PSYCHOMETRIC RESULTS

OF THE ARIZONA

HOMELESS DATA PILOT

Wayne D. Parker, Ph.D.

The Virginia G. Piper Charitable Trust

The Arizona Homeless Evaluation Project has had the goal of developing tools to evaluate the impact of various Arizona projects in improving the lives of their clients. An early step in the process was to have McKinney funded agencies submit evaluation tools they were currently using along with archived data generated by the instruments. Each instrument was psychometrically evaluated for internal reliability and construct validity.

Several agencies submitted what was called the Arizona Self Sufficiency Matrix. Although these instruments had the same name there were different domains and different descriptors for levels. It is likely that at one time this was a standard instrument, but that each agency adapted it over time so that several distinct versions were created. After reviewing all the submitted instruments, the Self Sufficiency Matrix utilized by UMOM (United Methodist Outreach Ministries) was clearly superior to the other submitted tools.

The archived data from the UMOM Self Sufficiency Matrix demonstrated a clear three-factor structure. One factor appeared to measure independence, another measured dysfunction, and the final measured child functioning. All three factors demonstrated good internal reliability. Based upon these results, the decision was made to pilot this instrument with active clients from a number of agencies. The pilot consisted of having agencies complete the instrument on current clients and retest at exit or at the end of 4 additional months of programming for those who had not exited.

For the data pilot 21 agencies agreed to participate. Of these, 13 agencies actually provided baseline data on a total of 597 clients. Of these clients, approximately one-quarter were in Emergency programs, one-half in Transitional programs, and the remaining quarter were in Permanent Supportive programming. One large agency had the tool filled out by clients while all other agencies had staff rate the level of functioning of each of their clients. The results of the partial data set in which clients filled out the instrument clearly indicated that the instrument should be filled out by staff rather than clients.

The internal reliability and construct validity was tested using the pre/baseline data. In the data pilot there were insufficient numbers of clients with their children, making the childcare and child education domains untestable within this sample. The parallel analysis criterion indicated that the remaining domains formed two factors, however, the factor structure was not the same as the one provided by the prepilot data found in the UMOM sample. The first factor from the UMOM data set, called independence, would be scored by summing the ratings (after imputation of missing data) of Income, Employment, Shelter, Food, Adult Ed, Health Care, Life Skills, Family Relations, Mobility, and Community Involvement. The second factor from UMOM, called dysfunction, would be measured by adding together the ratings on the Substance Abuse, Legal, and Mental Health domains. The total score would be the sum of the two factors.

In the data pilot sample, the Legal and Health domains did not load on a factor. This is likely due to a very restricted range on these variables. Restricted range was problematic but less evident in the Mental Health and Substance Abuse domains and may have been responsible for the different factor structures seen in the two different samples. In the data pilot the first factor, which could be referred to as self-sufficiency, would be scored by adding together the scores from Income, Shelter, Food, Life Skills, Substance Abuse, and Mobility. The second factor (community connectedness) would be scored by adding together Employment, Adult Ed, Mental Health, Family Relations, and Community Involvement. Again, the total score is the sum of these two sub scores.

Internal reliability was acceptable for both scoring solutions, but better in the UMOM sample. The factor solution was better for the UMOM sample as well. The most prudent approach, assuming the decision to go forward with a trial implementation of the matrix, is to score the instrument using both solutions and determine which solution is superior in the larger dataset. It would be important that training be conducted in the use of the matrix, especially by providing one or more thumbnail case studies that would then be scored according to the matrix. Reasonable gains in scores over time could be measured by using Cohen's standards for effect sizes based upon the standard deviations of baseline measures for types of programs and or subpopulations. Scoring of the changes from pre to post demonstrated that the instrument, given the short period of 4 months that was utilized, is sensitive to temporal changes.