Measurement Properties of the Short Multidimensional Observation Scale for Elderly Subjects (MOSES)

Manfred Diehl
University of Colorado at Colorado Springs
Diana L. Spore
Brown University
Michael A. Smyer
Boston College

This study evaluated the five-factor measurement model of the abbreviated Multidimensional Observation Scale for Elderly Subjects (MOSES), originally proposed by Pruchno, Kleban, and Resch in 1988. Modifications of the five-factor model were examined and evaluated with regard to their practical significance. A confirmatory second-order factor analysis was performed to examine whether the correlations among the first-order factors were adequately accounted for by a global dysfunction factor. Findings indicated that the proposed measurement model was replicated adequately. Although post hoc modifications resulted in significant improvements in overall model fit, the minor parameters had only a trivial influence on the major parameters of the baseline model. Results from the second-order factor analysis showed that a global dysfunction factor accounted adequately for the intercorrelations among the first-order factors.

Researchers and practitioners in long-term care have emphasized the need for comprehensive assessment of nursing home residents' functional status

AUTHORS' NOTE: This research was supported by a grant from the Retirement Research Foundation, principal investigator Michael A. Smyer. Ms. Spore was in part supported by a National Institute of Mental Health Training Grant awarded to the Pennsylvania State University and by an Agency for Health Care Policy and Research Training Grant awarded to Brown University. We gratefully acknowledge the thoughtful and constructive comments of John Nesselroade, Mike Rovine, and an anonymous reviewer on an earlier version of this article. Correspondence concerning this article should be addressed to Manfred Diehl, Department of Psychology, University of Colorado at Colorado Springs, 1420 Austin Bluffs Parkway, P.O. Box 7150; Colorado Springs, CO 80933-7150. Electronic mail may be sent via Internet to mdiehl@mail.uccs.edu.

The Journal of Applied Gerontology, Vol. 16 No. 4, December 1997 403-426

for a long time (R. L. Kane, 1990; R. A. Kane & Kane, 1981; Katz & Stroud, 1989). Issues such as resident screening, placement decisions, care plan development, or monitoring of resident status over time require multidimensional assessment of residents' functioning (Morris et al., 1990; Ouslander, 1994; Travis, 1988). Multidimensional functional assessment is also desirable with regard to the evaluation of individual- and program-level intervention programs in long-term care facilities (Smyer & Garfein, 1988). Indeed, the need for a system that provides comprehensive information about the functional potentials and limitations of nursing home residents led to a congressional mandate, as part of the Omnibus Budget Reconciliation Act (OBRA) of 1987, for the development of a Minimum Data Set (MDS) for Resident Assessment and Care Screening (Morris et al., 1990; Ouslander, 1994).

In the context of this larger development, the Multidimensional Observation Scale for Elderly Subjects (MOSES) (Helmes, Csapo, & Short, 1987) has merit because it avoids some of the shortcomings of the MDS and of the unidimensional assessment instruments (Lawton & Storandt, 1984) that are traditionally used. First, most multidimensional assessment procedures have been designed for use by specially trained professionals (e.g., clinical social workers, geriatric nurses, or psychologists) but not for use by the primary caregivers of long-term care residents, the nursing assistants. As a consequence, the knowledge base of the most informed day-to-day caregivers is often not incorporated into the resident assessment process (Ouslander, 1994). Unfortunately, this is also frequently the case with the MDS, which was designed and introduced in nursing homes with the intention of incorporating information from primary nursing staff (Ouslander, 1994). However, because of its difficult format and very elaborate structure, the MDS is usually not used by the primary caregivers, and the extent to which their knowledge about residents is incorporated into the MDS depends largely on the specific communication practices and the institutional climate of each individual nursing home.

Second, from a research perspective, there has been a dearth of rigorous examination and refinement of the psychometric properties of the existing measures (for an exception, see Mor et al., 1995; Morris et al., 1990). For example, evaluation studies with the MDS have so far focused on single dimensions of residents' functioning (see Hartmaier et al., 1995; Morris et al., 1994). Furthermore, confirmatory item factor analysis has rarely been used for scale development and validation of newly developed measures of resident functioning (for an exception, see Mor et al., 1995). The use of confirmatory item factor analysis has the great advantage in that it provides simultaneous estimates of the item factor loadings, the item measurement error, and an overall goodness-of-fit index for a specified measurement model. Thus, it is possible to test the hypothesized item factor structure directly without relying on the interpretation of exploratory factor analyses (see Hertzog, 1989).

With the MOSES, Helmes et al. (1987) wanted to offer an instrument that was sound with regard to its psychometric properties yet practical for use by primary caregivers. Helmes and his colleagues tried to accomplish this task by using plain language and by anchoring nursing staff's ratings using specific observable behaviors. In its original version, the MOSES is a 40-item questionnaire that focuses on five areas of functioning in institutionalized older adults: self-care ability, disoriented behavior, depressed/anxious mood, irritable behavior, and withdrawn behavior. For each of the five areas, higher scores indicate higher levels of impairment or dysfunctional behavior.

In developing the MOSES, Helmes et al. (1987) examined its psychometric properties in several ways. They showed that the instrument discriminated well across different institutional settings and that its scales had satisfactory internal consistencies as well as satisfactory interrater reliabilities. For smaller subsamples, Helmes et al. (1987) reported findings in support of the MOSES's convergent and discriminant validity. Furthermore, a principal components analysis of the items showed that they had their primary loadings on the hypothesized factors. Together, these findings led Helmes et al. (1987) to the conclusion that the MOSES provided "reliable, valid staff ratings of the major areas of clinical concern to health-care staff and researchers" (p. 400).

Pruchno, Kleban, and Resch (1988) extended this work and examined the factor structure of the MOSES using a sequence of exploratory and confirmatory factor analyses. Pruchno et al.'s (1988) exploratory work resulted in the elimination of 16 conceptually ambiguous items from the questionnaire. The reduced 24-item version was then examined via confirmatory item factor analysis, using an independent sample of nursing home residents. Findings from this analysis indicated a good fit between estimated and observed covariance matrices and good internal consistencies of the derived scales. Thus, based on these analyses, Pruchno et al. suggested that the 24-item MOSES was a parsimonious yet reliable alternative to the longer instrument. Pruchno et al. also demonstrated factorial invariance of the short MOSES across samples, a finding that awaits replication and extension with other independent samples of nursing home residents.

Building on this research, this study had two interrelated objectives. The first objective was to evaluate the factor structure of the 24-item MOSES using an independent cross-sectional sample of nursing home residents. The second objective was to examine whether a second-order dysfunction factor adequately accounted for the correlations among the first-order factors. With regard to the first objective, replication of findings from previous item factor analyses is often neglected in applied gerontological research (Hertzog, 1989). Several statisticians (Cohen, 1994), and psychometricians (Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991) have advocated that greater emphasis should be placed on replication. This emphasis is of particular importance in the area of measurement development and evaluation (Pedhazur & Schmelkin, 1991).

The second objective focused on the examination of a higher order dysfunction factor, an issue that has been alluded to by Pruchno et al. (1988) but has not been examined so far. Although one of the strengths of the MOSES is its ability to capture multiple dimensions of older adults' functioning, the findings from previous confirmatory factor analyses indicate that the first-order factors are substantially interrelated (Pruchno et al., 1988). These factor intercorrelations may be the consequence of a common second-order factor, which could be thought of as a global index of older adults' level of dysfunction. Substantively, the factor intercorrelations reported by Pruchno et al. (1988) suggest a single second-order factor solution rather than a multiple second-order solution because the correlations among the first-order factors fail to show the specific pattern that would justify a multiple factor solution.[1] Although the MOSES was specifically designed to assess physical, cognitive, social, and emotional components of older adults' functioning separately, as a complement to these domain-specific scores, the use of a global dysfunction score may serve some important pragmatic purposes in geriatric practice (e.g., facilitating communication among practitioners).

Method

Participants

Study participants were 273 nursing home residents from four Pennsylvania nursing homes that were part of a short-term longitudinal study on the effects of skills training on nursing staff's job performance and job satisfaction (see Smyer, Brannon, & Cohn, 1992). The sample for this study included all residents over age 50 with complete data on the MOSES at the first occasion of assessment.

Residents were, on average, 85 years old (SD = 7.7, range = 51-102) and had resided in the facilities, on average, for 27.2 months (SD = 27.2, range = .4-142.5). The majority of the residents were female (83%) and widowed

(78%). Comparison of the 273 residents with the total study sample yielded no significant differences with regard to key demographic variables such as age, gender, marital status, and residential duration.

Compared to population characteristics based on the 1985 National Nursing Home Survey (National Center for Health Statistics, 1989), the study sample included more residents who were female (83% vs. 72%) and widowed (78% vs. 61%) and fewer residents who were divorced/separated (1% vs. 8%) or who had never married (11% vs. 18%). On average, participants in this study were older (mean age: 85 vs. 79 years) and had not lived as long in the nursing home (mean duration of residency: 27 vs. 35 months) as older adults in the National Nursing Home Survey (National Center for Health Statistics, 1989).

Procedures

Administration of the MOSES. Ratings of residents' day-to-day behavior and functioning were collected from primary caregivers (i.e., nursing assistants) in group sessions at the study sites. During these sessions, nursing assistants responded to the 40-item MOSES and two other questionnaires concerning their situations as caregivers. All questionnaires were administered by a graduate research assistant, who was present during the testing sessions to answer questions and to provide assistance, if necessary. To ensure the accuracy of the information regarding residents' functioning, primary caregivers completed the MOSES only for residents for whom they had provided care during the week preceding the testing session.

Raters. Ratings were provided by 113 nursing assistants, most of them women (95%), who had, on average, a high-school education. About 62% of the nursing assistants reported that they had received nurse aide training beyond what was provided during in-services at the nursing homes. On average, raters were 35 years old (SD = 10.6, range = 19-63), had 68.8 months (SD = 67.2) of experience as nursing assistants, and had been employed at the site for 39.7 months (SD = 41.9). The majority (66%) worked 35 hours or more weekly, during the day or evening.

Scoring procedures. Given the more severely impaired nature of the study sample when compared to the standardization sample (Helmes et al., 1987), nursing assistants' responses were scored adopting the scoring procedures established by Pruchno et al. (1988). In particular, the rating *Does Not Apply* for the items Talking, Awareness of Time, Memory for Recent Events, and

Memory for Remote Events was scored to indicate the highest level of impairment. In contrast, the response *Does Not Apply* for the items Reporting Sadness and Depression, Sounding Sad and Depressed, Pessimism About the Future, Verbal Abuse of Staff, Physical Abuse of Others, and Friendships with Other Residents was scored as *Not At All.* Five items with anchors for rating a resident's behavior, exemplifying each of the scales, are provided in the appendix.

Statistical Procedures

Confirmatory factor analyses were performed using the maximum-likelihood algorithm of LISREL 8 (Jöreskog & Sörbom, 1993). The LISREL measurement model and the way in which it relates the observed/measured variables (i.e., items) to the latent variables (i.e., factors) have been described by Jöreskog and Sörbom (1989, 1993).

Modifications of the baseline measurement model were pursued in two ways: First, after examination of the estimated parameters and the overall goodness-of-fit indices, the modification indices provided by LISREL were consulted with regard to substantively meaningful modifications of the baseline model. These modifications were evaluated with regard to their practical significance using the approach suggested by Byrne, Shavelson, and Muthén (1989). Second, confirmatory second-order factor analysis was performed to examine whether the correlations among the first-order factors were adequately accounted for by a second-order factor indicative of participants' global dysfunction status (Liang & Bollen, 1983; Rindskopf & Rose, 1988).

Because recent research (see Bollen & Long, 1993b) on the evaluation of structural equation models has shown that different goodness-of-fit indices are differentially sensitive to the effects of sample size and to deviations from multivariate normality, we adopted the recommendations of Bollen and Long (1993a); Raykov, Tomer, and Nesselroade (1991); and Tanaka (1993) and inspected goodness-of-fit indices from different families. As Gerbing and Anderson (1993) have pointed out, none of these indices has been endorsed as the "best index" (p. 41), and it is important to examine them in concert. Specifically, five fit indices were consulted: (a) the likelihood ratio (χ^2) test statistic; (b) the relative likelihood ratio (χ^2/df) (Carmines & McIver, 1981); (c) the goodness-of-fit index provided by LISREL (GFI) (Jöreskog & Sörbom, 1993); (d) the Bentler and Bonett (1980) normed fit index (NFI), and (e) Bentler's (1990) comparative fit index (CFI).

A model that provides a good fit to the data has a low chi-square and high GFI, NFI, and CFI. Differences in chi-square were used to examine the changes in goodness of fit resulting from the successive modifications of the baseline model. For the purposes of this study, relative likelihood ratios (χ^2/df) between 2 to 1 and 3 to 1 were considered indicative of an acceptable model fit (Carmines & McIver, 1981). Similarly, a GFI, NFI, and CFI of .80 or higher was considered acceptable, whereas indices in excess of .90 were considered good. A relative fit index (RNFI) that reflects the degree of fit of the second-order model to the first-order covariance matrix was used to evaluate the fit of the second-order model (Hertzog, Van Alstine, Usala, Hultsch, & Dixon, 1990).

Results

Findings are reported in three major sections. First, results are presented with regard to the replication of the baseline measurement model of the 24-item MOSES. Second, findings derived from a sequence of modifications of the baseline model and an evaluation of the best fitting model with regard to the practical significance of the modifications are reported. Finally, findings of a confirmatory second-order factor analysis are reported.

Replication of the Measurement Model

LISREL 8 (Jöreskog & Sörbom, 1993) was used to replicate the measurement model established by Pruchno et al. (1988) for the abbreviated version of the MOSES. Items for each factor were estimated freely whereas the loadings of the remaining items were fixed to zero for the respective factor. The unit of measurement for each first-order factor was established by having the loading of the first item constrained to 1.0, and factors were allowed to be intercorrelated. The item covariance matrix was analyzed.

The estimated factor loadings derived from LISREL's standardized solution are shown in Table 1. The standardized solution provided by LISREL is the rescaled maximum likelihood solution in which the latent variables (i.e., factors) are scaled to have a variance of 1.0, whereas the observed variables remain in their original metric (Jöreskog & Sörbom, 1989). The standardized estimates are presented here to permit a direct comparison with the parameter estimates reported by Pruchno et al. (1988). To facilitate this comparison, the estimates obtained by Pruchno et al. (1988) are shown in parentheses. The

Table 1. Standardized Estimates of Factor Loadings

	Factor Loadings				
Items	*Self-Care*	*Disorientation*	*Depression*	*Irritability*	*Withdrawal*
Dressing	.88 (.89)				
Bathing	.78 (.65)				
Incontinence	.73 (.67)				
Using the Toilet	.91 (.91)				
Physical Mobility	.78 (.73)				
Understand Communication		.65 (.64)			
Talking		.73 (.77)			
Recognizing Staff		.81 (.78)			
Awareness of Time		.93 (.94)			
Memory for Recent Events		.93 (.87)			
Memory for Past Events		.88 (.89)			
Reporting Sadness			.86 (.79)		
Sounding Sad			.86 (.79)		
Looking Worried			.56 (.51)		
Pessimism About the Future			.63 (.79)		
Self-Concern			.49 (.54)		
Reactions to Frustration				.79 (.75)	
Verbal Abuse of Staff				.84 (.59)	
Physical Abuse of Others				.54 (.44)	
Initiate Social Contacts					.65 (.69)
Friendships With Residents					.64 (.58)
Interest in Daily Events					.78 (.82)
Interest in Outside Events					.74 (.75)
Keeping Occupied					.68 (.72)

NOTE: All factor loadings are significant at $p < .001$. Estimates reported by Pruchno, Kleban, and Resch (1988) are shown in parentheses.

factor intercorrelations and internal consistencies of the scales are displayed in Table 2.

All estimated factor loadings were significant ($p < .001$) and similar in magnitude to the loadings reported by Pruchno et al. (1988). Except for two intercorrelations (i.e., depression and disorientation, withdrawal and depression), the correlations among the first-order factors were all statistically significant ($p < .01$). The overall goodness-of-fit indices for this measurement model were: $\chi^2(242, N = 273) = 590.47$, $\chi^2/df = 2.44$, GFI = .84, NFI = .86,

Table 2. Factor Intercorrelations and Internal Consistencies of the Short Multidimensional Observation Scale for Elderly Subjects (MOSES) Scales

Factors	*1*	*2*	*3*	*4*	*5*
1. Self-Care	.93 (.90)				
2. Disorientation	.52 (.72)	.94 (.92)			
3. Depression	.18 (.16)	.08 (.04)	.86 (.75)		
4. Irritability	.27 (.27)	.44 (.27)	.27 (.45)	.86 (.62)	
5. Withdrawal	.56 (.64)	.75 (.78)	.14 (.11)	.46 (.33)	.86 (.84)

NOTE: Internal consistencies (Cronbach's alpha) for the derived scales are shown in the main diagonal. Estimates reported by Pruchno, Kleban, and Resch (1988) are shown in parentheses.

and CFI = .91. Although these indices were somewhat lower than the ones reported by Pruchno et al. (1988), they were in the acceptable range.

In general, the somewhat lower overall goodness-of-fit indices were most likely the result of some nonnormally distributed items (Bollen, 1989). For example, Hertzog (1989) has pointed out that the overall goodness-of-fit indices for item factor analyses are often attenuated because the assumptions of multivariate normality are violated in data consisting of ordinal rating scales. In particular, Huba and Harlow (1986) have shown that this violation does not greatly affect the individual LISREL parameter estimates, but it does affect the standard errors and the overall goodness-of-fit indices. In addition, Bollen and Long (1993a) have pointed out that overall goodness-of-fit indices should be used as *one* criterion only in assessing a model's goodness of fit and should be interpreted within the context of the individual parameter estimates. In this study, the individual parameter estimates were almost identical to the estimates reported by Pruchno et al. (1988) thus providing independent support for the previously established measurement model.

Post Hoc Modifications

Although the estimated parameters were very similar to the ones reported by Pruchno et al. (1988), the overall goodness-of-fit indices suggested that not all of the item covariances were adequately accounted for by the specified measurement model. Thus, the modification indices provided by LISREL were consulted to identify where the measurement model failed to fit the observed data. This inspection suggested four changes in the baseline model. First, the unique variances of the items Dressing and Bathing were suggested to be correlated. Second, the unique variances of the items Self-Concern and Pessimism About the Future were suggested to be correlated. Third, the item

Self-Concern was suggested to have a secondary loading on the factor Irritability, and fourth, the item Looking Sad and Anxious was suggested to have a secondary loading on the factor Irritability.

All of these suggested modifications seemed substantively meaningful, especially when the wording of the items was taken into account (Hertzog, 1989). Thus, post hoc modifications were pursued. One at a time, the suggested parameters were estimated, and both the magnitude of the parameters and the changes in overall goodness of fit were examined. The results of these modifications are summarized in Table 3.

As can be seen in Table 3, each subsequent modification of the baseline model resulted in a significant improvement in overall model fit. The overall goodness of fit of the best fitting model was $\chi^2(238, N = 273) = 456.81$, χ^2/df ratio = 1.92, GFI = .88, NFI = .89, and CFI = .95.

Of the newly estimated parameters, the standardized secondary loadings of the items Self-Concern and Looking Sad and Anxious on the factor Irritability were .29 and .41, respectively (both $p < .001$). The correlations between the unique variances of the items Dressing and Bathing and between Self-Concern and Pessimism About the Future were .20 and .18, respectively (both $p < .001$). Table 4 shows the standardized parameter estimates for the best fitting model.

Although the post hoc modifications resulted in significant improvements of overall model fit, whether the post hoc parameters were of practical significance was examined by correlating the major parameters (the λs and ϕs) in the baseline model with those in the best fitting post hoc model. As Byrne et al. (1989) have pointed out, coefficients close to 1 support the stability of the initial model and thus the triviality of the minor parameters in the post hoc model, whereas coefficients not close to 1 (e.g., less than .90) indicate that the major parameters were adversely affected and call for the inclusion of the post hoc parameters in the model.

The correlations between the estimated factor loadings and factor intercorrelations in the baseline model and the best fitting post hoc model were $r = .91$ and $r = .99$, respectively (both $p < .001$). This suggested that the minor parameters in the post hoc model had only a trivial influence on the major parameters of the baseline model. Thus, this sensitivity analysis further substantiated the stability of the baseline model established by Pruchno et al. (1988).

Confirmatory Second-Order Model

Given that the intercorrelations of the first-order factors were substantial both for the baseline and the best fitting model (see Tables 2 and 4), a

Table 3. Steps in Fitting Post Hoc Models

Competing Models	χ^2	df	$\Delta\chi^2$	Δdf	χ^2/df	*GFI*	*NFI*	*CFI*
1. Five-factor model, $\Theta_i\Theta_j = 0$	590.47	242			2.44	.84	.86	.91
2. $\Theta_{2,1}$, free	535.11	241	55.36*	1	2.22	.86	.88	.93
3. $\Theta_{2,1}$, $\lambda_{16,4}$, free	498.92	240	36.19*	1	2.08	.87	.88	.94
4. $\Theta_{2,1}$, $\lambda_{16,4}$, $\lambda_{14,4}$, free	475.06	239	23.86*	1	1.99	.87	.89	.94
5. $\Theta_{2,1}$, $\lambda_{16,4}$, $\lambda_{14,4}$, $\Theta_{16,15}$, free	456.81	238	18.25*	1	1.92	.88	.89	.95

NOTE: GFI = Jöreskog and Sörbom's (1989) goodness-of-fit index; NFI = Bentler and Bonett's (1980) normed fit index; CFI = Bentler's (1990) comparative fit index.
*$p < .001$.

confirmatory second-order factor analysis was performed to examine whether a global dysfunction factor adequately accounted for these correlations. Because the sensitivity analysis suggested that the post hoc modifications had only a trivial influence on the major parameters of the baseline model, the baseline model was specified for the higher order analysis. Figure 1 shows the simultaneous second-order model and the second-order factor loadings and unique variances from the completely standardized solution provided by LISREL. Table 5 displays the associated standardized first-order factor loadings and unique variances.

All of the factor loadings of the first-order factors on the global dysfunction factor were significant (see Figure 1), with the Depression factor loading being considerably lower than the loadings for Self-Care, Disorientation, Irritability, and Withdrawal. The goodness of fit of the second-order model was acceptable: $\chi^2(248, N = 273) = 625.36$, $\chi^2/df = 2.52$, GFI = .83, NFI = .85, CFI = .91. The RNFI was .88, suggesting that most of the information in the intercorrelations of the first-order factors was accounted for by the second-order factor. However, compared to the first-order baseline model, there was a statistically significant loss of fit, $\Delta\chi^2(6) = 34.89$, $p < .01$, suggesting that the combined first- and second-order model did not fit as well as the first-order model.

Discussion

This study sought to evaluate the measurement model for the short MOSES using an independent cross-sectional sample of nursing home residents and to examine whether a second-order factor adequately accounted for the intercorrelations among the first-order factors.

Table 4. Estimates for Best Fitting Post Hoc Model

	Factor Loadings					
Items	*Self-Care*	*Disorientation*	*Depression*	*Irritability*	*Withdrawal*	*Unique Variances*
Dressing	.83					.31
Bathing	.71					.50
Incontinence	.74					.46
Using the Toilet	.95					.10
Physical Mobility	.80					.36
Understand Communication		.65				.57
Talking		.73				.46
Recognizing Staff		.81				.34
Awareness of Time		.94				.12
Memory for Recent Events		.93				.13
Memory for Past Events		.88				.22
Reporting Sadness			.86			.26
Sounding Depressed			.88			.23
Looking Worried			.48	.29		.63
Pessimism About the Future			.62			.62
Self-Concern			.35	.41		.65
Reactions to Frustration				.81		.34
Verbal Abuse of Staff				.80		.36
Physical Abuse of Staff				.55		.70

Initiate Social Contacts	.64	.59
Friendships With Residents	.64	.59
Interest in Daily Events	.77	.40
Interest in Outside Events	.74	.45
Keeping Occupied	.68	.54

	Factor Intercorrelations				
Factors	*1*	*2*	*3*	*4*	*5*
1. Self-Care	1.00				
2. Disorientation	.48	1.00			
3. Depression	.15	.05	1.00		
4. Irritability	.24	.46	.22	1.00	
5. Withdrawal	.54	.76	.11	.47	1.00

NOTE: The estimates for the correlated errors/unique variances between the items Dressing and Bathing and between the items Self-Concern and Pessimism About the Future were .20 and .18, respectively.

With regard to the first objective, our findings provide additional support for the five-factor structure of the short MOSES. The parameter estimates obtained in this study were very similar to the estimates reported by Pruchno et al. (1988). Indeed, the correlations between the factor loadings (the λs) and the factor intercorrelations (the ϕs) across studies were .95 and .89, respectively (both $p < .001$), indicating almost identical estimates. In addition, the internal consistencies of the scales were all above .85, suggesting that the items form homogenous unidimensional scales.

Findings from this study also reflect positively on the work of Helmes et al. (1987), who sought to develop an instrument that could be used by primary caregivers without extensive training. To a great extent, this aim appears to have been achieved. We found that nursing assistants were able to complete the instrument with little difficulty. The items contained in the MOSES appear to be sufficiently concrete and behavior-specific so that nursing assistants can provide reliable accounts of the residents' functioning in the relevant domains. Reliabilities are particularly good for easily observable behaviors such as self-care activities and disoriented behavior and somewhat lower for behaviors indicative of depression, irritability, and withdrawal. Items on the latter scales often require some clinical knowledge and inferences on the part of the primary caregiver, which may lead to ambiguity and, hence, the somewhat lower reliabilities.

There was, however, also some indication that substantively meaningful modifications of the five-factor measurement model should be explored in future work with the MOSES. Four successive modifications of the baseline model resulted in statistically significant improvements in overall goodness of fit, suggesting that alternative measurement models should not be ruled out. For example, there was some evidence that some within-scale error variances (e.g., the unique variances of Dressing and Bathing) should be considered correlated rather than uncorrelated (see Liang & Bollen, 1983). Such correlated error variances may be the result of very similar item wording.

Although there was some evidence in support of a modified measurement model, at that point in time, we were hesitant to adopt the revised measurement model for primarily two reasons. First, we lacked an independent cross-validation sample, which should be used to test the modified measurement model, ruling out that the suggested modifications were simply an effect of sample-specific characteristics. Second, the sensitivity analysis of the best fitting post hoc model revealed that the additional parameters were of little practical significance.

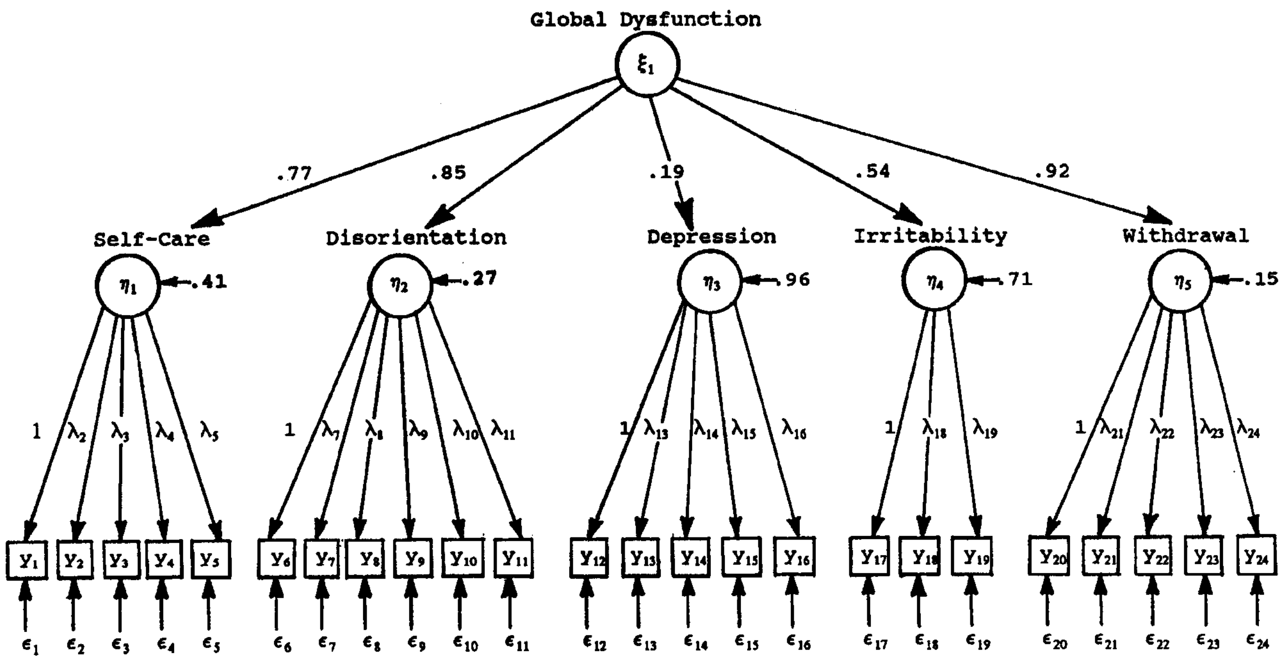

Figure 1. Second-order factor model for the short Multidimensional Observation Scale for Elderly Subjects (MOSES).

Table 5. Factor Loadings and Unique Variances from Simultaneous Second-Order Factor Analysis (see Figure 1)

	First-Order Factor Loadings					
Items	*Self-Care*	*Disorientation*	*Depression*	*Irritability*	*Withdrawal*	*Unique Variances*
y_1: Dressing	.92					.15
y_2: Bathing	.84					.30
y_3: Incontinence	.78					.39
y_4: Using Toilet	.92					.15
y_5: Physical Mobility	.81					.34
y_6: Understand Communication		.68				.53
y_7: Talking		.76				.43
y_8: Recognizing Staff		.83				.31
y_9: Awareness of Time		.94				.11
y_{10}: Memory for Recent Events		.94				.12
y_{11}: Memory for Past Events		.90				.20
y_{12}: Reporting Sadness			.85			.27
y_{13}: Sounding Depressed			.87			.25
y_{14}: Looking Anxious			.56			.68
y_{15}: Pessimism About the Future			.64			.60
y_{16}: Self-Concern			.49			.76
y_{17}: Reactions to Frustration				.78		.40
y_{18}: Verbal Abuse of Staff				.87		.25
y_{19}: Physical Abuse of Staff				.57		.68
y_{20}: Initiate Social Contacts					.68	.54
y_{21}: Friendships With Residents					.68	.54
y_{22}: Interest in Daily Events					.80	.37
y_{23}: Interest in Outside Events					.77	.41
y_{24}: Keeping Occupied					.71	.49

NOTE: Factor loadings and unique variances as provided by the completely standardized solution.

Thus, in combination with the results reported by Pruchno et al. (1988), our findings support the five-factor structure of the short MOSES and its applicability as a parsimonious instrument for the multidimensional assessment of long-term care residents. The ease of obtaining the information from the primary caregiver is particularly noteworthy because other multidimensional assessment procedures such as the MDS have been found to be too difficult for use by nursing assistants (Teresi & Holmes, 1992). Other features that make the short MOSES interesting to practitioners and researchers alike are its relative time- and cost-effectiveness.

With regard to the second objective, our findings provide some preliminary evidence in support of a second-order factor accounting for the observed correlations among the first-order factors. All of the factor loadings relating the first-order factors to the second-order factors were significant and, with the exception of the loading for Depression, of substantial magnitude. We propose that the second-order factor may best be labeled as a *global dysfunction factor*.

There was, however, also a significant loss in goodness of fit when we moved from a first-order to a simultaneous second-order model. This loss in goodness of fit in combination with the rather low second-order factor loading for Depression resulted in hesitance about accepting the revised measurement model. Thus, the evidence in favor of the second-order factor should be considered preliminary. Future work is needed to examine whether a global dysfunction factor can be replicated. Replication of the second-order factor solution with an independent sample would create the justification for the use of a total dysfunction score. Although such a global dysfunction score may serve some pragmatic purposes in applied gerontology, we do not advocate that it should be used instead of the domain-specific functioning scores. Rather, a global dysfunction score should serve as a complement to the domain-specific scores and not as a replacement.

Like all studies, this study has several limitations that need to be noted. First, although study participants were selected from four different nursing homes, the study sample only approximated the population characteristics defined by the 1985 National Nursing Home Survey (National Center for Health Statistics, 1989). Thus, it cannot be ruled out that some subtle selection effects (see Nesselroade, 1988) may have affected the caregivers' responses and, hence, the relations among the items.

Second, although this study used a sizable and diverse sample of nursing home residents, because of the numbers of free parameters estimated in the different models, it would have been desirable to have a somewhat larger study sample. In general, the required sample size increases as the model becomes more complex. Bentler (1985), for example, has recommended that the ratio between sample size and the number of estimated parameters should range from 5 to 10. In the present study, this ratio was somewhat below the lower bound and ranged from 4.7 to 4.4. Thus, replication of the analyses performed in this study with a larger independent sample seems to be warranted.

Third, it needs to be noted that in this study, the MOSES was administered in a slightly different way than was intended by the test developers. Because the amount of time that the nursing assistants could spend away from the floor was limited, we chose to administer the MOSES for each resident as a paper-and-pencil measure rather than in interview format. Although this procedure worked out extremely well, it cannot be completely ruled out that this mode of administration may have resulted in somewhat different responses than face-to-face interviews with the primary caregivers would have elicited. Future research, therefore, needs to address how different modes of administration may affect the obtained responses and may alter the instrument's psychometric properties.

Fourth, it needs to be mentioned that although the standardized solution makes the comparison of this study's findings with Pruchno et al.'s (1988) results easy, it may also obscure some differences between the two study samples. For example, it could well be that the variances of the latent variables in our study were notably different (i.e., larger or smaller) from the variances in the Pruchno et al. study. Only a comparison of the unstandardized maximum likelihood solutions from both studies could reveal whether the comparisons may be confounded by different sample variances and whether there may be any reason for concern.

Fifth, post hoc model fitting using the same sample has been criticized (Cliff, 1983; MacCallum, 1986) because it runs the risk of capitalizing on unique characteristics of the study sample. However, we chose to adopt the pragmatic position advocated by Byrne et al. (1989) and others (e.g., Jöreskog, 1982; Tanaka & Huba, 1984), who have argued "that as long as the researcher is fully cognizant of the exploratory nature of his or her analyses, the process can be substantively meaningful" (Byrne et al., 1989, p. 461).

Finally, it should be noted that this study was conducted before OBRA regulations mandated the nationwide implementation of a uniform resident assessment instrument, the MDS. Therefore, the issue of convergent and divergent validity of the MOSES with scales from the MDS cannot be addressed in this study but should be a major objective for future research. More important, future research with the MOSES should link nursing home residents' scores on the different scales of impairment to known causes and consequences of older adults' dysfunctions in order to further establish the construct as well as the predictive validity of the instrument.

In summary, this study replicated the five-factor structure of the 24-ite MOSES using an independent cross-sectional sample. In addition, a secor order factor analysis yielded preliminary evidence in support of a glol dysfunction factor determining the intercorrelations among the five fir order factors. Because there is a great need in applied gerontology and pub policy for reliable multidimensional assessment of older adults' functioni (see Morris et al., 1990), the short MOSES should be considered as one the instruments of choice. Its brevity and ease of administration together w its psychometric properties make it an attractive instrument for practition and researchers alike. However, there also remain some critically importa issues that need to be addressed in future work with the short MOSES. F example, its convergent and discriminant validity (Campbell & Fiske, 195 with other multidimensional measures, such as the MDS, needs to be exa ined systematically. Furthermore, the instrument's factorial invariance acro different populations of older adults and over time needs to be examin systematically (Cunningham, 1991; Horn & McArdle, 1992). Despite the outstanding issues, research with the MOSES to date suggests that it is useful assessment instrument for research and applied purposes.

Appendix

Selected Items From the Short Multidimensional Observation Scale for Elderly Subjects (MOSES)

1. *DRESSING*

On most days in the past week, the resident:

1. Initiated and completed dressing without staff supervision.
2. Dressed with only minor supervision.
3. Partly dressed him/herself, but needed frequent staff assistance.
4. Was either totally dressed by staff or remained in bedclothes.

9. *AWARENESS OF TIME*

Consider whether on most days in the past week the resident was awa of (a) the year (within 1), (b) the season, and (c) the approximate time day (for example, whether it was morning or after lunch or after suppe

1. S/he was aware of all three (year, season, and time of day).
2. S/he was aware of two of the three.
3. S/he was aware of one of the three.
4. S/he was confused about all three.
5. This information could not be obtained—the resident did not communica appropriately.

12. *REPORTING SADNESS AND DEPRESSION* (Talking about being sad or depressed or wanting to be somewhere else. Do not include complaints about the care. Also do not include talking about being worried.)

How often during the past week did the resident say (or write) something to indicate that s/he was sad or depressed:

1. Not at all.
2. Seldom (only one to three times during the week).
3. At times (either once or twice a day on more than three days, or several times a day on one to three days).
4. Often (several times a day on more than three days. Also include here any resident who specifically said s/he wanted to be dead).
5. Question does not apply—the resident did not speak (or write) in the past week.

18. *VERBAL ABUSE OF STAFF* (Include yelling at, swearing at, cursing, threatening.)

How often during the past week did the resident verbally abuse staff members:

1. Not at all.
2. Sometimes.
3. Frequently (at least once a day on more than three days) when asked to do something s/he didn't want to do.
4. Frequently (at least once a day on more than three days) with no apparent provocation or cause.
5. Question does not apply—the resident did not speak or make any sounds in the past week.

22. *INTEREST IN DAY-TO-DAY EVENTS* (For example, watching or listening and reacting to things going on around him or her.)

In the past week, how often did the resident pay active attention to the things happening around him or her:

1. Often (on more than three days for most of the day).
2. At times (either on more than three days for only short periods of time, or on one to three days for most of the day).
3. Seldom (on one to three days for only short periods of time).
4. Not at all.

Note

1. Conceptually, a plausible higher order solution with two factors would have been one in which the scales Self-Care, Disorientation, and Irritability would have loaded on one higher order factor, whereas the scales Depression and Withdrawal would have loaded on a second higher

order factor. Inspection of the factor intercorrelations in Table 2, however, shows, that Withdrawal had higher correlations with Self-Care, Disorientation, and Irritability than with Depression, suggesting that the data did not support this conceptually plausible model. Moreover, estimation of this conceptual model resulted in a significantly poorer goodness of fit than the single-factor model.

References

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program.* Los Angeles: BMDP Statistical Software Inc.

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,* 588-600.

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

Bollen, K. A., & Long, J. S. (1993a). Introduction. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 1-9). Newbury Park, CA: Sage.

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456-466.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 124-132.

Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt & E. F. Borgatta (Eds.), *Social measurement: Current issues* (pp. 65-115). Beverly Hills, CA: Sage.

Cliff, N. (1983). Some cautions of causal modeling methods. *Multivariate Behavioral Research, 18,* 115-126.

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

Cunningham, W. R. (1991). Issues in factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 106-114). Washington, DC: American Psychological Association.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.

Hartmaier, S. L., Sloane, P. D., Guess, H. A., Koch, G. G., Mitchell, C. M., & Phillips, C. D. (1995). Validation of the Minimum Data Set cognitive performance scale: Agreement with the Mini-Mental State Examination. *Journal of Gerontology: Medical Sciences, 50A,* M128-M133.

Helmes, E., Csapo, K. G., & Short, J. A. (1987). Standardization and validation of the Multidimensional Observation Scale for Elderly Subjects (MOSES). *Journal of Gerontology, 42,* 395-405.

Hertzog, C. (1989). Using confirmatory factor analysis for scale development and validation. In M. P. Lawton & A. R. Herzog (Eds.), *Special research methods for gerontology* (pp. 281-306). Amityville, NY: Baywood.

Hertzog, C., Van Alstine, J., Usala, P. D., Hultsch, D. F., & Dixon, R. A. (1990). Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 64-72.

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

Huba, G. J., & Harlow, L. L. (1986). Robust estimation for causal models: A comparison of methods in some developmental data sets. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 7, pp. 69-111). Hillsdale, NJ: Lawrence Erlbaum.

Jöreskog, K. G. (1982). Analysis of covariance structures. In C. Fornell (Ed.), *A second generation of multivariate analysis: Vol. 1. Methods* (pp. 200-242). New York: Praeger.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS Inc.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: User's reference guide*. Mooresville, IN: Scientific Software Inc.

Kane, R. A., & Kane, R. L. (1981). *Assessing the elderly: A practical guide to measurement.* Lexington, MA: Lexington Books.

Kane, R. L. (1990). Standardized assessment as a means rather than an end. *The Gerontologist, 30*, 291-292.

Katz, S., & Stroud, M. W. (1989). Functional assessment in geriatrics: A review of progress and direction. *Journal of the American Geriatrics Society, 37*, 721-727.

Lawton, M. P., & Storandt, M. (1984). Clinical and functional approaches to the assessment of older people. In P. McReynolds & G. J. Chelune (Eds.), *Advances in psychological assessment* (Vol. 6, pp. 236-276). San Francisco: Jossey-Bass.

Liang, J., & Bollen, K. A. (1983). The structure of the Philadelphia Geriatric Center Morale Scale: A reinterpretation. *Journal of Gerontology, 38*, 181-189.

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100*, 107-120.

Mor, V., Branco, K., Fleishman, J., Hawes, C., Phillips, C., Morris, J., & Fries, B. (1995). The structure of social engagement among nursing home residents. *Journal of Gerontology: Psychological Sciences, 50B*, P1-P8.

Morris, J. N., Fries, B. E., Mehr, D. R., Hawes, C., Phillips, C., Mor, V., & Lipsitz, L. A. (1994). MDS cognitive performance scale. *Journal of Gerontology: Medical Sciences, 49*, M174-M182.

Morris, J. N., Hawes, C., Fries, B. E., Phillips, C. D., Mor, V., Katz, S., Murphy, K., Drugovich, M. L., & Friedlob, A. S. (1990). Designing the National Resident Assessment Instrument for nursing homes. *The Gerontologist, 30*, 293-307.

National Center for Health Statistics. (1989). *The National Nursing Home Survey: 1985 summary for the United States* (Vital and Health Statistics, Series 13, No. 97, DHHS Pub. No. PHS 89-1758). Washington, DC: U.S. Government Printing Office.

Nesselroade, J. R. (1988). Sampling and generalizability: Adult development and aging research issues examined within the general methodological framework of selection. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 13-42). New York: Springer.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Ouslander, J. G. (1994). Maximizing the Minimum Data Set. *Journal of the American Geriatrics Society, 42*, 1212-1213.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Pruchno, R. A., Kleban, M. H., & Resch, N. L. (1988). Psychometric assessment of the Multidimensional Observation Scale for Elderly Subjects (MOSES). *Journal of Gerontology: Psychological Sciences, 43*, P164-169.

Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in *Psychology and Aging*: Some proposed guidelines. *Psychology and Aging, 6*, 499-503.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*, 51-67.

Smyer, M. A., Brannon, D., & Cohn, M. (1992). Improving nursing home care through training and job redesign. *The Gerontologist, 32*, 327-333.

Smyer, M. A., & Garfein, A. J. (1988). Effective assessment approaches. In M. A. Smyer, M. D. Cohn, & D. Brannon (Eds.), *Mental health consultation in nursing homes* (pp. 82-99). New York: New York University Press.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

Tanaka, J. S., & Huba, G. J. (1984). Confirmatory hierarchical factor analyses of psychological distress measures. *Journal of Personality and Social Psychology, 46*, 621-635.

Teresi, J. A., & Holmes, D. (1992). Should MDS data be used for research? *The Gerontologist, 32*, 148-149.

Travis, S. S. (1988). Observer-rated functional assessments for institutionalized elders. *Nursing Research, 37*(3), 138-143.

Article accepted April 17, 1997.

Manfred Diehl, Ph.D., is assistant professor of psychology at the University of Colorado at Colorado Springs. He received his master's in psychology from the Friedrich-Wilhelms-Universität, Bonn, Germany, and his doctorate from the Department of Human Development and Family Studies at the Pennsylvania State University. His research focuses on older adults' everyday problem-solving abilities and on personality development in adulthood. His most recent publications include "Everyday Problem-Solving in Older Adults: Observational Assessment and Cognitive Correlates" and "Age and Sex Differences in Strategies of Coping and Defense Across the Life Span," both published in Psychology and Aging.

Diana L. Spore, Ph.D., earned a master's degree in gerontological studies from Miami University and a doctorate in human development and family studies from the Pennsylvania State University. After completion of her doctoral work, she was an AHCPR-funded postdoctoral fellow at the Center for Gerontology and Health Care Research at Brown University. Her research interests focus on psychoactive drug use among residents of board and care facilities and nursing homes, innapropriate drug use by older adults, and functional and residential status transitions among nursing home residents. Her recent publications include "Regulatory Environment and Psychotropic Drug Use in Board and Care Facilities: Results From a 10-State Study" and "Inappropriate Drug Prescriptions for Elderly Residents of Board and Care Facilities," which have been published in the Journal of Gerontology: Medical Sciences *and the* American Journal of Public Health.

Michael A. Smyer, Ph.D., is dean of the Graduate School of Arts and Sciences and associate vice president for research at Boston College. He received his undergraduate degree in psychology from Yale University and his doctorate in personality and clinical psychology from Duke University. A licensed psychologist, he is past president of the Division of Adult Development and Aging of the American Psychological Association. His research has focused on the development, implementation, and evaluation of health-related interventions for older adults and their families. Among his recent publications are Older Adults' Decision Making and the Law, *coedited with K. W. Schaie and M. B. Kapp, and* Aging and Mental Health, *coauthored with S. H. Qualls.*