Psychological Services
2008, Vol. 5, No. 2, 161–168

1541-1559/08/$12.00 DOI: 10.1037/1541-1559.5.2.161

The Development of a Brief Version of the Nurse's Observation Scale for Inpatient Evaluation (NOSIE-30)

Nicole T. Jones and Brian Sheitman
Dorothea Dix Hospital and University of North Carolina, Chapel Hill

Dennis R. Combs
University of Texas at Tyler

David Penn
University of North Carolina, Chapel Hill

Mark Hazelrigg and Betty Paesler
Dorothea Dix Hospital

Psychiatric hospitals typically do not use reliable and valid clinical instruments when recording behavioral observations. This could be attributable to these instruments being perceived as too complicated and time consuming to routinely administer. The goal of the present study was to examine if a briefer version of the 30 item Nurses' Observation for Inpatient Evaluation (NOSIE); could be developed without compromising the reliability and the factorial validity of the scale. Confirmatory Factor Analysis (CFA) was performed with a combined sample of 67 persons (with 4,973 total numbers of observations) from two settings, by randomly assigning participant observations to either phase one (full-scale model, 30-item NOSIE) or to phase two (reduced-scale model, 12-item NOSIE). The results indicated that the reduced 12-item model with the correlated latent structures appears to be the best representation of the observed data. The potential clinical usefulness of a reduced item model of the NOSIE is discussed in light of these results.

Keywords: behavioral measurement, inpatient treatment, serious mental illness

Research on the inpatient treatment of persons with serious mental illnesses has highlighted the need for empirically supported clinical assessment procedures in the development and implementation of treatment programming (Paul & Menditto, 1992; Santos, Henggeler, Burns, Arana, & Meisler, 1995; Silverstein, Spaulding, & Menditto, 2006). In addition, two of the primary regulatory bodies for inpatient treatment have also cited the importance of routine and reliable clinical assessment. In its report to Congress (Thompson, 2002), the Centers for Medicare and Medicaid Services (CMS) reported inpatient assessment procedures should be "administratively simple, have a reasonable number of data elements, rely on commonly available data, and be uniformly collected" (p. 32). In 2003, the President's New Freedom Commission on Mental Health (2003) published, "*Achieving the Promise: Transforming Mental Health Care in America. Final Report*," which advocated for the improvement of the mental health system through the consistent use of scientifically sound, evidence based services. To achieve this, reliable and valid outcome instruments are a necessity. Building on this, the Joint Commission on Accreditation of Healthcare Organizations (2006) has now initiated the identification and implementation of core performance measures with assessment as one of the core seven domains to measure objectively as of January 1, 2007.

Although the inpatient psychiatric literature consistently cites the importance of ongoing clinical assessment, few hospitals routinely use structured data collection instruments that are reliable

Nicole T. Jones and Brian Sheitman, Dorothea Dix Hospital, Raleigh, North Carolina, and Department of Psychology, University of North Carolina, Chapel Hill; Dennis R. Combs, Department of Psychology and Counseling, University of Texas at Tyler; David Penn, Department of Psychology, University of North Carolina, Chapel Hill; Mark Hazelrigg and Betty Paesler, Dorothea Dix Hospital, Raleigh, North Carolina.

Dennis Combs was affiliated with the Department of Psychology, University of Tulsa, at the time of data collection.

Correspondence concerning this article should be addressed to Nicole Jones, c/o Brian Sheitman, Dorothea Dix Hospital, 3601 Mail Service Center, Raleigh, NC 27699. E-mail: nicole.jones@ncmail.net

and valid when assessing patient behavior. These types of measures would potentially offer better quality information and at the same time allow for a way to collect data that could then be used in the treatment planning process (Paul & Menditto, 1992). Despite the clinical advantages of utilizing behavioral assessments, frequent objections to their utilization are that they are too complicated, and time-consuming, and thus keep staff from interacting more with patients (Corrigan, Steiner, McCracken, Blaser, & Barr, 2001; Paul, 1986). While more complex and lengthier scales can contribute to higher reliability, efficiency in measurement should also be of concern (DeVillis, 1992).

One widely used research measure to assess behavioral change in persons who reside in inpatient psychiatric settings is the Nurses' Observation Scale for Inpatient Evaluation (NOSIE) (Honigfeld & Klett, 1965). The NOSIE is a 30-item rating scale that measures the frequency of patient behavior in six areas: social competence, social interest, personal neatness, irritability, psychoticism, and motor retardation. An "other" subscale consisting of four items was also identified but it is not hypothesized to be related to the positive or negative dimensions and does not contribute to the total score. The NOSIE was originally developed to assess therapeutic change in the behavior of persons with chronic schizophrenia who reside on inpatient psychiatric units. In a sample of 630 male inpatients with chronic schizophrenia, Honigfeld, Gillis, and Klett (1966) performed a principal components analysis of the NOSIE, with a final set of 30 items selected based on the highest factor loadings found at two different time periods. A two-dimension higher order latent structure was proposed with the three subscales of social competence, social interest, and personal neatness accounting for the positive dimension and the three subscales of irritability, psychoticism, and motor retardation accounting for the negative dimension. Previous research has supported the internal reliability of the NOSIE (Farrell & Mariotto, 1982; Honigfeld et al., 1966; Honigfeld & Klett, 1965; Lentz, Paul, & Calhoun, 1971; Ludwig, 1968; Lyall, Hawley, & Scott, 2004; Margari et al., 2005; McMordie, 1975; McMordie & Swint, 1979; Philip, 1973) and validity (Hafkenscheid, 1991; Honigfeld et al., 1966; Lentz et al., 1971; McMordie & Swint, 1979; Pattison & Rhodes, 1974). While previous studies have examined the factor structure of the NOSIE, most studies have not reported substantial information about item loadings and reliability estimates (Dingemans, Bleeker, & Frohm-deWinter, 1984; Hafkenscheid, 1991; Philip, 1977). No studies have examined the factor structure of the NOSIE with a more robust Confirmatory Factor Analysis (CFA) procedure (Van Prooijen & Van Der Kloot, 2001).

While previous research with the NOSIE has cited its usefulness in measuring behavioral changes in persons who reside in inpatient psychiatric settings (Hafkenscheid, 1991; Honigfeld, 1974; Swett & Mills, 1997; Volavka et al., 2005), its length (currently 30 items) and the original design of being a joint-rated scale may be prohibitive for routine ongoing clinical use in hospital settings. For any situation in which data is routinely collected on more than a few patients, completing a 30-item instrument, however simple, becomes a substantial investment in time and resources. The purpose of this study was to examine if a reduced item version of the NOSIE (the NOSIE-Shortened Form; NOSIE-SF) could be developed without compromising the internal consistency and factorial validity of the original longer version.

Method

Participants

The sample was drawn from individuals who resided in inpatient psychiatric facilities from two settings. The first group of participants consisted of 18 male patients who resided on the management section of a long-term, inpatient psychiatric facility in North Carolina. The second sample was comprised of 49 inpatients (37 male, 12 female) from psychiatric facilities in Oklahoma. See Table 1 for a description of the participants. Some participants did not have complete data sets because they were out of the unit on scheduled visits or medical appointments or had transferred onto the unit after the start of data collection. Data was available for 99.98% of the possible data points for participants in the first sample and for 99.97% of the possible data points in the second sample.

Prior to data analysis, approval to review medical records and analyze the data was received by the University of North Carolina,

Table 1
Characteristics of Participants by Setting

	Sample 1 (N = 18)	Sample 2 (N = 49)
Age	43.06 (SD = 10.51)	41.1 (SD = 12.3)
Days hospitalized	1633.78 (SD = 2345.14)	1728.02 (SD = 2091.82)
Gender		
Male	18	37
Female	0	12
Race		
Caucasian	9	24
African American	9	20
Other	0	5
Diagnosis		
Schizophrenia spectrum	13	49
Dementia related	2	0
Psychosis not otherwise specified	2	0
Impulse control due to brain injury	1	0

Note. Sample 1 = participants from North Carolina; Sample 2 = participants from Oklahoma; Days hospitalized = average number of days hospitalized at time of first observation.

Chapel Hill, and the University of Tulsa's institutional review boards.

Measure

The NOSIE (Honigfeld & Klett, 1965) is a 30-item behavioral observation scale used to assess ongoing changes in the behaviors of persons who reside in inpatient, psychiatric settings. Items are rated on a five-point scale based on the frequency of occurrence. A global score can be obtained as a function of both positive and negative dimensions. The four items that constitute the "other" subscale do not usually contribute to the summary score and, thus, were not included in our analyses.

Procedure

Staff training on the NOSIE was conducted at both settings prior to the start of the study. Training was conducted by psychology and nursing staff, all of whom had been trained on the NOSIE. Trainees/raters consisted of health care technicians (nurse's aides) who had been assigned to this unit. Training consisted of didactic instruction and training scenarios in which staff rated specific behaviors within each scenario. For the first group of participants, the NOSIE was completed on each patient one time each shift (every eight hours) by the person's assigned clinical care staff member as part of ongoing, routine clinical observation and care. For the second group of participants, at least two staff raters completed the NOSIE for each participant. Interrater reliability estimates found a high level of agreement between rater pairs (average ICC [interclass correlation coefficient] = .83, with a range of .89–90). Raters then completed the NOSIE one week later for each participant. Scores obtained by the first rater were utilized for data analyses. All raters observed participants for at least 72 hours prior to making their ratings.

Prior to combining the two samples for data analysis, a maximum likelihood estimation via Mplus (Muthen & Muthen, 2006) was used to assess if the factor structure differed for our two inpatient samples. Thus, this addressed the comparability between our two samples. The standard group error variance in the second group was too small, though, to conduct a test of the equality of covariances, due to the number of cases. However, when the CFA was conducted on the two groups separately, comparable fit indices were found (sample one comparative fit index [CFI] = .901 and sample two CFI = .863). This reflected similar factor structures across both samples. Thus, the two samples were combined. The observations were then randomly assigned by the computer to sample one (phase one: full-scale model) and sample two (phase two:

reduced-scale model), with 2,492 and 2,481 observations, respectively.

The goal of the second phase was to determine whether the number of items on the NOSIE could be reduced without compromising the relative goodness-of-fit. Items from the NOSIE were retained based on: (1) independent ratings by the authors, who have both clinical and research experience with the NOSIE, as to which two items best described the theoretical construct of that factor; and, (2) NOSIE factor loadings from the CFA conducted during the first phase of the study. Discrepancies between the author ratings and the factor loading occurred only for the second item for four of the six subscales. These four items were then reexamined and then chosen, again, based on independent ratings by the authors as to which two items best represented the behavioral domain of that subscale. Eleven of the 12 items chosen corresponded to the two items with the highest factor loadings for that subscale. The items retained are found in Table 5. This process resulted in 12 items being retained (two items per latent factor).

Statistical Analysis

Confirmatory maximum-likelihood factor analysis via Mplus (Muthen & Muthen, 2006) was used to assess whether the NOSIE factor structure would provide a good fit to the data and to assess the relative goodness-of-fit of the model. Due to the large number of observations from the 49 participants, multilevel modeling with a sandwich estimator was utilized to obtain standard errors that account for stratification and clustering (Binder, 1983). This procedure takes into account within subject correlation in the estimation process (Kalton, 1983). Model fit was examined using the following criterion: corrected χ^2 (Satorra & Bentler, 1988), the CFI, the Tucker-Lewis Index (TLI), and the root mean square error of approximation (RMSEA). Although there is no agreement on specific cutoff values that indicate an acceptable model fit, Hu & Bentler (1999) suggest using a combination of: (1) values .95 and above for the relative fit indices, and (2) an RMSEA of .06 or lower for model selection.

CFA can also be utilized to examine the overall fit of the data to a proposed scale model and to assess if the individual items load on those hypothesized latent factors (Tabachnick & Fidell, 1996). In addition, when latent factors are expected to be correlated and connected to a higher order factor, confirmatory factor analysis can estimate the relationship between the latent factors and assess whether the latent factors load on a higher order factor.

Results

Phase One: CFA of the NOSIE

The data were first examined to ensure that the variable distributions did not violate statistical assumptions of the planned analyses; no substantial violations were found. Furthermore, the sample size exceeded the recommended number of observations needed to estimate the factor structure of the scale (parameter ratio of 5 to 10:1) (Bentler, 1990). For the full scale NOSIE, the internal consistency reliabilities (alpha coefficients) are presented in Table 2. Results showed the reliability estimates ranged from .92 for the irritability subscale to as low as .59 for the motor retardation subscale.

The full-scale model of the NOSIE structure was first examined assuming no correlations among factors (Full-scale Model Uncorrelated). The goodness-of-fit indices associated with this model are presented in Table 3. Neither of the goodness of fit indices was greater than .95 and the RMSEA value was slightly above the recommended cutoff of .06.

The second step was to try and improve model fit by allowing the latent factors to correlate. Subscale correlations of the NOSIE-30

Table 2
Means, SDs, and Coefficient Alphas for Nurse's Observation Scale for Inpatient Evaluation (NOSIE) and Nurse's Observation Scale for Inpatient Evaluation–Short Form (NOSIE–SF)

	NOSIE Sample 1 (N = 2492)		NOSIE–SF Sample 2 (N = 2481)	
	Means (*SD*)	α	Means (*SD*)	α
Social competence	3.33 (0.52)	.79	3.13 (0.62)	.84
Social interest	0.91 (0.56)	.83	0.98 (0.63)	.77
Neatness	2.23 (0.51)	.67	1.01 (0.65)	.83
Irritability	0.51 (0.62)	.92	0.51 (0.6)	.85
Motor retardation	0.69 (0.52)	.59	0.76 (0.56)	.54
Psychoticism	0.46 (0.60)	.82	0.55 (0.68)	.68

Table 3
Goodness-of-Fit Indicators for Nurses Observation Scale for Inpatient Evaluation (NOSIE)–Full Scale and NOSIE–Short Form

Model	χ^2	*df*	CFI	TLI	RMSEA
Full (NOSIE-30)					
Uncorrelated	18349.69***	325	.813	.797	.067
Correlated (1)	5610.20***	284	.705	.662	.087
Correlated (2)	16147.74***	293	.120	.024	.147
Short-Form (NOSIE–SF)					
Uncorrelated†					
Correlated (1)	88.98***	39	.982	.970	.023
Correlated (2)	247.58***	48	.930	.904	.041

Note. Full (Sample 1) $N = 2492$; Short Form (Sample 2) $N = 2481$; Correlated (1) = All subscales correlated; Correlated (2) = Positive Dimensions (Social Competence, Social Interest and Neatness) correlated and Negative Dimensions (Irritability, Motor Retardation, and Psychoticism) correlated; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean squared error of approximation.
*** $p < .0001$.
† model not identified.

are presented in Table 4. The results showed that the correlated factor structure did not provide a better fit to the data. Again, neither the CFI nor TLI indices reached a value of .95 or more and the RMSEA value was .087, indicating a poor fit of the data to this model as well.

The third step was to assess the full scale model with the hypothesized correlated latent factors added in. The goodness-of-fit indices for the Full Scale Model Correlated (2) are presented in Table 3. The results indicated the model with the specified latent correlations of a positive and negative dimension provided the worst fit to the data, with the CFI and TLI indices both below .150 and an RMSEA of .147, indicating poor model fit.

While the uncorrelated model provided the best fit to the data of our three full-scale (NOSIE-30) models, neither the CFI or TLI indices reached a value of .95 or more (CFI = .813; TLI = .797), and the RMSEA value was .067. The results of this CFA did not support the theoretical structure of a "positive" and "negative" dimension. In addition, while the uncorrelated model provided a more satisfactory account of the data, it did not meet the recommended guidelines for acceptable model fit (Hu & Bentler, 1999). This was initially unexpected and would suggest that researchers who utilize the 30-item NOSIE conduct their own CFA.

Phase Two: Testing a Reduced Factor Model of the NOSIE

Confirmatory maximum-likelihood factor analysis with the sandwich estimator via Mplus

Table 4
Correlations Among Positive and Negative Dimensional Subscales on the Full-Scale Nurses Observation Scale for Inpatient Evaluation (NOSIE)-30

Subscales	1	2	3	4	5	6
Positive dimensions						
1. Social competence	—					
2. Social interest	−.02	—				
3. Neatness	.40*	.35*	—			
Negative dimensions						
4. Irritability	−.49*	−.07*	−.29*	—		
5. Motor retardation	−.42*	−.06*	−.31*	.28	—	
6. Psychoticism	−.30*	.003	−.15*	.21	.18	—

* $p < .01$.

Table 5
Nurses Observation Scale for Inpatient Evaluation–Short Form (NOSIE–SF) Items

Subscale/item
Positive dimension
Social competence
1. Has to be reminded what to do.
2. Has to be told to follow hospital routine.
Social interest
3. Tries to be friendly with others.
4. Starts conversation with others.
Neatness
5. Keeps clothes neat.
6. Keeps self clean.
Negative dimension
Irritability
7. Becomes upset easily if something does not suit him.
8. Is irritable/grouchy.
Motor retardation
9. Sits, unless directed into activity.
10. Sleeps, unless directed into activity.
Subscale/item
Psychoticism
11. Hears things that are not there.
12. Talks, mutters, or mumbles to self.

(Muthen & Muthen, 2006) was again used to assess if the NOSIE factor structure would provide a good fit to the data and to assess the relative goodness-of-fit of the reduced item model. Coefficient alpha reliabilities for the reduced item model are shown in Table 2. Results showed the reliability estimates ranged from .85 for the irritability subscale to as low as .54 for the motor retardation subscale.

As in phase one, the reduced item model was first examined assuming all latent factors were uncorrelated (Reduced Model Uncorrelated). The model was not identified, and respecification of the model with the correlations among the latent factors added in was needed to achieve identification. Thus, as in phase one, the second step was to assess the reduced item model with the latent factors allowed to correlate. The goodness-of-fit indices are presented in Table 3. The correlated, 12-item NOSIE (Reduced Model Correlated resulted in superior goodness-of-fit indices, with the CFI = .982 and the TLI = .970 (both above the recommended .95) and an RMSEA estimate of .023 (below the recommended .06). Finally, the reduced item model was reexamined to corroborate the existence of two major dimensions, positive and negative. The goodness-of-fit indices are presented in Table 3. The RMSEA was lower than .06 (.041), however, neither of the goodness of fit indices were above .95 (CFI = .930; TLI = .904).

Thus, these results show that the reduced 12 item, correlated model was able to provide a more adequate representation of this data than the full 30-item scale model, most likely due to the removal of items with low factor loadings. It was the only model tested that met the desired outcome criteria as outlined earlier.

Discussion

The primary goal of this study was to examine whether a shortened version of the NOSIE (the NOSIE–SF) could be developed without weakening the quality of the original instrument. A theoretical as well as statistical approach to item reduction was utilized, in an attempt to maintain content domain coverage of the six factors. The results of the confirmatory factor analyses provide initial evidence that the number of items can be reduced to 12 while maintaining adequate scale properties.

These results suggest that a reduced form of the NOSIE may offer a practical alternative to the standard, 30-item version and would offer the possibility that the NOSIE could be routinely used for clinical purposes. Due to its shorter length, the NOSIE-SF has been incorporated into the already existing paperwork completed by staff in one of the settings for this study. Adoption by other inpatient units should not be difficult, given the brevity of the scale. In addition, some of the six factors could be incorporated into treatment plans as short-term goals. The NOSIE-SF data could then be used to determine in a measurable way whether the goal has been achieved or not. Finally, follow-up validation studies that assess if the NOSIE–SF is related to other measures of clinical functioning may help to increase the clinical utility of this measure.

While there are methodological strengths of this study, primarily the large number of behavioral observation points, there are also limitations. The study was conducted only with persons who resided on long-term inpatient units with most of the subjects being male. A second limitation includes the lack of interrater reliability data collected with the first sample. However, the second sample found high interrater agreement (average ICC = .83) and previous

research found that raters with minimal training in the administration of the NOSIE-30 scale had markedly similar results to Positive And Negative Syndrome Scale (PANSS) ratings done by doctoral level persons with rigorous interrater reliability training (Volavka et al., 2005). These limitations notwithstanding, the results suggest that the factor structure of a short form of the NOSIE is psychometrically sound. Future research that utilizes confirmatory factor analysis should be used to replicate the NOSIE–SF with other clinical samples. In addition, follow-up studies would be necessary to assess if the NOSIE–SF exhibits adequate predictive validity and is sensitive to behavioral change over time.

References

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

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review, 51,* 279–292.

Corrigan, P. W., Steiner, L., McCracken, S. G., Blaser, B., & Barr, M. (2001). Strategies for disseminating evidence-based practices to staff who treat people with serious mental illness. *Psychiatric Services, 52,* 1598–1606.

DeVillis, R. F. (1992). *Scale development: Theory and application.* Newbury Park, CA: Sage.

Dingemans, P. M., Bleeker, J. A. C., & FrohmdeWinter, M. L. (1984). A cross-cultural study of the reliability and factorial dimensions of the nurses' observation scale for inpatient evaluation (NOSIE). *Journal of Clinical Psychology, 40,* 169–171.

Farrell, A. D., & Mariotto, M. J. (1982). A multimethod validation of two psychiatric rating scales. *Journal of Consulting and Clinical Psychology, 50,* 273–280.

Hafkenscheid, A. (1991). Psychometric evaluation of the nurses observation scale for inpatient evaluation in the Netherlands. *Acta Psychiatrica Scandinavica, 83,* 46–52.

Honigfeld, G. (1974). NOSIE-30: History and current status of its use in pharmacopsychiatric research. *Psychological Measurements in Psychopharmacology, 7,* 238–263.

Honigfeld, G., Gillis, R. D., & Klett, J. (1966). The NOSIE-30: A treatment sensitive ward behavior rating scale. *Psychological Reports, 19,* 180–182.

Honigfeld, G., & Klett, J. (1965). The nurses observation scale for inpatient evaluation: A new scale for measuring improvement in chronic schizophrenia. *Journal of Clinical Psychology, 21,* 65–71.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Joint Commission on Accreditation of Healthcare Organizations (2006). *Specifications manual for National Hospital Quality Measures-Hospital-Based Inpatient Psychiatric Services (HBIPS) test set.* Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations.

Kalton, G. (1983). *Introduction to Survey Sampling*. Newbury Park, CA: Sage.

Lentz, R. J., Paul, G. L., & Calhoun, J. F. (1971). Reliability and validity of three measures of functioning with "hard-core" chronic mental patients. *Journal of Abnormal Psychology, 78,* 69–76.

Ludwig, A. M. (1968). The influence of nonspecific healing techniques with chronic schizophrenics. *American Journal of Psychotherapy, 22,* 382–404.

Lyall, D., Hawley, C., & Scott, K. (2004). Nurses' observation scale for inpatient evaluation: Reliability update. *Journal of Advanced Nursing, 46,* 390–394.

Margari, F., Matarazzo, R., Casacchia, M., Roncone, R., Dieci, M., Safran, S., et al. (2005). Italian validation of MOAS and NOSIE: A useful package for psychiatric assessment and monitoring of aggressive behaviors. *International Journal of Methods in Psychiatric Research, 14,* 109–118.

McMordie, W. R. (1975). An aide to screening patients for new programs. *VA Newsletter for Research in Mental Health and the Behavioral Sciences, 17,* 8–9.

McMordie, W. R., & Swint, E. B. (1979). Predictive utility, sex of rater differences, and interrater reliabilities of the NOSIE-30. *Journal of Clinical Psychology, 35,* 773–775.

Muthen, L. K., & Muthen, B. O. (2006). *Mplus user's guide: Fourth edition.* Los Angeles: Muthen and Muthen.

Pattison, E. M., & Rhodes, R. J. (1974). Clinical prediction with the NOSIE-30 scale. *Journal of Clinical Psychology, 30,* 200–201.

Paul, G. L. (Ed.). (1986). *Assessment in residential treatment settings (part one*). Champaign, IL: Research Press.

Paul, G. L., & Menditto, A. A. (1992). Effectiveness of inpatient treatment programs for mentally ill adults in public psychiatric facilities. *Applied and Preventative Psychology, 1,* 41–63.

Philip, A. E. (1973). A notes of the Nurses Observation Scale for Inpatient Evaluation (NOSIE). *British Journal of Psychiatry, 122,* 595–596.

Philip, A. E. (1977). Cross cultural study of the factorial dimensions of the NOSIE. *Journal of Clinical Psychology, 33,* 467–468.

President's New Freedom Commission on Mental Health. (2003). *Achieving the Promise: Trans-*

forming Mental Health Care in America. Final Report (DHHS Publication No. SMA-03-3832). Rockville, MD: United States Department of Health and Human Services.

Santos, A. B., Henggeler, S. W., Burns, B. J., Arana, G. W., & Meisler, N. (1995). Research on field-based services: Models for reform in the delivery of mental health care to populations with complex clinical problems. *American Journal of Psychiatry, 152,* 1111–1123.

Satorra, A., & Bentler, P. M. (1988). *Scaling corrections for chi-square statistics in covariance structure analysis.* ASA 1988 *Proceedings of the Business and Economic Statistics Section*, 308–313, Alexandria, VA: American Statistical Association.

Silverstein, S. M., Spaulding, W. D., & Menditto, A. A. (2006). *Schizophrenia, (Vol. 5).* Cambridge, MA: Hogrefe & Huber.

Swett, C., & Mills, T. (1997). Use of NOSIE to predict assaults among acute psychiatric patients. *Psychiatric Services, 48,* 1177–1180.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins Publishers.

Thompson, T. (2002). *Report to Congress: Prospective payment system for inpatient services in psychiatric hospitals and exempt units*. Washington, DC: Department of Health and Human Services.

Van Prooijen, J-W., & Van Der Kloot, W. A. (2001). Confirmatory factor analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61,* 777–792.

Volavka, J., Nolan, K. A., Kline, L., Czobor, P., Citrome, L., Sheitman, B., et al. (2005). Efficacy of clozapine, olanzapine, risperidone, and haloperidol in schizophrenia and schizoaffective disorder assessed with nurses observation scale for inpatient evaluation. *Schizophrenia Research, 76,* 127–129.

Received September 21, 2007
Revision received January 7, 2007
Accepted January 17, 2007 ■