

Downtown Emergency Service Center's Vulnerability Assessment Tool for Individuals Coping with Chronic Homelessness:

A Psychometric Analysis

March 15, 2010
Report of Pilot Testing

Prepared by
Joshua Aaron Ginzler, Ph.D.
Maria Monroe-DeVita, Ph.D.



The Washington Institute for
Mental Health Research & Training
The University of Washington
Division of Public Behavioral Health & Justice Policy

Purpose

This study is a secondary analysis of data collected by the Downtown Emergency Service Center (DESC) during a psychometric study of its internally developed Vulnerability Assessment Tool (VAT) for chronically homeless persons. In 2009, the Washington Institute for Mental Health Research and Training (WIMHRT) at the University of Washington was contracted by DESC to conduct a psychometric assessment of the VAT instrument based upon 277 interviews and 171 follow-up interviews with new or continuing clients. The purpose of these interviews was to determine the current level of an individual's vulnerability to continued instability.

DESC developed the VAT in 2003 in an effort to objectively prioritize homeless clients for limited shelter beds. The agency supported the value that survival services like shelter should first go to the most vulnerable people and needed a way to measure vulnerability based on individual characteristics and behaviors. Subsequently the tool has been used as the principal method for assuring that DESC's permanent supportive housing units are offered to the most vulnerable chronically homeless individuals.

Another method for determining placement into supportive housing is based on service utilization; particularly acute care health services (e.g., ER & hospital services, EMS, sobering and detoxifications services) and criminal justice system involvement (e.g., arrests, jail/prison stays, etc.). In this approach housing is targeted to individuals with the highest utilization and cost profiles, under the theory that living situation stabilization will reduce excessive use of other, more expensive public services. The primary objective is to reduce overall public expenditures, and policymakers have used the successes of this approach as justification for future housing units to be allocated in the same way. While service utilization may correlate strongly with vulnerability, there are other highly vulnerable individuals who do not extensively use services, and thus would be missed by a housing placement system focused solely on systems utilization.

The DESC VAT offers another method to assure that the assessment outcomes are not based solely on externally indirect indicators (such as service utilization) of an individual's vulnerability and need for assistance. This instrument includes more direct measures of an individual's actual well-being across a number of factors to calculate the acuteness of that person's vulnerability to prolonged instability and potentially deteriorating mental and physical health.

The intent of using the VAT is to determine who is at greatest risk if left on the streets. The importance of making such determinations is evidenced by the growing number of requests for copies of the tool DESC has received from homeless housing and service providers throughout North America. This level of interest combined with the paucity of similar measures in the field point to the need for reliability and validity analyses of the VAT. If the results help establish confidence in the tool, its use might be expanded and the underlying principle of a non-service based vulnerability factor perpetuated.

Methods

Design. In late summer 2009, six DESC staff, recognized for their high level of assessment skills set, were reassigned to serve as dedicated assessors for a two week period. Another staff member managed all the recruitment and scheduling of participants across the period. Follow-up interviews were conducted one month later. All participants were currently homeless at the time

of assessment. Following each interview phase, the project director provided Dr. Joshua Ginzler, who was contracted by WIMHRT, with an electronic file of participant responses to the VAT, as well as some minimal amount of key individual de-identified demographic information. Table 1 includes basic information regarding gender, ethnicity/race, sexual orientation, income, and veteran status.

Table 1. Sample demographics

<u>Ethnicity (N=259)</u>	
Caucasian	44%
African American	26%
Mixed Ethnicity	13%
All Others	17%
<u>Sexual Orientation (N=252) ^a</u>	
Heterosexual	91%
Bisexual	4%
Gay/Lesbian	3%
Other	2%
Sexual Minority ^b	17%
<u>Gender (N=259) ^c</u>	
Male	77%
Female	22%
<u>Veteran Status (N=264)</u>	
Vietnam	11%
Other Military Conflicts/Experience	7%
<u>Primary Income Sources</u>	
Social Security Benefits	52%
General Assistance	17%
Veteran Affairs	3%
None	25%
<u>Primary Income(N=254; M=\$492 [\$409.33]) ^d</u>	
\$0	27 th %ile
\$580	50 th %ile
\$694	75 th %ile

Note: ^a = first four categories sum to 100%.

^b = formulated across numerous items such that sexual identity, gender identity, and attraction to the same sex could identify someone as sexual minority.

^c = three participants identified as either transgender, intersex, or otherwise not classifiable as male/female.

^d = cumulative percentile.

Analytic Plan. One-way ANOVA tests were conducted using individual items and full scale composite score as the independent variables and ethnicity/race, sexual minority status, gender identification, veteran status, and primary income level as the dependent variables to test for the impact of specific demographic variables.

Reliability was tested across three domains to evaluate the consistency of the VAT a) within itself, b) across repeated measurements at two separate time points, and c) across the codings of

multiple assessors. The combined findings from the first three analyses suggest the upper limits of the validity of the instrument given its current state and instruction. The final domain of analysis, a test of concurrence with simultaneously collected qualitative measurements of vulnerability, is a more direct evaluation of the current instrument validity.

Inter-rater reliability was quantified using kappa (Cohen, 1960), a reliability statistic that calculates the level of agreement while correcting for the level of agreement expected by chance. To thoroughly examine the VAT the entire group of six assessors (i.e., interviewers) rated ten randomly selected interviews across the first wave of the study. The assessors met as a group and viewed the videotaped recording of the assessment selected, and completed assessment sheets during their viewing. Cohen's Kappas were calculated on agreement between the 11 individual item codings, thus resulting in as many as 5 scores per assessment (the original assessor did not score the video). Kappas were then averaged for each assessment across all symmetric pairings.

Finally, the critical examination of the VAT reliability was the test-retest analysis of the instrument reliability. Using a Spearman-Brown Correlation Method for equal weightings across both time points, a coefficient was derived.

The VAT validity was evaluated by examining assessor narratives against item scores. The assessment team prepared a narrative of their observations for each participant following each interview. This narrative included any pertinent details of the individual's current state in a manner that mirrored the VAT items. Thus, it was a more subjective accounting of what the assessors believed was the current state for the person, regardless of the raw score they gave the participant on the VAT. The premise was that if the assessors are well trained both to understand the nature of the issues their clients faced as well as how to score them on the VAT based upon the intended purpose of the instrument then the narrative and VAT should be highly related. Therefore, to test the concurrent convergent validity of the instrument, a member of the assessment team and another staff member trained on the VAT were directed to score a random set of narratives ($n = 166$) across all interviewers. The scoring method was generally in the same manner as the VAT scoring was done, but all facets were assessed on a three-point scale instead of a five-point scale (as was the VAT). The assessor involved in the coding never scored any of her own assessments. Only assessments from the baseline collection period were used. The narrative scores were then compiled in the same manner as the VAT scores and a correlation matrix was created using Spearman bivariate correlations.

Results

Generally speaking, the sample was predominately heterosexual men who were disproportionately twice as representative of ethnic minorities than the general population of the Seattle area. An apparently large number of the participants were veterans, primarily from the Vietnam era. Finally, the participants subsisted on little to no monthly income, and most sources that they reported were from public assistance programs of one sort or another. See Table 1.

Test for Covariate Effects. None of the ANOVA F -tests were statistically significant, revealing no main effects due to any specific demographic information. This suggests that the variance in the instrument will be due to some factor(s) less easily tapped than these intrapersonal demographics. Thus, these variables were not considered any further in subsequent analyses.

Reliability Analyses. We evaluated the item-total correlations and coefficient alphas (Cronbach,

1951) for the VAT. Data presented in Tables 2 reveal adequate internal consistency with all items contributing to the scale. While all items reliably contribute to the VAT, two items, if removed, would offer slightly nominal improvements to the overall α of scale (i.e., Medical Risks & Substance Use - B).

Table 2. Descriptive information and coefficients regarding Chronbach's α analysis for scale internal consistency of AVCHP at baseline.

Item	Mean (SD)	Item-total r	R ²	α if item deleted
Survival Skills	1.9 (0.95)	.36	.34	.59
Basic Needs	1.7 (0.82)	.50	.37	.56
Indicated Mortality Risks	1.7 (0.94)	.25	.24	.61
Medical Risks	2.4 (1.21)	.15	.23	.63
Organization/Orientation	1.8 (0.96)	.41	.53	.58
Mental Health	2.9 (0.94)	.21	.19	.62
Substance Use - A	2.2 (1.07)	.32	.66	.59
Substance Use - B	2.7 (1.45)	.15	.67	.65
Communication	1.7 (0.87)	.36	.47	.59
Social Behaviors	2.1 (0.94)	.40	.29	.58
Homelessness	2.6 (0.60)	.21	.08	.61
Full Scale total	23.43 (5.02)	-	-	-
Alpha		.62		
Standardized Item α		.66		

Note: N = 277; The Standardized Item α is bolded as it is the appropriate coefficient to accurately depict the internal consistency of the scale; The " α if item deleted" column is in reference to the original alpha score, and italicized scores indicate that items would raise the coefficient if removed.

The inter-rater reliability was quantified using the Cohen coefficient of agreement methodology (Cohen, 1960), a reliability statistic called kappa that calculates the level of agreement between coders while correcting for the level of agreement expected by chance. Table 3 shows mean kappas and the full-scale scores for each assessor across the multiply rated interviews. The kappa means ranged from .31 to .86. Typically, the heuristic is that any kappa below .40 is poor, coefficients between .40 - .75 are considered good, and those kappa coefficients above .75 are excellent.

The Spearman-Brown Correlation coefficient for test-retest instrument reliability ($\rho = .891$ [$p < .001$]) suggests a strong consistency across time, and thus, reveal the strongest findings of these analyses. With such a strong consistency in the temporal nature of the instrument, all other moderate findings are tempered in a more positive direction.

Validity Analyses. The Spearman bivariate correlation matrix, presented in Table 4, reveals along the diagonal a consistently high relationship between each item on the VAT and its corresponding score on the coded narratives (.54-.83) and the total scaled scores correlated strongly (.83) as well. This suggests that there does exist strong evidence of convergent and concurrent validity.

Table 3. Analyses for Inter-rater Reliability: Cohen's Kappa and comparisons of scale total scores.

Interview #	k_{avg}	Total Scale Scores					
		T1 Assessment	Assessment scores of video-taped interviews*				
1	.58	24	21	21	21	22	23
2	.65	23	24	24	26	29	~
3	.79	25	23	23	24	24	26
4	.31	23	21	22	22	23	23
5	.64	21	23	26	27	27	27
6	.56	24	21	24	24	24	24
7	.76	25	25	27	27	28	28
8	.54	21	23	26	28	29	~
9	.69	25	20	21	23	23	24
10	.86	23	21	22	23	24	~
K_{avg}	.67						

Note: k_{avg} = the mean of Kappa scores for all possible pairings of scores based on viewing of the video-taped interview (does not include pairings with actual T1 assessment).

K_{avg} = the mean of the Kappa means for each individual assessment. This does not include the grayed-out Kappa means given that those were either based upon only one kappa score given asymmetrical crosstabs, or because the k_{avg} were the most extreme scores, which we removed.

* = scores are placed in ascending order and assessors are not represented by individual columns.

~ = one less assessor scored video-taped interviews on three occasions due to absences.

Table 4. Correlation matrix demonstrating convergent validity between VAT scores and coded scores for corresponding facets of the assessor narrative reports.

VAT Item	Corresponding Coded Narrative Facets											Total
	SS	BN	IMR	MR	OO	MH	SUA	SUB	COM	SB	HL	
Survival Skills	.64**	.33**	.10	.07	.39**	.24**	-.01	-.08	.47**	.13	.02	.43**
Basic Needs	.27**	.61**	.05	.15	.49**	.23**	-.03	-.04	.39**	.26**	.07	.45**
Indicated Mortality Risk	.23**	.13	.67**	.49**	.01	-.02	.24**	.15	.02	.06	.10	.39**
Medical Risk	.07	.19*	.41**	.68**	-.01	-.09	.06	.05	-.02	-.02	.06	.26**
Organization/Orientation	.24**	.33**	-.02	-.02	.73**	.36**	-.05	-.11	.61**	.18*	.13	.44**
Mental Health	.04	.06	-.17*	-.18*	.19*	.68**	-.03	-.09	.26**	.23**	.02	.20*
Substance Use – A	.19*	.19*	.25**	.16*	.02	.00	.73**	.72**	-.10	.11	.12	.53**
Substance Use – B	-.02	.03	.21**	.10	-.07	-.01	.73**	.79**	-.17*	.05	.15	.42**
Communication	.23**	.25**	.07	-.07	.46**	.20*	-.19*	-.19*	.83**	.12	.05	.30**
Social Behaviors	.20**	.22**	-.00	-.04	.35**	.36**	.02	.03	.36**	.54**	.11	.42**
Homelessness	.07	.07	.07	.11	.13	.03	.26**	.22**	.03	.06	.74**	.39**
Total Scaled Score	.40**	.45**	.37**	.33**	.47**	.36**	.41**	.37**	.46**	.32**	.27**	.83**

Note. * $p < .01$; ** $p < .001$

Conclusions

The psychometric analyses of the properties of the VAT reveal a promising outlook for the use of the measure in the field. Three basic sets of psychometric analyses were conducted that suggest adequate to strong reliability within the instrument. The key finding is that when used across time, the VAT shows a strong consistency. This is not a simple finding when you consider the population that is being assessed. Given the extreme and acute nature of the lives of people coping with chronic homelessness, we would expect that their true vulnerability would be in high fluctuation, with the worst cases regressing back to the mean and newcomers to the street potentially moving in the opposite direction as they living situations have become even more unstable. That the analysis produced a strong & significant coefficient of .891 is a tremendous accomplishment and bodes well for the future temporal state of the instrument after improvements are made both to the internal consistency of the measure and to the assessor training to strengthen coding agreement.

The reliability tests of internal consistency amongst the individual items and the inter-rater agreement are more tempered, but still encouraging. The internal consistency of a measure not only suggests how well the items each relate to the remainder of the scale as a whole, but also how strong a facsimile the scale is of the true focal construct, which in this case is vulnerability to continued instability and homelessness. The resultant indicators of internal consistency (Cronbach's α of .66), the test of how well the individual questions seem to work together as a scale, is a thoroughly positive outcome for a novel instrument in this initial testing phase. The findings do demonstrate, though, a need for adjustments to bring the consistency into the range of established and vaunted instruments. Similarly, the Kappas reassure us that on a 5-point Likert scale assessors can be trained to exhibit a decent level of agreement in their ratings, especially for most of the items. A range of Kappas, of which one is of "poor" status and one is of "excellent" status, with most clustering in the "good" range, is not troubling. Here again, with such an innovative instrument and its first stringent test, we find these results quite promising. This is especially true when you consider the easily deciphered areas for adjustment, as we will address below. In final conclusion, the VAT, as a tool for assessing vulnerability to homelessness and all of the impacts associated with such instability, can enthusiastically be considered useful for its intended purpose.

Recommendations

Two main recommendations are made, based upon the previously discussed findings. First, the training program for assessors needs enhancement in the area of substance use interviewing. The addition of the substance use item-B has added a broader spectrum to the depiction of this issue and its impact on the larger definition of vulnerability. Assessor training should be augmented to include a more comprehensive informational session and continued discussion throughout the training so that the assessment team can find a greater comfort and fluency with the factors that lie beneath the primary question itself regarding the individual's level of substance use, their impairment due to that substance use, and their relationship to treatment.

We also suggest combining the substance use items together entirely, mainly because, as a second scale examining substance use problems in addition to the substance use item in the tool originally, it possibly weights the total score higher for those individuals who exhibit significant substance use disorders.

References

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.