Archives of Assessment Psychology, Vol. 1, No. 1, (pp)

SEM-Based Composite Reliability Estimates of the Crisis Acuity Rating Scale with Children and Adolescents

Steven Thurber, Ph.D. and Michael R. Bonynge, B.S.

Abstract

On-call clinicians completed a Likert-type scale of crisis acuity rating 227 children and adolescents admitted consecutively to a youth crisis unit of a rural community mental health center. Traditional methods of estimating reliability for the scale were precluded. In particular, the scale was not found to be unidimensional when subjected to confirmatory factor analysis. This rendered Cronbach's α coefficient inappropriate. A six-factor oblique model corresponding to a priori subscales was supported. SEM-based reliability estimates for this confirmed factor structure using a phantom composite was detailed. The results yielded a point estimate of 0.845 with a bootstrap (1000 samples) 90% confidence interval between 0.819 and 0.867.

Introduction

Youth seen in community mental health crisis units typically present acute conditions that are unlikely to persist over time, and they may experience confusion and other cognitive conditions that preclude accurate information on self-report inventories (see Bonynge & Thurber, 2010). Clinicians working with children and adolescents in crisis may benefit from psychometrically sound rating scales of crisis symptoms.

Bonynge recognized a need for a psychometrically sound measuring instrument for crisis patients (Bonynge, Thurber, & Hoffman, 2004). The setting for the development of the Bonynge Crisis Acuity Scale (BCAC) was a rural community mental health center that provided a variety of services for persons in crisis, including a hot-line, urgent care, mobile crisis, short-term residential beds, and walk-in counseling. Part of these services involved mental health professionals' face-to-face contact with adult patients entering the urgent care or crisis center. Salient symptom characteristics of persons in crisis were identified on the basis of consensus among mental health professionals experienced in direct crisis services (see Bonynge & Thurber, 2008). The characteristics delineated were (1) danger to self, (2) danger to others, (3) functional decline, (4) confusion, and (5) depression. Likert-type scales were developed for these symptoms, to be rated by clinicians from 0 (not present) to 4 (extreme). Subsequently, several psychometric properties were evaluated, and the subscales were modified. There was an elaboration of the five symptom areas with new items representing sub-components. For example, "lethality of plan" and "detailed plan for self harm" were later subsumed under the "danger to self" rubric. In addition, a sixth scale was added related to "intervention resistance." A revised 24-item crisis acuity scale obtained an internal consistency (α) coefficient of 0.82; additionally, it was also found to be factorially complex (six factors). Validity of the BCAC was evaluated via relationships with the clinical dispositions for crisis patients (i.e., higher, more severe symptom scores were associated with more restrictive inpatient treatment recommendations; lower scores were associated with outpatient dispositions) (Thurber & Bonynge, 2008; Bonynge & Thurber, 2010).

Reliability Considerations

A child in a crisis situation reflects a unique, single event that cannot be repeated in vivo conducive to a test-retest reliability evaluation. A crisis is an environmental-individual interaction condition that cannot be recreated by investigators; a test-retest reliability evaluation is therefore precluded. Further, a child in crisis cannot ethically be evaluated sequentially by more than one rater to assess inter-judge reliability. As both test-retest and inter-rater reliability are untenable, the main recourse is internal consistency of items, traditionally estimated by Cronbach's alpha (α) coefficient (Cronbach, 1951). However, as is becoming well-known, the use of alpha should be done with circumspection. In brief, coefficient alpha is only appropriate with test items evincing tau-equivalence. From the perspective of latent variable research, a tau-equivalent test is unidimensional with equal loadings from all manifest variables (Raykov, 2004). More explicitly, a tau-equivalent test assumes all items measure the same latent variable, on the same scale, with the same degree of precision, with all true scores being equal (Graham, 2006). When tau-equivalence does not obtain, Cronbach's alpha cannot be viewed as a dependable estimate of reliability; it will over- or underestimate (more often the latter) the population value (Raykov, 2004).

An oft-cited goal of test construction is to produce an internally consistent measuring instrument with homogeneous items selected from the same domain of content (see Nunnally, 1978). But if the nature of the phenomenon to be measured is in itself multidimensional and heterogeneous, construct validity may demand variation and non-homogeneity in test items. Indeed, it can be reasonably argued that homogeneous items will not capture the nature of variegated crisis-related symptomatology (e.g., suicidal and homicidal proclivities, cognitive and affective maladaptations). There is no reason to expect a valid measure of crisis acuity to have homogeneous items or tau-equivalence (see Edwards & Bagozzi, 2000). Moreover, as mentioned, the BCAS with adults was not found to be unidimensional.

In classical test theory, reliability is conceptualized as the ratio of true to observed variance (Bollen, 1989). But because true score variance is unknown, this ratio has to be estimated using familiar methods that include the test-retest and internal consistency measures mentioned above. An alternative to coefficient α involves structural equation modeling (SEM) with composite variables. A composite is a variable based on the sum of at least two other variables. As elaborated by Raykov (1997), one SEM composite is termed "phantom" and comprises the added variances of the individual observed variables; i.e., it receives unitary paths from each observed variable or test item (see figure 1). The other composite is a latent variable labeled "G" in figure 1. It is a hierarchical general factor stemming from a fitted oblique factor structure. A phantom summation or composite provides for an implied correlation with the general latent variable. This equals the composite reliability index; the squared value is the reliability estimate symbolized r*ph.*g by Gignic (2007) (see also Graham, 2006).

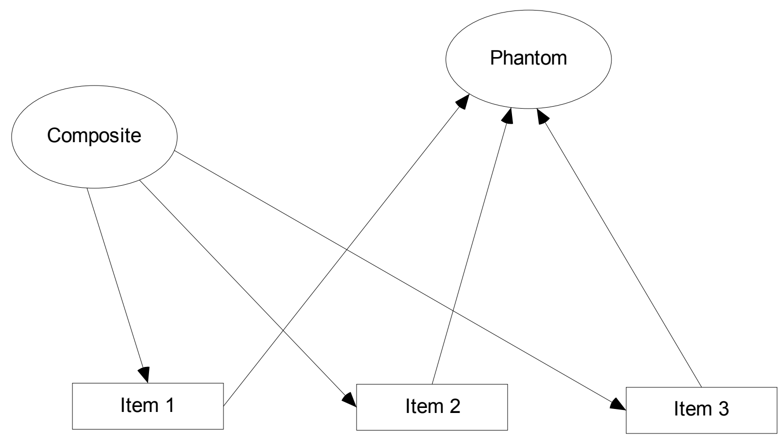

Figure 1. Basic Composite Reliability Model

In the current study, we investigated a slightly modified version of the adult rating scale (i.e., in item 17, "vocational/educational functioning," the word "vocational" was deleted). The aim was to ascertain whether the BCAC is an adequately reliable rating scale for children and adolescents in crisis. It was designed to provide an index of measurement accuracy to inform clinicians about bias due to measurement error when used with younger individuals.

Method

Participants. The participants were 227 children and adolescents admitted consecutively to a rural community mental health center. They were showing sufficiently maladaptive reactions to warrant placement in a crisis unit. There were 103 boys and 124 girls. The large majority (217) was Caucasian, with five African Americans and five Native Americans. Thirty-four individuals were Latino in ethnicity. They ranged in age from 7 to 17 with a mean age of 14 years old. Admitted patients were then interviewed by an on-call clinician who also collected any available information regarding the crisis situation. The study included 15 mental health professionals who were alternately on-call during the approximate 13 months of the study. They were all state-licensed and included five doctors of psychology, two psychologists with Master's degrees, five social workers, and three marriage and family counselors.

Measure. Following information gathering and interview, the clinician completed the BCAC, discussed above. As mentioned, the scale had been used in the past but only with adult crisis patients. The adult scale was slightly modified for use with young people (above).

Procedure. Following data collection, a preliminary item analysis was conducted, including item distributions and assessment of multivariate normality. Exploratory factor analysis preceded SEM for confirmatory factor analysis, which was then used to evaluate the unidimensionality of the rating scale, to ascertain if α or r*ph.g* constituted the appropriate approach for reliability estimation.

Results

Item content of the Crisis Acuity Scale is presented in table 1, together with psychometric information pertaining to item means, variability, and distributions. Seven items evinced skew of 2 and beyond. The Mardia coefficient of multivariate kurtosis was also elevated (252.741, z = 56.145, $p < .00001$). Therefore, methods for dealing with multivariate non-normal data were deemed necessary. Several indices suggested that the instrument did not consist of homogeneous items (e.g., mean correlation among the *a priori* scales or average item-total correlation). The α coefficient was 0.768.

Table 1, Item Content, Means, Standard Deviations, Skew, and Kurtosis

Item (Scale in bold)	M	SD	Skew	Kurtosis
Danger to self				
1. Detailed plan for self-harm	.56	1.01	1.74	2.11
2. Lethality of plan	.48	.95	2.09	3.69
3. Impulsivity	1.91	1.01	-.08	-.46
4. History of suicide attempts	.50	.86	1.89	3.40
Danger to others				
5. Auditory hallucinations	.11	.53	5.69	34.54
6. History of aggression	1.00	1.19	.97	-.06
7. Plan for aggression	.22	.61	3.23	11.51
8. Temperament	1.26	1.10	.54	-.45
9. Hopelessness	1.32	1.15	.47	-.74
Depression				
10. Vegetative signs	1.26	1.22	.49	-.93

Item (Scale in bold)	M	SD	Skew	Kurtosis
11. Motivation level	1.49	1.18	.34	-.75
12. Anhedonia	1.19	1.10	.75	-.12
Confusion				
13. Impaired logic	.29	.61	2.06	3.44
14. Orientation	.03	.25	10.28	111.13
15. Thought content	.18	.47	2.90	9.30
16. Decision-making	1.26	1.04	.42	-.52
Functional impairment				
17. Educational functioning	1.26	1.22	.09	-1.02
18. Therapeutic interventions past three months	1.27	1.22	.63	-.67
19. Achievement inconsistent with estimated intelligence	1.44	1.04	.16	-.78
20. Social/familial functioning	2.43	1.08	-.38	-.49
Intervention resistance				
21. Willingness to answer questions	.36	.73	2.12	3.76
22. Willingness to provide information	.48	.84	1.68	2.03
23. Agreeable with recommendations	.65	.96	1.59	2.15
24. Support network	.54	.87	1.39	.73

Non-normal Data. There are several techniques available for analysis of data that violate assumptions of multivariate normality. We used Principal axis extraction in exploratory factor analysis, and scale-free estimation in confirmatory factor analysis together with the "bootstrapping" technique afforded by Amos 16.0 (Arbuckle, 2007) for such violations. This involves a resampling from the empirical distribution in which multiple subsamples of the same size are drawn at random with replacement. The method allows for the assessment of the stability and accuracy of parameter estimates; specifically, bootstrapping facilitates the non-normal empirical distribution to approximate the theoretical distributional assumptions of the statistic (Cheung, 2009; see also West, Finch, & Curran, 1995). The next analyses concern tau-equivalence and appropriateness of α as the reliability estimate.

Tau Equivalence and Structural Equation Modeling. As mentioned, a tau-equivalent test can be inferred from factor analysis, obtaining a single factor with equivalent loadings on all test items. A principal axis extraction approach was selected for exploring the dimensionality of the scale because it is the method of choice when data violate multivariate normality (Costello & Osborne, 2005). The correlation matrix was deemed adequate for conducting factor analyses, with a Kaiser-Meyer-Olkin measure of sampling adequacy of 0.673. The results did not support a unifactor model; rather, the eigenvalue 1 and scree test criteria suggested a six factor structure with correlated dimensions (oblique rotation).

Factor confirmation using Amos 16.0 (Arbuckle, J. L. 2007) was computed. Again, because of multivariate normality concerns, scale-free estimation was used. The results for a general factor and a six factor oblique solution are presented in table 2, with the fit measures SRMR and LTI. The use of these measures is based on the work of Hu and Bentler (1999), who recommended the standardized root-mean-square-residual (SRMR) as a generally unbiased measure. Further, they recommend the use of the SRMR with one other fit measure to evaluate the adequacy of a structural equation model such as the Tucker Lewis Index (TLI). Values of SRMR >0.08 correspond to a well-fitted model; values of TLI beyond 0.90 suggest an acceptable fit (Marsh, Balla, & McDonald, 1996). As presented in table 2, the unifactor model is not supported by the values of SRMR and TLI. This suggests that the acuity scale lacks tau-equivalence, and α is not the appropriate estimate of scale reliability. On the other hand, those indices do indicate that the six factor oblique model has acceptable fit. This further supports the use of a composite reliability estimate.

Composite Reliability

The implied correlation between the phantom composite with variance summation from all items of the crisis acuity scale, and the composite based on the six-factor variance composite, was 0.919; this is the reliability index. Its square yields $r_{ph.g} = 0.845$. This composite reliability is elevated in comparison to the alpha coefficient of 0.768 (above). Next, we resampled from the original data 1000 times, yielding an approximate standard error and 90% confidence interval for the composite reliability (see Raykov, 1998; Raykov & Shrout, 2002; Graham, 2006). The results were a mean point estimate of reliability across samples together with estimated confidence interval. Specifically, these data indicate with relative certainty that the population composite reliability lies between 0.819 and 0.867 (standard error = 0.008), with an optimal estimate of 0.845.

Table 2

Fit Indices

Model	SRMR	TLI
General factor	.141	.477
6 factor oblique	.075	.926

Note: SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index

Discussion

This study represents an initial attempt to gather reliability data on a rating scale purporting to measure crisis acuity, adapted for children and adolescents. The current scale as constituted might serve as a reliable foundation for future modifications and refinements. Refinement might include investigation of a rating format with more categories. Additionally, the dimensions (factors and subscales) of the acuity scale were developed more through input from practitioners in the field than from an empirically-justified theory regarding the specific nature of crisis situations in interaction with a person's characteristics. The BCAC scale is close to the practical world while distant from theoretical space. Any extant scale of crisis symptomatology could benefit from modifications related to theory, leading to a refined domain of content for evaluation of crisis responses. Furthermore, that proposed theory might benefit from integration with developmental data germane to crisis reactions.

The platitude regarding reliability as a necessary but not sufficient condition for validity remains apt. Data from our investigation do inform us that the validity of a crisis acuity scale likely must reflect the multifarious nature of the crisis response.

We used several methods to deal with non-normality: Principal Axis extraction, scale-free estimation, and bootstrapping. Endorsement rates for acuity items will likely result in items with non-normal distributions in future studies. Other methods for adjusting non-normal data, including parceling and scale transformation, should be considered.

Finally, our main psychometric message to practitioners is to be aware that the α coefficient assumes unidimensionality; failure to attain this may result in misleading reliability estimation. SEM provides a reasonable composite reliability option when tau-equivalence is not extant.

About the Authors

Steven Thurber, Ph.D. is with the Department of Psychology, Woodland Centers. Correspondence concerning this article should be addressed to Steven Thurber, Department of Psychology, Woodland Centers, 1125 SE Sixth Street, Willmar, Minnesota 56201.
E-mail: *steven_thurber@yahoo.com*

Michael R. Bonynge, B.S. is with the Department of Psychology, Woodland Centers Willmar, Minnesota

References

Arbuckle, J. L. (2007). *Amos 16.0*. Spring House, PA: Amos Development Corporation.

Bollen, I. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bonynge, E. R., Thurber, S., & Hoffman, H. J. (2004). Development of a scale to assist in assessment and treatment planning: Psychometric properties and factorial composition. *Counseling and Clinical Psychology Journal*, 1, 153–163.

Bonynge, E. R., & author (2008). Development of clinical ratings for crisis assessment in community mental health. *Brief Treatment and Crisis Intervention*, 8, 1–9.

Bonynge, E. R., & author. (2010). Development of a measure for mental health crisis assessment: Preliminary psychometrics of the Crisis Acuity Scale. *Psychology Journal*, 7, 15–30.

Bonynge, E. R., Lee, R. L., & author. (2005). A profile of mental health crisis response in a rural setting. *Community Mental Health Journal*, 41, 675–685.

Cheung, M. W. L. (2009). *Structural Equation Modeling*, 16, 267–294.

Costello, A. B., & Osborne, J. W. (2005). *Practical Assessment, Research and Evaluation*, 10, 1–9.

Cronbach, L. J. (1951). *Psychometrica*, 16, 297–334.

Edwards, J., & Bagozzi, R. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.

Ellis, T., & Dickey, T. (1998). Procedures surrounding the suicide of a trainee's patient: A national survey of psychology internships and psychiatry residency programs. *Professional Psychology: Research and Practice*, 29, 492–297.

Gignic, G. E. (2007) Working memory and fluid intelligence are both identical to g?! Reanalysis and critical evaluation. *Psychological Science*, 49, 187–207.

Graham, J. M. (2006). Congeneric and essentially tau-equivalent estimates of score reliability:

What they are and how to use them. *Educational and Psychological Measurement*, 66, 930–944.

Hu, Li-Tze, & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453.

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

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1996). Goodness-of-Fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391–410.

Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173–184.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated non-homogeneous items. *Applied Psychological Measurement*, 22, 375–385.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavioral Therapy*, 35, 299–331.

Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point estimation using a structural equation modeling approach. *Structural Equation Modeling*, 9, 195–212.

Thurber, S., & Bonynge, E. R. (2008). Symptom ratings and clinical decision-making in a rural crisis-response unit. *Psychology Journal*, 5, 5–9.

West, S. G., Finch, H. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In RH Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 56–75). Thousand Oaks, CA: Sage.