

Conceptual and Empirical Approaches to Developing Family-Based Assessment Procedures: Resolving the Case of the Family Environment Scale

RUDOLF H. MOOS, Ph.D.^a

^aSocial Ecology Laboratory and Far West Health Services Field Program, Stanford University and Department of Veterans Affairs Medical Centers, Palo Alto CA 94304.

This article focuses on the reliability and validity of the Family Environment Scale (FES). The FES subscales generally show adequate internal consistency reliability and stability over time when applied in samples that are diverse; the items also have good content and face validity. An extensive body of research supports the construct, concurrent, and predictive validity of the FES. More generally, reliability and validity are a joint function of scale items and response formats and of the characteristics and diversity of specific samples. To contribute to further advances in family assessment, researchers need to use both conceptual and psychometric criteria rather than rely too heavily on the pursuit of internal consistency reliability and factor analytic approaches to scale construction and validation.

Over the last decade, researchers have developed an impressive array of assessment procedures that focus on family characteristics. As informed scholars and clinicians, how should we evaluate the growing body of information about the reliability, validity, and clinical utility of these assessment procedures? How can we develop realistic standards for reliability and validity that allow for a diversity of conceptual and empirical approaches to test construction and help to advance the field of family assessment? I focus on these issues here primarily with reference to the Family Environment Scale (FES). However, the issues are general ones that apply to other family-focused assessment procedures, as well as to many procedures that assess individual personality, attitudinal, and behavioral characteristics.

In pursuing these issues, I comment on Roosa and Beals' (27) discussion of measurement issues in family assessment and correct some apparent misconceptions about the development of the FES, the meaning and value of internal consistency reliability, and the use of factor analysis in scale construction and validation. More specifically, I describe the approach we used to develop the FES, and present new information on the internal consistency and stability of the five FES subscales on which Roosa and Beals focus. Following that, I discuss the content, construct, concurrent, and predictive validity of the FES. Finally, I consider some general issues and trade-offs involved in different approaches to test construction, emphasize the importance of conceptual analyses and broadband measures in family assessment, and describe some guidelines researchers and journal editors can adopt to promote the development and appropriate use of family-based assessment procedures.

CONCEPTUAL AND EMPIRICAL DEVELOPMENT OF THE FES

I have described my approach to scale development both in relation to the FES (18, 19, 23, 24) and, more generally, with respect to the Social Climate Scales (20). These publications present a detailed picture of the conceptual and empirical steps involved in the development of the FES.

Conceptual Item Formulation Procedures

In brief, the initial choice and wording of FES items were guided by information obtained from observations and interviews with families, and by a conceptual formulation of three general areas and a number of specific dimensions that might differentiate among families. Each of the 200 items chosen for pilot testing identified an aspect of the family environment that reflects the quality of interpersonal relationships (such as the degree of cohesion), the focus on an area of personal growth (such as the degree of achievement or moral-religious emphasis), or the emphasis on system maintenance (such as family organization). For example, family cohesion is inferred from items such as "Family members really help and support one another," and "There is a of togetherness in our family." Family organization is inferred from items such as "Activities in our family are carefully planned," and "We are generally very neat and orderly."

Thus, we built content and face validity into the FES subscales at the outset by defining constructs such as cohesion and organization, preparing items to fit the construct definitions, and selecting items that were conceptually related to a dimension.

Empirical Scale Construction Procedures

Our initial assignment of items to a set of 12 hypothesized dimensions was based on conceptual considerations and was preliminary. Next, we developed empirical information on a diverse sample of more than 1,000 individuals in 285 families. We recruited families from newspaper advertisements, church groups, local high school students, and people in crisis situations who were undergoing treatment in psychiatric and correctional facilities. We included adolescents as well as adults in these families. To broaden the sample, our Black and Hispanic research assistants recruited a group of Black and Hispanic families.

We then applied several empirical criteria to select the final set of items and develop the ten FES subscales. Specifically, we tried to select items that (a) had a reasonable response distribution, that is, were not answered in one direction (true or false) by more than 80% of the respondents; (b) discriminated significantly among families; (c) were positively correlated with other items on their subscale; and (d) correlated more highly with their subscale than with any other subscale. Thus, the selected items met empirical criteria in addition to a conceptual criterion of "fit" with the dimension to which they were assigned. The resulting FES subscales had acceptable internal consistencies in this sample.

INTERNAL CONSISTENCY AND STABILITY OF THE FES SUBSCALES

Roosa and Beals (27) correctly point out that it is appropriate to cross-validate the psychometric characteristics of a new assessment procedure and that I should have followed this approach in the development of the FES. Given the need for a diverse and heterogeneous sample, however, this is not such an easy task. The psychometric characteristics of a scale should not be "cross-validated" on a narrow sample with restricted item or subscale variability. Although we have not obtained information on other samples as varied as the FES normative sample, my colleagues and I are using the FES in three ongoing projects.

One project focuses on depressed patients at treatment intake and at 1- and 4-year followups. These patients and their spouses are compared with matched case controls and their spouses (for descriptions of the samples and measures, see 3, 4, 13, 14, and 17). The second project focuses on alcoholic patients and their spouses as compared to matched case controls and their spouses (21, 22). The third project involves a study of family-based psychosocial factors linked to the course and outcome of juvenile rheumatic disorders (9).

Representative FES Subscale Internal Consistencies

Roosa and Beals focus on five of the FES subscales. Table 1 shows the internal consistencies of these five FES subscales in these samples. The average alphas for cohesion and conflict are essentially identical to those reported in the FES Manual (23, 24). The average alphas for expressiveness, organization, and control are somewhat lower than those reported in the Manual. Overall, there is little "shrinkage" in the alphas, as one would expect with scales whose development is conceptually based (1). Moreover, these samples are not as diverse as the original normative sample; specifically, the new samples include fewer individuals of low socioeconomic status, fewer minority individuals, fewer adolescents, and fewer currently distressed families. Given the reduced variability in these new samples, the FES subscale internal consistencies are quite acceptable.^{1,2} In addition, they are all higher than those Roosa and Beals obtained in their samples.

Table 1
Family Environment Scale Subscale Internal Consistencies for Different Samples

Sample	Subscales					
	Cohesion	Expressiveness	Conflict	Organization	Control	FRP ^a
Depressed and control families (Time 1; N = 904)	.79	.63	.78	.70	.58	.84
Depressed and control families (Time 2; N = 791)	.78	.66	.76	.69	.60	.84
Depressed and control families (Time 3; N = 740)	.76	.62	.75	.71	.63	.83
Alcoholic and control families (N = 356)	.76	.60	.72	.71	— ^b	.81
Families of children with rheumatic disease (Time 1; N = 386)	.77	.58	.74	.60	.58	.83
Average alpha	.77	.62	.75	.68	.60	.83
Alpha reported in FES Manual	.78	.69	.75	.76	.67	.89 ^c

Note: The number of cases for each subscale varies somewhat due to missing data.

^a Family Relationships Index

^b The FES Control subscale was not used in this sample.

^c The alpha for the Family Relationships Index is based on a representative community sample as originally reported by Holahan and Moos (sec 12).

Sample Characteristics and Internal Consistency

Roosa and Beals point out that subscale internal consistencies may vary from sample to sample, but they do not consider why such variations occur or apply potential explanations to their samples. In general, the internal consistency of a subscale is higher in more diverse samples, that is, samples in which there is more item and subscale variance (28).³ The mean and standard deviation of a subscale are a joint property of the items and response formats and of the specific sample in which they are applied. The same conclusion holds for internal consistency and test-retest reliability. Thus, to interpret reliability statistics on a sample, we need specific information about the item and subscale means and variances in that sample.

Roosa and Beals may have obtained relatively low internal consistencies for the five FES subscales because of the restricted nature of their samples. They state that their combined sample is comparable to the initial FES normative sample, but this is not correct. Their sample is much less diverse; it is composed wholly of adults (we included a substantial number of adolescents), 86% women (we included many more men), 87% Caucasian (we included more ethnic minority respondents), and primarily well-educated and middle- to upper-middle-class individuals.

Group Comparisons and the Pursuit of Internal Consistency

Roosa and Beals state that a measure cannot be used to make meaningful comparisons between samples when the measure has either low or different internal consistencies in those samples. This is not correct. Consider a researcher who wants to find a way to discriminate reliably between the families of depressed patients and those of normal case controls. After reviewing available measures, the researcher selects the FES and administers it to a large group of families. Assume that respondents in families with a depressed family member all obtain a score between zero and four on the FES Cohesion subscale, whereas respondents in the normal control families all obtain scores ranging between five and nine. Due to the restricted range of within-group scores, the internal consistency of the Cohesion subscale is likely to be quite low in both groups, but the subscale discriminates perfectly between the groups! Moreover, the alpha for the Cohesion subscale in the combined group is likely to be high.

We ran an example on the Cohesion subscale to illustrate this point. We used the depressed patients and case controls at time 1 ($N = 904$; see Table 1) and compared subsamples of respondents who scored between zero and four and those who scored between five and nine. The low-cohesion group had a mean cohesion score of 2.43 ($SD = 1.37$); the subscale alpha was .10. The high-cohesion group had a mean cohesion score of 7.57 ($SD = 1.27$). The alpha was .34. As shown in Table 1, the alpha is .79 for the overall group. Thus, researchers must be careful not to overlook validity (reliable group discrimination in this example) in an unrealistic pursuit of reliability.

Stability of the FES Subscales and Profile

The test-retest reliability and longer-term stability of an assessment procedure in different samples are also important psychometric characteristics. In this respect, the FES subscales have good 8-week test-retest reliabilities (varying from .73 to .86 for the five subscales we focus on here) and 4-month stabilities (varying from .66 to .78 for these five subscales). The 12-month subscale stabilities for averaged family subscale means vary from .63 for cohesion to .81 for organization. The overall FES profiles are also reasonably stable over these time intervals (for more details, see 24).

We examined the subscale stabilities over 12-month, 36-month, and 48-month intervals for the 676 individuals assessed at all three intervals in the sample of depressed and control families shown in Table 1. The 12-month subscale stabilities varied from .59 for cohesion to .67 for conflict and control (mean for the five subscales = .63). The stabilities ranged from .47 to .58 over the 36-month interval (mean = .56) and from .45 to .54 over the 48-month interval (mean = .53). Thus, FES subscale scores may be quite stable over intervals of as long as 4 years.

Just as with internal consistency, however, these stabilities are likely to vary in different groups of families, depending on whether families are in crisis situations, are experiencing changes in family composition or structure, or are in counseling or treatment. In this respect, the FES sub-scales reflect changes that occur in family environments over time.

CONTENT VALIDITY OF THE FES

As noted earlier, the FES items were initially assigned to potential dimensions on the basis of their item content and conceptual connection to specific family constructs. These decisions were supported empirically by selecting items that were more highly correlated with their FES subscale than with any other subscale.

Roosa and Beals asked twelve psychology graduate students to assign 45 of the FES items to the correct five subscales on the basis of the very brief subscale descriptions given in the FES Manual (23). These descriptions provide an overall idea about each dimension; they are not intended to provide sufficient information for raters to accurately place items on specific dimensions. Moreover, Roosa and Beals instructed raters to place items in a discard pile unless they were certain about the correct dimension. Using this procedure, the majority (67%) of the panelists were *certain* (and correct) about placement of 24 of the 45 items. Thus, these raters did moderately well in placing the FES items on their correct dimensions even when they had little information about the dimension and were asked to make judgments at a high level of certainty.

We have found that untrained raters can place most of the FES items on the correct dimension when they are given reasonably adequate information about the conceptual content of the dimension and are allowed to provide a "probable" judgment. To address this point, we asked nine raters to assign each of the 45 FES items to one of the five subscales if they were (a) reasonably certain or thought it was (b) probable that the item belonged to that subscale.⁴ A total of 39 of the 45 items were categorized correctly by at least six of the nine (67%) raters; more specifically, 67% or more of the raters categorized 30 items as "certainly" belonging on the correct dimension and 9 more items as "probably" belonging on the correct dimension. Thus, Roosa and Beals's relatively modest results probably were due to the paucity of information they provided and the high level of certainty they required. Our findings show that the FES items have good content and face validity.

CONSTRUCT, CONCURRENT, AND PREDICTIVE VALIDITY OF THE FES

Space constraints permit only a brief comment on the extensive evidence of the construct, concurrent, and predictive validity of the FES subscales. With respect to construct validity, for example, FES cohesion is positively related to measures of dyadic and marital adjustment as well as to reports of support from other family members. FES conflict is positively associated with family arguments, and FES organization and control are linked to reliance on predictable and regular family routines.

The FES dimensions tend to be predictably related to external criteria in both concurrent and predictive studies. For example, aspects of the family environment, as measured by the FES subscales, are associated with adaptation to pregnancy and parenthood, childhood and adolescent adjustment to parental divorce, adaptation to chronic childhood illness and other life stressors, children's cognitive and social development, adjustment among families of psychiatric and medical patients, and the outcome of treatment for alcoholism, depression, and other psychiatric and medical disorders (for an integrated overview of more than 150 studies that focus on these validity issues, see 24, pp. 24-48).

CONCEPTUAL AND PSYCHOMETRIC CRITERIA: SOME CHOICES AND TRADE-OFFS

Roosa and Beals raise some important test construction and psychometric issues. Their conclusions about the FES are incorrect; a considerable body of evidence shows that the FES is reasonably internally consistent, stable, and valid when applied to moderately diverse samples. Nevertheless, researchers can and should improve on the FES, as well as on other currently available family assessment procedures. The problem is to specify the test construction criteria that are likely to maximize advances in the field.

I advocate the use of conceptual considerations as the primary initial guidelines for test construction. I also believe that a focus on relatively broad constructs (such as cohesion and organization) is consistent with the current state of conceptualization in the field of family assessment. As knowledge about broad constructs accumulates, we may be able to accurately define and assess more specific constructs (aspects of cohesion and organization, for example) and to develop more internally reliable subscales. Of course, such scales may or may not be as stable and valid as scales that assess broadband constructs. To examine some of the choices and trade-offs that may arise in developing an assessment procedure, I briefly describe three sets of decisions involved in constructing the FES.

Item Content and Internal Consistency

Because we wanted to measure relatively broad family constructs (such as cohesion, expressiveness, and conflict), we selected items with some diversity in content. Thus, for example, if two items were highly intercorrelated, we dropped one in order to reduce item redundancy and broaden the content of the final item set. If we had selected less diverse items, we could have developed more internally consistent sub-scales. But we would have measured a narrower construct and probably sacrificed some generalizability and construct validity.

Subscale Length and Internal Consistency

The length of a subscale also affects its internal consistency. We wanted to develop a family-based screening procedure that could provide a quick "snapshot" of some of the major dimensions that differentiate families. In prior work we found

that such a screening procedure needs to be limited to between 80 and 100 items. Given ten dimensions, we thought each dimension should have a maximum of nine or ten items. However, when the breadth of item content (or, more specifically, the average item intercorrelations) is held constant, longer subscales are likely to be more internally consistent.

To illustrate the effect of subscale length (number of items), we obtained internal consistencies for the Family Relationships Index (FRI), which assesses the overall quality of family relationships. The FRI is a 27-item subscale that is a combination of the three FES Relationship dimensions: cohesion, expressiveness, and conflict, which is reverse scored (11, 12). As shown in Table 1, the alpha for the 27-item FRI is higher than the alphas for the separate 9-item Cohesion, Expressiveness, and Conflict subscales.

Response Format and Internal Consistency

A third decision that affects internal consistency is whether to use dichotomous (true/false) or 3- or 4-point response options. We opted for a dichotomous response format because of its simplicity and to make the FES more applicable in clinical situations for cognitively impaired individuals (for the rationale, see 20). However, a multipoint response format is likely to result in a more internally consistent subscale. In this respect, Roosa and Beals cite Bloom (5), who used a 4-point response format, as providing data on the internal consistency of the FES subscales, but they fail to emphasize that he obtained alphas varying from .65 to .85 for the 9-item version of the five FES subscales in question here, and alphas varying from .64 to .84 for 5-item versions of these subscales.

THE PROMISE AND PERILS OF FACTOR ANALYSIS

Roosa and Beals conducted two confirmatory factor analyses (CFA) of their FES data. One analysis used 45 items while the other focused on the 24 items their raters placed correctly on one of the five sub-scales. The five-subscale model did not quite fit the data adequately in the analysis on all 45 items, although the Goodness of Fit Index (GFI) indicates that the model accounted for a considerable proportion of the variation in the data. The model did fit the data quite well in the analysis on the reduced set of 24 items.

Diagnostic Indicators and Improving Factor Models

More important, the five-factor model may fit the 45-item data set quite well when modifications are made to the model *other than* by dropping items or reassigning items to different dimensions. In this respect, the computer output generated by CFA provides diagnostic indicators (modification indices) that suggest how the fit between the model and the data could be improved. Roosa and Beals might have been able to improve the fit by changing their assumptions about fixed parameters, adjusting the allowance for correlated dimensions and error terms, and so on. Decisions about the adequacy of factor models also may vary depending on the specific form of the input data (raw data, correlation matrix, covariance matrices), the size of the sample, the choice of which of more than 25 alternative, goodness-of-fit indicators should be used, and what value of an index constitutes an acceptable fit (2, 6, 16). A related point involves the question of whether the LISREL model is appropriate for dichotomous data and whether the resulting standard errors and GFIs can be judged against standards developed under quite different distributional assumptions (15).

Factorial Fidelity and Conceptual Breadth

The fundamental approach typically used in factor analysis presents a more serious problem. Researchers typically use factor-analytic procedures to try to create relatively circumscribed and highly internally consistent dimensions. Our intent, however, was to create conceptually broad subscales composed of a diverse set of items. A broadband measure is likely to have more stability over time and greater validity, although it may be somewhat less internally consistent (for a discussion of the bandwidth versus fidelity issue, see 8).

A related point is that factor-analytic solutions depend on conceptual considerations, aspects of the specific sample (such as its diversity), and a range of decisions about statistical procedures, factoring criteria, goodness-of-fit indices, and so on. Thus, for example, investigators who have examined the subscale factor structure of the FES in different samples have obtained widely varying solutions ranging from two through six factors (for the references, see 24, p. 22). Three item-level factor analyses have each identified eight factors in the FES, but the specific item composition of the eight factors has varied (10, 25, 26). It is reasonable to try to identify a consistent set of constructs to characterize family social environments, but there are serious problems in looking for "the" factor structure of an assessment procedure. Like the mean, standard deviation, and internal consistency of specific subscales, the factor structure of an assessment procedure depends largely on the sample used in the analysis. In general, more factors are likely to emerge in more heterogeneous samples.

PROMOTING ADVANCES IN FAMILY ASSESSMENT

Roosa and Beals recommend that family journals require researchers to report the internal consistency of the assessment procedures they use in their specific samples. Although I agree with this recommendation, there is danger in a narrow focus on specific psychometric criteria. To interpret the meaning of an internal consistency statistic in a particular sample, researchers need to report more specific information about that sample, such as the context in which the sample was recruited, family composition and demographic characteristics, and the subscale means and standard deviations.

More important, researchers and journal editors need to understand that any assessment procedure may have relatively low internal consistency and test-retest reliability, and a uni- or bidimensional factor structure, when it is applied in homogeneous samples. As we saw earlier, a measure that discriminates reliably between two subgroups may show low within-group variation and internal consistency reliability in each of these subgroups. Thus, we need to temper our emphasis on reliability in light of related information on the validity of an assessment procedure in different samples and the context in which the procedure will be used.

From a broader perspective, we should find better ways in which new and promising assessment procedures can develop and evolve over time. Family assessment procedures, such as the FES, need to be adapted and updated to keep abreast of changing times, family compositions, and cultural and value contexts. A lively debate on how to achieve this aim can contribute to advances in the field of family assessment and, ultimately, to the effective planning and evaluation of prevention and intervention programs.

REFERENCES

1. Anastasi, A., Evolving concepts of test validation. In M.R. Rosenzweig & L.W. Porter (eds.), *Annual review of psychology* (Vol. 37). Palo Alto CA: Annual Reviews, Inc., 1986.
2. Bentler, P. and Bonett, D. G., Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606, 1980.
3. Billings, A., Cronkite, R. and Moos, R., Social environmental factors in unipolar depression: Comparisons of depressed patients and nondepressed controls. *Journal of Abnormal Psychology*, 92, 119-133, 1983.
4. Billings, A. and Moos, R., Psychosocial processes of remission in unipolar depression: Comparing depressed patients with matched community controls. *Journal of Consulting and Clinical Psychology*, 53, 314-325, 1985.
5. Bloom, B. L., A factor analysis of self-report measures of family functioning. *Family Process*, 24, 225-239, 1985.
6. Breckenridge, J. N., Structural equation models for depression prevention research. In R.F. Munoz (ed.), *Depression prevention: Research directions*. New York: Hemisphere Publishing, 1987.
7. Cronbach, L., Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334, 1951.
8. Cronbach, L., *Essentials of psychological testing* (5th ed.). New York: Harper & Row, 1990.
9. Daniels, D., Moos, R., Billings, A. and Miller, J., Psychosocial risk and resistance factors among children with chronic illness, healthy siblings, and healthy controls. *Journal of Abnormal Child Psychology*, 15, 295-308, 1987.
10. Garfinkle, A. S., Genetic and environmental influences on the development of Piagetian logico-mathematical concepts and other specific cognitive abilities: A twin study. *Acta Geneticae Medicae Gemellologiae*, 31, 10-61, 1982.
11. Holahan, C. J. and Moos, R., Social support and adjustment: Predictive benefits of social climate indices. *American Journal of Community Psychology*, 10, 403-415, 1982.
12. Holahan, C. J. and Moos, R., The quality of social support: Measures of family and work relationships. *British Journal of Clinical Psychology*, 22, 157-162, 1983.
13. Holahan, C. J. and Moos, R., The personal and contextual determinants of coping strategies. *Journal of Personality and Social Psychology*, 52, 946-955, 1987.
14. Holahan, C. J. and Moos, R., Risk, resistance, and psychological distress: A longitudinal analysis with adults and children. *Journal of Abnormal Psychology*, 96, 3-13, 1987.
15. Joreskog, K. G. and Sorbom, D., *LISREL VII: A guide to the program and applications*. Chicago: SPSS Inc., 1988.
16. Marsh, H. W., Balla, J. R. and McDonald, R. P., Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410, 1988.
17. Mitchell, R., Cronkite, R. and Moos, R., Stress, coping, and depression among married couples. *Journal of Abnormal Psychology*, 92, 433-448, 1983.
18. Moos, R., *Family Environment Scale preliminary manual*. Palo Alto CA: Consulting Psychologists Press, 1974.
19. Moos, R., *Evaluating correctional and community settings*. New York: John Wiley & Sons, 1975.
20. Moos, R., *The Social Climate Scales: A user's guide*. Palo Alto CA: Consulting Psychologists Press, 1987.
21. Moos, R., Finney, J. and Chan, D., The process of recovery from alcoholism: I. Comparing alcoholic patients and

-
- matched community controls. *Journal of Studies on Alcohol*, 42, 383-402, 1981.
22. Moos, R., Finney, J. and Gamble, W., The process of recovery from alcoholism: II. Comparing spouses of alcoholic patients and spouses of matched community controls. *Journal of Studies on Alcohol*, 43, 888-909, 1982.
 23. Moos, R. and Moos, B., *Family Environment Scale manual*. Palo Alto CA: Consulting Psychologists Press, 1981.
 24. Moos, R. and Moos, B., *Family Environment Scale manual: Second edition*. Palo Alto CA: Consulting Psychologists Press, 1986.
 25. Oliver, J. M., Handal, P. J., Enos, D. M. and May, M. J., Factor structure of the Family Environment Scale: Factors based on items and subscales. *Educational and Psychological Measurement*, 48, 469-477, 1988.
 26. Robertson, D. and Hyde, J., The factorial validity of the Family Environment Scale. *Educational and Psychological Measurement*, 42, 1233-1241, 1982.
 27. Roosa, M. W. and Beals, J., Measurement issues in family assessment: The case of the Family Environment Scale. *Family Process*, 29, 191-198, 1990.
 28. Zeller, R. A. and Carmines, E. G., *Measurement in the social sciences: The link between theory and data*. New York: Cambridge University Press, 1980.

Manuscript received June 6, 1989; Accepted June 6, 1989.

¹Roosa and Beals state that internal consistency reliability may be inflated when more than one person per family is included in a sample. This is not correct. For example, if each person's response were used twice in a reliability analysis, independence would be violated but the alpha coefficient would remain the same. In any case, the FES subscale alphas for subsamples in which only one family member is included (for example, either depressed patients or their spouses) are closely comparable to those shown in Table 1 for the overall samples.

²FES subscale means and standard deviations for the samples shown in Table 1 are available on request. As an aside, the internal consistencies of four of the other five FES subscales in these samples are also comparable but slightly lower than those reported in the FES Manual. The Independence subscale sometimes shows relatively low internal consistency.

³The formula for Cronbach's alpha (7) is

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum_{i=1}^N V_i}{V_s} \right)$$

where N = the number of items,

$\sum_{i=1}^N V_i$ = the sum of the item variances, and

V_s = the variance of the subscale.

As the formula shows, the primary components of Cronbach's alpha are the item and subscale variances. Thus, alphas necessarily vary in line with the characteristics of specific samples.

⁴Copies of the subscale descriptions used in the rating task are available upon request.
