

**An Examination of the Cultural Competence
and Inter-Rater Reliability of the
North Carolina Family Assessment Scales
(NCFAS-R and NCFAS-G+R)**

Prepared by

Raymond S Kirk, PhD, Director
Division of Research and Program Evaluation
Independent Living Resources, Inc.

and

Priscilla Martens, MSW, Executive Director
National Family Preservation Network

December 2012

Abstract

Objective

This study examined the cultural competence of the North Carolina Family Assessment Scales (R and G+R versions) with respect to the domains of family functioning comprising the scales. It also examined the inter-rater reliability of the Scales.

Methods

A group of White and Alaskan Native social workers collaborated to write a 2-part case scenario to be assessed by volunteer social workers prior to (intake) and after (closure) services. Volunteers were 55 social workers who varied by gender, race, age, experience, education, and location of case practice (3 states and 1 international country).

Results

Volunteer social workers who assigned ratings on family functioning were very similar regardless of demographic characteristics of the workers. The NCFAS Scales were found to be applicable to the Alaska Native case, when all volunteers' ratings were compared to the case author's standard. Demographic variables did not contribute in any meaningful way to assigned ratings. A very large majority of workers' assigned ratings were within one scale increment of the case standard set by the scenario author.

Conclusions

The findings contribute to increasing evidence that the NCFAS Scales are applicable across various racial, cultural and ethnically identified groups, and that the purpose of the NCFAS Scales (to provide an organizing framework to conduct a comprehensive family assessment) contributes to that property of racial/cultural/ethnic relevance and applicability. Evidence suggests good inter-rater reliability among divergent social workers using the same instrument and the same case scenario.

Introduction

General Overview

This report presents the findings of a study designed to examine the racial/cultural appropriateness of the North Carolina Family Assessment Scales with respect to indigenous populations; specifically, using Alaskan indigenous tribes as the focus of the study. The report also examines inter-rater reliability of raters from various racial/ethnic groups within and outside of Alaska when using the NCFAS-R or NCFAS-G+R when assessing a hypothetical case study based on an Alaskan native family.

The impetus for the study was a response to questions about the racial/cultural appropriateness of the NCFAS scale content, especially in relation to Alaskan native cultures. The NCFAS Scales have been used by programs and social workers serving families representing many racial and ethnic identifies, including White, African American, North American Indian (contiguous lower 48 states), Alaskan Indian, New Zealand Maori, Australian Aboriginal, American Hispanic, international Hispanic, and Pacific Island indigenous groups. The Scales appear to have been successfully used across these groups, in part because during scale development the authors of the scales strived to avoid clinical jargon or scale defining language that was dominant culture-centric. However, this study specifically examines the NCFAS family functioning ratings assigned by workers of various ethnicities/cultures, on the same case scenario, with that case being based on an indigenous Alaska Native family.

The research questions addressed included:

- Do workers representing different demographics vary in the way they rate the family in the case scenario at the time of Intake?
- Do workers representing different demographics vary in the way they rate the families progress (or lack of progress) of the family in the case scenario at the time of case Closure?
- Do the domains retain their reliability (as measured by Cronbach's alpha) when this varied group of workers is examined?
- Do the scales exhibit the psychometric property of inter-rater reliability?

The analytic strategy included testing the groups from the four volunteer sites to see if there are meaningful differences between them on the demographic variables, and whether they behave differently when rating the scenarios. If there is good group equivalency, then there will be strength in the comparability of data from the groups. And, if the groups rate the scenario similarly (i.e., no significant differences on scale ratings) using the case study based on input of the Alaskan social workers and based on an indigenous Alaska Native family, this will be interpreted as additional evidence that the Scales are, in fact, applicable across racial and cultural groups.

The design of the study called for a group of Alaskan social workers, including Alaskan Native social workers, and the National Family Preservation Network to construct a two-part family scenario, based on the cultural and practice wisdom of the Alaskan workers. The two parts of the case scenario represented the pre-service and post-service situation of an Alaska Native family involved with the Alaskan child welfare system, in accordance with traditional use of the NCFAS Scales, which is at the stage of intake (pre-service) and again at closure (post-service).

This construction of the design is based on the presumption (and intention in actual practice) that the results of the intake ratings from the NCFAS Scales are used to identify areas of family functioning in which the family is either exhibiting strengths of varying degrees (mild or clear strengths), functioning adequately but not in the strengths range, or experiencing problems of varying degrees (mild, moderate or serious). The intake ratings are used to develop a service plan designed to mobilize family strengths (protective factors) and mitigate family problems (risk factors) so that risk is reduced and protective factors prevail. The closure ratings provide a pre-post comparison of the family's level of functioning on the scale domains, reflecting the impact of the service plan on overall family functioning.

Implementation of the design and collection of data required these steps to be followed by all volunteer social workers from the four participating programs:

- Log onto a website and provide demographic information (with confidentiality assured),
- Read the pre-service family case scenario,
- Complete the intake ratings on the NCFAS-R, or NCFAS-G+R (depending on which NCFAS Scale the volunteers' home agency was using at the time),

- Read the post-service case scenario, and
- Complete the closure ratings on the NCFAS-R or NCFAS-G+R.

It should be noted that the NCFAS-R and the NCFAS-G+R share 7 domains in common. These include Environment, Parental Capabilities, Family Interactions, Family Safety, Child Well-Being, Caregiver/Child Ambivalence, and Readiness for Reunification. The NCFAS-G+R has three additional domains (it is a later version and development, compared to the NCFAS-R), including: Social and Community Life, Self-Sufficiency, and Family Health. The responses from the two groups using the NCFAS-R were compared to the responses from the two groups using the NCFAS-G+R, and no differences were observed on the domain ratings. (These results are discussed further in the Discussion Section, and detailed statistics are presented in Appendix A.) For remaining analyses, the domains that are shared by the two scales are combined thereby increasing the sample size on the common domains, and the subset of domains unique to the NCFAS-G+R is subjected to the same analyses but the results are based on smaller sample sizes.

Alaskan Case Scenario

The case scenario used in this study was developed by a panel of Native and non-Native social workers at the Cook Inlet Tribal Council Child and Family Services Department in Anchorage, AK. The Director of the participating Alaskan child welfare program and eight social workers collaborated with the Executive Director of the National Family Preservation Network (under whose auspices the study was conducted) to develop the two part scenario. Among the social workers who contributed to this process, four were Alaska Natives representing four different tribes, one was bi-racial (White/Alaska Native), and four were White. The Executive Director of NFPN (who is White) drafted the 2-part scenario after receiving input from the Alaska group. Some members of the Alaskan group subsequently reviewed the scenario after it was written and provided feedback, and some suggested revisions were incorporated into the scenario. The goal for this process was to develop a case scenario that was deemed to be representative of a typical family served by the Cook Inlet Tribal Council Child and Family Services Department. The Cook Inlet Tribal program uses the NCFAS-G+R version of the NCFAS Scales. Some social workers from the Cook Inlet program participated in the study as volunteer participants.

Case scenarios are frequently used for training and research purposes. The scenario used in this study contains a great deal of information related to the domains and subscales comprising the NCFAS domains of family functioning. However, in order to provide sufficient information for all subscales of all domains to be confidently rated by trainees (or, in this case, volunteer social worker research participants) the scenario would have required many pages to present and would have become far too cumbersome and time consuming to sustain a purely voluntary group of subjects. Therefore, many volunteers did not feel they had sufficient information to assign ratings to specific subscales and selected “Unknown” or “Not Applicable” rating options for one or more subscales. However, in all but a very few cases, the volunteers did assign ratings to the domains, even if one or two of the domains subscales were not rated. This situation is the result of the trade-offs made to recruit volunteers and conduct the study, as compared to a more real-world experience in which workers would be encouraged to pursue additional information from a variety of sources in order to obtain sufficient information to confidently assign ratings to all subscales prior to assigning the overarching domain ratings.

A byproduct of this situation is that when conducting the Cronbach’s alpha reliability analyses, any volunteer’s information that includes “Unknown” or “Does Not Apply” responses to any subscale caused their data to be eliminated from the Cronbach’s alpha computations relating to that particular domain. This happened with sufficient frequency during the Intake portion of the study that some of the computations of Alphas suffered from small sample sizes. That said, the resulting Alphas for the domain ratings are quite respectable, and are in line with multiple large-sample studies conducted previously. Furthermore, during the Closure rating part of the study, workers had much more information about the family (having read both the intake and closure portions of the case scenario). As a result, many more volunteers assigned ratings to all of the subscales as well as the domains, and the sample size for the closure rating reliability statistics increased substantially. The reliability statistics are discussed in greater detail in the Results section.

Volunteer Participants

Volunteers were recruited from child welfare programs in three states and one international country. The states included Alaska (the state of origin of the impetus for the study), from which

13 volunteers were recruited; New Jersey, from which 16 volunteers were recruited; New South Wales, Australia, from which 11 volunteers were recruited; and Washington State, from which 14 volunteers were recruited.

The volunteers represent a broad range of demographics (including White and non-White workers at each site), but all had experience using the NCFAS Scales; and although experience varied among workers, none was a neophyte with respect to child protection social work.

The demographic information provided by each volunteer included:

- Race,
- Gender,
- Age,
- Education,
- State or Country (where practicing), and
- Years of Experience (in their present job)

Across all groups, the combined sample was 84% female, and 16% male.

Ages of volunteers were collapsed into 5 groups: 18 years to 29 years (18%); 30 years to 39 years (27%); 40 years to 49 years (20%), 50 years to 64 years (29%) and 65+ years (6%).

The total sample was 67% White, and 33% non-White.

The 4 sites were fairly evenly represented: Alaska (24% of sample); New Jersey (29% of sample); Washington State (27% of sample); and New South Wales, Australia (20% of sample).

The total sample represents a well-educated group: 9% had at least some college, 42% were college graduates, and 49% had a post graduate degree.

The total sample was also experienced in their current position: only 16% had less than 1 year of experience, 22% had 1 to 3 years of experience, 27% had 4 to 6 years, 11% had 7 to 9 years, and 27% had 10 or more years of experience.

The sample was also reasonably well experienced using the NCFAS Scales, with only 18% having less than 1 year of experience with the NCFAS, 20% having 1 to 2 years of experience, 24% having 4 to 6 years of experience, and 35% having 5 or more years of experience.

Results

Volunteer Group Comparability

Because the study specifically focuses on possible differences in family assessment ratings as a function of race, it was important to determine if the groups were comparable with respect to racial makeup, defined as White or non-White, where non-white could include African Americans, Alaskan Natives, Hispanics, African Australians, or Australian Aboriginal workers. A Chi² analysis of the categorical race data determined that there were no differences among the groups with respect to the distribution of White and non-White volunteer social workers (Chi² = 3.833, df = 3, p = .280).

Although race was a variable of primary importance, group equivalency on all demographic variables is desirable, so the demographic composition of groups was tested on all remaining variables using Chi² analyses of each demographic variable. There were no differences among groups with respect to gender make-up of the groups (Chi² = 1.742, df = 3, p = .63). There were no differences among groups with respect to the distribution of workers of various ages (Chi² = 17.388, df = 12, p = .136). There were no differences among groups with respect to the number of years of experience held by group members (Chi² = 16.465, df = 12, p = .171).

There was a slight trend for workers in Washington State and New Jersey to hold post graduate degrees (Chi² = 15.631, df = 6, p < .05); and there was a slight trend among social workers in Washington State and Alaska to have more experience using the NCFAS scales (manifested in the highest category of 5+ years of experience) (Chi² = 17.797, df = 9, p < .05). However, it is apparent from subsequent analyses that neither of these small differences in group composition affected ratings on the NCFAS scales in relation to the Alaska Native family case scenario.

Instrument Equivalency

The two versions of the NCFAS scales used in this study vary slightly. The NCFAS-R (Reunification) and the NCFAS-G+R (General Services and Reunification) share seven domains in common. The NCFAS-G+R has 3 additional domains. Some of the scenario content related to

the 3 domains on the NCFAS-G+R that two of the program sites were not using. Since this information might possibly have affected the way workers using the NCFAS-R rated the commonly held domains, the mean domain rating for each domain on the two instruments were tested for equivalency both at intake and at closure, using t-tests of group means. There were no differences between the domain ratings at either point in time (intake/closure) on any domain. The t-statistic values ranged from -0.960 to 1.111 (dfs from 35 to 53), and group mean domain rating differences ranging from -.127 to 0.329, on the 6-point rating scale used on the NCFASes. The p-values ranged from .921 to .109, indicating very small, insignificant differences. (See Appendix B for a complete table of t-test results on these analyses.)

As a result of these findings, the data from all groups were combined for testing the primary research questions, with the sample sizes varying slightly within analyses depending on whether the 3 unique NCFAS-G+R domains were present in the analysis.

Instrument Reliability

As mentioned in the Introduction, the case scenario, while rich in content, was not complete with respect to every subscale. The result of having insufficient information in actual case practice is to pursue additional information so that ratings on all subscales can be made confidently by social workers conducting the assessments. In actual practice, response selections of Unknown or Does Not Apply are discouraged (although there are legitimate reasons for selecting them from time to time). However, the unfortunate result of having insufficient information in a training setting or a research setting using a case scenario is that there are no additional sources of information to pursue. As a result, a number of volunteers selected “U/K” or “N/A” when entering their responses to some subscales into the automated database used in the study. The analytic algorithms for computing Cronbach’s alpha require complete information on all subscales, or that volunteer’s data are excluded from that particular analysis.

On the intake portion of the case scenario there were five instances where fewer than 8 subjects completed all subscales on the NCFAS-G+R and one instance where this occurred on the NCFAS-R. On both Scales, the domain of Family Safety was affected, and on the NCFAS-G+R, the domains of Parental Capacities, Social/Community Life, Ambivalence and Readiness for Reunification were also affected. On remaining domains (where sufficient Ns were available

on all subscales) the Cronbach's alpha statistics ranged from .580 to .962, which are acceptable-to-high.

On the closure portion of the case scenario workers had information from both portions of the scenario and were evidently much more confident to assign ratings across most subscales. Only one domain (Family Safety, in both cases) was affected. And the Cronbach's alphas ranged from .866 to .968. These are high to very high, and more in keeping with the larger-sample reliability studies previously conducted. (A complete table of Cronbach's alpha statistics and their contributing Ns is presented in Appendix B.)

Upon reviewing the case scenario, it is evident that the information needed to address certain elements of Family Safety was, indeed, lacking, so workers' reluctance to assign some subscale ratings is understandable.

Influences of Demographics on the Scenario Rating Process

It has already been stated that there were no differences among groups with respect to demographics except for two slight trends: a few more social workers in Washington State and New Jersey hold post graduate degrees than do the social workers in other sites, a few social workers in Washington State and Alaska have 5 or more years of experience using the NCFAS Scales than do social workers in other sites. However, the influence, if any, of demographic variable on actual assignment of ratings on the NCFAS, in relation to the Alaskan Native scenario, must also be examined if the research questions relating to the race/culture of the assessed family in the scenario are to be isolated. In order to accomplish this examination, a series of Chi² analyses was performed on each demographic variable, cross-tabulated with each domain rating on both the NCFAS-R and the NCFAS-G+R, on both intake ratings and closure ratings. Thus, each demographic variable was subjected to 20 cross-tabulations (10 domain ratings at intake, and 10 domain ratings at closure, for a total of 120 cross-tabulations). (A complete set of analysis results and descriptions of the cross-tabulations can be found in Appendices C1 through C6.)

The results of these analyses are strikingly devoid of significant effects, which indicate that demographics, per se, have virtually no influence on individual volunteers' ratings. The very few

(8 out of 120 comparisons) that were statistically significant represent very small effects, and three of the eight disappear from intake to closure, likely the result of additional information being interpreted more in line with the larger group's thinking on the measure. Results of the individual analyses follow.

There was no effect of gender of the social workers (male versus female) on NCFAS domain ratings at intake or closure. (See Appendix C1.) This suggests that the gender of the social worker does not inordinately affect his or her interpretation of the substance of the case or the facts presented therein, as compared to social workers of the other gender, when using the NCFAS Scales to guide the assessment.

There was no effect of the volunteers' affiliation (i.e., the state or country where the participating program was located) on NCFAS domain ratings at intake or closure. This suggests that social workers from different states, even from different countries (using the English language versions of the NCFAS Scales), interpret the substance of the case or the facts presented therein in similarly, when using the NCFAS Scales to guide the assessment.

There was no effect of social workers' experience (tenure in present job) on NCFAS domain ratings at intake. There was a slight trend at closure for workers with more experience in their present job to be less willing to rate the family in the scenario to be Ready for Reunification ($\text{Chi}^2 = 28.512$, $\text{df} = 16$, $p < .05$). However, the grand group mean for reunification at closure was 2.65 (where 2 = Mild Strength and 3 = Baseline), and the trend was slight, such that those few very experienced social workers less willing to rate Readiness for Reunification as a strength were likely rating it as baseline, or perhaps mild problem. This suggests that by and large duration of work experience (assuming basic competence in the worker's position) has little if any practical impact on their ratings, with the possible exception of being slightly more conservative with respect to recommending reunification when using the NCFAS Scales to guide the assessment. (See Appendix C3.)

There was no effect of social workers' education (some college, college degree, post-graduate degree) on NCFAS domain ratings at intake. There was a slight trend at closure for workers with more education to rate the family in the scenario as more problematic with respect to Parental Capabilities ($\text{Chi}^2 = 15.638$, $\text{df} = 8$, $p = .05$). [Note that the "p" value is equal to .05,

not less than .05, so it could be argued that this trend is not significant. However, it is on the cusp of significant and deserves notice.] However, the grand group mean for Parental Capabilities at closure falls between baseline and mild problem ($Mean_G = 3.25$, where 3 = Baseline and 4 = Mild Problem), and the trend was very slight. Even assuming that this marginal finding is significant, the practical impact is negligible, in that only the most highly educated of workers are even slightly more conservative with respect to assessing parental capabilities when using the NCFAS Scales to guide the assessment. (See Appendix C4.)

Findings relating to the age of the volunteer social workers are interesting. There were three modest trends observed at intake such that the oldest workers (50 years of age and older) were more likely to rate Family Safety as slightly more problematic than younger workers ($Chi^2 = 29.534$, $df = 16$, $p < .05$); for that same group of workers to rate Child Well-Being as more problematic than younger workers ($Chi^2 = 32.712$, $df = 12$, $p < .05$); and for that same group to rate Social/Community Life as more problematic than younger workers ($Chi^2 = 45.517$, $df = 12$, $p < .05$). Each of these domains was rated (using the grand mean as the statistic) in the mild to moderate problem range at intake, and each was rated in the baseline to mild strength range at closure. And, at the time of closure, two of the three trends had disappeared, leaving only Child Well-Being at closure where the oldest workers remained slightly less willing to rate Child well-Being as baseline or in the strength range ($Chi^2 = 26.965$, $df = 16$, $p < .05$). One can only speculate as to the reasons these differences, as small as they are, occurred. Given the small spread of ratings about the grand means (discussion to follow) on virtually all of the domains, the practical impact of these findings is probably negligible. (See Appendix C5.)

Finally, the variable of race of the social workers was examined, this being the demographic variable of most interest with respect to this study. In fact, there were no differences on the way White and non-White social workers rated the family in the Alaska Native family scenario, with one exception: non-White workers were very slightly more inclined to rate the Environment as being more problematic than were White social workers. This finding was observed both at intake ($Chi^2 = 8.807$, $df = 3$, $p < .05$), and at closure ($Chi^2 = 11.586$, $df = 5$, $p < .05$). As with other demographic trends, this one is slight. The grand mean for Environment at intake was 5.13, falling between moderate and serious problem. This trend suggests that non-White social workers were slightly more likely to rate Environment as a serious problem in the scenario

(assigning it a “6”), compared to White workers, who were slightly more likely to rate Environment as a moderate problem at intake (Assigning it a “5”). Similarly, at closure, the grand mean was 3.46, about equidistant between baseline and mild problem; non-White workers were slightly more likely to rate Environment as a mild problem (assigning it a “4”), compared to White workers who were slightly more likely to rate it as baseline (assigning it a “3”). This finding suggests that non-White social workers may be very slightly more critical of Environmental issues, or see them as more serious, than their White fellow social workers, when using the NCFAS Scales to guide the assessment. (See Appendix C6.)

It should be noted that the criterion level of alpha for these comparisons was set at .05 for this series of analyses. That is a very liberal level for alpha. In fact, within a set of 120 comparisons and alpha set at .05, one could expect up to 6 false-positive findings (apparent trends due solely to random chance). In this case, there were 8 such findings, and only two transcended intake and closure ratings, suggesting that the other 6 were likely due to chance. Had a more conservative alpha been set, a priori (say, .001, or even .01) none of the apparent trends would have risen to the level of statistical significance. However, in the interest of closely examining these variables and any possible influence they might have on ratings by these varied social workers when assessing an Alaskan Native family, a very liberal level was chosen. This analytic strategy, coupled with the very small magnitudes of these trends, suggests that none of them has any practical impact on the rating process or the overall assessment of the family in the case scenario.

Group Mean Ratings and Inter-Rater Reliability

The most important findings from this study are those relating to how all of the volunteer social workers compared to the “case standard” established for this family by those who developed the case scenario, and how similar they were to one another when assigning their ratings. These findings are presented in Table 1, on the following page.

The table presents a very positive set of findings with respect to the initial research questions. The case standard ratings (i.e., the ratings that the developers of the scenario expected to elicit from the volunteer social workers during their processing of the information in the scenario and their assignment of ratings) can be found in the column labeled Case Rating Standard. The group

grand mean responses of all volunteer social workers from all participating programs (across all demographic variables) appear in the adjacent column, labeled Volunteer Group Grand Means.

Examination of these means in relation to the case standard show that at intake, the group mean is within one-half of a scale point from the standard on six of the domains. This is true for Parental Capabilities, Family Interactions, Family Safety (even though the case was perhaps weakest on this domain), Self-Sufficiency, Family Health, and Ambivalence. On three of the remaining four, the group grand mean is within one scale point of the standard (Child Well-Being, Social/Community Life, and Readiness for Reunification). On only one domain, Environment, did the group mean (5.13) exceed this distance from the standard (4). Recall that during the analyses of demographic trends, there was a slight tendency for non-White workers to rate environment as more problematic than White workers given the same information.

Table 1. Pre-Service and post-service group mean domain ratings (compared to case scenario author’s standard) and volunteers’ bracketing of group mean ratings.

Intake Domains	Case Rating Standard	Group Mean	Group SD	Group % Above/Below Mean*
Environment	4 (Mild Problem)	5.13	0.795	82%
Parental Capabilities	5 (Moderate Problem)	5.07	0.790	77%
Family Interactions	4 (Mild Problem)	4.29	1.126	62%
Family Safety	5 (Moderate Problem)	4.63	1.121	54%
Child Well-Being	6 (Serious Problem)	5.13	0.818	82%
Social/Community Life	3 (Baseline/Adequate)	4.00	1.155	89%
Self-Sufficiency	5 (Moderate Problem)	4.56	0.917	72%
Family Health	5 (Moderate Problem)	4.68	0.988	64%
Ambivalence	5 (Moderate Problem)	4.55	0.879	75%
Readiness/Reunification	4 (Mild problem)	4.57	1.015	62%

Closure Domains	Case Rating Standard	Group Mean	Group SD	Group % Above/Below Mean*
Environment	3 (Baseline/Adequate)	3.46	1.094	76%
Parental Capabilities	3 (Baseline/Adequate)	3.25	0.821	80%
Family Interactions	3 (Baseline/Adequate)	2.88	0.922	81%
Family Safety	3 (Baseline/Adequate)	2.80	0.810	80%
Child Well-Being	3 (Baseline/Adequate)	3.04	0.726	83%
Social/Community Life	3 (Baseline/Adequate)	2.72	1.100	72%
Self-Sufficiency	4 (Mild Problem)	2.96	0.790	72%
Family Health	3 (Baseline/Adequate)	3.12	0.971	76%

Cultural Competence and Inter-Rater Reliability of the NCFAS Scales

Closure Domains	Case Rating Standard	Group Mean	Group SD	Group % Above/Below Mean*
Ambivalence	3 (Baseline/Adequate)	2.91	0.741	89%
Readiness/Reunification	3 (Baseline/Adequate)	2.65	0.883	81%

* This statistic represents the combined proportions of raters whose rating bracketed the group mean (i.e. if a group mean = 3.41, the proportions representing ratings of 3 and 4 were summed for the table entry).

This may or may not be a true finding, given both the very slight demographic trend and the very slight excursion of the group grand mean on this variable, but within this study and within this group of workers, it may explain the slight variance.

Interestingly, all four of those more extreme mean ratings (i.e., those greater than one-half scale increment difference) disappear at closure, after the volunteers have read the second part of the family Scenario and have additional information. In fact, nine out of ten group means are within one-half scale point of the standard, and six of these are within one quarter scale point. Those within one-half point include Environment, Parental Capabilities, Family Interactions, Family Safety, Child Well-Being, Social/Community Life, Family Health, Ambivalence, and Readiness for Reunification. Only Self-Sufficiency exceeded one scale point difference between the volunteers' group mean and the standard: the standard is 4 (Mild Problem), and the group was slightly more willing to rate the domain as being a 3 (Baseline), with a group mean of 2.96.

Taken as a whole, these results indicate very good “standard to group means” alignment at intake, and excellent alignment at closure. It is not surprising that the closure ratings are in closer alignment due to the additional information and evidence of progress (or lack of it) achieved by the family between intake and closure.

NCFAS Scale users, or potential users, sometimes ask whether the NCFAS Scales exhibit good inter-rater reliability. Not only is this a legitimate question based solely on the heuristic value of this psychometric property, but it is often asked in response to the intended use of the Scales, and the underlying reasons for their development. The NCFAS Scales were not developed to be diagnostic instruments, or instruments that yield an index of family functioning. Those accustomed to receiving “indices of risk” or “indices of mental impairment,” for example, come to understand that the developers of the NCFAS Scales envisioned the Scales as providing an organizing framework for a comprehensive family assessment across multiple domains, where

community or practice standards may vary from place to place (and therefore potentially influence ratings), and an instrument where social workers are encouraged to exercise their judgment, apply their practice wisdom, and in other ways to “make the Scale their own.” Historically, we have talked about this intention and scale property to be “intra-rater reliability.” To be sure, the 7 to 10 domain ratings from intake are intended to be used to design a service plan, and the same domains, when rated at closure, provide an indication of the progress made (or not made) by the family as a result of the service plan. But the Scales provide latitude for workers that many highly structured instruments do not. So, if the same worker completes both intake and closure ratings, “intra-rater reliability” is maintained, as the same rating strategies are highly likely to be applied at the time of each assessment.

However, agency administrators and academic researchers legitimately inquire about inter-rater reliability, and it is a desirable property of assessment instruments. This study provides an excellent opportunity to examine the volunteers’ rating behavior as a group, acknowledging that the group represents multiple agencies (even international agencies) and that the volunteers are quite diverse with respect to the 6 demographic variables studied, and who have provided the authors with information from a single, two-part scenario of an Alaskan Native family. Examining the right-most column in Table 1 provides valuable information about the similarity of ratings and consistent rating behavior of all volunteers, regardless of their individual constellation of gender, age, experience, race, education and practice location. The numbers in the column represent the proportion of volunteer workers whose assigned ratings fall within one scale increment above or below the grand mean of the group. For example, if a group mean = 3.41, the proportions representing ratings of 3 and 4 were summed for the table entry. Of course, individual volunteers (or any other Scale user) cannot assign a scale rating that falls between the interval ratings provided on the instrument, itself. A worker cannot assign, for example, a 3.5, if they feel that a situation falls between baseline (3) and mild problem (4). The user must decide whether, all things considered, the situation warrants a 3 or a 4. So, even with due deference to encouraging workers to exercise their own judgment, the “bracketing” statistic represented by the data in the table column under discussion is a good proxy for the consistency with which the volunteer workers rate the family within one increment of the group mean. The more workers that do so, the better is the inter-rater reliability.

At intake, between 54% and 89% of volunteer social workers assigned ratings that were within one increment of the mean. [The 54% is associated with Family Safety, the domain judged to be weakest in the scenario; if it is eliminated, the statistics improve to 62% to 89%.] Bear in mind that the nature of the use of case scenarios in training and research frequently results in less than adequate information on some scale domains, so these findings are considered to be quite positive.

The inter-rater consistency (reliability) increases considerably at closure, after more information is available to the volunteers. On all ten domains, the “bracketing statistic” ranges from 72% to 89%. These are quite high, and compare favorably with many more highly standardized instruments that discourage individual exercise of judgment by workers (or technicians) using the other instruments.

Discussion

This study was conducted to examine the capability of the NCFAS Scales to apply directly to an Alaska Native family case (as an examination of the cultural competence of the Scales), and to examine whether workers representing varying demographic and geographic locations (practice settings) would rate the family similarly on the domains of family functioning comprising the NCFAS Scales. We found no significant differences (of a practical nature) in the way that volunteer workers in this study assessed the family at intake and closure. The very few differences found on the relationship of various demographics and one or two domain ratings are very small, and as likely due to chance as to substance. These findings contribute to the growing body of evidence that the NCFAS Scales are applicable across various racial, cultural and ethnically identified groups, and that the purpose of the NCFAS Scales (to provide an organizing framework to conduct a comprehensive family assessment) contributes to that property of racial/cultural/ethnic relevance and applicability.

We found that when minimum sample sizes were obtained (a difficulty brought about by analytic algorithms as much as by the volunteers’ reticence to assign ratings without inadequate information) the NCFAS Scales retain their reliability as measured by Cronbach’s alpha.

We also found that workers of different races, ages, experience, gender, education and practice location assigned very similar ratings on the NCFAS domains, suggesting that the Scales exhibit good inter-rater reliability.

Appendix A

Values of between-groups t- test statistic and group mean differences on domain ratings of 7 domains held in common by the NCFAS-R and NCFAS-G+R

Intake Domains	t value	df	p (2-tailed)	Group Mean Difference
Environment	-.960	53	.342	.215
Parental Capabilities	.278	53	.782	.216
Family Interactions	-.251	50	.290	.301
Family Safety	-.100	50	.921	.316
Child Well-Being	-.932	53	.355	.222
Ambivalence	-.294	49	.770	.251
Readiness/Reunification	-1.64	35	.109	.329
Closure Domains	t value	df	p (2-tailed)	Group Mean Difference
Environment	.276	52	.784	.083
Parental Capabilities	1.111	53	.271	.274
Family Interactions	.633	50	.529	.163
Family Safety	.711	52	.481	.158
Child Well-Being	1.102	52	.275	.218
Ambivalence	-.063	51	.950	-.031
Readiness/Reunification	-.510	50	.612	-.127

**Levene's test of equivalency of variances produced no significant results, so t-tests are conducted on the basis of the assumption of equal variances.*

Appendix B

Reliability of NCFAS-G+R Domain ratings at intake and closure (as measured by internal consistency; expressed as Cronbach’s alpha).

Domains Ratings NCFAS-G+R (Total N = 25)	N Contributing at Intake*	Cronbach’s Alpha at Intake	N Contributing at Closure*	Cronbach’s Alpha at Closure
Environment	24	.929	19	.866
Parental Capabilities	—	—	9	.946
Family Interactions	10	.881	16	.928
Family Safety	—	—	—	—
Child Well-Being	16	.824	19	.894
Social/Community Life	—	—	8	.939
Self-Sufficiency	22	.761	20	.866
Family Health	10	.580	11	.907
Ambivalence	—	—	14	.878
Readiness/Reunification	—	—	9	.913
Domain Ratings NCFAS-R (Total N = 30)	*N Contributing at Intake	Cronbach’s Alpha at Intake	*N Contributing at Closure	Cronbach’s Alpha at Closure
Environment	24	.837	21	.943
Parental Capabilities	13	.714	16	.951
Family Interactions	11	.782	17	.877
Family Safety	—	—	—	—
Child Well-Being	19	.709	23	.923
Ambivalence	16	.780	21	.891
Readiness/Reunification	9	.962	16	.968

**During the study, total of 25 social workers from Alaska and Australia were using the NCFAS-G+R, and a total of 30 social workers from New Jersey and Washington State were using the NCFAS-G).*

— Empty cells result from Ns being 7 or fewer, and not considered sufficient to satisfy the assumptions of the Cronbach’s alpha analysis. Populated cells are considered to have sufficient Ns for reliable computation.

Appendix C1

Demographics: Gender of volunteer social workers cross-tabulated with intake and closure domain ratings on the NCFAS Scales (Male, Female).

Intake Domains	Chi² Value	DF*	P Value	Interpretation
Environment	0.802	3	.85	Not significant
Parental Capabilities	0.199	3	.98	Not significant
Family Interactions	6.325	5	.28	Not significant
Family Safety	1.325	4	.86	Not significant
Child Well-Being	2.194	3	.53	Not significant
Social/Community Life	2.265	4	.69	Not significant
Self-Sufficiency	1.850	3	.60	Not significant
Family Health	2.369	3	.50	Not significant
Ambivalence	1.536	3	.67	Not significant
Readiness/Reunification	4.380	3	.22	Not significant
Closure Domains	Chi² Value	DF*	P Value	Interpretation
Environment	2.772	5	.74	Not significant
Parental Capabilities	2.746	4	.60	Not significant
Family Interactions	2.100	5	.84	Not significant
Family Safety	2.516	4	.64	Not significant
Child Well-Being	1.220	4	.88	Not significant
Social/Community Life	1.881	4	.76	Not significant
Self-Sufficiency	3.144	3	.37	Not significant
Family Health	3.740	4	.44	Not significant
Ambivalence	1.457	3	.69	Not significant
Readiness/Reunification	2.758	4	.60	Not significant

*Degrees of Freedom

Appendix C2

Demographics: State or country of residence of volunteer social workers cross-tabulated with intake and closure domain ratings on the NCFAS Scales (Alaska; New Jersey; Washington State; New South Wales, Australia)

Intake Domains	Chi² Value	DF*	P Value	Interpretation
Environment	15.719	12	.20	Not significant
Parental Capabilities	8.802	12	.72	Not significant
Family Interactions	21.306	20	.38	Not significant
Family Safety	13.735	16	.62	Not significant
Child Well-Being	13.069	12	.36	Not significant
Social/Community Life	3.594	8	.89	Not significant
Self-Sufficiency	7.265	6	.30	Not significant
Family Health	3.715	6	.72	Not significant
Ambivalence	8.700	12	.73	Not significant
Readiness/Reunification	17.946	12	.12	Not significant
Closure Domains	Chi² Value	DF*	P Value	Interpretation
Environment	11.838	20	.92	Not significant
Parental Capabilities	13.186	16	.66	Not significant
Family Interactions	15.129	20	.77	Not significant
Family Safety	11.620	16	.77	Not significant
Child Well-Being	17.467	16	.36	Not significant
Social/Community Life	8.591	8	.38	Not significant
Self-Sufficiency	9.580	6	.14	Not significant
Family Health	3.007	8	.93	Not significant
Ambivalence	19.370	12	.08	Not significant
Readiness/Reunification	8.454	16	.93	Not significant

*Degrees of Freedom

Appendix C3

Demographics: Job experience of volunteer social workers cross-tabulated with intake and closure domain ratings on the NCFAS Scales (Years in current job collapsed as follows: <1, 1–3, 4–6, 7–9, 10+).

Intake Domains	Chi² Value	DF*	P Value	Interpretation
Environment	8.460	12	.75	Not significant
Parental Capabilities	11.019	12	.53	Not significant
Family Interactions	21.322	20	.38	Not significant
Family Safety	22.853	16	.12	Not significant
Child Well-Being	8.065	12	.78	Not significant
Social/Community Life	23.623	16	.10	Not significant
Self-Sufficiency	12.940	12	.37	Not significant
Family Health	16.823	12	.16	Not significant
Ambivalence	15.511	12	.22	Not significant
Readiness/Reunification	8.654	12	.73	Not significant
Closure Domains	Chi² Value	DF*	P Value	Interpretation
Environment	21.451	20	.37	Not significant
Parental Capabilities	15.522	16	.49	Not significant
Family Interactions	27.915	20	.11	Not significant
Family Safety	20.963	16	.18	Not significant
Child Well-Being	19.155	16	.26	Not significant
Social/Community Life	16.280	16	.43	Not significant
Self-Sufficiency	11.370	12	.50	Not significant
Family Health	16.458	16	.42	Not significant
Ambivalence	17.397	12	.14	Not significant
Readiness/Reunification	28.512	16	P < .05	Slight trend for more experienced workers to rate Readiness for Reunification as more problematic

*Degrees of Freedom

Appendix C4

Demographics: Education of volunteer social workers cross-tabulated with intake and closure domain ratings on the NCFAS Scales (Some College or Less / College Degree / Post Grad Degree).

Intake Domains	Chi² Value	DF*	P Value	Interpretation
Environment	8.798	6	.19	Not significant
Parental Capabilities	6.424	6	.38	Not significant
Family Interactions	10.173	10	.43	Not significant
Family Safety	9.076	8	.34	Not significant
Child Well-Being	8.206	6	.22	Not significant
Social/Community Life	4.601	8	.80	Not significant
Self-Sufficiency	2.902	6	.82	Not significant
Family Health	5.060	6	.54	Not significant
Ambivalence	3.507	6	.74	Not significant
Readiness/Reunification	9.109	6	.17	Not significant
Closure Domains	Chi² Value	DF*	P Value	Interpretation
Environment	10.038	10	.44	Not significant
Parental Capabilities	15.638	8	p = .05	Slight trend for more educated to rate Parent. Capab. as more problematic
Family Interactions	13.899	10	.18	Not significant
Family Safety	14.231	8	.08	Not significant
Child Well-Being	4.594	8	.80	Not significant
Social/Community Life	8.897	8	.35	Not significant
Self-Sufficiency	11.596	6	.07	Not significant
Family Health	14.298	8	.07	Not significant
Ambivalence	6.663	6	.35	Not significant
Readiness/Reunification	7.567	8	.48	Not significant

*Degrees of Freedom

Appendix C5

Demographics: Age of volunteer social workers cross-tabulated with intake and closure domain ratings on the NCFAS Scales (ages collapsed as follows: 18–29, 30–39, 40–49, 50–64, 65+).

Intake Domains	Chi² Value	DF*	P Value	Interpretation
Environment	17.942	12	.12	Not significant
Parental Capabilities	11.915	12	.45	Not significant
Family Interactions	13.433	20	.86	Not significant
Family Safety	29.534	16	.02	Slight trend for older workers to rate Safety as more problematic
Child Well-Being	32.712	12	.01	Slight trend for older workers to rate CWB as more problematic
Social/Community Life	45.517	16	.01	Slight trend for older workers to rate S/CL as more problematic
Self-Sufficiency	12.292	12	.42	Not significant
Family Health	11.667	12	.47	Not significant
Ambivalence	11.240	12	.51	Not significant
Readiness/Reunification	12.779	12	.39	Not significant
Closure Domains	Chi² Value	DF*	P Value	Interpretation
Environment	17.492	20	.62	Not significant
Parental Capabilities	13.757	16	.62	Not significant
Family Interactions	27.627	20	.12	Not significant
Family Safety	19.245	16	.26	Not significant
Child Well-Being	26.965	16	.04	Slight trend for older workers to rate CWB as more problematic
Social/Community Life	14.375	16	.57	Not significant
Self-Sufficiency	11.969	12	.45	Not significant
Family Health	22.018	16	.14	Not significant
Ambivalence	7.295	12	.84	Not significant
Readiness/Reunification	16.086	16	.45	Not significant

*Degrees of Freedom

Appendix C6

Demographics: Race of volunteer social workers cross-tabulated with intake and closure domain ratings on the NCFAS Scales (White, and non-White; including Alaskan Native, African American, Aboriginal).

Intake Domains	Chi² Value	DF*	P Value	Interpretation
Environment	8.807	3	p < .05	Slight trend for non-Whites to rate Environment as more problematic
Parental Capabilities	1.126	3	.77	Not significant
Family Interactions	7.898	5	.16	Not significant
Family Safety	3.436	4	.49	Not significant
Child Well-Being	0.679	3	.88	Not significant
Social/Community Life	1.490	4	.83	Not significant
Self-Sufficiency	2.455	3	.48	Not significant
Family Health	3.660	3	.30	Not significant
Ambivalence	3.178	3	.37	Not significant
Readiness/Reunification	2.746	3	.43	Not significant
Closure Domains	Chi² Value	DF*	P Value	Interpretation
Environment	11.586	5	p < .05	Slight trend for non-Whites to rate Environment as more problematic
Parental Capabilities	2.698	4	.61	Not significant
Family Interactions	5.484	5	.36	Not significant
Family Safety	3.712	4	.45	Not significant
Child Well-Being	4.752	4	.31	Not significant
Social/Community Life	1.435	4	.84	Not significant
Self-Sufficiency	0.911	3	.82	Not significant
Family Health	1.500	4	.83	Not significant
Ambivalence	1.776	3	.62	Not significant
Readiness/Reunification	3.155	4	.53	Not significant

*Degrees of Freedom