Organizational Effectiveness Program
Packard Grant # 99-4830

October 15, 1999

Submitted by the Institute for Community Collaborative Studies
California State University, Monterey Bay
by
Jerry Endres, M.S.W. Principle Investigator, Institute for Community Collaborative
Studies
Brad Richardson, Ph.D. Research Consultant, National Resource Center for Family
Centered Practice
Judi Sherman, M.A. Research Assistant, Valley Resource Center

TESTING THE RELIABILITY/VALIDITY OF THE CALIFORNIA MATRIX

Grantee

California State University Monterey Bay (CSUMB) Institute for Community
Collaborative Studies (ICCS) is located on the Central Coast of California in Monterey
County on the former Fort Ord. The primary role for ICCS as a university/community
partnership is to facilitate collaborative problem solving and build capacity to develop
integrated human service delivery systems within the tri-county region of Monterey, San
Benito and Santa Cruz counties. The Institute examines how human service delivery
systems can collaboratively function to serve the whole child, the whole family, and the
whole community. A tri-county advisory board made up of community leaders in health,
human services, probation, recreation, education, United Way, libraries and others,
provides leadership, advocacy, and guidance for the activities and direction of ICCS.
Since its inception in 1995, ICCS has provided regional leadership in establishing a
strong academic program in Collaborative Human Services well as securing funding for
innovative collaborative projects in the areas of technology for human services, welfare
reform and community empowerment, developing interprofessional outcomes for human
services education, and evaluation of human service programs.

Background for this Project

Human service agencies are now routinely required to account for their funding through
measurable outcomes of their service delivery. This form of evaluation has proven to be
challenging for service providers who are comfortable and used to measuring their
service delivery success by outputs and what typically represents the type of work the
agency performs - whether or not it actually helps families and responds to needs. The

Organizational Effectiveness Program
Packard Grant # 99-4830

California Matrix Model (AKA California Family, Agency and Community Development Scales) is a scales model that measures outcome interventions at the family, agency and community levels.
The development of the model came in response to the passage of the federal Government Performance and Results Act and has been developed in coordination with the federal department of Health and Human Services, Office of Community Services, Monitoring and Assessment Task force (HHS OCS MATF) and is compatible with the six national goals adopted by the MATF for family self sufficiency. This model is being used in varying degrees by Community Action Agencies, Healthy Start programs, Family Preservation and Support Programs, the California Mentoring program and the REACH program, Family Resource Centers and other interested community based agencies. In our tri-county region, the model is in use by at least 50 different public and private nonprofit agencies. Most agencies using the Matrix Model are using the Family Matrix/Scale, which has gained considerable acceptance for case management practices. The community development matrix/scale is used less often. The agency matrix/scale appears to be the least used at this time.

In the Monterey Bay region, the demand for reliability and validity standards and criteria in using this model has increased as more human service agencies adapt to the use of outcomes as a measure of accountability. ICCS convened a Matrix Design Group in 1997 to address the demand for training and technical assistance, testing reliability and validity as well as to incorporate best practices for human service evaluation. Towards this end, ICCS has:

   Convened numerous Matrix Design Group meetings and provided a series of
   regional workshops in using the model;
   Developed a web site for workshop tutorials and information on using the Matrix
   Model that is downloadable for all Matrix users;
   Provided technical assistance to the Family Resource Center Network made up of 19
   nonprofit agencies in Santa Cruz County to develop a model for testing reliability;
   Provided presentations at various seminars on the model including the recent Ca.
   State Fullerton Outcomes Conference and a Cowell/Packard foundation joint meeting
   on the California Matrix model;
   Provided technical assistance to interested nonprofits within the region in adapting
   the Family and Community models. Examples include the Girl Scouts of America,
   Monterey Bay and Resources for Families and Children in Santa Clara County;
   Participated in California meetings of the California Outcomes Symposium
   sponsored by the California Department of Community Services and Development.

Organizational Effectiveness Program
Packard Grant # 99-4830

ICCS works with approximately 70 regional members of the Matrix Design Group to identify areas to improve the California Matrix model as an assessment, case management and outcomes tool to improve services with children, families and communities. We adopt a research and development mode addressing issues of common concern. That is, ICCS seeks funding to develop and test the Matrix matrices using a participatory approach emphasizing involvement by Design Group members. As a result, member agencies receive accurate and timely information and are prepared to implement the research findings.

Project Activities

After receiving an announcement of the grant on April 2, 1999 the following activities were generated to accomplish the project objectives:

April 12	ICCS hired Brad Richardson as research consultant. Presentation by Jerry Endres, Glenn Chee and Brad Richardson at a Cowell / Packard Foundation sponsored meeting of foundation and family services agency representatives on the California Matrix. Discussions of reliability and validity and the need for research were established.
April 13	Meeting with Brad Richardson and the ICCS Matrix Design Group to establish study methods for testing Family Matrix reliability. Decision was made to focus on intra- and inter-reliability using case scenarios with participation of Design Group member agencies.
May 1	ICCS contracted with Valley Resource Center to hire Judi Sherman, VRC Assistant Director, to design 10 case scenarios, help participating agencies respond to these scenarios, and provide their responses to the research consultant, Brad Richardson. Ms. Sherman recruited all study participants, developed data entry rating forms, and provided technical assistance to participants on their role in the study.
July 20	ICCS Matrix Design Group met with Brad Richardson to review preliminary results of the study. Also, began to design research for testing reliability and validity of the Family Matrix using actual data from clients from participating agencies.

September 22   Meeting of Matrix Design Group with Brad Richardson to review
                      results of study and complete future research plans. Also, a focus
                      group discussion of questions related to the use of the Family
                      Matrix with clients and suggestions for improving reliability within
                      specific Matrix categories. This meeting also represented
                      Reliability and Validity Standards and Criteria Workshop for
                      regional human service agencies.

October/November   This report is written and will be placed on the ICCS web site for
dissemination.

Project Objectives, Design, Testing and Findings

The Design Team recognized in May, 1997 that the model was not perfect and the
fundamental issue of reliability and validity needed to be addressed. Last year, ICCS
wrote to the Packard Foundation requesting funding to secure a credentialed researcher to
design a research methodology to test the validity and reliability of the Family Matrix and
address these objectives:

Questions addressed by the Matrix Design Group included the following:
          a) Why should standards and criteria for reliability and validity be applied to the
          Matrix Scales model of evaluation?
          Standards and criteria are the foundation of designing a research methodology to
          test reliability and validity of a tool like the Family Matrix.   Issues such as
          balancing the science of outcome measurement with the art of service delivery
          and the capacities and resources of local agencies were working assumptions of
          this study.   Testing results of reliability and validity results in improved usage of
          the model and stronger community stakeholder buy in. We want to strengthen the
          outcomes system for members of the Matrix Design Group and help make the
          Matrix Model valuable for agencies and communities and the clients who use it.

          b) In what ways is the Family Matrix Scale reliable and valid for measuring
          outcomes?
          We determined to examine an existing Family Matrix model developed by ICCS
          with the Family Resource Center Network of Santa Cruz County. This Family
          Matrix model represents an assessment and case management tool developed with

a sub-committee to be used by the FRCN to establish client outcome data. The
use of an existing model allowed us to immediately implement recommendations
from the study into the model.

Study Procedure for Objective #1: Develop standards and criteria for scales models such as the California Matrix Model

Quasi-experimental Study of reliability

Ten narratives were written based on actual case histories of program participants from
agencies participating in the study (see attachment). Twenty-five workers from agencies
participating in the study volunteered to serve as raters using the FRCN Family
Development Matrix. The workers completed a rater profile and score sheet on which
they rated the narratives using the 11-outcome categories of the Matrix. The rater profile
included individual participant information: last six social security digits, age, agency
years in present position, years social service experience, whether or not the participant
has used the matrix before the study, whether or not the rater received training on the
Matrix model.

Upon arriving, the test administrator gave an overview regarding the purpose of the
study. Workers were instructed on how to complete the rater profile, the purpose of
which is to provide information needed to analyze the data among raters and between
points in time for the same raters. Workers received a brief review on how to use the
Matrix to rate the case scenarios. The test administrator reviewed the rating scale and the
status levels listed on the bottom of the rating form. The workers were instructed to score
numbers (1) in crisis; (2) at risk; (3) stable; (4) safe/self-sufficient; (5) thriving; and were
allowed to use half-points if they determined that a situation fell between two status
levels. The test administrator used an example from the outcome category of "Shelter,"
stating that if the case study describes a family who can pay the rent each month but has
no savings, participants would score "Shelter" a 2.5, between at risk and stable. They
were directed to read the "Shelter" category, and apply this example.

For each case scenario, workers were instructed to use the corresponding rating form
numbered 1 through 10. Below each outcome category, raters were to write the number
of the category within the scale that corresponds to each situation. Any questions from
the participants were answered at this time.

Participants were requested to rate as accurately as possible and were told that this experiment was to determine how the Matrix works as a reliable instrument. The test administrator emphasized that this is not a test of the workers or how well they interpret the case scenarios.

Raters spent between two and three hours on the study, while one rater spent more than four hours. Upon completion of the 10 sets of ratings, raters placed the forms and their rater profiles in an envelope. The test administrator mailed the envelope to the study researcher at the University of Iowa.

Approximately two weeks following the completion of the first study, workers met again to complete the Matrix for the same 10 case scenarios. Workers were instructed to complete the rater profile using the same identification code, and to complete the ten rating forms corresponding to the case scenarios. The test administrator then placed the forms and the rater profiles in an envelope and mailed the envelope to the analyst.

Reliability Assessment

Analysis of reliability, the degree to which ratings are consistent for similar events, was approached in three ways. First, to assess the how consistently raters were in their ratings of families described in the narratives (intra-rater reliability), paired comparisons (t-tests and Wilcoxon signed rank tests) were performed. The paired comparisons were used to assess each pair of ratings on the California Matrix between the first and second administrations of the narratives.

Second, we examined inter-rater reliability from the perspective of the degree to which ratings were the same or similar on each individual item for each narrative. Results from this analysis provide information on which domains produce the most consistent, or discrepant, ratings. Along with feedback from the subsequent focus group discussions, this information suggests where improvements in definition and sensitivity can be made.

Third, we examined exogenous factors such as training, age, and experience that may impact ratings. By controlling for these effects we can determine if extraneous factors may have an impact on the domain ratings.

Finally, we examined the consistency of ratings from the perspective of the narratives to determine if particular case descriptions adversely affected the results; this approach is

similar to item analysis used in test construction. By examining each set of domain scores (e.g. ratings on shelter across all narratives) we can assess variation across narratives to determine if raters consistently received the information from the narratives as indicated by their scores across domains. This approach can indicate where information contained in the narratives may have compromised reliability.

Intra-Rater Reliability

Table 1 presents the domains of the California Family Matrix Model, and, for each domain across the ten narratives, specifies which domains were marginally or significantly different on ratings from the first to the second administration.

Of the eleven domains, three showed no significant differences between administrations: Health/Safety, Social/Emotional Health and Family Relations were each found to have no marginally or significantly different scores between the two administrations. Shelter had only one marginally significant difference. Food/Clothing, Community Relations, Adult Education/Employment, Children's Education/Development each had two marginally or significantly different scores between administrations. Finances and Immigration/Resettlement obtained three marginally or significantly different scores, and Transportation/Mobility obtained four marginally or significantly different scores.

Table 1:
Differences Within Raters on Narratives
Using the California Matrix Model Domains

Narrative No. ------------->	1	2	3	4	5	6	7	8	9	10
Domain										
1 Shelter	*									
2. Food/Clothing		**					**			
3. Transportation/Mobility	**				*		*		*	
4. Health/Safety										
5. Social/Emotional Health										
6. Finances		**			*		**			
7. Family Relations										
8. Community Relations		**							*	
9. Adult Education/Employment	**	**								
10. Children's							**		**	

Organizational Effectiveness Program
Packard Grant # 99-4830

Education/Development										
11. Immigration/Resettlement		*	*						*	
Total	3	5	1	0	2	0	4	0	4	0

** p. < .05
* p. < .10

There are ten scores that differ between the two administrations of the tests at a statistically significant level. There are nine differences, which are marginally significant. This means that of the 110 paired comparisons, 91 of the paired comparisons showed consistent ratings by raters at the two points in time of administration.

Transportation/Mobility, Finances, and Immigration/Resettlement appear to require the most refinement. These three domains represent more than half (10 out of 19) of the marginally or significantly different scores. This is encouraging since it means that the remaining 8 domains appear to be relatively robust with only 9 marginally or significantly difference scores among them.

It should also be noted that narratives 2, 7 and 9 account for 13 of the 19 significantly different scores. Removing these three narratives from the analysis would result in only two significantly different ratings (both of which appear in narrative 1) and four marginally significant ratings (in narratives 1, 3 and 5 which has two marginally significant ratings).

Examining only significant differences (Table 2) shows that removing narratives 2, 7, 9 and 1 would yield a Matrix with no significant differences in ratings between the two points in time of administration of the narratives. However, concluding that all unreliability is due to the information contained in these four narratives would be fallacious at best.

Table 2:
Significant Differences Within Raters on Narratives
Using the California Matrix Model Domains

Narrative No. ------------->	1	2	3	4	5	6	7	8	9	10	T
Domain											
1 Shelter											
2. Food/Clothing		**					**				
3. Transportation/Mobility	**										
4. Health/Safety											
5. Social/Emotional Health											
6. Finances		**					**				
7. Family Relations											
8. Community Relations		**									
9. Adult Education/Employment	**	**									
10. Children's Education/Development							**		**		
11. Immigration/Resettlement											
Total	2	4	0	0	0	0	3	0	1	0	

Table 3 presents the direction of significant or marginally significant differences (from Table 1) between the first and second administrations.

Differences in Immigration/Resettlement scores are attributable to more raters indicating that the domain did not apply during the first rating than the second rating. The only other positive change was for Children's Education/Development in Narrative 7, and that too was the result of an increase in the number indicating that the domain did not apply to the case. All other differences were negative. A negative difference score indicates that the raters scored higher during the second administration than they did during the first administration.

Table 3:
Direction of Differences within Raters on Narratives
Using the California Matrix Model Domains

Narrative No. ------------->	1	2	3	4	5	6	7	8	9	10 T
Domain										
1 Shelter	-									
2. Food/Clothing		-					-			
3. Transportation/Mobility	-				-		-		-	
4. Health/Safety										
5. Social/Emotional Health										
6. Finances		-			-		-			
7. Family Relations										
8. Community Relations		-							-	
9. Adult Education/Employment	-	-								
10. Children's Education/Development							+		-	
11. Immigration/Resettlement		+	+						+	
Total	3	5	1	0	2	0	4	0	4	0

Inter Rater Reliability

Inter-rater reliability is perhaps the most intuitively appealing of our approaches to
establishing reliability. Simply stated, the question of inter-rater reliability is how many
raters wrote the same response. While not sufficient, establishment of reliability is a
necessary condition for validation.

Table 4 presents the percent rating the modal response (that response which occurs most
frequently) at the second administration (Time 2). Table 5 presents the percent rating the
modal response for Time 1. In each table, highlighted percentages indicate where inter-
rater reliability requires improvement (i.e., less than 60 percent agreement). During the
second administration, 32 of the 110 items resulted in less than 60 percent agreement
compared to the first administration where 50 of the 110 items resulted in less than 60
percent agreement.

Table 6 presents the differences between inter-rater agreement at Time1 and Time2 (i.e., Time 2 minus Time 1). This is an indication of the fluctuation in agreement occurring between the first and second administrations of the instruments. Numeric percentages indicate the magnitude of the change, while the sign of the percentage change indicates which administration had greater agreement (negative scores indicate where Time 1 agreement percentages were greater, positive percentages indicate where Time 2 agreement was greater). Agreement appears to have improved at Time 2; 87.3 percent of the scores are positive indicating more agreement at Time 2. Highlighted percentages in Table 6 indicate where a gain of nearly one-third of raters was realized at Time 2.

Table 4:
Percent Agreement for Raters on Narratives
Using the California Matrix Model Domains (Time 2)

Narrative No. ------------->	1	2	3	4	5	6	7	8	9
Domain									
1 Shelter	61.1	100.0	72.3	50.0	66.7	50.0	55.6	66.7	66.7
2. Food/Clothing	88.9	50.0	55.6	61.1	61.2	38.9	72.2	77.8	72.2
3. Transportation/Mobility	77.8	66.6	50.0	66.7	61.1	72.2	94.4	72.2	77.8
4. Health/Safety	61.1	55.6	50.0	61.1	55.6	72.2	61.1	72.2	72.2
5. Social/Emotional Health 5x	55.6	72.2	88.8	55.6	66.7	55.6	66.7	55.6	61.2
6. Finances	77.3	66.7	72.2	72.7	66.7	50.0	72.2	89.0	83.4
7. Family Relations 7x	55.6	61.1	72.2	50.0	55.6	55.6	38.9	38.9	61.1
8. Community Relations 6x	55.6	55.6	66.6	50.0	55.6	38.9	55.6	72.2	72.2
9. Adult Education/Employment 4x	88.9	88.9	83.3	50.0	55.6	66.7	61.1	66.7	83.3
10. Children's Education/Development	77.8	88.9	55.6	44.4	77.8	55.6	77.8	55.6	61.1
11. Immigration/Resettlement	77.8	83.3	88.9	83.3	50.0	66.7	88.9	83.3	94.4
Percent ± .5 of the mode									

Table 5:

Percent Agreement for Raters on Narratives

Using the California Matrix Model Domains (Time 1)

Narrative No. ------------->	1	2	3	4	5	6	7	8	
Domain									
1 Shelter	55.6	88.9	66.7	55.6	61.1	61.1	55.6	50.0	7?
2. Food/Clothing 6x	55.6	61.1	38.9	55.6	55.6	55.6	72.3	77.8	66
3. Transportation/Mobility	55.6	61.1	44.4	66.7	55.6	72.2	72.3	72.2	94
4. Health/Safety	38.9	50.0	44.4	61.1	55.6	61.1	44.5	61.1	6]
5. Social/Emotional Health	55.6	61.1	66.7	55.6	55.6	72.2	72.3	44.5	6]
6. Finances	55.6	66.7	61.1	61.1	83.3	38.9	77.8	72.2	66
7. Family Relations 7x	55.6	55.6	61.1	50.0	72.2	66.7	55.6	44.5	5(
8. Community Relations	44.5	61.1	66.7	50.0	61.1	66.7	61.1	55.6	5:
9. Adult Education/Employment	55.6	61.1	66.7	50.0	50.0	55.6	61.1	72.2	66
10. Children's Educ./Development 7x	44.4	61.1	55.6	44.5	88.9	44.5	61.1	33.3	5(
11. Immigration/Resettlement	66.7	66.7	66.7	66.7	50.0	50.0	77.8	61.1	7?
Percent $\pm$ .5 of the mode									

Table 6:

Percent Change in Agreement for Raters on Narratives

Using the California Matrix Model Domains

(Time 2 minus Time 1*)

Narrative No. ------------->	1	2	3	4	5	6	7	8	
Domain									
1 Shelter	5.5	11.1	5.6	-5.6	5.6	-11.1	0	16.7	-1]
2. Food/Clothing	33.3	-11.1	16.7	5.5	5.6	-16.7	-0.1	0	:
3. Transportation/Mobility	22.2	5.5	5.6	0	5.5	0	22.1	0	-1(
4. Health/Safety	22.2	5.5	5.6	0	0	11.1	16.6	11.1	1]
5. Social/Emotional Health	0	11.1	22.1	0	11.1	-16.7	-5.6	11.1	(
6. Finances	21.7	0	11.1	11.6	-16.6	11.2	-5.6	16.8	1(
7. Family Relations	0	5.5	11.1	0	-16.6	-11.1	-16.7	-5.6	1]
8. Community Relations	11.1	-5.5	-0.1	0	-5.5	-27.8	-4.5	16.6	1(
9. Adult Education/Employment	33.3	27.8	16.6	0	5.6	11.1	0	-5.5	1(
10. Children's Education/Development	33.4	27.8	0	-0.1	-11.1	11.1	16.7	22.3	1]
11. Immigration/Resettlement	11.1	16.6	22.2	16.6	0	16.7	11.1	22.2	1(

*Note: Negative scores indicate Time 1 is greater than Time 2

Rater Characteristics and Ratings

The average (mean) age of raters was 41 years. Average number of years in social service was 10 with an average of 3 to 4 years of employment in the current position at the agency. Seventy-eight percent (78%) reported that they had been trained to use the California Matrix, and 78 percent also reported that they had experience using the Matrix. Forty-four percent (44%) of participants were employed in two agencies that volunteered to participate; not more than two raters participated from any of the other agencies. Significant correlations between ratings and rater characteristics were not found.

Test Analysis

Examining the internal consistency of ratings over each domain may indicate where narratives used in the experiment may not have provided adequate information consequently having an adverse effect on the results. By examining each set of domain scores (e.g. ratings on shelter across all narratives) we can assess variation across narratives to determine if raters consistently received the information from the narratives as indicated by their scores across domains. Assuming that raters have varying but constant thresholds for rating using the matrix, this approach can also indicate which domains appear to have inconsistencies in ratings over the series of narratives.

Though less sophisticated than procedures drawing from work in item response theory, this approach is similar to item analysis used in test construction. Table 8, below presents coefficient alpha for each domain at each administration (Time1 and Time2). Interestingly, coefficient alpha was quite high for Time1, however, Time2 results were much less consistent. This suggests that some test effects may have been present. In an attempt to fully understand the Time2 results, narratives that were indicated to compromise reliability were removed from the analysis. Asterisks indicate which narratives were removed from the analysis of domains for the Time 2 data to obtain these higher reliabilities. With the deletion of narratives from the analysis, resultant increases in the alpha scores are reported as adjusted Time2 scores (see the column headed: Time2adj).

Organizational Effectiveness Program
Packard Grant # 99-4830

These results indicate that narratives 2 and 5 reduced alpha – narrative 2 reduced alpha in 7 domains and narrative 5 reduced alpha in 8 domains. Narratives 3 and 9 reduced alpha in 6 domains. However, this was the case only during the second administration of the narratives.

Examination of domains across narratives reveals that Food/Clothing is inconsistently scored by raters in 6 of the 10 narratives (i.e., raters did not score consistently low or consistently high on this domain). Shelter was scored inconsistently by raters in 5 of the 10 narratives, as was Adult Education/Employment. Transportation/Mobility, Health/Safety, Social Emotional Health, and Community Relations were inconsistent in four of the 10 narratives. While Immigration/Resettlement was not shown to be inconsistently rated, there are indications that the use of this domain requires some refinement. The statistical data may not be sensitive enough to capture this since many of the narratives did not include issues of immigration and resettlement.

Table 8:
Alpha Reliability for Domains and Narratives
Using the California Matrix Model

Narrative No. ------------->	1	2	3	4	5	6	7	8	9	10	Coefficient A Time1	Time 2
Domain												
1 Shelter		*	*		*		*		*		0.88	0.18
2. Food/Clothing		*	*	*	*	*				*	0.88	0.43
3. Transportation/Mobility		*	*	*				*		*	0.91	0.43
4. Health/Safety		*	*			*			*		0.90	0.18
5. Social/Emotional Health		*			*				*	*	0.87	0.51
6. Finances				*	*	*					0.88	0.60
7. Family Relations			*	*	*					*	0.77	0.30
8. Community Relations					*	*		*	*	*	0.88	0.70
9. Adult Ed./Dev.		*	*		*			*	*		0.87	0.00
10. Children's Ed./Dev.		*					*				0.85	0.70
11. Immigration/Resettlmnt					*			*	*		0.93	0.87
Time1 alpha (across domains)	0.83	0.82	0.83	0.91	0.82	0.89	0.78	0.85	0.84	0.91		
Time2 alpha (across domains)	0.16	0.74	0.75	0.66	0.55	0.66	0.19	0.00	0.58	0.69		

× Adult Education/Employment achieved this level of reliability with two separate sets of narratives: 1,4, and 10, and 6 and 7.

Findings and Recommendations

The analyses show that there is generally good reliability, and that there are specific areas where improvements could be made to increase reliability. Time 2 scores tended to increase and the analysis suggested that some test effects occurred. Fatigue may have also play a part in the ratings; raters generally spent two to three hours reading the narratives and marking the rating instruments. Finally, we must be mindful of the fact that the present assessment was conducted as in a laboratory experiment and not in the setting in which the instrument is generally used. Therefore, artificial limitations were placed on the information raters were able to obtain in order to make their ratings.

Following the second testing period, the test administrator conducted a debriefing with participants in the study. Participants reported the following observations about their experience in the experiment:

- It took a long time to read ten cases.
- It was tiring do work for three hours.
- I wasn't sure how to rate Adult Education and/or child development when there were more than one adult or child in a family scenario.
- Transportation was problematic because it is focused on having a car. Some people do fine without a car.
- Some situations needed follow-up questions to clarify. I would have asked questions if I were working with a real family.
- The case scenarios were very realistic.

Observations by the test administrator:

- During the second phase, some participants said they tried to remember how they'd answered in the first phase.
- Several people who use the Matrix in their work read the scenarios, marked the score sheets, then went back to review the scores with the Matrix.

Organizational Effectiveness Program
Packard Grant # 99-4830

Recommendations: standards and criteria for scales models, such as the California Matrix Model

There has been growing concern that traditional experimental approaches to scales development do not allow the flexibility needed for effective program use and evaluation, and that those in the field find the benefits of these research efforts to be exceeded by their cost. Below we describe four areas that are critical in the validation of scales models:

1. Maintain contact with a Matrix Design Team or Similar Group
Our experience has been that working with the Matrix Design Team, or a similar group of field experts, is valuable for creating and revising scales. Such groups are important in the initial design, and after implementation their expertise can be invaluable asset to the process of continual improvement. In our research, empirical results were presented to the Matrix Design Team who provided feedback and qualitative data upon which further improvements in the definitions of the domains could be made.

2. Monitor Reliability
Establishing the initial reliability of scales models is important. This is a first step to achieving the conventional 90 percent agreement for both inter and intra rater reliability. Our initial reliability analyses indicated the performance levels for each of the domains, and working with the Matrix Design Team we believe that the reliability may be improved to 90 percent. Once initial reliability is established, the refinement process can take place resulting in a revised instrument that may then be subjected to the validation process.

3. Engage in a process of validation:

3a. Determine concurrent measures
Initial validation of the instrument has occurred to some extent through the efforts of the Matrix Design Team (i.e., face and content validity). Once reliability estimates are maximized, however, appropriate measures must be found with which to establish concurrent and predictive (see below) validity. Concurrent measures may be other similar instruments which could be simultaneously taken, extant data on similar constructs, or other means by which to measure the domains of the matrix (cf. Campbell and Fiske, 1959, "Convergent and discriminant validity by the multimethod multitrait matrix").

3b. Include a time period, which will allow for measurement and testing of predictive validity

For predictive validity, a time period that will allow for measurement and testing at a follow-up interval is necessary. For scales model such as the Matrix Model we suggest at least a two year measurement period, with at least two measurement points within each of those two years, and a follow-up measurement in the third year. These time periods are based on the experience of the present research and the authors' knowledge of other recent and similar validation studies (e.g., North Carolina Family Assessment Scale; Developmental Assets; Automated Assessment of Family Progress).

4. Provide Results of the Analysis of Data and Encourage Feedback

Careful analysis, and interaction of the program evaluation and research team with the users of the model, both clients and workers, along with input from the design team is essential for validation and utilization of scales models. Analysis time must be built into the validation process in order to obtain the full benefit of outcome measures systems such as the California Matrix Model. It is important to adopt a longitudinal design that incorporates concurrent and predictive measures for validating family functioning domains, along with service data that can be correlated with the effects. There is currently a dearth of instruments which are strengths-based and developmental, such as the California Matrix Model, that have undergone a rigorous validation process such as we describe and also respond to the present needs and concerns of users in the field. Finally, results must be presented in an ongoing manner to users of the instruments so that ongoing feedback demonstrates the value of the measures for practitioners and families.

Study Procedure for Objective #2: Design a research methodology for testing the reliability and validity of the Matrix using data from the agencies that use the model.

Based on discussions with agency administrators and staff it was determined that the first phase of the validation process would focus on reliability testing and improving the reliability of the Family Development Matrix. A similar approach could then be adopted for the Agency and Community Matrices. To accomplish this, the methodology agreed upon began with an experimental design requiring participants (test subject volunteers from agencies) to rate narratives that were written based on actual agency cases. Utilizing the results from the experimental design, and input from the Matrix Design Team, refinement of the domains of the California Family Matrix is now the next step.

Working with counties to gather extant data on domain-related indicators will provide concurrent validation of the domain scores, especially when analyzed over time. It is necessary to track initial population measures and client measures to establish comparative baselines. In this manner, change can be separately attributed to program, agency, or community influences. Through a longitudinal design, validation can be achieved seriatim, first concurrently and then predictively.

We will propose a two-year study focusing on both reliability and validity utilizing data from 7 agencies serving families and children in four counties (Monterey, San Benito, Santa Clara and Santa Cruz). Each agency will receive training on using appropriate domain-related indicators, a standard operating procedure to collect data and will provide ICCS with the aggregated data in a timed schedule. In addition to Family Matrix outcome measures, services provided will be tracked so that linkages between services and outcomes may be obtained. The research consultant will complete the data analysis and provide feedback to the agencies. At this time there is a high interest for agencies to participate especially with some seed funding for their involvement.

We already have in place a Matrix Design Team and reliability estimates for the domains. Poised for bringing together the Design Team to revise and refine the domain category definitions, the work schedule for implementation of the validation process is streamlined enormously.

Simultaneous with the Design Teamwork we will begin work with counties to gather appropriate data for concurrent validation of the domain scores. Tracking measures (time series data) will ensure validation and ongoing reporting to programs and the Design Team will provide continuous quality improvement and maximize utilization of the scales. Through the analysis we will also be able to track change and attribute outcomes to sources in the programs, agencies, and communities.

Study Procedure for Objective # 3: <u>Design an evaluation of a project using the Matrix Model.</u>

We propose to evaluate a project using the Matrix Model Community Scaling Tool for measuring outcomes based upon standards of reliability and validity. Community Matrices can be validated using this methodological approach by correlating family development domains and county, state and local data; where aggregated client data from the Family Development Matrix are available. The effort is strengthened because the

scales can be used in a manner similar to that described by Patton (1980) as "triangulation." Consonant with our recommendations for criteria and standards for scales models of 2 to 3 years of measurements at 4 to 6 month intervals, our method for testing validity and reliability of the Family, Community and Agency matrices is one which suggests longitudinal progress tracking at specified measurement intervals.

The Santa Clara County-based Resources for Families and Children agency has agreed to participate in a study to track and evaluate outcomes associated with services provided to their clients - 50 neighborhood grassroots groups and organizations serving multi-cultural community settings and populations throughout Santa Clara County.

Data have been collected over the past year; however, some changes in domains as a result of the reliability analysis may mitigate the use of existing data. The evaluation of the project will incorporate the strategy recommended above for testing the reliability and validity of the Matrix Model.

Future
It is our intent to return to your Foundation with another grant proposal to carry out Phase Two, anticipated to be two to three years. The objectives developed by the Matrix Design Group for continuing the reliability and validity study to Phase 2 include:

1) Improve the levels of indicator reliability especially in 4-5 selected categories (Food/Clothing, Finances, Adult Education/Employment, Children's Education/ Development and Immigration/Resettlement) as identified in the analysis of the experimental data on the Family Matrix model. The Tools Committee of FRCN working with ICCS will adjust indicators identified by current study. Make sure we capture client participant entry dates. (We currently don't do this and it effects how aggregated data looks. We will also work to improve the comfort level, vocabulary, and transitions from one status level to another.

2) In the field (7 agencies in 4 counties) implement and test the Family Development Matrix as an outcome measure system that looks at family progress in association with services. Establish a working group of agencies participating in the study. Keep that group together for the life of the study to obtain direct and immediate feedback about how the study progresses. Look at: a) which interventions move people to self-sufficiency; b) our understanding the model so we know what we are measuring; and c) Matrix model as both a case management and evaluation tool. Develop research study guidelines (how we'll use the Matrix with clients, how often we expect to Matrix the clients, how to gather the data regarding service provision), and request

that all agencies participating agree to use the guidelines. Develop methods to track services provided or referrals made. Match to status level change over time.

3) Establish the beginning of a clearinghouse to help agencies aggregate and analyze data. Help workers/agencies to set up Excel and use SPSS worksheets to enter data for categories and conduct simple training on how to do data entry and aggregate it.

4) Provide training and technical assistance to enable participating agencies to serve as mentors to help with agency adoption of the Family Matrix model. Develop clear guidelines for training staff to use Matrix. An example is how to rate a family when there's more than one child or adult or, when a family fits in two status levels.

1) Provide research and technical assistance to one agency (RFC in Santa Clara County) to test in the field the reliability /validity of the Community Matrix. Develop replicable procedures for transfer to other community-based programs.

Closure

There is considerable work occurring on the national level and throughout the country in the area of developing and measuring community outcomes. Scaling models are also being used in varying degrees. The extent to which we can support the use of a valid and reliable model that increases agency capacity to help families in this era of welfare reform, is imperative. We found it an efficient use of resources to study this model as it adds to the work in progress at the local, state and national levels in the field of measuring human service outcomes. For example, this is the first quasi-experimental study of the Family Development Matrix Model. The present work has provided information upon which improvements in the domains of the Family Development Matrix can be made and has provided us with a strategy for Phase 2 of the study for testing validity with client data. This effort will require an investment of time and additional resources. There are many agencies willing to participate in Phase 2 as long as the expense related to their investment of staff time is compensated to some degree.

The challenges of this project were with the short time (5-6 months) to complete a quasi-experimental study. Therefore we were unable to completely develop the details for a research design for Phase 2. These will follow in a subsequent proposal. On the other hand, the value of this project was the already functioning Matrix Design Group, and the resource work of Judi Sherman and Brad Richardson to conduct the study and the data analysis. We have completed the first test of the reliability of the Family development Matrix. ICCS with the tools committee of the FRCN and RFC family advocates will utilize these findings to strengthen the FRCN and RFC Matrix Model. ICCS and RFC staff and community group members will further develop the Community Matrix for in the field testing. With this collaborative team together for Phase 2, ICCS and the Matrix

Organizational Effectiveness Program
Packard Grant # 99-4830

Design Group will be able to test both reliability and validity for the Family and
Community Matrix.