence and divergence of multiple indicators of each dimension but also whether measures of the multiple dimensions of a construct are more strongly related to one another than to measures of other constructs. These divergent practices not only create confusion as to the manner in which construct validity is established (e.g., the relative importance of convergent and discriminant validity), they also bear little clear correspondence to the use of individual items versus aggregations of items as indicators. To compound matters, the construct depth often shifts from study to study, such that potentially misleading comparisons are made between constructs based on individual items and those based on aggregations of items.

One construct depth is not necessarily superior to another, nor is a unidimensional conceptualization of a construct necessarily better than a multidimensional one. Much depends on the nature of the research question under study and the researcher's purposes. However, in any empirical study, it is essential to be specific as to the depth and dimensionality of constructs and their measures if meaningful results are to be obtained. Although often overlooked, these issues are as fundamental as those pertaining to levels of analysis and the aggregation of observations (House, Rousseau, & Thomas-Hunt, 1995; James, 1982; Klein, Dansereau, & Hall, 1994; Ostroff, 1993; Rousseau, 1985), in that the depth and dimensionality of a measure directly affect the theoretical generalizations that may be drawn from the measure and the criteria used to establish its construct validity. Until now, we have lacked a framework for specifying measures and interpreting construct validity across both depths of measurement and dimensional specificity.

The purpose of this article is to develop a general approach to representing constructs in organizational research that incorporates the ideas of construct depth and dimensionality and to illustrate this approach using relevant data. We begin with a framework for representing constructs that brings issues of construct depth and dimensionality into clearer focus. Next, we consider the particular method and models used to investigate construct validation in our empirical study. We then present the results of application of the models to the measurement of work values as operationalized by the Work Aspect Preference Scale (WAPS; Pryor, 1983a). We chose work values based on their relevance in organizational research (e.g., Dawis, 1991; Judge & Bretz, 1992; Locke, 1976; Nord, Brief, Atieh, & Doherty, 1990) and on the conceptual premise that values are multidimensional and hierarchically arranged and thus can be meaningfully measured at different degrees of construct depth (e.g., Gati, 1991; Ravlin & Meglino, 1989; Schwartz & Bilsky, 1987). Finally, the interpretation of our findings and broader construct validation issues are addressed. This includes issues of convergent and discriminant validity, generalizability, and tests of hypotheses on means of constructs. Our ideas extend previous work in the psychometric literature (e.g., Bagozzi & Heatherton, 1994; Hull, Lehn, & Tedlie, 1991; Marsh & Hocevar, 1985) by integrating construct breadth and depth, dimensionality, and construct validation within a single, unified framework for representing constructs in organizational research. The psychometric literature has not addressed construct depth, the use of structured means, and examination of the MTMM matrix within the context of hierarchical rating scales the way we have herein, nor has it treated construct breadth as fully as herein.

General Framework For Representing Constructs

The general framework presented here is organized in terms of constructs, operationalizations, and their hierarchical arrangement. This framework is displayed in Table 1, which shows four depths for indicators (i.e., individual items, subsets of items, individual scales, and subsets of scales) and three depths for constructs (i.e., factors interpreted as either components, facets, or global constructs). The framework shows that, for each indicator depth, factors can be also depicted at various depths and, in most cases, hierarchical structures relating factors to one another can be specified. Our discussion of this framework will follow the four depths for indicators shown in Table 1,

Table 1
General Framework for Representing Constructs

	Indicator Depth				
	Individual Items	Item Subsets	Individual Scales	Scale Subsets	
			Construct Depth		
Model			Components	Facets	Global
Total disaggregation					
First order	M_1		$L_{1,1}$		
	M_2			$L_{2,1}$	
	M_3				$L_{3,1}$
Second order	M_4		$L_{4,1}$	$L_{4,2}$	
	M_5		$L_{5,1}$		$L_{5,2}$
	M_6			$L_{6,1}$	$L_{6,2}$
Third order	M_7		$L_{7,1}$	$L_{7,2}$	$L_{7,3}$
Partial disaggregation					
First order		M_8	$L_{8,1}$		
		M_9		$L_{9,1}$	
		M_{10}			$L_{10,1}$
Second order		M_{11}	$L_{11,1}$	$L_{11,2}$	
		M_{12}	$L_{12,1}$		$L_{12,2}$
		M_{13}		$L_{13,1}$	$L_{13,2}$
Third order		M_{14}	$L_{14,1}$	$L_{14,2}$	$L_{14,3}$
Partial aggregation					
First order			M_{15}	$L_{15,1}$	
			M_{16}		$L_{16,1}$
Second order			M_{17}	$L_{17,1}$	$L_{17,2}$
Total aggregation					
First order				M_{18}	$L_{18,1}$

Note. M = manifest variable; L = latent variable. For both manifest and latent variables, the first subscript identifies the model; for latent variables, the second subscript indicates whether the variable is a first-order, second-order, or third-order factor.

which generate four general models for specifying relations among items and factors. There are 18 different cases under these models (each row in Table 1 represents one of the cases).

To facilitate the discussion of our framework, we use the WAPS as an illustration. The WAPS consists of 52 items and measures work-related values, operationally defined as "the qualities of work that people consider important" (Pryor, 1983a, p. 1). In its original version, the WAPS consisted of 13 subscales, wherein each subscale was formed as the sum of responses to four items (Pryor, 1983a, p. 4). Pryor (1987) factor analyzed the 13 sums for various samples and found three factors: nonwork orientation, freedom, and human/personal concern. Thus, the WAPS depicts work values as a multidimensional construct, represents values at multiple depths, and aggregates across both items and subscales. As we shall see, this formulation has implications for construct validation and the interpretation of the scale. Moreover, alternative formulations are meaningful under certain conditions and entail lesser and greater degrees of abstraction than considered by Pryor (1987).

For scales such as the WAPS, it is useful to think of items creating components, components forming facets, and facets composing the entire or global scale. This permits a specification of the following organizing schema. When items are treated individually and factors are represented at the component, facet, or global level (or some combination thereof), we term this the *total disaggregation model*. When factors are represented in the same manner but items are combined (e.g., summed or averaged) into subsets, we term this the *partial disaggregation model*. When sums or averages of all items within a component are treated as indicators and factors are represented at the facet level, global level, or both, we term this the *partial aggregation model*. Finally, when facets are used as indicators and a single, global factor is represented (e.g., overall job satisfaction), we term this the *total aggregation model*. Note that this scheme excludes models in which items and factors are treated at the same depth, given that such models necessarily contain one item per factor and, hence, do not permit estimation of the relationship between factors and items or the full evaluation of construct validity.

Before we develop the particular application of the framework to the WAPS scale, which is rather complex, it is helpful to present a simple example for pedagogical purposes. Figure 1 summarizes four ways that 16 items hypothesized to measure four components of a scale can be represented. Under the total disaggregation model, each component (C_i) is shown as a first-order factor with four items (c_{ij}) directly connected to it. The model is said to be totally disaggregated because each indicator is an individual item, which constitutes the most concrete operationalization for a component of a scale. When items within components are aggregated (e.g., summed or averaged), the aggregates can be used as indicators of the components (see the partial disaggregation model in Figure 1). We call this a partial disaggregation to reflect the fact that items for a component are split into sets of aggregated items. When all items for a component are aggregated and the aggregations serve as indicators of a factor, itself defined as a facet accounting for two or more components, we term this the partial aggregation model (see Figure 1). It is a partial aggregation in the sense that the total scale is represented as composed of multiple facets, each operationalized with indicators formed as the aggregation of items. We can think of the two facets shown under the partial aggregation model as constituting relatively abstract dimensions of the scale. Indeed, a conceptually equivalent representation is possible by introducing two second-order factors into the total or partial disaggregation models, wherein C_1 and C_2 indicate or load on a second-order F_1, and C_3 and C_4 indicate a second-order F_2 (not shown in Figure 1). Under the total aggregation model, which is the most abstract representation of a scale or construct, a single global factor (G) is hypothesized to account for variation in all measures, except for measurement error. The indicators for G could conceivably be formed in any one of a number of ways, depending on the context and the purposes of the researcher. In Figure 1, we have taken each indicator to be an aggregation of all items representing its corresponding facet. But other aggregations are possible (e.g., all 16 items could be randomly grouped into two or more categories and then aggregated within a category). Notice that a special case of the total aggregation model occurs when one sums all items in a scale and uses the sum as the representation of the global construct. Overall job satisfaction is sometimes treated this way. We turn now to a specific application of the framework to the WAPS scale.
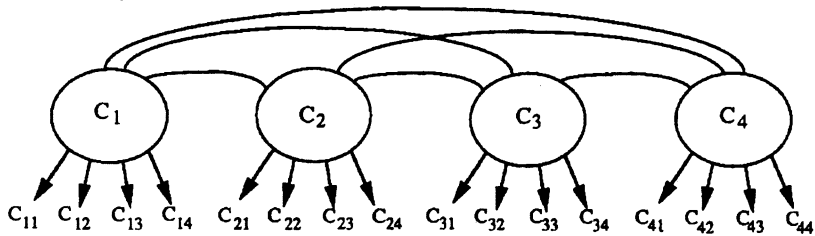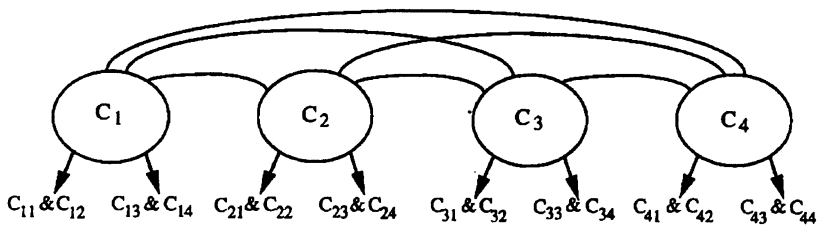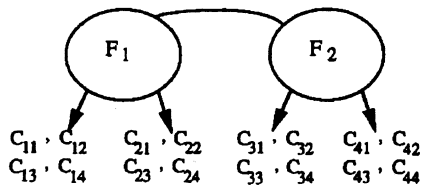
A. <u>Total disaggregation model</u>

B. <u>Partial disaggregation model</u>

C. <u>Partial aggregation model</u> (assumes the pairs C_1 and C_2 and C_3 and C_4 are highly correlated)

D. <u>Total aggregation model</u> (assumes F_1 and F_2 are highly correlated)
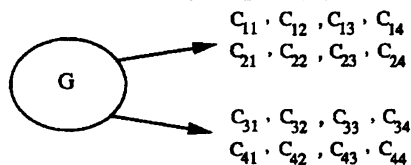
Figure 1: Four Ways of Representing Scales as Applied to the Case of 16 Items and Four Dimensions of the Scale

Note. C_i = component of scale; C_{ij} = measure j of component i; F_k = facet of scale; G = overall scale.

Total Disaggregation Model

The total disaggregation model is the most concrete or *atomistic* representation of construct. In this model, each item is treated as a separate indicator of a component, facet, or global factor. A familiar version of this model is represented by the combina-

tion of M_1 and $L_{1,1}$ (see Table 1), which is a first-order factor model wherein items load on components and the correlations among components are left unanalyzed (i.e., components are allowed to freely correlate). Figure 2 illustrates this model for the WAPS, where we have restricted the presentation to eight components for reasons detailed in the method section below. Following the convention established in the literature for confirmatory factor analysis (CFA) models, latent variables are drawn as ellipses and designated as ξs, measurements are depicted as boxes (for the total disaggregation model, these are individual items from the entire scale; for other models introduced below, they are composites of items), correlations among latent variables are shown as curved lines, the arrows connecting latent variables to measurements and labeled with λs stand for factor loadings, and arrows without origin and terminating at the measurements are disturbances and marked as δs.

The principal advantage of the total disaggregation model is that it gives the most detailed level of analysis of all approaches discussed herein, in that statistical properties are evaluated for each individual item. Estimates of factor loadings indicate the degree of correspondence between each item and its respective factor. The square of each standardized loading shows the amount of variance in each item accounted for by the factor. Furthermore, as shown in the Method section, the factor loadings on items for a component and the error variances for these items can be used to compute composite measures of reliability for each component, analogous to Cronbach's alpha.

Another advantage of the total disaggregation model, and one shared by the partial disaggregation model discussed below, is the ability to specify and test the distinctiveness of multiple components. This is done by examining the correlations between components. To the extent that the correlations are significantly less than 1.00, the components are distinct. The degree of practical distinctiveness depends on the absolute magnitudes of the correlations and the nature of the components under examination. The smaller the correlations, the greater the distinctiveness between components.

The first three specifications of the total disaggregation model shown in Table 1 (i.e., M_1 with $L_{1,1}$, M_2 with $L_{2,1}$, M_3 with $L_{3,1}$) can be characterized as first-order CFA models. One drawback of first-order CFA models is that they confound specific with common variance. Distinct estimates of common variance, specific variance, and error can be obtained by use of a second-order CFA model, whereby the first-order factors load on one or more second-order factors (e.g., M_4 with $L_{4,1}$ and $L_{4,2}$, M_5 with $L_{5,1}$ and $L_{5,2}$, M_6 with $L_{6,1}$ and $L_{6,2}$). The second-order factors parallel those found by Pryor (1987) in his exploratory factor analyses. Unlike Pryor's (1987) approach, however, the CFA models considered herein yield a partitioning of variance of the indicators into error, specific, and common variance. The CFA models also provide estimates of correlations among the higher order factors that are corrected for attenuation. These correlations can be inspected to ascertain the degree of uniqueness achieved among the hypothesized higher order dimensions (i.e., facets) of the WAPS. We should add that this approach presumes that the constructs of interest are not at the component level but instead are at the facet level (i.e., M_4 with $L_{4,1}$ and $L_{4,2}$) or global level (i.e., M_5 with $L_{5,1}$ and $L_{5,2}$, M_6 with $L_{6,1}$ and $L_{6,2}$), given that the first-order factors are treated as latent "indicators" of second-order factors. Specific variance can be further separated by estimating a third-order factor model (i.e., M_7 with $L_{7,1}$, $L_{7,2}$, and $L_{7,3}$), for which the focus is on a single construct at the global level.
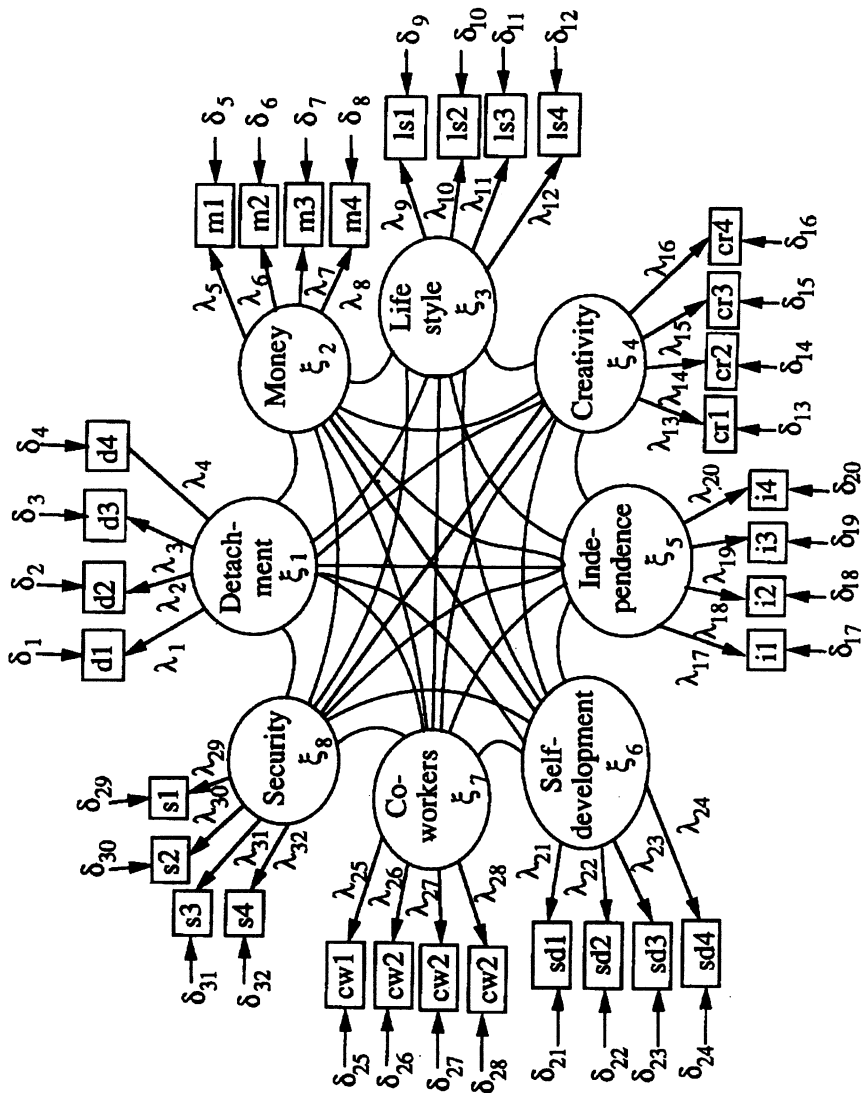
Figure 2: First-Order Total Disaggregation Model

52

The primary disadvantage of the total disaggregation model is that, in comparison with the other models discussed below, measures of factors tend to exhibit greater amounts of measurement error, and correlations among measures tend to be less proportional (i.e., large discrepancies sometimes exist between correlations of items within and across factors). As a consequence, the likelihood of achieving poorly fitting models is greater (e.g., chi-square values will be larger). Indeed, as the number of items per factor increases and sample sizes increase, it is likely that many total disaggregation models will fail to fit the data satisfactorily. Moreover, because the total disaggregation model has the highest number of parameters to be estimated of all models considered herein, it requires a larger sample size to achieve a reasonable ratio of cases to parameter estimates (some researchers recommend that at least 5 cases per parameter should be available (Bentler, 1989, p. 6), and even for well-behaved data, one can seldom perform analyses with confidence with fewer than 3 cases per parameter). From a practical standpoint, the total disaggregation model is likely to be applied successfully in terms of satisfactory fits and stable solutions only when about four or five measures per factor (or fewer) are used. Nevertheless, even with more than five measures per factor, the total disaggregation model can be useful in scale development, item analysis, and the modeling of method effects.

Partial Disaggregation Model

The partial disaggregation model is structurally similar to the total disaggregation model but differs in the way items are handled and indicators are formed. Figure 3 shows a first-order partial disaggregation model applied to the WAPS (corresponding to M_8 with $L_{8,1}$ in Table 1), in which we see that two indicators per component are shown instead of four (cf. Figure 2). For this particular example, each indicator is constructed as the sum (or average) of responses to two of the items intended to measure each component. With more items, three or more could be summed (or averaged) to form indicators and/or three or more indicators per component could be used. Thus, the partial disaggregation model is distinguished by the aggregation of subsets of items and, hence, is a *molecular* approach in contrast to the *atomistic* total disaggregation model.

The principal advantages of the partial disaggregation model are that it reduces the number of parameters to be estimated and, at the same time, tends to decrease measurement error. Thus, in comparison to the total disaggregation model, smaller sample sizes are required and the model generally yields better fit (i.e., the likelihood is less that discrepancies will be found between observed and implied covariances in the computation of chi-square indices). Like the total disaggregation model, the partial disaggregation model also facilitates the evaluation of the distinctiveness of factors (as represented by the magnitudes of the interfactor correlations), the estimation of the reliability of each indicator and of each component treated as a composite of its respective indicators (see the method section), and the specification of second- and third-order factor models to separate common variance, specific variance, and measurement error (see Table 1). Figure 4 presents a second-order model applied to the WAPS and designated at M_{11} with $L_{11,1}$ and $L_{11,2}$ in Table 1. By removing ϕ_{31}, ϕ_{21}, and ϕ_{32} from Figure 4 and adding a single global factor at the top, with $\xi_1 - \xi_3$ loading on it, we would have a third-order factor model (designated M_{14} with $L_{14,1}$, $L_{14,2}$, and $L_{14,3}$ in Table 1).
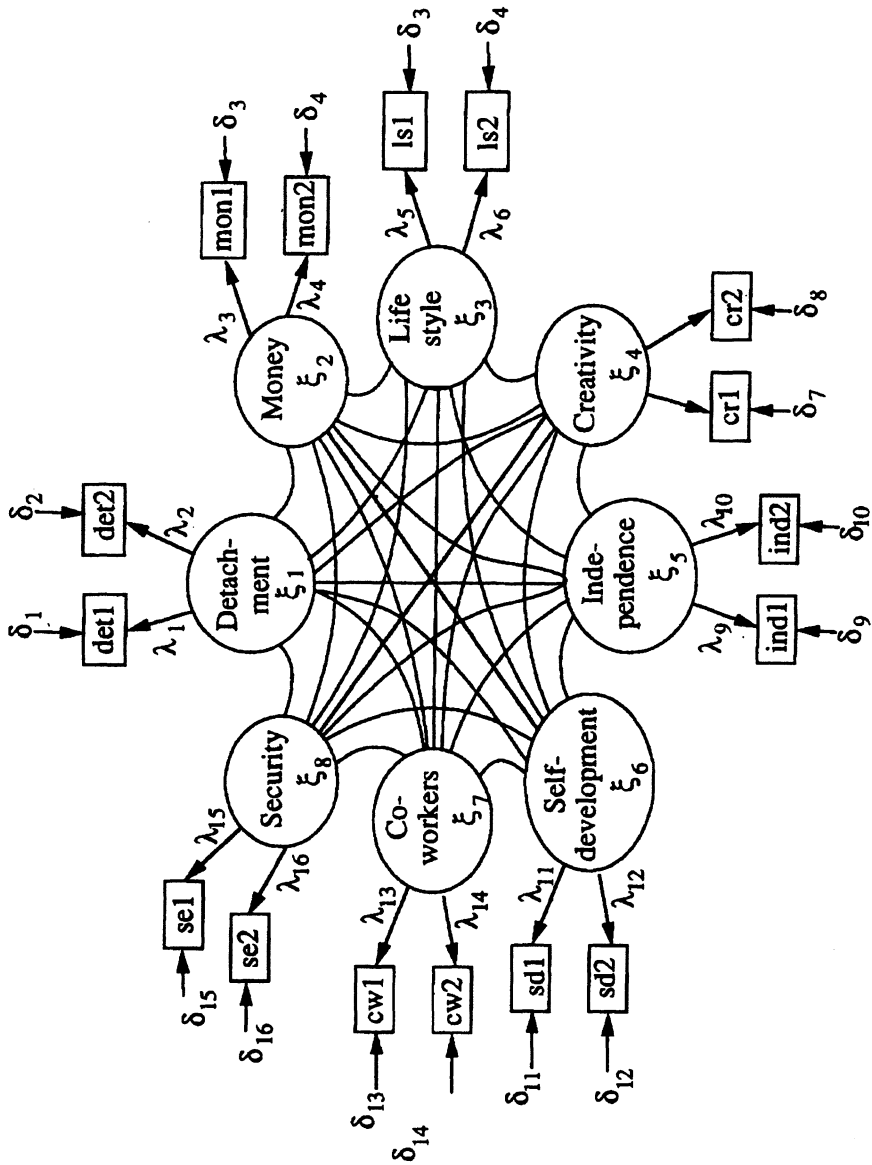
Figure 3: First-Order Partial Disaggregation Model

54

The main disadvantage with the partial disaggregation model lies in the possibility of forming invalid indicators for components. To the extent that the hypothesized items for a component either share more variation with items from other components or fail to load highly and uniformly on the proper component, any combination of items may be misleading. As a result, it is important that items are developed carefully and that exploratory factor analyses at the individual item level support any planned aggregation into indicators before the partial disaggregation model is investigated. Of course, the above caution applies to any aggregation of items or subscales, including those discussed below.

Partial Aggregation Model

The partial aggregation model is more abstract than the partial disaggregation model and might be termed a *molar* level of analysis. Figure 5 depicts a first-order partial aggregation model applied to the WAPS (i.e., M_{15} with $L_{15,1}$). The three factors (i.e., facets) match those found by Pryor (1987) in his exploratory factor analyses. The indicators correspond to the subscales found by Pryor to load highly on the factors. Each indicator is an aggregation (e.g., sum or average) of items from the respective WAPS subscale.

The disaggregation models discussed earlier are particularly suited to fine-grained analyses from which one desires to examine components (e.g., detachment, creativity, security) and obtain detailed information. For example, a researcher concerned about the nature, antecedents, and consequences of specific work values could scrutinize the disaggregation models. In contrast, if one's purposes are more general, such as the examination of overall patterns of work values and their organization into a small number of basic categories (i.e., facets), the partial aggregation model would be appropriate. By loading scales together on a single factor (i.e., M_{16} with $L_{16,1}$) or introducing higher order factors (i.e., M_{17} with $L_{17,1}$ and $L_{17,2}$), it is also possible to investigate constructs at a global level.

The partial aggregation model can be used to investigate the distinctiveness of abstract constructs. For the three facets shown in Figure 5 for the WAPS, this means examining the estimated values for factor correlations to determine whether they are significantly less than 1.00. Furthermore, measures of reliability can be computed for each individual component, as well as for the composites formed by multiple components loading on their respective factors (see method). Moreover, as with the disaggregate models, distinct estimates of error, specific, and common variance can be obtained by use of a second-order partial aggregation model (M_{17} with $L_{17,1}$ and $L_{17,2}$), for which the focus is on a single construct at the global level. In contrast to the disaggregate models for the WAPS, which used three second-order factors, the partial aggregation model requires only a single second-order factor. The number of higher order factors will, of course, depend on the particular scale under study.

The main disadvantage with the partial aggregation model is that it obscures any distinctiveness among components within a facet. Rather, focus is on the facet or global level, which occurs at higher levels of abstraction than components. Again, the appropriateness of levels of analysis depends on one's purposes. With the partial aggregation model, one gives up specificity both to capitalize on increased reliability achieved when components are assembled into wholes and to explore the properties
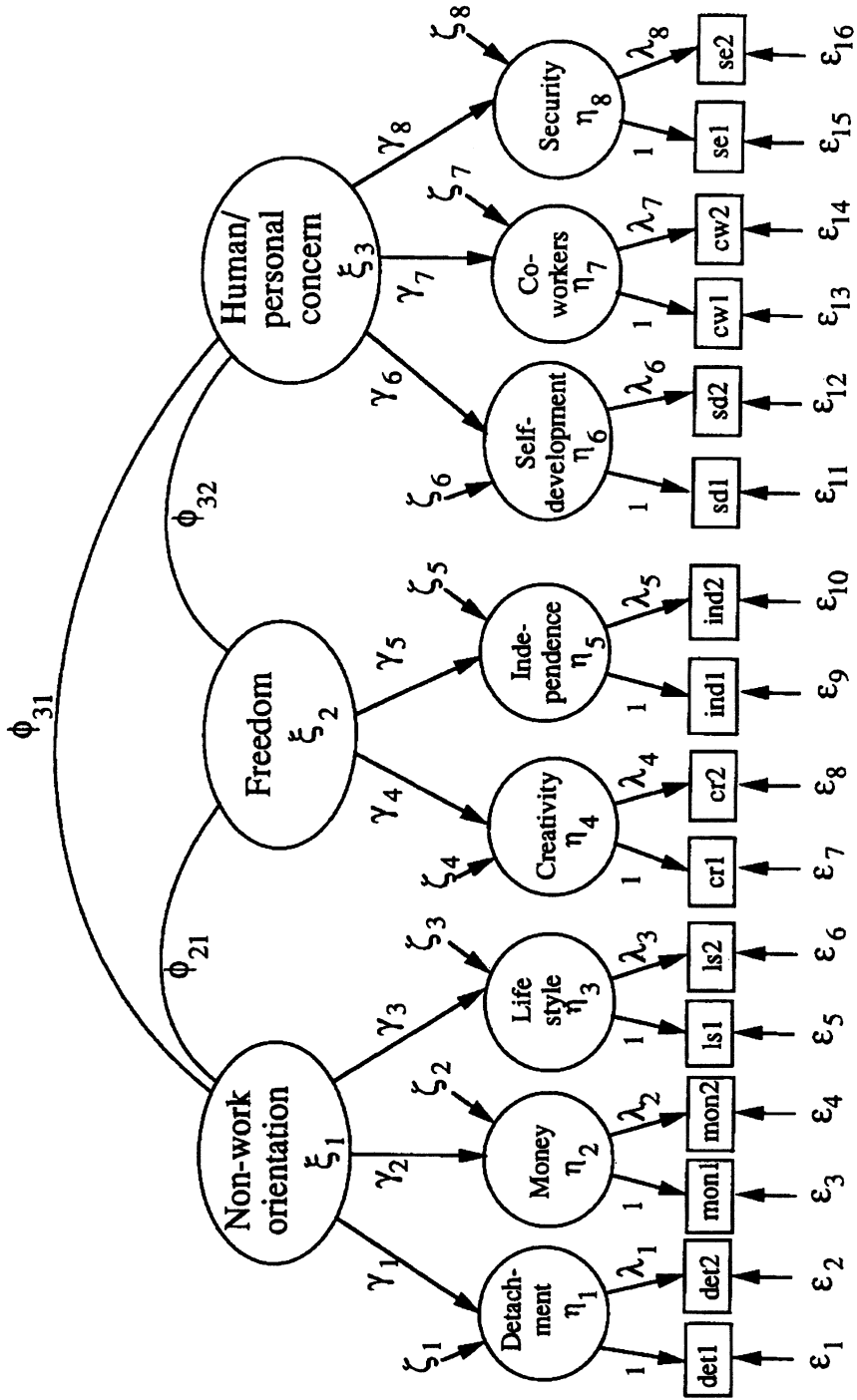
Figure 4: Second-Order Partial Disaggregation Model

of integrated facets of a scale. To date, the WAPS has been used exclusively at the partial aggregation level (e.g., Macnab & Fitzsimmons, 1987; Pryor, 1987).

Total Aggregation Model

On occasion, one might desire abridged or condensed representations of a construct. For instance, inquiry into overall preferences for work or overall job satisfaction might be appropriate for studies of vocational choice (Holland, 1985) or comparisons of work and nonwork domains (Lobel, 1991). When more global or unqualified characterizations of a construct are required, dimensions of a construct can be collapsed into composites of facets or even into a singular entity in what we term the total aggregation model.

For complex scales with many components and items, it may be possible to aggregate across items and components within facets and use the composites so formed for each facet as indicators of one or more global latent variables (M_{18} with $L_{18,1}$). Figure 6 illustrates this conceptualization of total aggregation as applied to the WAPS, depicting each aggregated facet as an indicator of a single work value factor. This model implies that the facets share sufficient common variance to be considered indicators of a singular work values concept. Note that, with only three facets as indicators, the model will fit the data perfectly (i.e., the model is exactly identified). However, this model can be used to advantage in at least two ways. When employed as a predictor or predicted variable, the latent variable representing total aggregation can be used to examine the integrity of the measures of the latent variable—as well as the dependence or implications of the latent variable—when it is part of a larger theory. Second, even with only three indicators, one can investigate the fit of the model to data, as well as examine measurement properties, if certain constraints are imposed. For example, either factor loadings or error variances can be assumed to be equal in some instances (see Bagozzi & Heatherton, 1994, pp. 48-51). Or the reliability of each indicator can be used to estimate the error variances (in which reliability for each sum of items comprising a facet could be computed as an alpha). With more than three facets as indicators, the model will be overidentified, and no additional assumptions would be needed.

The main advantages of the total aggregation model are its simplicity and ability to capture the essence of the underlying meaning of a scale. However, these advantages accrue at the expense of understanding the underlying structure of a scale. Likewise, to the degree that shared variance is high across components or facets, it is possible that a reliable scale can be achieved but be invalid. That is, the construct formed through aggregation may create an artificial entity and be inherently ambiguous (Gerbing & Anderson, 1988; Hattie, 1985; Wolins, 1982).

Additional Issues

Generalizability. One way to augment the hierarchical framework proposed herein is to compare the statistical properties from one sample to the properties from one or more other samples on the same model. To the extent that factor patterns, loadings, error variances, and variances and covariances of factors generalize, the soundness of the framework will be enhanced for the particular context(s) at hand. The benefits of doing this are analogous to those obtained in cross-validations. In our study, we use
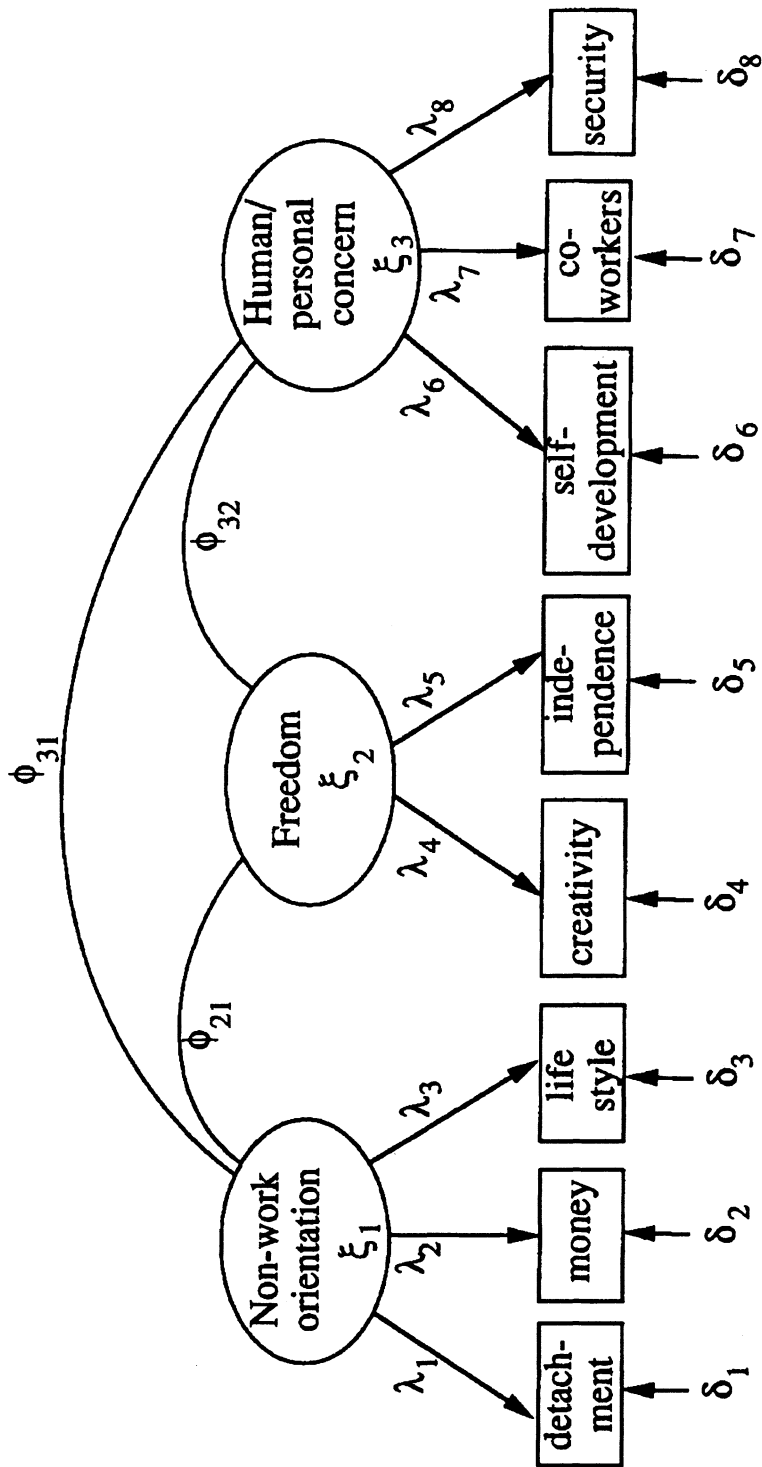
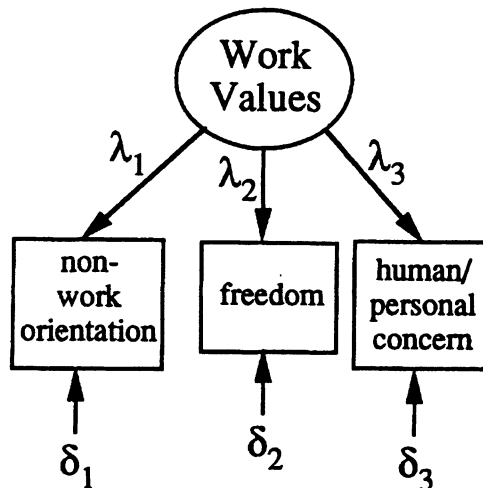**Figure 5:** First-Order Partial Aggregation Model

Figure 6:   Total Aggregation Model

gender as the venue for examining generalizability. Nevertheless, it is important to recognize that other bases for investigating generalizability are available. For data collected at the level of individuals, generalizability might be investigated by use of differences in age, education, profession, position in the organization, or ethnicity. For data collected at the level of small groups or organizations, generalizability might be scrutinized by use of differences in task, technology, industries, geographical location, key processes, or structural factors.

Tests of generalizability are useful in establishing boundary conditions for any construct. That is to say, the validity of a scale or its parts can be tested across contexts or samples of interest, thereby establishing its sense and range of applicability of the constructs. Generalizability refers herein to the degree of equivalence of the representation of a construct across meaningful groups. Some parameters establish the quality of equivalence in a measurement sense (e.g., factor loadings, error variances, variances of factors), whereas other parameters establish equivalence in a relational sense among factors (e.g., factor covariances, regressions of one factor on another). Relational equivalence may be in some research contexts a way to test generalizability, whereas in others it may be part of a hypothesis testing in which differences are expected.

We should note that some aspects of a multidimensional construct may not be expected to generalize and may even be expected to differ across groups. For example, the association between components may be expected to differ systematically across groups according to theoretical criteria. The procedures described herein to test for generalizability can be used to test hypotheses of differences in parameters across groups.

Structured means. Once satisfactory measurement properties of one or more models in the hierarchical framework have been confirmed and invariance of key parameters has been achieved across samples under comparison, it is meaningful to consider differences on average scores for components, facets, or a global scale. This can be done by computation of structured means (e.g., Jöreskog & Sörbom, 1989) and is illustrated in the present study by comparing work values for men and women at different levels of abstraction. This permits the tests of hypotheses on differences between groups on the overall scores on factors, in that mean levels of scores are often expected, depending on the substantive context. The investigation of structured means also establishes boundary conditions of a scale in the sense that generalizability of factor means can be ascertained and related to meaningful, theoretical criteria associated with the samples chosen.

Convergent and discriminant validity. Constructs and rating scales are typically represented with measures derived from a single method. The hierarchical framework developed above and the investigation of generalizability and structured means address issues of measurement and prediction within their monomethod context. However, to address issues of convergent and discriminant validity, it is necessary to employ two or more methods for each construct. The representations depicted in the hierarchical framework proposed herein can be used as a basis for MTMM matrix designs in this regard. We illustrate this extension of the framework below with the WAPS.

The Present Study

We investigated the WAPS from the point of view of the general framework for representing constructs. This permitted examination of work values at different depths and from the perspective of alternative structures. Furthermore, in addition to various measures of internal consistency, we inspected the stability of work values over time. Generalizability of key psychometric parameters and gender differences in overall work values were also explored. Finally, a MTMM matrix analysis was performed on the most promising representation of the WAPS.

Method

Subjects

Three samples were investigated. One sample consisted of 311 men and 126 women MBA students and was used to examine alternative representations of the WAPS. A second sample consisted of 123 men and women MBA students selected from the first sample and followed over time. This provided panel data for investigation of stability of measures of the WAPS over a 1-month period. Finally, a third sample of 259 men and women MBA students was obtained for implementation of the MTMM matrix.

Measures

For Samples 1 and 2, all subjects were administered the 52 items of the WAPS (Pryor, 1983a). We selected subscales based on Pryor's (1987) factor analysis of

Table 2
Description of the Eight WAPS Subscales

Subscale	Description
Detachment	A concern for being able to separate work and its influence from other parts of one's life.
Money	A concern for obtaining large financial rewards from one's work.
Lifestyle	A concern for the effect that employment may have on where and how one lives.
Creativity	A concern for developing something original through one's work.
Independence	A concern for being free from imposed constraints in the work environment.
Self-development	A concern for developing and using one's skills and abilities.
Coworkers	A concern for friendship and understanding from those with whom one works.
Security	A concern for being able to maintain one's job.

Note. WAPS = Work Aspect Preference Scale.

responses by adults. Our selection rule, applied to Pryor's findings, was to choose subscales with loadings of .60 or greater on the three factors and not loading highly on two or more factors (see Table 4, p. 429, in Pryor, 1987). This yielded the subscales (i.e., components) corresponding to detachment, money, lifestyle, creativity, independence, self-development, coworkers, and security shown in Figures 2 through 5. Table 2 provides definitions of the subscales. For the third sample, subjects completed the 52-item WAPS, 45-item Work Values Inventory (WVI, Super, 1970, 1973), and 100-item Minnesota Importance Questionnaire (MIQ, Gay, Weiss, Hendel, Dawis, & Lofquist, 1971). The WAPS, WVI, and MIQ provided the methods for the MTMM matrix analyses. Six traits were common to the three scales and were selected for the MTMM matrix: money, coworkers, creativity, independence, security, and self-development from the WAPS; economic returns, associates, creativity, independence, security, and intellectual stimulation from the WVI; and compensation, coworkers, creativity, independence, security, and ability utilization from the MIQ. Detachment is a subscale only for the WAPS, and no lifestyle subscale exists for the MIQ; hence, these two traits could not be implemented in the MTMM matrix analyses. All responses were recorded on 7-point items, for which subjects were asked to relate how important (not at all to very) each value was for them personally. Each WAPS, WVI, and MIQ subscale consists of four, three, and five items, respectively. Within each scale, items were interspersed on the questionnaire across subscales; for example, item 1 for independence was followed by item 1 for coworkers, then item 1 for self-development, and so forth on the WAPS. The list of items used in this study can be obtained by writing the authors.

Statistical Criteria

All models were fit and hypotheses tested by use of structural equation models (Bentler, 1989; Bollen, 1989; Jöreskog & Sörbom, 1989). The LISREL8 program was used (Jöreskog & Sörbom, 1994).

One measure of fit that was employed is the likelihood ratio χ^2 statistic, which can be used to test the null hypothesis that a specified model reproduces the population covariance matrix of the observed variables. By convention, an acceptable model is

one in which the p value is greater than or equal to .05. Reliance on the χ^2 test as the sole measure of fit is not recommended because of its dependence on sample size. For example in large samples, even trivial deviations of a hypothesized model from a true model can lead to rejection of the hypothesized model; or for very small samples, large deviations of a hypothesized model from a true model may go undetected. Therefore, it is desirable to examine other measures of fit not as sensitive to sample size (Gerbing & Anderson, 1992). Another drawback with the χ^2 test is that it does not directly provide an indication of the degree of fit such as is available with indices normed from 0 to 1.

Another approach to the assessment of goodness of fit is to use an index that is based on the comparison of the fit of a hypothesized model to the fit of a baseline model, such as a null model in which all variables are uncorrelated (i.e., only error variances are estimated). Such comparisons can be used to calculate incremental fit indices, which contrast the fit of a hypothesized model to that of a more restricted, nested model. Currently, the most highly recommended fit index is the Comparative Fit Index (CFI) developed by Bentler (1990; see also McDonald and Marsh, 1990):

$$\text{CFI} = \frac{(\chi_0^2 - df_0) - (\chi_f^2 - df_f)}{\chi_0^2 - df_0}, \tag{1}$$

where χ_0^2 and χ_f^2 are for the null and focal models, respectively, and df stands for degrees of freedom. The CFI is normed in the population and thus has values bounded by 0 and 1, except when $\chi_f^2 < df_f$, in which case one can follow the convention established by Bentler (1990) and report the CFI = 1.00. The CFI provides an unbiased estimate of its corresponding population value, and therefore it should be independent of sample size. Monte Carlo studies show that the CFI performed well for sample sizes varying from 50 to 1,600, in the sense of producing estimates that were unbiased and low in variability (e.g., Bentler, 1990).

From an intuitive perspective, the CFI can be thought of as a measure of how much variation in measures is accounted for from a practical standpoint. A rough rule of thumb is that the CFI should be greater than or equal to about .90, where values less than .90 suggest that significant amounts of variance remain to be explained, and values greater than or equal to .90 imply that further relaxation of parameter constraints is not warranted and might lead to overfitting (Bentler & Bonett, 1980).

Although there are no universal guidelines for evaluating alternative models, we have chosen to use the general approach recently proposed by Marsh (1990, pp. 678-679). Marsh suggested that three steps be employed when evaluating the goodness of fit of any model. First, one should ascertain whether the solutions are well defined. In particular, the iterative procedures in LISREL should converge to a proper solution, parameter estimates should lie within their permissible ranges, and standard errors of parameter estimates should not be too large. Second, parameter estimates should be examined in relation to the substantive, a priori model and common sense. Finally, the χ^2 test and the index mentioned above should be evaluated and compared to alternative models where appropriate. If one desires further to reward model parsimony or penalize model complexity, either the Tucker and Lewis or root mean squared error of approximation can be computed. Newcomb and Bentler (1988) provide additional perspective on assessing model fit.

Tests of hypotheses. Chi-square difference tests were used to test hypotheses concerning the equivalence of models and parameters across gender (e.g., Anderson &

Gerbing, 1988; see also Meredith, 1993). A sequence of hypotheses was examined to explore generalizability in this sense. Marsh (1994) notes that "there is no clear consensus in recommendations about the ordering" of hypotheses concerning invariance constraints and the "choice of a particular ordering . . . must be evaluated in relation to the aims of a particular study" (p. 14). There is consensus, however, on the first two steps: One should begin with a test of invariance of variance-covariance matrices and, then, if the matrices are found to differ, tests of invariance of the factor pattern and the factor loadings should be performed. Marsh (1994) points out that "the minimal condition for 'factorial invariance' is the equivalence of all factor loadings in the multiple groups" (p. 11). For the subsequent tests of invariance, we next examine whether error variances are equal across groups, and then we investigate the invariance of factor variances and covariances among factors.

More specifically, the following sequence of invariance constraints is examined. First, a test is performed of the equality of variance-covariance matrices across samples. A failure to reject this hypothesis means that the relations among the measures do not differ across samples. The data could be pooled, and no further cross-sample analyses are warranted on the differences between psychometric properties of the scales (the means of variables can differ even if the psychometric properties of scales are invariant, however). Rejection of the hypothesis of the equality of variance-covariance matrices sets the stage for investigation of specific differences between the samples. The next hypothesis is to scrutinize whether the same factor structure exists for the samples. This is done by doing a multiple sample analysis on the models (e.g., Jöreskog & Sörbom, 1989). A satisfactory fit implies that the identical factors are reasonable representations of the data in both samples. Given equal factor patterns, we turn to a test of whether the factor loadings are equal across samples. Invariant factor loadings show that the measures indicate the same factors and do so in an equal way. In other words, the evidence supports the conclusion that the hypothesized relationships between measures and constructs generalize across samples. This is tested with a chi-square difference test comparing the chi-squares for the equal factor loadings model to the equal factor pattern model. Given equal factor loadings, it is meaningful to look at the equality of error variances. This is done by comparing the chi-square for the model with both factor loadings and error variances constrained to be equal across samples to the chi-square for the model in which only factor loadings are constrained to be equal. Next, a test can be performed on the invariance of factor variances and covariances. This is accomplished by constraining the ϕ_{ii}s and/or ϕ_{ij}s to be equal across samples and comparing the chi-square test so obtained to the chi-square for the model in which factor loadings and error variances are constrained to be equal across samples.

The aforementioned hypotheses examine the generalizability of psychometric properties. The tests must be performed on the covariance matrices as input (see Cudeck, 1989).

A brief digression is in order with respect to tests of invariance. On occasion, the test of invariance for factor loadings treated as a full set will be rejected. This implies that one or more factor loadings differ across groups. When this occurs, it is desirable to test for "partial invariance" of factor loadings by examining the invariance of individual loadings to pinpoint which ones are invariant and which ones are not (e.g., Byrne, 1994). At least one factor loading per factor, preferably more, should be invariant to conclude that a reasonable degree of equivalence of correspondence

between factors and measures is achieved across groups. This is necessary to rule out rival hypotheses as accounting for any differences found in subsequent tests of invariance on one or more other categories of parameters. For example, if one performed a test of equivalence on the covariance between two factors across groups and found that the covariance differed across the groups, this difference would be interpretable only if the test of invariance for factor loadings showed at least partial invariance for the loadings on the two factors. If no loadings were found to be invariant on one or both factors, the difference for the covariance across groups could be due to lack of correspondence between measures and factors, a true difference in covariances, or both. Hence, it is important to establish at least partial invariance of factor loadings before testing for invariance of parameters relating latent variables to each other.

Structured means. The aforementioned tests of hypotheses address goodness of fit of the models and the psychometric properties of key parameters and their gener- alizability. The parameters so considered are relational criteria. It is also interesting to examine mean differences between men and women on factors (e.g., on components or facets). In the past, it was common to perform univariate or multivariate tests directly on measurements. However, such tests are confounded with the psychometric properties of a scale. Observed differences in means could be a function, for example, of true differences and/or group differences in means. As a result, it is advisable to perform tests of differences in factor means after the invariance of key psychometric properties has been established across groups. The structured means procedure is useful in this regard and is described in Jöreskog and Sörbom (1989, pp. 245-253). Briefly, by introducing intercepts for each measurement, it is possible to estimate differences in means for factors across groups.

Reliability. Internal consistency reliabilities can be computed for the composite formed by measures of a factor by use of the following formula:

$$\rho = \frac{(\Sigma\lambda_i)^2 var\,(\xi)}{(\Sigma\lambda_i)^2 var\,(\xi) + \Sigma(\theta_{\delta i})} , \qquad (2)$$

in which λ_i refers to the ith factor loading, ξ is the factor in question, and $\theta_{\delta i}$ is the error variance for the ith measure of ξ. This formula can be used to compute reliabilities of composites of measures for (a) components in the disaggregate models, (b) facets in the partial aggregation model, and (c) global values in the total aggregation model. As indicated earlier, internal consistency reliability can also be ascertained at the individual measure level within the context of the CFA model; that is, the square of standardized factor loadings gives the reliability for the indicator corresponding to the loading. Thus, we can obtain reliabilities for (a) individual items within components under the total disaggregation model, (b) aggregations of subsets of items within components for the partial disaggregation model, (c) aggregations of all items for each facet under the partial aggregation model, and (d) aggregations of all items across facets under the total aggregation model. These are in addition to the reliabilities computed for the composite of indicators for each case (a) through (d). It should be stressed that all the aforementioned reliabilities are lower bounds.

Test-retest reliability or stability can be ascertained for subscales by use of structural equation models. Specifically, each CFA factor for a subscale at Time 2 can be specified as an endogenous latent variable, and its corresponding factor at Time 1 can be

specified as a latent exogenous variable. The standardized coefficient linking these latent variables yields a measure of stability for the latent variables for which measurement error in the indicators has been corrected. Equivalently, one can investigate a two-factor CFA model for constructs measured at two points in time and examine the correlation between factors.

MTMM Matrix analyses. With one exception, we followed the multistep strategy for analyzing MTMM matrix data proposed by Bagozzi (1993) and recently discussed by Becker and Vance (1993). That is, we performed a CFA with and without method effects, a correlated uniqueness (CU) representation of method effects, and a direct product (DP) model in which traits and methods interact (see the discussion section below). In the interests of brevity, we have chosen not to present the classic MTMM matrix analysis due to Campbell and Fiske (1959). The procedures noted above overcome limitations inherent in the classic method (Widaman, 1985).

For the CFA with and without method factors, each trait (i.e., component) is modeled as a factor, with measures from each scale—the WAPS, WVI, and MIQ— serving as indicators. The model without method factors is called the *trait-only* model and hypothesizes that variation in measures can be explained by the underlying traits plus error. The model with method factors—the *trait-method* model—hypothesizes that variation in measures can be explained by the underlying traits plus method and error. Traits and methods are modeled as additive effects. In sum, for the CFA models, we have six traits (i.e., money, coworkers, creativity, independence, security, and self-development) and three methods (i.e., the WAPS, WVI, and MIQ). It may be informative to perform the CFA trait-method model once with methods correlated and once with methods uncorrelated.

In practice, CFA models of MTMM data sometimes yield ill-defined solutions and may be inappropriate. Marsh (1989) identified four types of ill-defined solutions occurring in CFA investigations of MTMM data:

> underidentified or empirically underidentified models . . . , failures in the convergence of the iterative procedure used to estimate parameters, parameter estimates that are outside their permissible range values (e.g., negative variance estimates called Heywood cases), or standard errors of parameter estimates that are excessively large. (p. 339)

One study even suggests that ill-defined solutions might be expected in a majority of cases (Marsh & Bailey, 1991). As a remedy, Marsh proposed the CU model (see also Kenny, 1976). The CU model consists of the trait-only CFA model plus correlations among error terms corresponding to common method biases. One assumption of the CU model that may pose a problem in some studies is the requirement that methods be uncorrelated.

The CFA and CU models both assume that trait and method variance are additive. To test for the possibility that traits and methods interact, the DP model is useful (e.g., Bagozzi, 1993; Becker & Vance, 1993; Browne, 1984, 1989). It is beyond the scope of this article to discuss the DP model in detail, but fortunately, a user-friendly program is available called MUTMUM (Browne, 1990; see also Wothke & Browne, 1990). We also caution that a number of subtle differences exist across the CFA, CU, and DP models, and that the nature of traits and methods should be taken into account when choosing one model or the other and interpreting findings. A number of issues remain

unresolved at this time, and you are referred to the references cited above (see also Bagozzi, Yi, & Nassen, in press; Wothke, 1996).

Results

Before we present the results for our empirical illustrations, we wish to note that it may not be feasible or even necessary to examine all models summarized in Table 1. Sometimes one's theory will be formed only at a particular level of analysis, and therefore one may desire to test this specification directly. Other specifications at different levels of analysis may or may not be meaningful, depending on the researcher's purposes. Also, when a more detailed model is found to fit satisfactorily, it may not be interesting or informative to investigate a more abstract model. For example, if the total disaggregation model is found to fit satisfactorily, little would be gained by examining the partial disaggregation model. We present a relatively full summary of analyses for the WAPS below for purposes of illustration but caution that not all analyses would be necessary in many instances in the general case.

Comparison of Covariance Matrices for Men and Women

Covariance matrices used as input for the total disaggregation, partial disaggregation, partial aggregation, and total aggregation models were compared for men and women, using a chi-square difference test. Results indicated that the covariance matrices for men and women differed for the total disaggregation model but not for the partial disaggregation model or the partial or total aggregation models (the findings are discussed later under headings for the appropriate models). To simplify the presentation of results, findings for the aggregation models are presented only for the pooled samples. To compare findings on structured means for men and women under the total disaggregation model to the partial disaggregation model, we analyze men and women separately for the latter model, even though the data could be pooled here for analyses of measurement properties of the model.

Reliability

Internal consistency reliability was determined for each component by use of Cronbach's alpha. Table 3 shows the reliabilities for men and women in the cross section and for the panel data at both points in time. All reliabilities are generally quite high with the exception of measures of lifestyle in the cross section for women, where $\alpha = .61$.

Test-retest reliability, or stability, over the 1-month period was determined two ways (see right panel of Table 3). The Pearson product-moment correlations show relatively high stability. Again, measures of lifestyle were the least stable ($r = .65$). Nevertheless, the correlations corrected for attenuation indicate very high stability indeed, with a range from .77 to .89.

Total Disaggregation Model

The seven total disaggregation models shown in Table 1 were estimated separately for men and women. The fit of each model was evaluated in an absolute sense and by

Table 3
Reliability of Measures for
Dimensions of the Work Aspect Preference Scale

| | Internal Consistency Reliability | | Test-Retest Reliability (N = 123) | | | |
| | Cross-Sectional Data | | Panel Data (N = 123) | | | |
Subscale	Men (N = 311)	Women (N = 126)	Time 1	Time 2	Pearson Correlations	Correlations Corrected for Attenuation
Detachment	.92	.93	.95	.96	.74	.77
Money	.83	.81	.88	.89	.76	.89
Lifestyle	.75	.61	.83	.85	.65	.82
Creativity	.83	.87	.94	.89	.81	.88
Independence	.75	.76	.89	.87	.70	.83
Self-development	.73	.78	.84	.85	.70	.84
Coworkers	.84	.86	.91	.90	.74	.82
Security	.94	.95	.94	.93	.78	.84

Note. All correlations are significant at the $p < .001$ level or better.

conducting chi-square difference tests for nested models. Nested models were identified using two principles. First, if the factors in a model are collapsed to yield a higher level of abstraction, then the latter model is nested within the former model, given that collapsing factors is equivalent to fixing their correlations to unity. Thus, examining the models in Table 1 shows that M_3 with $L_{3,1}$ is nested within M_2 with $L_{2,1}$, which in turn is nested within M_1 with $L_{1,1}$ (note that transitivity holds for nested models, such that M_3 with $L_{3,1}$ is also nested within M_1 with $L_{1,1}$). Likewise, M_5 with $L_{5,1}$ and $L_{5,2}$ is nested within M_6 with $L_{6,1}$ and $L_{6,2}$. Second, if a model is respecified by imposing higher order factors to account for the correlations among a set of lower order factors, then the latter model is nested within the former model. This is based on the notion that using a factor to account for a set of interitem correlations effectively imposes proportionality constraints on the correlations among the items, which may be expressed as a set of vanishing tetrads (cf. Bollen & Ting, 1993). Thus, M_4 with $L_{4,1}$ and $L_{4,2}$, and M_5 with $L_{5,1}$ and $L_{5,2}$ are both nested within M_1 with $L_{1,1}$; M_6 with $L_{6,1}$ and $L_{6,2}$ is nested within M_2 with $L_{2,1}$; and M_7 with $L_{7,1}$ and $L_{7,2}$ is nested within M_4 with $L_{4,1}$ (the latter two pairs of models will yield identical fit, given that the higher order factor uses three loadings to account for three correlations among the lower order factors). Results of these analyses indicated that, for both men and women, M_1 with $L_{1,1}$ fit the data significantly better than the remaining six models (all $p < .05$). Therefore, the following discussion will focus on M_1 with $L_{1,1}$ for men and women (see Tables 4 and 5). Looking first at the results for women, we see that, although the chi-square goodness of fit is significant ($\chi^2(436, N = 126) = 646.21, p \cong .00$), CFI = .90, and thus the model fits the data satisfactorily as a practical matter. Factor loadings are generally high and error variances are low to moderate in most cases. Hence, true score variance is satisfactory. Correlations among factors are low to moderate and thus reveal that the components are unique.

Looking next at the findings for men, we again see that, although the chi-square goodness of fit is significant ($\chi^2(436, N = 311) = 913.54, p \cong .00$), the model fits the data satisfactorily from a practical point of view (CFI = .90). All factor loadings, save

Table 4
**Findings for the First-Order Factor Model for the Total Disaggregation
Model of the Work Aspect Preference Scale (Women, N = 126)**

Model	Goodness of Fit
Null model	χ^2(496, N = 126) = 2508.26; $p \cong$ .00
Eight factor model	χ^2(436, N = 126) = 646.21; $p \cong$ .17; CFI = .90

Detachment	Money	Lifestyle	Creativity	Independence	Self-Development	Co-workers	Security
				Factor loadings			
.75 (.08)	.88 (.08)	.50 (.09)	.78 (.08)	.53 (.10)	.87 (.08)	.70 (.09)	.91 (.07)
.83 (.07)	.62 (.09)	.54 (.09)	.82 (.08)	.67 (.09)	.73 (.08)	.78 (.08)	.91 (.07)
.96 (.07)	.75 (.08)	.73 (.09)	.79 (.08)	.51 (.10)	.71 (.08)	.63 (.09)	.95 (.07)
.80 (.08)	.60 (.09)	.85 (.08)	.77 (.08)	.73 (.09)	.55 (.09)	.75 (.08)	.77 (.08)
				Error variances			
.44 (.06)	.23 (.07)	.75 (.10)	.40 (.06)	.72 (.10)	.25 (.07)	.51 (.08)	.17 (.03)
.31 (.05)	.61 (.09)	.71 (.10)	.32 (.06)	.56 (.10)	.46 (.07)	.39 (.08)	.18 (.03)
.07 (.03)	.44 (.07)	.46 (.08)	.38 (.06)	.74 (.11)	.50 (.08)	.60 (.09)	.10 (.02)
.36 (.05)	.64 (.09)	.29 (.08)	.40 (.06)	.46 (.10)	.70 (.10)	.44 (.08)	.40 (.06)

Factor correlations

Detachment	1.00						
Money	.23 (.09)	1.00					
Lifestyle	.40 (.09)	.02 (.11)	1.00				
Creativity	−.05 (.10)	.18 (.10)	−.18 (.10)	1.00			
Independence	.16 (.11)	.25 (.11)	.24 (.11)	.40 (.10)	1.00		
Self-development	.08 (.10)	.30 (.10)	−.11 (.11)	.49 (.08)	.31 (.11)	1.00	
Coworkers	.20 (.10)	.05 (.11)	.22 (.11)	.18 (.10)	.09 (.11)	.20 (.10)	1.00
Security	.34 (.08)	.35 (.09)	.31 (.09)	−.01 (.10)	.15 (.11)	.21 (.10)	.39 (.09) 1.00

Note. Standard errors are in parentheses.

one (λ = .37 for the first measure of independence), are generally high, and error variances are low to moderate in most instances (except for the first measure of independence, where θ_ϵ = .86). Therefore, true score variance is satisfactory. Correlations among factors are low to moderate and show similar values and patterns in comparison to the findings for women. The components are distinct.

Partial Disaggregation Model

A series of nested model tests were performed for the seven partial disaggregation models, analogous to those conducted for the total disaggregation models. These tests indicated that M_8 with $L_{8,1}$ fit the data significantly better than the other six models. Thus, we focused primarily on this model (see Tables 6 and 7) but also report the results for the best-fitting, second-order model (M_{11} with $L_{11,1}$ and $L_{11,1}$) for illustrative purposes.

Table 6 presents the results for women, where it can be seen that the partial disaggregation model (see Figure 2) fits the data well: χ^2(76, N = 126) = 87.69, $p \cong$

Table 5
Findings for the First-Order Factor Model for the Total Disaggregation
Model of the Work Aspect Preference Scale (Men, $N = 311$)

Model	Goodness of Fit
Null model	$\chi^2(496, N = 311) = 5434.11; p \cong .00$
Eight factor model	$\chi^2(496, N = 311) = 913.54; p \cong .00; CFI = .90$

Detachment	Money	Lifestyle	Creativity	Independence	Self-Development	Co-workers	Security
				Factor loadings			
.73 (.05)	.87 (.05)	.59 (.06)	.73 (.05)	.37 (.06)	.83 (.05)	.74 (.05)	.90 (.04)
.82 (.05)	.47 (.06)	.58 (.06)	.82 (.05)	.80 (.06)	.71 (.05)	.66 (.05)	.93 (.04)
.92 (.04)	.71 (.05)	.76 (.05)	.68 (.05)	.50 (.06)	.71 (.05)	.65 (.05)	.93 (.04)
.83 (.05)	.72 (.05)	.84 (.05)	.77 (.05)	.76 (.06)	.49 (.06)	.85 (.05)	.78 (.05)
				Error variances			
.46 (.04)	.24 (.05)	.65 (.06)	.47 (.05)	.86 (.07)	.31 (.05)	.46 (.05)	.18 (.02)
.32 (.03)	.78 (.07)	.66 (.06)	.32 (.04)	.36 (.06)	.49 (.05)	.56 (.05)	.14 (.02)
.15 (.02)	.50 (.05)	.42 (.05)	.53 (.05)	.75 (.07)	.50 (.05)	.57 (.05)	.13 (.02)
.31 (.03)	.49 (.05)	.29 (.05)	.41 (.04)	.42 (.06)	.76 (.06)	.29 (.04)	.40 (.03)

				Factor correlations			
Detachment	1.00						
Money	.16 (.06)	1.00					
Lifestyle	.50 (.05)	.17 (.07)	1.00				
Creativity	.13 (.06)	.19 (.07)	.20 (.07)	1.00			
Independence	.25 (.06)	.22 (.07)	.28 (.07)	.44 (.06)	1.00		
Self-development	.04 (.07)	.23 (.07)	.10 (.07)	.53 (.05)	.11 (.07)	1.00	
Coworkers	.33 (.06)	.09 (.07)	.26 (.06)	.13 (.07)	.12 (.07)	.30 (.06)	1.00
Security	.48 (.05)	.31 (.06)	.45 (.05)	−.02 (.06)	.11 (.07)	.10 (.06)	.27 (.06) 1.00

Note. Standard errors in parentheses.

.17, CFI = .99. All factor loadings are high, and error variances are generally low. The correlations among the eight factors are low to moderate in value and thus reveal uniqueness among the components.

Table 7 shows the findings for men. Although the chi-square goodness of fit is significant ($\chi^2(76, N = 311) = 136.75, p \cong .00$), the measure of practical relevance suggests a satisfactory model (i.e., CFI = .97). Factor loadings are high, and error variances are generally low. The correlations among factors are low to moderate and point to the conclusion of distinct components. The pattern and magnitude of correlations among components are similar for men and women. Moreover, the components show similar intercorrelations whether we look at the total or partial disaggregation models.

Table 8 summarizes the results for the second-order partial disaggregation model (see Figure 4). Although the chi-square test was significant ($\chi^2(93, N = 437) = 297.63, p \cong .00$), the CFI demonstrated that the model fits well as a practical matter (i.e., CFI = .94). Inspection of Table 8 reveals that random error is low, except for the second

Table 6
Findings for the First-Order Factor Model for the Partial Disaggregation Model of the
Work Aspect Preference Scale (Women, $N = 126$)

Model	Goodness of Fit
Null model	$\chi^2(120, N = 126) = 10688.50; p \cong .00$
Eight factor model	$\chi^2(76, N = 126) = 87.69; p \cong .17; CFI = .99$

Detachment	Money	Lifestyle	Creativity	Independence	Self-Development	Co-workers	Security
				Factor loadings			
.95 (.08)	.87 (.10)	.69 (.10)	.82 (.08)	.82 (.10)	.80 (.09)	.86 (.09)	.89 (.07)
.91 (.08)	.78 (.10)	.64 (.10)	.93 (.08)	.74 (.10)	.80 (.09)	.87 (.09)	1.00 (.07)
				Error variances			
.09 (.08)	.24 (.12)	.52 (.11)	.33 (.08)	.33 (.13)	.36 (.10)	.26 (.11)	.21 (.05)
.18 (.08)	.40 (.11)	.59 (.11)	.14 (.09)	.45 (.11)	.37 (.10)	.24 (.11)	.00 (.06)

Factor correlations

Detachment	1.00						
Money	.25 (.10)	1.00					
Lifestyle	.48 (.10)	.13 (.12)	1.00				
Creativity	.01 (.10)	.25 (.10)	−.15 (.12)	1.00			
Independence	.28 (.10)	.24 (.11)	.46 (.11)	.30 (.10)	1.00		
Self-development	.03 (.10)	.34 (.10)	−.11 (.13)	.54 (.08)	.20 (.11)	1.00	
Coworkers	.23 (.10)	.06 (.11)	.27 (.12)	.18 (.10)	.10 (.11)	.19 (.11)	1.00
Security	.35 (.08)	.36 (.09)	.40 (.10)	−.03 (.09)	.20 (.10)	.23 (.10)	.40 (.08) 1.00

Note. Standard errors in parentheses.

measure of lifestyle ($\theta_\delta = .55$) and the first measure of self-development ($\theta_\delta = .67$), which exhibit moderate levels of error. The measures of detachment, lifestyle, and security have approximately equal amounts of specific and common variance. The measures of money, creativity, and coworkers have high amounts of specific and low amounts of common variance. The measures of independence have high amounts of common and low amounts of specific variance. The second measure of self-development has very high specific variance and low common variance. Inspection of the factor intercorrelations in Table 8 reveals that nonwork orientation and human/ personal concern are not distinct ($\phi = .91$, $SE = .06$); nonwork orientation and freedom are distinct ($\phi = .47$, $SE = .10$), as are freedom and human/personal concern ($\phi = .25$, $SE = .08$).

· Similar results were found for the total disaggregation model in terms of error, specific, and common variance and in terms of factor intercorrelations. However, the total disaggregation models demonstrated poor fits for the second-order factor models and, therefore, the results are not presented here in the interests of brevity.

Partial Aggregation Model

The three partial aggregation models shown in Table 1 were estimated. Because only three facets were specified, M_{15} with $L_{15,1}$ yielded the same fit as M_{17} with $L_{17,1}$

Table 7
Findings for the First-Order Factor Model for the Partial Disaggregation Model of the
Work Aspect Preference Scale (Men, $N = 311$)

Model	Goodness of Fit
Null model	$\chi^2(120, N = 311) = 2510.96;\ p \cong .00$
Eight factor model	$\chi^2(76, N = 311) = 136.75;\ p \cong .00;\ CFI = .97$

Detachment	Money	Lifestyle	Creativity	Independence	Self-Development	Co-workers	Security
				Factor loadings			
.92 (.05)	.76 (.06)	.86 (.06)	.75 (.06)	.78 (.06)	.68 (.06)	.80 (.06)	.92 (.05)
.93 (.05)	.92 (.07)	.68 (.06)	.93 (.06)	.78 (.06)	.83 (.06)	.90 (.06)	.95 (.05)
				Error variances			
.15 (.04)	.43 (.07)	.25 (.07)	.44 (.06)	.40 (.07)	.54 (.06)	.36 (.07)	.14 (.04)
.14 (.04)	.15 (.10)	.53 (.06)	.14 (.07)	.39 (.07)	.31 (.07)	.18 (.08)	.09 (.04)

				Factor correlations			
Detachment	1.00						
Money	.19 (.06)	1.00					
Lifestyle	.57 (.05)	.29 (.06)	1.00				
Creativity	.12 (.06)	.22 (.06)	.21 (.07)	1.00			
Independence	.29 (.06)	.27 (.07)	.38 (.07)	.43 (.06)	1.00		
Self-development	.13 (.07)	.31 (.07)	.24 (.07)	.57 (.06)	.14 (.07)	1.00	
Coworkers	.36 (.06)	.12 (.06)	.32 (.06)	.10 (.06)	.18 (.07)	.32 (.07)	1.00
Security	.48 (.05)	.32 (.06)	.47 (.05)	.02 (.06)	.13 (.07)	.19 (.07)	.28 (.06)	1.00

Note. Standard errors in parentheses.

and $L_{17,2}$. Therefore, only a single nested model test was performed, comparing M_{16} with $L_{16,1}$ to M_{15} with $L_{15,1}$. This test showed that M_{15} with $L_{15,1}$ yielded a significantly better fit than M_{16} with $L_{16,1}$ ($p < .05$). Thus, we focused our analyses on the former model (see Figure 5). A key finding from this model concerns the intercorrelations among factors, where it was found that nonwork orientation and human/personal concern were not distinct ($\phi = .88$, $SE = .07$), but nonwork orientation and freedom ($\phi = .44$, $SE = .11$) and freedom and human/personal concern ($\phi = .23$, $SE = .08$) were distinct. These findings are similar to those found for the second-order CFA model applied to the disaggregation models. Nevertheless, based on the goodness-of-fit indices, one must reject the partial aggregation model for these data ($\chi^2(17, N = 437) = 150.52$, $p \cong .00$, CFI = .75).

Total Aggregation Model (Figure 6)

Because only three manifest variables were used, the model fit the data perfectly. Therefore, to achieve overidentification, the three factor loadings were constrained to be equal. Results indicated that the fit of the model was poor: $\chi^2(2, N = 437) = 66.78$, $p \cong .00$, CFI = .62.

Table 8
Variance Components for Measures of Partial Disaggregation
Model of the Work Aspect Preference Scale

Measure	Variance		
	Error	Specific	Common
Detachment 1	.14	.47	.40
Detachment 2	.13	.47	.40
Money 1	.37	.54	.09
Money 2	.23	.67	.11
Lifestyle 1	.28	.32	.41
Lifestyle 2	.55	.20	.25
Creativity 1	.45	.49	.06
Creativity 2	.07	.84	.10
Independence 1	.42	.00	.58
Independence 2	.37	.00	.62
Self-development 1	.67	.31	.03
Self-development 2	.00	.92	.08
Coworkers 1	.37	.48	.14
Coworkers 2	.15	.65	.20
Security 1	.11	.44	.45
Security 2	.12	.43	.45
Correlations among factors	1.00		
	.47 (.10)	1.00	
	.91 (.06)	.25 (.08)	1.00

Note. Standard errors in parentheses. Variances do not sum to 1.00 because of rounding error.

Generalizability

Table 9 shows the findings for tests of invariance of parameters across men and women under the total disaggregation model. The test of equality of variance-covariance matrices reveals that $\chi^2(528, N^M = 311, N^W = 126) = 624.95, p \cong .00$. Thus, we must reject the hypothesis that the matrices are equivalent for men and women. The second row in Table 9 indicates that the factor pattern is similar for men and women. That is, the eight factors shown in Figure 1 fit the data satisfactorily for men and women: $\chi^2(872, N^M = 311, N^W = 126) = 1559.75, p \cong .00$, CFI = .90. Next, it can be seen that the hypothesis of equal factor loadings cannot be rejected, $\chi_d^2(32) = 30.82, p > .30$ (see third row in Table 9). Likewise, we cannot reject the hypothesis that error variances are equal, $\chi_d^2(32) = 39.57, p > .10$. Last, the hypothesis of equal correlations among factors cannot be rejected, $\chi_d^2(28) = 35.45, p > .10$.

Table 10 summarizes the results for tests of invariance of parameters across men and women under the partial disaggregation model. The test of equality of variance-covariance matrices shows that $\chi^2(136, N^M = 311, N^W = 126) = 131.12, p \cong .60$. Therefore, we cannot reject the hypothesis that the matrices are equivalent for men and women. The second row in Table 10 demonstrates that the factor pattern is similar for men and women: $\chi^2(152, N^M = 311, N^W = 126) = 224.44, p \cong .60$, CFI = .98. Next, we see that the hypothesis of equal factor loadings cannot be rejected, $\chi_d^2(16) = 12.90, p > .50$. Further, we cannot reject the hypothesis that error variances are equal, $\chi_d^2(16) = 6.69, p > .97$. Finally, the hypothesis of equal correlations among factors cannot be rejected, $\chi_d^2(28) = 35.47, p > .10$.

Table 9
Findings for Multiple-Group Analyses of the Work Aspect
Preference Scale: Tests of Invariance for Men and Women
Under the Total Disaggregation Model

Model	Goodness of Fit	Test of Hypotheses
M_1: Equal variance-covariance matrices	$\chi^2(528, N^M = 311, N^W = 126) = 624.95, p \cong .00$	—
M_2: Baseline (equal factor pattern)	$\chi^2(872, N^M = 311, N^W = 126) = 1559.75, p \cong .00, CFI = .90$	—
M_3: Factor loadings invariant	$\chi^2(904, N^M = 311, N^W = 126) = 1590.57, p \cong .00, CFI = .90$	$M_3 - M_2$: $\chi^2_d(32) = 30.82, p > .30$
M_4: Factor loadings and error variances invariant	$\chi^2(936, N^M = 311, N^W = 126) = 1630.14, p \cong .00, CFI = .90$	$M_4 - M_3$: $\chi^2_d(32) = 39.57, p > .10$
M_5: Factor loadings, error variances, and factor correlations invariant	$\chi^2(964, N^M = 311, N^W = 126) = 1665.59, p \cong .00, CFI = .90$	$M_5 - M_4$: $\chi^2_d(28) = 35.45, p > .10$

Because the aggregation models fit poorly, we did not examine generalizability for these cases.

Structured Means

Given that factorial invariance has been established for the total and partial disaggregation models, it is meaningful to examine gender differences in the level of work values. Table 11 summarizes the results for the total disaggregation model. Here we find that women value detachment, lifestyle, independence, and coworkers more than men. No statistical differences were found between men and women for the value of money, creativity, self-development, and security. Table 12 presents the findings for the partial disaggregation model. The results are similar to the total disaggregation with three exceptions. As before, women value detachment, lifestyle, and independence more than men, and no differences were found for the value of creativity and security. However, women were found to value money less, self-development more, and coworkers equally in comparison to men. These three inconsistencies between the total and partial disaggregation models are modest, however, in that all gender differences here were in the same direction and were fairly close to significance at the .05 level.

MTMM Matrix Analyses

MTMM analyses can be applied to any of the four levels in the hierarchy of models described herein. Obviously, this should be done only for those levels shown to satisfactorily describe the data. For the WAPS data, only the disaggregated models

Table 10
Findings for Multiple-Group Analyses of the Work
Aspect Preference Scale: Tests of Invariance for Men
and Women under the Partial Disaggregation Model

Model	Goodness of Fit	Test of Hypotheses
M_1: Equal variance-covariance matrices	$\chi^2(136, N^M = 311, N^W = 126) = 131.12, p \cong .60$	—
M_2: Baseline (equal factor pattern)	$\chi^2(152, N^M = 311, N^W = 126) = 224.44, p \cong .00, CFI = .98$	—
M_3: Factor loadings invariant	$\chi^2(168, N^M = 311, N^W = 126) = 237.34, p \cong .00, CFI = .98$	$M_3 - M_2$: $\chi^2_d(16) = 12.90, p > .50$
M_4: Factor loadings and error variances invariant	$\chi^2(184, N^M = 311, N^W = 126) = 244.03, p \cong .00, CFI = .98$	$M_4 - M_3$: $\chi^2_d(16) = 6.69, p > .97$
M_5: Factor loadings, error variances, and factor correlations invariant	$\chi^2(212, N^M = 311, N^W = 126) = 279.50, p \cong .00, CFI = .98$	$M_5 - M_4$: $\chi^2_d(28) = 35.47, p > .10$

achieved satisfactory psychometric properties, and therefore, MTMM analyses are only meaningful at these levels.

We limited our MTMM analyses to the partial disaggregate model because, with a sample of $N = 259$, too many parameters had to be estimated for the total disaggregate model (e.g., 225 parameters must be estimated for the CU MTMM analysis). As the measurement properties of the two disaggregate models proved to be quite similar, this should not comprise the MTMM analyses. For tests of hypotheses, items for each subscale were aggregated, and the WAPS, WVI, and MIQ constituted the methods for each subscale.

We applied the CFA model to the MTMM data. The trait-only model fit poorly: $\chi^2(120, N = 259) = 539.42, p \cong .00, CFI = .83$. This result suggests that a significant amount of variance remains unexplained in the data and that the introduction of method effects is warranted, as a hypothesis. Application of LISREL8 to the data and the trait-method model yielded ill-defined solutions, and in fact, the program failed to converge. Therefore, the CU model was tried. Here, a satisfactory fit was obtained: $\chi^2(75, N = 259) = 180.2, p \cong .00, CFI = .96$.

Table 13 summarizes the findings for factor loadings and correlations among factors under the CU model. Trait variance is high for all measures of money, coworkers, creativity, and security, and for the first measure of independence and the first and third measures of self-development. Trait variance is moderate for the third measure of independence and the second measure of self-development. Trait variance is low for the second measure of independence. However, as the factor loading for the second measure of independence is statistically significant, the minimal criterion for conver-

Table 11
Mean Differences for Components of the Work Aspect
Preference Scale: Total Disaggregation Model

Factor	Change in Mean[a]
Detachment	.36 (.11)
Money	−.15 (.12)
Lifestyle	.51 (.12)
Creativity	−15 (.11)
Independence	.30 (.12)
Self-development	.22 (.12)
Coworkers	.27 (.12)
Security	.18 (.11)

Note. Standard errors in parentheses.
a. Means of factors for men fixed to zero.

Table 12
Mean Differences for Components of the Work Aspect
Preference Scale: Partial Disaggregation Model

Factor	Change in Mean[a]
Detachment	.38 (.11)
Money	−.24 (.12)
Lifestyle	.57 (.13)
Creativity	−.15 (.11)
Independence	.28 (.12)
Self-development	.26 (.12)
Coworkers	.20 (.12)
Security	.14 (.11)

Note. Standard errors in parentheses.
a. Means of factors for men fixed to zero.

gent validity is met even here. Hence, convergent validity is established for all measures.

The correlations among factors reveal that discriminant validity has been achieved among all six traits. The highest correlations occur between creativity and self-development ($\phi = .71$, $SE = .06$), creativity and independence ($\phi = .60$, $SE = .06$), and independence and self-development ($\phi = .52$, $SE = .08$). The remaining 12 intertrait correlations are smaller and range from −.13 to .40.

Table 14 presents the variance-covariance matrix for disturbances. Although many covariances are significant, they are generally low in magnitude. Overall, most of the variation in measures is due to traits, and measurement error is relatively low.

To test whether traits and methods interact, the DP model was applied to the data. The findings demonstrate that this model must be rejected: $\chi^2(127, N = 259) = 1070.19$, $p \cong .00$, and CFI = .61.

Table 13
Factor Loadings and Factor Correlations for
Correlated Uniqueness Model (N = 259)

	Factor Loadings					
Measure	Money ξ_1	Coworkers ξ_2	Creativity ξ_3	Independence ξ_4	Security ξ_5	Self-Development ξ_6
WAPS$_1$	.77 (.06)					
WVI$_1$	.77 (.06)					
MIQ$_1$	.63 (.06)					
WAPS$_2$		.83 (.05)				
WVI$_2$		.88 (.05)				
MIQ$_2$		.87 (.05)				
WAPS$_3$			.84 (.05)			
WVI$_3$			.86 (.05)			
MIQ$_3$			.81 (.05)			
WAPS$_4$				.92 (.08)		
WVI$_4$				.28 (.07)		
MIQ$_4$				.46 (.07)		
WAPS$_5$					.87 (.05)	
WVI$_5$					.85 (.05)	
MIQ$_5$					.92 (.05)	
WAPS$_6$						.58 (.07)
WVI$_6$						.42 (.07)
MIQ$_6$						.60 (.07)
	Factor Correlations					
ξ_1	1.00					
ξ_2	.20 (.07)	1.00				
ξ_3	.02 (.07)	.15 (.07)	1.00			
ξ_4	.16 (.07)	.05 (.07)	.60 (.06)	1.00		
ξ_5	.37 (.06)	.40 (.06)	−.13 (.07)	−.09 (.07)	1.00	
ξ_6	.38 (.08)	.30 (.08)	.71 (.06)	.52 (.08)	.07 (.09)	1.00

Note. Standard errors in parentheses. WAPS = Work Aspect Preference Scale; WVI = Work Values Inventory; MIQ = Minnesota Importance Questionnaire.

Discussion

Comments on the WAPS

Overall, the findings show that the WAPS can be represented in a statistically sound way and that its measures achieve convergent and discriminant validity. However, these conclusions depend on the depth of analysis and dimensional characterization of the scale used to depict work values.

The most compelling evidence supports the disaggregation models, where each subscale is modeled as a latent variable and is measured directly by either individual items or aggregations of item pairs. At this relatively low level of abstraction, people were found to hold work values in eight distinct, yet intercorrelated, cognitive categories: detachment, money, lifestyle, creativity, independence, self-development, coworkers, and security. This finding must be tempered, however, by the realization that a common method was used to measure each item. With a single method, it is

Table 14
Variance-Covariance Matrix for Disturbances of
Correlated Uniqueness Model ($N = 259$)

	$WAPS_1$	WVI_1	MIQ_1	$WAPS_2$	WVI_2	MIQ_2	$WAPS_3$	WVI_3	MIQ_3	$WAPS_4$	WVI_4	MIQ_4	$WAPS_5$	WVI_5	MIQ_5	$WAPS_6$	WVI_6	MIQ_6
$WAPS_1$	.46*																	
WVI_1		.42*																
MIQ_1			.59*															
$WAPS_2$	.04			.30*														
WVI_2		.16*			.25*													
MIQ_2			.00			.23*												
$WAPS_3$	.12*			.02			.30*											
WVI_3		.12*			.07*			.29*										
MIQ_3			.10*			.02			.32*									
$WAPS_4$	.12*			.04			.09*			.16								
WVI_4		.09			.00			.11*			.92*							
MIQ_4			.10*			.03			.08*			.77*						
$WAPS_5$	.15*			.04			.10*			.12*			.23*					
WVI_5		.17*			.01			.14*			.05			.30*				
MIQ_5			.06*			.01			.04			.08*			.13*			
$WAPS_6$	.25*			-.02			.13*			.14*			.07*			.66*		
WVI_6		.16*			.10*			.14*			.06			.15*			.83*	
MIQ_6			.04			.03			.07			.08			.02			.63*

Note. WAPS = Work Aspect Preference Scale; WVI = Work Values Inventory; MIQ = Minnesota Importance Questionnaire.
* $p < .05$.

impossible to disentangle trait variance from method variance. On the other hand, when the MTMM analyses were performed on those values common to the WAPS, WVI, and MIQ (i.e., money, creativity, independence, self-development, coworkers, and creativity), convergent and discriminant validity was achieved. In these latter analyses, explicit estimates of variation in traits and measurement error were obtained.

The MTMM analyses provide a relatively weak basis for examination of convergent validity, however, because the WAPS, WVI, and MIQ use similar methods (i.e., self-reports). Similar method might inflate shared variation among items of each subscale. All the same, the MTMM analyses actually provide a strong basis for investigation of discriminant validity, because the likelihood of shared method biases across items and components makes it more difficult to attain discriminant validity than had maximally dissimilar methods been employed. Evidence for distinct work values in the face of likely common method biases is encouraging, indeed. Nevertheless, it is recommended that different methods be employed as well in MTMM analyses whenever possible.

The disaggregate models constitute representations of work values in categories at the most concrete level of abstraction. The values are cognitive categories presumably grounded in experience. The question remains whether values are organized in memory in a hierarchical way. Here, the second-order factor analyses and the aggregation models provide some insight.

The findings generally support a hierarchical organization of work values. Three abstract values were found to capture variation in the eight subscales: nonwork orientation (attachment, money, lifestyle), freedom (creativity, independence), and human/personal concern (self-development, coworkers, security). But it should be stressed that, although freedom was distinct from nonwork orientation and human/personal concern, nonwork orientation and human/personal concern were not distinct at this relatively high level of abstraction. Moreover, although error variance is typically low, the amounts of common and specific variance generally vary for measures of subscales within their respective second-order factors. This implies that an organization of work values at an intermediate level of abstraction has some support, but more research is needed.

In contrast, evidence is weak for representation of the WAPS at higher levels of abstraction. The pattern of findings for the partial aggregation model is quite similar to that found in the second-order CFA representations of work values under the disaggregation models. However, the partial aggregation model fit the data poorly overall. Likewise, the total aggregation model fit the data poorly. A more comprehensive test of the WAPS in particular and work values in general should include measures of values designed to directly measure facets and higher order factors rather than relying only on aggregations of items designed to measure components.

In sum, the most sound representation of the WAPS was obtained for the total and partial disaggregation models. The partial and total aggregation models received little support, and their use for the WAPS is not recommended. At least, there is little to recommend that individual items be aggregated to measure facets or the global scale.

Further support for the disaggregate models can be seen in the investigation of generalizability. Under the total disaggregation model, the factor structures for work values of men and women were identical and, in addition, showed equally reliable measures. Even the correlations among subscales were the same for men and women, although this may or may not be expected in general, depending on the context or scale

under consideration. Similar findings resulted for the partial disaggregation model but with one added outcome: The variance-covariance matrices of measures of work values did not differ between men and women. This latter finding is probably a consequence of the smoothing out of measurement error when items are combined and implies that the measures are equivalent across gender. Pryor (1983b) also found that the pattern of work values in his cluster analyses were similar for men and women, although unlike herein, he did not perform explicit tests of generalizability. In short, the disaggregation models were remarkably robust as representations of the WAPS.

Despite the equivalence of the WAPS for men and women in terms of measurement properties, interesting differences were found between men and women with respect to the mean levels of work values for the components. Focusing on the partial disaggregation model, we find that women more than men valued detachment (i.e., separation of work from the rest of one's life), lifestyle (i.e., the ability to live where one wants to live), independence (i.e., freedom from constraints in the work environment), and self-development (i.e., concern for growth and use of one's skills and abilities). Women also placed less value on money (i.e., financial rewards from work). No gender differences were found for creativity (i.e., concern for developing something original through work), coworkers (i.e., friendship and understanding), and security (i.e., ability to maintain one's job). Under the total disaggregation model, however, women did value coworkers, and to a lesser extent security, more than men.

What could account for these gender differences? Before we speculate on the reasons, we should acknowledge that very little research exists on gender differences in work values, and the research that we did find is frequently inconclusive and at times contradictory (e.g., Chusmir & Parker, 1991, 1992; Jensen, White, & Singh, 1990; Mottaz, 1986). Two factors possibly explain gender differences in work values. One of these is the role of peer socialization, particularly for young children into adolescence (e.g., Hall, 1987; Paley, 1984). Boys tend to be socialized to be aggressive, competitive, achievement oriented, individualistic, and to function in larger, status-oriented groups. Girls, on the other hand, tend to be socialized to be cooperative, empathic, other oriented, and to function in smaller, intimate groups. Another factor possibly accounting for gender differences is family socialization (e.g., Eagly, 1987; Gleason, 1989). Expectations and encouragement for schooling and career tend to follow the pattern described above for peer socialization. Differences in socialization are consistent with the gender differences found for detachment, lifestyle, coworkers, and money on the WAPS. The gender differences found for independence and self-development and the lack of difference found for security are less easy to explain. Overall, the findings for gender differences are remarkable, given that all respondents were MBA students and would be expected to have strong managerial values and aspirations. The presence of gender differences shows the power of socialization forces and stereotypes indeed.

Comments on the General Framework

The central concern of this article has been to develop a framework for representing and testing constructs, or more generally, variables in organizational research. For quite some time, the most prevalent approach has been to conceive of constructs in unidimensional terms and to operationalization these with either the summation of items in a scale or, on occasion, individual items treated as indicators by use of factor

analysis or structural equation models. This practice has been adequate for much of contemporary empirical work, but as theoretical constructs become more complex (e.g., multidimensional, hierarchical) and as greater demands are placed on researchers for showing the reliability, validity, and generalizability of the measures of these constructs, the traditional approaches to theory representation and testing will become less diagnostic and powerful for the tasks at hand.

We attempted to develop a framework that reconceptualizes the way constructs and their measures are represented and related to each other. To do this, we proposed that constructs and their measures can be fruitfully conceived through specification of three characteristics that capture their meaning: depth, breadth, and dimensionality.

The first is *depth* and can be seen to apply to both measures and constructs, each framed at different levels of abstraction. Measures (e.g., items from rating scales, entrees from archival reports, observations by experts) can be represented at four levels, no matter what unit of analysis they are targeted on. This type of meaning is labeled *indicator depth* in Table 1. Measures that directly indicate a phenomenon and undergo no aggregation, but rather are kept at the particular concrete level expressed by the measurement procedure used to generate them, are termed *individual items* in Table 1. When individual items are consolidated and the consolidations are intended to operationalize the same constructs as implied by the individual item case, we term this depth of measurement *item subsets* in Table 1 to convey the idea that individual items within a factor can be grouped as sums or averages of subsets of items, and these then become the empirical indicators of the hypothesized constructs.

An important issue to address is the rationale for grouping items to form subsets. We suggest a number of conceptual and empirical rules in this regard. First, the items grouped within a subset should each be linked in a sound way to the construct they are intended to measure. Effort should be made to formulate correspondence rules that specify (a) how the items can be deduced from the constructs they are purported to measure, (b) in what sense items are special cases of these constructs, or (c) why and to what extent items are reflections of these constructs. Second, the items forming a subset should be at comparable levels of specificity and should constitute independent empirical observations. Third, items in one subset of items that indicate a construct should be at comparable levels of specificity as, and constitute independent observations from, the items in any other subset of items used to indicate the same construct. The above conceptual rules are intended to increase the precision of measurement and at the same time avoid inclusion of measures of other constructs (e.g., measures of causes, effects, or covarying variables) or of subordinate or superordinate aspects of a focal construct. A fourth rule concerns the empirical basis for grouping items. We suggest that items be grouped into subsets only if all items (across and within subsets) can be shown to be unidimensional. The strong form of this rule requires that all items entering subsets should load highly on a single factor produced by CFA. The items loading on a good-fitting CFA with one factor can then be selected at random and grouped into subsets. A weaker empirical rule is to perform an exploratory factor analysis and to select items loading highly on a factor to form subsets, again by use of random assignment to subsets. The selection for a subset of items that load on separate factors would not only violate the conceptual integrity of the hypothesized construct and the correspondence rules linking construct to indicators but would interject spurious associations between factors and introduce capitalization on chance, thereby

bringing into doubt construct validity and contributing to errors in inference or prediction when constructs are used to test hypotheses in any substantive context.

When all items for a component are aggregated to form a single indicator, we term this type of depth *individual subscales* (see Table 1). And when items are combined across components to form indicators of an entire scale (i.e., items from individual subscales are combined), we call this type of depth *scale subsets.*

Depth also varies at the construct or latent variable level and, in combination with indicator depth, defines the four classes of models shown in the left-hand column of Table 1. Three types of construct depth are specified. The least abstract is termed the *component* level and is intended to represent a construct or variable at the lowest level possible for a theoretical entity. To represent concepts at intermediate levels of abstraction, we use the term *facet*. Finally, the highest level of abstraction is termed the *global* level. The above levels correspond roughly to the categories used in cognitive psychology to express three tiers that concepts are represented in human memory: subordinate, basic, and superordinate categories of thought (e.g., Rosch, 1978).

Orthogonal to the notion of depth is the *breadth* of measures and constructs. The greater the number of measures of a construct or the greater the number of constructs, the greater the breadth. Consider measure breadth. If we were to specify indicators of organizational affective commitment, for example, we might use two indicators (e.g., the degree of attachment and happiness with the organization) or, alternatively, many indicators (e.g., the amount of felt love for, joy from, enthusiasm toward, excitement with, worry about, disappointment in, and anger toward the organization). Thus, relatively narrow or wide breadth can be achieved simply with the number of indicators used. This has conceptual and empirical implications for measurement. The more the number of indicators, the more detailed and comprehensive the specification of a construct. But the greater the number of indicators, the more difficult it is to demonstrate validity and thus the more valid a construct's measures are when such validity is achieved. Consider next construct breadth in terms of the number of dimensions specified. The greater the number of components or facets of a construct, the richer are the predictions that can be made using the construct as both a predictor or predicted variable. Work values represent a relatively broad construct in that the construct validated herein (i.e., the WAPS) has eight components. We might expect each of the components to both share and have unique antecedents and to potentially differentially influence multiple behaviors. Finally, we should point out that breadth can refer to qualitative aspects of a construct at a single depth. For instance, affective commitment might be defined as degree of attachment only (e.g., liking for, affection toward, or fondness for an organization), degree of positive affect felt as a result of working for the organization (e.g., happiness, joy, pride), both attachment and positive affect, or even positive and negative affect, where the latter is a form of anticommitment.

We thus see that measures and constructs achieve dimensionality through the breadth of things that constitute them but in a horizontal sense. Dimensionality also arises hierarchically to the extent that constructs can be related in vertical patterns. For example, superordinate and subordinate constructs can be modeled in relation to a focal construct(s) with complex functional, causal, or associational relations linking constructs within and across levels.

In general, any hierarchical arrangement of constructs can be expressed in one of two ways: as a top-down or bottom-up structure. Although the empirical findings were

mixed, we found that work values might be depicted in a two-tier hierarchy. At the top of the hierarchy are three relatively abstract values: nonwork orientation, freedom, and human/personal concern. The eight components of the WAPS were found to be reflections of these three values, and therefore, the hierarchy is basically top-down. As an example of a bottom-up hierarchy of constructs, job satisfaction may be considered. Overall job satisfaction, a global construct at the top, might be thought to be produced by satisfaction with such specific aspects of the job at the bottom of the hierarchy as satisfaction with pay, coworkers, one's task, supervision, and the opportunity for promotion.

There is one important limitation to the tests of our framework performed herein that we wish to point out, which was a consequence of data limitations. All levels of depth for constructs were, by necessity, operationalized by use of items that were formed at narrow levels of specificity. These were the only data available for our study of the WAPS. Thus, whether components, facets, or global constructs were operationalized (see Table 1), either individual items at the level of components or aggregations of these items were used as indicators. This is a reasonable strategy for many hierarchical representations of variables, but we wish to stress that other possibilities should be acknowledged. Rather than aggregating items designed originally to measure concrete constructs, to operationalize more abstract ones, it may be appropriate to design direct measures of each level in the hierarchy. For example, we might generate measures of facets and of global constructs in a direct manner. An example of a measure of the freedom facet for the WAPS might be, "Working at Company X gives me the freedom to express my ingenuity and talents with a minimum of constraints." The operationalizations used in the current study employed more indirect indicators composed of aggregations of such specific indicators as "work in which you can start and finish your work when you like" and "work in which you experiment with different ways of doing things." Global job satisfaction might be directly operationalized with such items as "Overall, I am very happy at Company Y" and "My very feelings toward Company Y can be best summarized as being positive"; these direct global measures or the abstract latent variable they indicate could then be shown to be functions of more specific dimensions of job satisfaction.

Conclusions

The point of view taken in this article is that it is useful to view rating scales from the perspective of a four-tiered hierarchy. When global hypotheses are under scrutiny, such as the study of overall job satisfaction or work commitment, the total or partial aggregation models might be appropriate. The former is the least specific representation and focuses on ensembles of items. Concern is with a scale or construct in its entirety. Few criteria exist to ascertain the integrity and validity of measures in the total aggregative sense. About all one can do is scrutinize measures of reliability and the results of factor analyses. The partial aggregation model perhaps has more currency and sacrifices only a small amount of breadth to achieve a useful level of generality. It is particularly applicable to cases in which a scale exhibits a small number of abstract facets. When measures are aggregated to form facets and facets are used as indicators of a latent variable(s), interpreted as an abstract concept, a number of diagnostics are available based on the CFA model. This approach has been fruitful in such areas of

personality research as self-esteem, self-monitoring, and self-consciousness and potentially can be used to advantage in organization research. When hypotheses regarding specific subscales or components are under study, such as the differential impact of multiple work values on commitment or effort, the disaggregation models come to the fore. These models promote the fullest investigation of validity, yet, via the introduction of higher order factors, permit representation of a certain degree of abstractness and hierarchical organization of subscales. Hence, the choice among the four models contained within the hierarchy described herein should be based on the purposes of a particular study, combined with CFA to determine whether the hierarchical level chosen is adequately represented by the measures at hand.

In sum, any representation of a scale or theoretical variable entails a complex decision process whereby dimensionality and construct and indicator breadth and depth must be specified adequately (see Table 1). The present article attempted to provide a conceptualization and guidelines for making such decisions. The ideas developed herein can be used to help researchers better specify and operationalize theories in construct validation, as well as aid more generally in hypothesis testing in theoretical or substantive research.

References

Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance, and normative commitment to the organization. *Journal of Occupational Psychology, 63*, 1-18.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411-423.

Bagozzi, R. P. (1993). Assessing construct validity in personality research: Applications to measures of self-esteem. *Journal of Research in Personality, 27*, 49-87.

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1*, 35-67.

Bagozzi, R. P., & Phillips, L. W. (1982). Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly, 27*, 459-489.

Bagozzi, R. P., Yi, Y., & Nassen, K. D. (in press). Representation of measurement error in marketing variables: Review of approaches and extension to three-facet designs. *Journal of Econometrics*.

Becker, T. E., & Vance, R. J. (1993). Construct validity of three types of organizational citizenship behavior: An Illustration of the direct product model with refinements. *Journal of Management, 19*, 663-682.

Bentler, P. M. (1989). *EQS: Structural equations program manual* [Computer program manual]. Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K. A., & Ting, K. (1993). Confirmatory tetrad analysis. In P. V. Marsden (Ed.), *Sociological methodology 1993* (pp. 147-175). Washington, DC: American Sociological Association.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology, 37*, 1-21.

Browne, M. W. (1989). Relationships between an additive model and a multiplicative model for multitrait-multimethod matrices. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 507-520). Amsterdam, The Netherlands: North-Holland.

Browne, M. W. (1990). *MUTMUM pc user's guide.* Unpublished manuscript, Ohio State University.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows.* Thousand Oaks, CA: Sage.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Chusmir, L. H., & Parker, B. (1991). Gender and situational differences in managers' values: A look at work and home lives. *Journal of Business Research, 23*, 325-335.

Chusmir, L. H., & Parker, B. (1992). Success strivings and their relationship to affective work behaviors: Gender differences. *Journal of Social Psychology, 132*, 87-99.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field setting.* Chicago: Rand McNally.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317-327.

Dawis, R. V. (1991). Vocational interests, values, and preferences. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 833-872). Palo Alto, CA: Consulting Psychologists Press.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Gati, I. (1991). The structure of vocational interests. *Psychological Bulletin, 109*, 309-324.

Gay, E. G., Weiss, D. J., Hendel, D. D., Dawis, R. V., & Lofquist, L. H. (1971). Manual for the Minnesota importance questionnaire. *Minnesota Studies in Vocational Rehabilitation, 28*, 1-83.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25*, 186-192.

Gerbing, D. W., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Research, 21*, 132-160.

Gleason, J. (1989). Sex differences in parent-child interaction. In S. Philips, S. Steele, & C. Tanz (Eds.), *Language, gender, and sex in comparative perspective* (pp. 189-199). Cambridge, UK: Cambridge University Press.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology, 60*, 159-170.

Hall, J. A. (1987). On explaining gender differences: The case of nonverbal communication. *Review of Personality and Social Psychology, 7*, 177-200.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139-164.

Holland, J. L. (1985). *Making vocational choices* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

House, R., Rousseau, D. M., & Thomas-Hunt, M. (1995). The meso paradigm: A framework for the integration of micro and macro organizational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 17, pp. 71-114). Greenwich, CT: JAI.

Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology, 61*, 932-945.

Ironson, G. H., Smith, P. C., Brannick, M. T., Gibson, W. M., & Paul, K. B. (1989). Construction of a job in general scale: A comparison of global, composite, and specific measures. *Journal of Applied Psychology, 74*, 193-200.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*, 219-229.

Jensen, T. D., White, D. D., & Singh, R. (1990). Impact of gender, hierarchical position, and leadership styles on work-related values. *Journal of Business Research, 20*, 145-152.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL7 user's reference guide* (1st ed.). Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1994). *LISREL8: Structural equation modeling with the SIMPLIS command language.* Chicago: Scientific Software.

Judge, T. A., & Bretz, R. D. Jr. (1992). Effects of work values on job choice decisions. *Journal of Applied Psychology, 77*, 261-271.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology, 12*, 247-252.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review, 19*, 195-229.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72-107.

Lobel, S. A. (1991). Allocation of investment in work and family roles: Alternative theories and implications for research. *Academy of Management Review, 16*, 507-521.

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297-1350). Chicago: Rand McNally.

Macnab, D., & Fitzsimmons, G. W. (1987). A multitrait-multimethod study of work-related needs, values, and preferences. *Journal of Vocational Behavior, 30*, 1-15.

Mael, F. A., & Ashforth, B. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior, 13*, 103-123.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335-361.

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*, 623-636.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling, 1*, 5-34.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of the behavior of alternative models. *Applied Psychological Measurement, 15*, 47-70.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across groups. *Psychological Bulletin, 97*, 562-582.

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*, 107-117.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*, 247-255.

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

Mottaz, C. (1986). Gender differences in work satisfaction, work-related rewards and values, and the determinants of work satisfaction. *Human Relations, 39*, 359-378.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior, 14*, 224-247.

Newcomb, M. D., & Bentler, P. M. (1988). *Consequences of adolescent drug use.* Newbury Park, CA: Sage.

Nord, W. R., Brief, A. P., Atieh, J. M., & Doherty, E. M. (1988). Work values and the conduct of organizational behavior. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 10, pp. 1-42). Greenwich, CT: JAI.

Organ, D. W., & Konovsky, M. (1989). Cognitive versus affective determinants of organizational citizenship behavior. *Journal of Applied Psychology, 74,* 157-164.

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology, 78,* 569-582.

Paley, V. G. (1984). *Boys and girls: Superheroes in the doll corner.* Chicago: University of Chicago Press.

Pierce, J. L., McTavish, D. G., & Knudsen, K. R. (1986). The measurement of job characteristics: A content and contextual analytic look at scale validity. *Journal of Occupational Behavior, 7,* 299-314.

Pryor, R.G.L. (1983a). *Manual for the Work Aspect Preference Scale.* Melbourne: Australian Council for Education Research.

Pryor, R.G.L. (1983b). Sex differences in the levels of generality of values/preferences related to work. *Journal of Vocational Behavior, 23,* 233-241.

Pryor, R.G.L. (1987). Differences among differences: In search of general work preference dimensions. *Journal of Applied Psychology, 72,* 426-433.

Ravlin, E. C., & Meglino, B. M. (1987). Effect of values on perception and decision making: A study of alternative work values measures. *Journal of Applied Psychology, 72,* 666-673.

Ravlin, E. C., & Meglino, B. M. (1989). The transitivity of work values: Hierarchical preference ordering of socially desirable stimuli. *Organizational Behavior and Human Decision Process, 44,* 494-508.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds.). (1991). *Measures of personality and social psychological attitudes.* New York: Academic Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-71). Hillsdale, NJ: Lawrence Erlbaum.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 7, pp. 1-37). Greenwich, CT: JAI.

Schwab, D. P. (1980). Construct validity in organizational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 2, pp. 3-43). Greenwich, CT: JAI.

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology, 53,* 550-562.

Super, D. E. (1970). *Work values inventory manual.* Boston: Houghton Mifflin.

Super, D. E. (1973). The work values inventory. In D. G. Zytowski (Ed.), *Contemporary approaches to interest measurement.* Minneapolis: University of Minnesota Press.

Vance, R. J., MacCallum, R. C., Coovert, M. D., & Hedge, J. W. (1988). Construct validity of multiple job performance measures using confirmatory factor analysis. *Journal of Applied Psychology, 73,* 74-80.

Vandenberg, R. J, & Self, R. M. (1993). Assessing newcomers' changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology, 78,* 557-568.

Warr, P. B., Cook, J. D., & Wall, T. D. (1979). Scales for the measurement of some work attitudes and aspects of psychological well-being. *Journal of Occupational Psychology, 52,* 129-148.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9,* 1-26.

Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology*, *74*, 462-468.

Wolins, L. (1982). *Research mistakes in the social and behavioral sciences*. Ames: Iowa State University Press.

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 7-56). Mahwah, NJ: Lawrence Erlbaum.

Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika*, *55*, 255-262.

Richard P. Bagozzi is the Dwight F. Benton professor of behavioral science in management at the University of Michigan Business School. He earned his Ph.D. at Northwestern University. His research interests span structural equation models, the theory of action, emotions, and goal-directed behaviors.

Jeffrey R. Edwards is a professor of management at the Kenan-Flagler Business School at the University of North Carolina. He earned his Ph.D. at the Graduate School of Industrial Administration at Carnegie Mellon University. His current research interests include stress, coping, and well-being in organizations; person-environment fit; work and nonwork values; measurement and construct validation; and methodological issues in congruence research.