

Title Page

1. Title of the Article: Reliability of the Colorado Family Support Assessment: A Self-Sufficiency Matrix for Families.
2. Corresponding Author: Melissa K. Richmond, Director of Research and Evaluation, OMNI Institute, 899 Logan St., Suite 600, Denver, CO 80203. Email: mrichmond@omni.org, phone: 303-839-9422; fax: 303-839-9420
3. Co-authors and Affiliations: Fred C. Pampel, University of Colorado, Boulder; Flavia Zarcula, OMNI Institute; Virginia Howey, Family Resource Center Association; and Brenda McChesney, Family Resource Center Association.
4. Key words: self-sufficiency matrix, reliability, family support, case management

Author Note

This study was funded by the Family Resource Center Association. We thank Jerry Endres, MSW, Family Development Matrix, for supplying the California case studies, and Kristy Beachy-Quick, MPP, for her work on early versions of the CFSA 2.0. We also thank the family advocates for their participation in the study and their dedication to supporting Colorado families.

Manuscript accepted for publication in *Research on Social Work Practice (RSWP)*, July 2015, DOI: 10.1177/1049731515596072

Abstract

Purpose: Family support programs commonly use self-sufficiency matrices (SSMs) to measure family outcomes; however, validation research on SSMs is sparse. This study examined the reliability of the Colorado Family Support Assessment 2.0 (CFSA 2.0) to measure family self-reliance across 14 domains (e.g., employment).

Methods: Ten written case studies were developed. Family advocates (n=24) independently coded each case study on each domain of the CFSA 2.0. Intra-class correlation (ICC) coefficients were calculated to determine the level of agreement between participants. Rating accuracy was calculated by subtracting the correct score for each domain and case study from the score given by the raters.

Results: ICCs ranged from .79 to .96. Across all domains and case studies, family advocates were 84.4% accurate. Allowing for minor deviations (off by 1 on the 5-point scale) shows 96% accuracy.

Discussion: A well-constructed self-sufficiency matrix can be used to collect reliable and objective data on family self-sufficiency.

Reliability of the Colorado Family Support Assessment: A Self-Sufficiency Matrix for use in
Family Support Programs

Economic hardship, family stress, poor parenting, and family instability can have detrimental effects on children's development (Benedetti, 2012; Sandstrom & Huerta, 2013). Significant adversity and toxic stress in a child's early years can affect the developing brain and can have lasting implications for the mastery of cognitive, language, and social skills (Shonkoff, 2010). In 2013, 20% of children in the United States lived in poverty (U.S. Census Bureau, 2014), and social service institutions are struggling to cope with the rising levels of poverty and inequality (Hout et al., 2014). Many families are struggling to make ends meet, which can interfere with their ability to provide the supportive, nurturing environments in which children thrive. In low-income families, stressors are often compounded as there are rarely excess resources to draw from in times of crisis (Mills & Amick, 2010). Fortunately, well-implemented prevention programs that have a two-generation approach targeting young children, parents, and families can result in positive outcomes many years later (Manning, Homel, & Smith, 2009). Two-generation approaches build human capital across generations by combining education or job training for adults with early childhood education for their children (Chase-Lansdale & Brooks-Gunn, 2014). Interventions are likely to be most successful when they consider the broader family, neighborhood, cultural, and community contexts (Daro & Dodge, 2009). Thus, community-based interventions that are able to strengthen vulnerable families and protect children from the negative impacts of extreme and prolonged stress have the potential to dramatically improve the wellbeing of children, families, and communities.

As a means to strengthen families and build healthy communities, several states support Family Resource Centers (FRCs) as a model of service provision in which families can access

resources and services in a family-centered manner to address complex issues (Hubble, 2010). The model is designed to reduce fragmentation in social service delivery by providing coordinated, multi-service care. FRCs use a philosophy of strengths-based family development, in which a family advocate acts as a coach and systems navigator to support the family as they set and meet their own goals. There are a core set of principles and premises that guide family support programs, including the premise that empowering families to build on their own strengths supports the healthy development of children (Harper Browne, 2014). Strengthening family assets leads to growth and development, instills confidence in one's own skills, and fosters a sense of empowerment (Fernandez, 2004, 2007). This holistic process contributes to the ultimate goal of creating internal motivation for change (Walker, 2011). While a focus on deficits tends to discourage participants, positivity in relationships with family support workers and a positive outlook helps to maintain high family participation and the effort needed to overcome the inevitable obstacles in making life changes (O'Brien et al., 2012). Furthermore, family development specialists are well positioned to provide customized services that help families make long-term life changes.

As society has shifted towards increased accountability for service organizations, family support programs are being asked to demonstrate their value through outcome-based standards. The use of measures that objectively assess family self-reliance are necessary to support valid and accurate measurement of family outcomes. However, in the field, successful measurement implementation depends on tools that are consistent with a program's underlying philosophy; that are effectively integrated into service delivery; and that do not create unnecessary administrative burden for workers or families. For these reasons, family support organizations have been attracted to self-sufficiency matrices (SSMs) as a means to measure family

functioning and stability. SSMs are designed to create a full picture of circumstances for diverse clientele and they provide dual purpose, serving as both a case management and outcome tool (Snohomish County Self-Sufficiency Taskforce, 2004). In practice, family advocates use the SSM to identify and discuss family assets and areas of need. The information is used to document initial functioning (baseline assessment) and to help guide family goal setting, resource and referral provision, and multi-service delivery. After goal setting, advocates follow-up with families and re-administer the SSM to assess progress towards changes in self-reliance, and further modify, as needed, family goals and the provision of services and supports. However, despite increasing use of SSMs as outcome measures (Culhane et al., 2008), rigorous testing of the use of these tools in this capacity lags behind. The goal of the current study was to address this gap in the literature by testing the reliability of a well-constructed SSM-type family assessment.

Research on Self-Sufficiency Matrices

To our knowledge, only one peer-reviewed article has documented evidence of reliability and validity of a self-sufficiency matrix: The SSM-D, a Dutch-language SSM adapted from the Utah and Arizona SSM versions, was examined for its use with clients receiving mental health services (Fassaert et al., 2013). In this study, researchers compared scores on the SSM-D to the Health of the Nation Outcome Scale (HoNOS) and the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS). Social workers, psychiatric nurses, and case managers completed the SSM-D and different raters (psychologists or research associates) completed the criterion measures (the CANSA or HoNOS) on the same clients. Through a factor analysis of the 11 domains tested, the researchers found that the SSM-D measured a single construct of self-sufficiency. Furthermore, scores on the SSM-D correlated as expected with the HoNOS and the

CANSAS, demonstrating that clients who had higher self-reliance scores on the SSM-D had lower scores on the nationally-recognized measures of 'need for care'. The Dutch study, however, focused on clients receiving intensive mental health services and only examined 11 domains assessing individual-level, rather than family-level, functioning.

In addition to the peer-reviewed Dutch study, there are two publically-available reports examining the reliability of SSMs. Specifically, Parker (2008) prepared a report for the Arizona Department of Economic Security Office of Community Partnerships and Innovative Practices, in which he presented findings of a factor analysis on a version of an SSM used in Arizona. He found a two-factor structure: one labeled economic self-sufficiency and the other labeled social-emotional self-sufficiency. The study also found significant improvements in self-sufficiency over time for families receiving services. However, the study did not test the tool on its ability to achieve consistent ratings across different program case managers (inter-rater reliability).

In another report, in California, Endres and colleagues (1999) tested the reliability of the Family Development Matrix (FDM) model. They developed ten written narratives based on case histories of program participants and had 25 workers score the narratives using the FDM at two points in time. Results identified some domains that were reliably administered and some that required revisions to improve reliability. This study was not submitted for dissemination in peer-review and thus lacked critical methodological information to assess its merits. Nonetheless, it provides an important standard to which assessments can be tested for reliability.

Considering the dearth of scientific studies testing family support assessments, additional research is needed to demonstrate the utility of SSM-type measurement as a reliable and valid method of assessing family outcomes. To accurately measure family change over time, research

needs to demonstrate that SSMs can be administered in a way that obtains accurate data across multiple family workers.

Development of the Colorado Family Support Assessment 2.0

The Colorado Family Support Assessment 2.0 (CFSA 2.0), was developed to provide a strengths-based, objective index of family functioning for use by the Family Resource Center Association (FRCA) member-FRCs. It is based on the concept of a scaled outcome matrix that documents family functioning across several areas that contribute to self-reliance (e.g., income, housing, employment, transportation, debt management). Each domain area is rated on a scale from ‘in-crisis’ to ‘thriving’, using indicators at each level to guide ratings on the scale (Snohomish County Self-Sufficiency Taskforce, 2004; Richmond & Valentin, 2012). Matrices are frequently developed by local groups to reflect community-based indicators of self-sufficiency in each area. To develop the CFSA 2.0, researchers and program staff at the state and local level engaged in a collaborative process.

To begin the development process, researchers reviewed the scant research literature on reliability and validity studies of self-sufficiency matrices; examined the many publically available SSMs to learn how other communities were assessing family outcomes; obtained feedback from FRC staff on the version of the CFSA that was in use prior to the development of the CFSA 2.0; and reviewed national organizations’ materials with deep knowledge of domain areas (e.g., USDA definitions of food security). This information was used to develop a draft matrix that sought to maximize objectivity with what could realistically be obtained during an interview with a family. The draft tool was vetted and revised with input from FRC directors, family advocates and other stakeholders to ensure relevancy and utility in the field.

The collaborative process led to a family assessment that consists of three components: Part A, the domain matrix; Part B, the Protective Factors Survey (FRIENDS National Resource Center, 2011); and Part C, family readiness to change and goal setting. Part A, the domain matrix, includes 14 locally-relevant categories considered critical to family self-reliance (income, employment, housing, transportation, food security, child care, child education, adult education, cash savings, debt management, health coverage, physical health, mental health, and substance abuse). Each domain is rated on a scale from 1 to 5, with higher scores indicating greater self-reliance. For each domain, each indicator is defined for scoring (e.g., for employment, 1 = *all employable adults in the family are not employed*, 2 = *at least one adult in the family has temporary or part-time employment and no adult has full-time employment*, 3 = *at least one adult in the family is employed full-time and no adult has stable employment*, 4 = *at least one adult has full-time stable employment*, and 5 = *at least one adult has full-time stable employment and access to employer-based benefits*). To facilitate accurate scoring, key definitions are provided at the beginning of domains (e.g., for employment, definitions are provided for *employable*, *stable employment*, *benefits*, and *full-time status*). In addition, each domain includes a not enough information (N/I) option for instances in which the family advocate is unable to obtain the required level of information to accurately code the domain. Three domains also include a not applicable (N/A) option: Employment is not applicable when there are no employable adults in the family (e.g., a single parent on disability); child care is not applicable when there are no children under 12 years old or the family is adequately able to care for children and does not need child care (e.g., by choice, one parent stays home to care for the children); and child education is not applicable when all children in the family are not school-

aged or have earned a GED or high school diploma. The N/I category was not tested as a part of this study.

Parenting and social support, although commonly included in SSMs, were not included in the domain matrix because the CFSA 2.0 development committee members believed that these areas would be too challenging to assess objectively through the matrix format. Rather, the PFS (FRIENDS National Resource Center, 2011) was included in the tool in its entirety in Part B, along with a single-item question on relationship safety. The PFS assess five factors that protect against child abuse and neglect (family functioning/resiliency, social support, concrete support, nurturing and attachment, and knowledge of parenting/child development). Because the PFS has demonstrated reliability and validity (Counts et al., 2010), the current study only examined the reliability of the domain matrix, Part A, of the CFSA 2.0.

The Current Study and Hypotheses

The current study was designed to fill a gap in the research literature by testing whether a well-constructed SSM can objectively and reliably measure family self-reliance, which is a necessary condition to support its use as an outcome tool. Current SSMs have not been adequately tested to determine whether two different family workers would score the same family in the same way. To test the objectivity of the tool's construction, we used Endres and colleagues (1999) evaluation as a guide, and designed a study in which family advocates scored written case studies on each domain of Part A of the CFSA 2.0. This approach ensures that each family advocate receives the exact same information when using the tool. High reliability estimates provide evidence that the domain matrix is sufficiently objective to collect consistent information across family advocates. Evidence of high reliability in a controlled environment

will lay the groundwork needed in the future for a more resource-intensive study examining the psychometrics of the tool when used in the field.

We tested the following hypotheses: 1) family advocates will rate case studies similarly to each other on each domain of the CFSA 2.0, and 2) family advocates will be accurate in their ratings, compared to the advisory group who developed the tool. We also explored whether family advocate characteristics predicted rating accuracy.

Method

Study Population and Recruitment

The study population consisted of staff providing family development services in Colorado FRCs. In state statute (Colorado General Assembly, 2009), FRCs are tasked with ‘providing comprehensive, intensive, integrated, and collaborative state and community-based services’. Each of the 24 FRCs that are members of the FRCA provided a list of staff currently providing support services to families. Using this list, researchers sent an email invitation to approximately 120 FRC staff. The invitation provided a brief description of the study, and requested that participants be available at two pre-set times for data collection. Some individuals indicated that they were unavailable during the study window and others indicated that they would not have time with their current work schedules. In the end, 31 individuals expressed interest in participating and 24 consented and completed study protocols.

The 24 participants represented 12 FRCs across urban and rural regions of the state. Table 1 provides descriptive information on participant socio-demographic characteristics. In brief, most participants were female, more than one-third had a Bachelor’s or Master’s degree, about one-fifth were of Hispanic/Latina origin, most were white, and about 15% spoke Spanish as their primary language. Participants were between 30 and 72 years old, with a mean age of 42

(SD =12.5). There was variability in length of employment at the FRC, and length of time working with families. Similarly, participants varied in the number of CFSA's (Version 1.0) that they administered in the month prior to participation. Due to lack of information on non-participants, we were unable to identify the degree to which participants reflect the larger population of family support workers.

Most (71%) participants had received their Strengthening Families Certificate (California Network of Family Strengthening Networks, 2014) and more than two-thirds had their Family Development Credential (Forest, 2003). The Standards of Quality for Family Strengthening & Support Training Certificate prepares social service providers with an understanding of each protective factor that helps build parental resilience and strengthen family support systems, while offering concrete programmatic strategies for implementation. The Family Development Credential is a research- and evidence-based family-support training and credentialing program developed in 1996 at Cornell University and has expanded to 17 states, and Washington D.C. The 90 hour training covers core competency areas for social service providers such as effective communication skills, strengths-based case management, cultural responsiveness, goal setting, home visitation, and facilitated family meetings. Both trainings help to equip family support staff with the tools, techniques, and resources needed to work with families in a strengths-based family-centered approach.

Development of Written Case Studies

To support a rigorous test of the CFSA 2.0 domains, we sought to develop case studies that 1) provided sufficient information for accurate coding of each of the domains, 2) contained situations that were realistic and consistent with families seeking services in Colorado FRCs, and

3) captured sufficient variation so that each case study represents a different combination of problems and assets, and defines a unique case for evaluation by participants.

To achieve these objectives, we first obtained the original ten case studies used in California evaluation (Endres et al., 1999). The ten case studies were developed from authentic case histories of program participants from agencies in California participating in the study, and contained a diverse set of family circumstances. FRCA staff reviewed the 10 case studies to verify that each scenario reflected the situation of a family that would be served by a Colorado FRC. FRCA staff determined that nine of the 10 studies reflected situations similar to those of Colorado families receiving FRC services. One case study was deemed unsuitable for the study because it was a high-need individual who was not caring for children. To find a replacement, one of the advisory team members created a new case study based on a family previously served at a Colorado FRC.

The nine California cases studies were adapted to meet study goals. Specifically, a team of two FRCA staff, two researchers, and one family advocate reviewed and revised case studies using the following process: 1) each case study was organized to mirror the CFSA 2.0 interview progression so that the case studies provided information in the same order it would be obtained during the family interview; 2) California-specific language was replaced with Colorado-specific language (e.g., references to public assistance programs); 3) extraneous information that was not relevant to CFSA 2.0 coding was removed; 4) additional content was developed when key information was missing (e.g., ‘...*Cathy borrowed some money from the older children’s father that she is unable to pay back*’ was created for a case study in which no information on debt management was provided); 5) case study language was reviewed to ensure sufficient separation from CFSA 2.0 descriptors (e.g., one of the employment indicators in the CFSA 2.0 states ‘at

least one adult in the family has temporary or part-time employment'; thus, in one case study, we changed original wording of 'temporary employment' to '*Jerry works construction jobs when they are available, but it can be slow during the winter months*'; and 6) as needed, case studies were revised so that each domain rating was captured at least once in a case study.

After the ten case studies were developed, researchers examined the coding of each case study across each domain to determine whether the set of studies contained sufficient variability for a strong test of each of the domains. First, researchers calculated the percent of responses across all domains and cases with domain codes of 1 (25%), 2 (18%), 3 (22%), 4 (11%), 5 (20%), and N/A (4%). Maximum variability would be 20% for each category (with no N/A ratings). Similarly, an even distribution of scores of 1-5 would yield a mean of 3 and a standard deviation of about 1.5. Looking across case studies and all domains, the mean was 2.85 and the standard deviation was 1.49. Finally, we correlated scores for each of the domains with scores for the other domains across all case studies (N=10). Some domains had high correlations, but we determined that this reflects realistic variability as problems occur in clusters in real life – fully randomizing characteristics within a case study would be unrealistic. In addition, given that high, positive correlations with some domains were balanced by negative correlations in other domains, the case studies avoided simple scenarios in which all domains were scored very high or very low. Based on these checks, we determined that the degree of variability was sufficient weighed against further revisions to the case studies.

As a final test, two family advocates and one FRCA staff member, who were not involved in case study development, reviewed the case studies. One advocate reviewed each case study and rated the degree to which she perceived the case study to be realistic (very, somewhat, not at all), and why. The other two individuals each coded three case studies to test the data collection

protocols. Few changes were needed at this point; some minor edits were made to further refine and improve case study relevance, without adjusting the rating codes. These three individuals were excluded from further participation in the study.

Data Collection

Researchers hosted two web-based data collection sessions, one week apart in September 2014, during which participants independently read and scored 10 case studies (four case studies were coded in the first webinar, six were coded in the second). Participants were asked to complete a demographic survey prior to the first webinar. During the first webinar session, researchers provided a brief training (approximately 30 minutes) on scoring the CFSA 2.0. After the training, participants read and scored four case studies while researchers remained available on the webinar to answer questions. The second webinar session was similar to the first, with the exception of the training, and included coding the six remaining case studies.

Participants received a \$100 gift card at study completion. The study received Exempt status from the University of Colorado, Institutional Review Board.

Analysis

Intra-class correlation (ICC) coefficients were calculated to identify the degree to which family advocates agreed with each other on case study ratings. For the ICC calculations, the data file was structured so that the 24 raters of the 10 case studies were compared separately for each of the 14 domains. ICC values of .60 - .74 are considered good, and values between .75 and 1.0 are considered excellent (Hallgren, 2012). For the three domains that include the N/A option (Employment, Child Care, Child Education), we excluded N/A responses from the ICC calculations because the code of N/A does not fit the 1-5 scale. However, we separately examined the appropriate and inappropriate use of the N/A code.

The advisory committee identified the correct score on each domain rating for each of the 10 case studies. Participant accuracy ratings were calculated by subtracting the correct score for each domain and case study from the score given by the participants. Using this method, the most accurate value is zero – no difference between the family advocate rating and the correct rating. A positive value means the rater scored the case study too high on the domain, whereas a negative value means the rater scored the case too low on the domain. Values of 1 or -1 reflect small error but values of 2 or more or -2 or less reflect more serious deviations from the correct score. Given 14 domains by 10 case studies, each rater provides 140 scores. With 24 raters and 140 scores, the maximum sample size for the accuracy measure equals 3,360. When a case study did not apply to a domain, such as when the correct response for Child Care or Child Education was N/A, we present a separate accuracy analysis.

Finally, to investigate whether family worker characteristics predict accuracy, we first computed the mean error for each participant ($N = 24$). Then, the absolute value of the error served as the outcome variable. We used the absolute value because the direction of the error is not relevant; a score of 0 is best and each higher score reflects greater error. The analysis then used the mean absolute error to calculate bivariate correlations and multiple regression coefficients for the sociodemographic variables (with listwise deletion of missing data on the sociodemographic variables, the N drops to 22).

Results

Table 2 lists the ICC coefficients for each of the 14 domains, in order from highest to lowest. The results show high inter-rater reliabilities that are all in the excellent range. The case study N s are reduced for the three domains in which the N/A code was assigned (for results of N/A analyses, see below). The domains with the lowest reliabilities – although still in the

excellent range – are transportation, adult education, housing, food security, and physical health. Domains with the highest reliabilities include debt management, cash savings, and health coverage.

Accuracy analyses are presented in Table 3. Focusing on scored ratings of 1-5 across the 14 domains, 84.4% of the ratings are correct (error score of 0), and 95.8% are within one category of the correct score (-1 or 1). Only 4% of the ratings err by two or more categories. The errors are symmetric rather than systematic, with positive errors about as common as negative errors. The maximum number of ratings is 3,360, but excluding those assigning an N/A, the N falls to 3,215.

Separate analysis of the use of the N/A code shows similarly high accuracy. The N/A code can be used correctly or incorrectly. It is used correctly when the case study has no relevance to the domain (e.g., the child is too young for school). Incorrect usage of the N/A code occurs when it is wrongly used for a case study with information to assign a score or when it is not used for a case study with no information to assign a score. Of the 154 ratings that involved use of the N/A response, 83.1% were used correctly and 17% were used incorrectly. The accuracy rate is quite similar to the 84.4% accuracy for the scores.

Examining errors across domains in Table 4 shows that raters did poorest in correctly scoring mental health (70% correct), transportation (71% correct), and physical health (77% correct). All other domains were 80% accurate or higher. Scores on transportation tended to be too high, scores on mental health tended to be too low, and scores on physical health were balanced between too low and too high. Note that the accuracy of the scores for each domain does not correspond exactly to the ICC measures of inter-rater reliability. For example, the ICC for adult education is comparatively low, while the accuracy rate is relatively high. In this case,

the errors are few but are dissimilar across raters. Conversely, the ICC for mental health is relatively high while the accuracy is relatively low. In this case, it appears that the raters are similar to each other in the errors they make.

Table 5 first shows the correlations between sociodemographic characteristics and accuracy for variables with sufficient variation to be included in the analyses (there are too few males, minorities, and Spanish speakers to include gender, race, ethnicity, and language). Most correlations do not reach statistical significance ($p < .05$), but those with a FD credential have significantly lower errors, whereas those with a high count of CFSAAs in the last month show greater error. However, the positive correlation with the count variable stems from a single outlier and disappears when deleting the outlier (results not presented). The multiple regression results listed in Table 5, which exclude the measure of the length of time working with families because it is nearly identical to the length of time of FRCA employment ($r = .96$), show much the same as the correlations. FD Credential lowers error, while the count of CFSAAs increases the error due to a single outlier. In addition, a higher level of education reduces error, when examined with other variables in the model.

Discussion and Applications to Practice

This study subjected a locally-developed self-sufficiency matrix to a systematic test of inter-rater reliability. Reliability is a necessary requirement of measurement validation, and this study suggests that self-sufficiency matrices can be constructed to capture reliable information across a wide range of family support workers. These findings are important to the social work field as program funders are increasingly demanding evidence of program impact. Measures that are locally relevant, that can be realistically implemented in the field, and that have strong

psychometric properties are critical for family support organizations to measure and document family-level program outcomes.

As hypothesized, results of this study indicate that the CFSA 2.0 is constructed so that family advocates will obtain consistent and objective ratings of family self-sufficiency across the 14 areas examined. We examined both the consistency across coders (ICCs) and the accuracy of coders compared to a consensus rating determined by an advisory group. Across both tests, the CFSA 2.0 achieved strong psychometric properties – all ICCs were in the excellent range, supporting agreement across family support workers, and workers achieved an exact accuracy rate of 84% overall. Allowing for some deviation (off by 1 on the 5-point scale) shows 96% accuracy, suggesting that, provided sufficient information, workers will achieve accuracy when using the tool. Thus, findings support the reliability of the CFSA 2.0.

Findings also suggest the value of using case studies as a training mechanism to ‘certify’ family advocates after service training and prior to administration of the CFSA 2.0 with families. In the study, a brief web-based training was sufficient for accurate coding of the CFSA 2.0 – most family advocates achieved high accuracy compared to the advisory group ratings. This supports the idea that the CFSA 2.0 is objective, easy to administer, and does not rely on extensive training to achieve accurate ratings. However, supplemental analyses indicate that three workers did not achieve an overall 80% accuracy rate, and highlight the potential of using the case studies as a means to check whether family advocates are demonstrating a sufficient understanding of how to use the tool before using it with families. In addition, researchers received anecdotal feedback from family advocates that scoring the case studies was a beneficial exercise and would be a useful component of training protocols.

Training in providing family-centered services may also support strong implementation of the assessment. In particular, results of socio-demographic predictors indicated that having an FD credential was associated with more accurate coding. Thus, FD credential status may not only be important to providing strong family support services, but it also may help with coding of the CFSA 2.0. In addition, we found that one participant who reported administering a high number of the prior version of the CFSA had high error; when this participant was removed from analyses, the association between the number of prior CFSA administered and coding error disappeared. Thus, we do not consider this finding robust, and it is not clear, at this stage, whether the amount of experience using a prior tool is a hindrance to accurate coding. Still, it is worth considering the possibility, as is often the case with use of instruments like this one, that raters become more mechanical, and potentially less careful, after long-term use. In addition, it may be that those with heavy caseloads are less careful when administering assessments. Monitoring the number of families advocates are serving, and supporting reasonable case loads, may also help ensure careful administration of assessments.

Despite strong results, findings from domain-specific analyses and anecdotal feedback offered opportunities to further improve the tool. Specifically, compared to the advisory group rating, three domains achieved lower than 80% accuracy - transportation, mental health, and physical health. The advisory committee reconvened, reviewed these three domains in particular as well as reviewed study participant feedback. Using this information, the committee recommended minor modifications to the tool, including additional definitions to provide further clarity and guidance.

A next step in the CFSA 2.0 validation process is to test the psychometric properties of the tool when used by family advocates in their work with families. Specifically, the current

study was unable to examine family advocates' ability to obtain, during the family interview, the necessary information for accurate coding. Training on how to ask guiding questions in a supportive manner that facilitates the provision of accurate information will be critical for successful implementation of the tool. In addition, while care was taken to develop case studies that reflect complex and real-world family conditions, family circumstances often are multifaceted and nuanced, and family members may describe their circumstances in sometimes vague or inconsistent ways. The tool needs to be tested in its ability to measure family situations across a broader range of circumstances and in real-life situations than the ten provided in the case studies. Future tests should also examine the sensitivity of the CFSA 2.0 to measure changes in family functioning over time. Nonetheless, this study demonstrates that the domain index is well-constructed and sufficiently objective to obtain reliable information, thereby laying the groundwork for the next validation phase of field testing.

References

- Benedetti, G. (2012). *Innovations in the field of child abuse and neglect prevention: A review of the literature*. Chicago, IL: Chapin Hall at the University of Chicago.
- California Network of Family Strengthening Networks. (2014). *Standards of quality for family strengthening and support*. Retrieved from http://www.slocap.org/wp-content/uploads/2011/12/Standards-of-Quality_February-5-2014.pdf
- Chase-Lansdale, L. P., & Brooks-Gunn, J. (2014). Two generation programs in the twenty first century. *The Future of Children, 24*, 13-40. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1029028.pdf>
- Colorado General Assembly. (2009). *Senate Bill 09-055*. Chapter 48. Retrieved from http://www.state.co.us/gov_dir/leg_dir/olls/sl2009a/sl_48.htm
- Counts, J. M., Buffington, E. S., Chang-Rios, K., Rasmussen, H. N., & Preacher, K. J. (2010). The development and validation of the protective factors survey: A self-report measure of protective factors against child maltreatment. *Child Abuse and Neglect, 34*, 762-772. doi: 10.1016/j.chiabu.2010.03.003
- Culhane, D. P., Parker, W. D., Poppe, B., Gross, K. S., & Sykes, E. (2008). Accountability, cost-effectiveness, and program performance: Progress since 1998. *National Symposium on Homelessness Research*. Retrieved from http://repository.upenn.edu/cgi/viewcontent.cgi?article=1116&context=spp_papers
- Daro, D., & Dodge, K. A. (2009). Creating community responsibility for child protection: Possibilities and challenges. *Future Child, 19*(2), 67-93. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3730282/>

- Endres, J., Richardson, B., & Sherman, J. (1999). *Testing the reliability/validity of the California Matrix*. Retrieved from Matrix Outcomes Model website:
<http://www.matrixoutcomesmodel.com>.
- Fassaert, T., Lauriks, S., van de Weerd, S., Theunissen, J., Kikkert, M., Dekker, J., ... de Wit, M. (2013). Psychometric properties of the Dutch version of the Self-Sufficiency Matrix (SSM-D). *Community Mental Health Journal, 50*, 583-590. doi:10.1007/s10597-013-9683-6
- Fernandez, E. (2004). Effective interventions to promote child and family wellness: a study of outcomes of intervention through children's family centres. *Child and Family Social Work, 9*, 91-104. doi: 10.1111/j.1365-2206.2004.00315.x
- Fernandez, E. (2007). Supporting children and responding to their families: Capturing the evidence on family support. *Children and Youth Services Review, 29*, 1368-1394. doi:10.1016/j.childyouth.2007.05.012
- Forest, C. (2003). *Empowerment skills for family workers: A worker handbook, the comprehensive curriculum of the Cornell Family Development Credential*. Ithaca, NY: Cornell Family Development Press.
- FRIENDS National Resource Center for Community Based Child Abuse Prevention. (2011). *The Protective Factors Survey user's manual*. Retrieved from FRIENDS National Resource Center website: <http://friendsnrc.org>.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials for Quantitative Methods in Psychology, 8*, 23-34. Retrieved from <http://www.tqmp.org/Content/vol08-1/p023/p023.pdf>

Harper Browne, C. (2014). *The Strengthening Families Approach and Protective Factors*

Framework: Branching out and reaching deeper. Washington, DC: Center for the Study of Social Policy.

Hubble, C. (2010). *Alabama Network of Family Resource Centers (ANFRC) impact study report*.

Retrieved from the Alabama Network of Family Resource Centers website:

<http://anfrconline.homestead.com/about.html>

Hout, M., Cumberworth, E., Danziger, S., Wimer, C., Jusko, K., Weisshaar, K., ... Reardon, S.

(2014). *State of the Union: The Poverty and Inequality Report 2014*. Retrieved from

Stanford University, The Stanford Center on Poverty and Inequality website:

http://web.stanford.edu/group/scspi/sotu/SOTU_2014_CPI.pdf

Manning, M., Homel, R., & Smith, C. (2010). A meta-analysis of the effects of early

developmental prevention programs in at-risk populations on non-health outcomes in adolescence. *Children and Youth Services Review*, 32, 506-519.

doi:10.1016/j.childyouth.2009.11.003

Mills, G. & Amick, J. (2010). *Can Savings Help Overcome Income Instability?* (Brief #18).

Washington, DC: The Urban Institute.

O'Brien-Strain, M., Gunther, K., Rosenberger, A., & Theobald, N. (2012). *San Francisco family*

resource center initiative: Year 2 evaluation: FY10/11. San Francisco, CA: Mission

Analytics Group.

Parker, W. (2008). *Family Connections Program two-year evaluation report*. Report prepared

for the Arizona Department of Economic Security, Office of Community Partnerships

and Innovative Practices (CPIC). Retrieved from

https://www.azdes.gov/InternetFiles/Reports/pdf/family_connections_evaluation_2008.pdf

Sandstrom, H., & Huerta, S. (2013). *The negative effects of instability on child development: A research synthesis*. Retrieved from the Urban Institute website:

<http://www.urban.org/UploadedPDF/412899-The-Negative-Effects-of-Instability-on-Child-Development.pdf>

Richmond, F., & Valentin, M. (2012). *The accountable case manager: Principles and practices for effective case management, a guide for community action agencies and CSBG Funded eligible entities* (2nd ed.). Camp Hill, PA: The Center for Applied Management Practices, Inc.

Shonkoff, J. P. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development, 81*, 357-367. doi: 10.1111/j.1467-8624.2009.01399.x

Snohomish County Self-Sufficiency Taskforce. (2004). *Self-Sufficiency Matrix: An assessment and measurement tool created through a collaborative partnership of the human services community in Snohomish County*. Retrieved from http://www.performwell.org/index.php?option=com_mtree&task=att_download&link_id=48&cf_id=24.

Walker, J. (2008). How and why, does wraparound work: A theory of change. In E. J. Bruns & J. S. Walker (Eds.), *The resource guide to wraparound*. Portland, OR: National Wraparound Initiative, Research and Training Center for Family Support and children's Mental Health.

U.S. Census Bureau. (2014). *Income and Poverty in the United States: 2013*. Retrieved from the

U.S. Census Bureau website:

<http://www.census.gov/content/dam/Census/library/publications/2014/demo/p60-249.pdf>

Table 1

Participant Demographic Characteristics as a Percentage of the Sample

Characteristic	Percent (n=24)
Sex	
Male	9
Female	91
Highest Degree Achieved	
Master's Degree	16
Bachelor's Degree	21
Associate Degree	25
Some College	38
Ethnicity	
Hispanic or Latino	21
Not Hispanic or Latino	79
Race	
White	91
Black or African American	9
Primary Language	
English	85
Spanish	15
Length of Family Resource Center Employment	
0 to 4 years	29
5 to 9 years	21
10 to 14 years	29
over 15 years	13
Missing	8
Length of time working with families	
0 to 4 years	21
5 to 9 years	25
10 to 14 years	29
over 15 years	25
Participant has a Strengthening Families Certificate	
Yes	71
No	29
Participant has an FDC Credential	
Yes	67
No	33

Table 2

Inter-rater reliability by domain

Domain	N	ICC
Debt Management	10	.96
Cash Savings	10	.94
Child Education	4	.93
Employment	8	.92
Child Care	5	.90
Health Coverage	10	.90
Substance Abuse	10	.89
Income	10	.88
Mental Health	10	.87
Physical Health	10	.84
Food Security	10	.83
Housing	10	.80
Adult Education	10	.80
Transportation	10	.79

Table 3

Distribution of Errors in Case Study Ratings

Score	Count	Percent
-4	4	0.1
-3	11	0.3
-2	43	1.3
-1	194	6.0
0	2714	84.4
1	175	5.4
2	50	1.6
3	23	0.7
4	1	0.0
Total	3215	100

Table 4

Distribution of Errors by Domain

Domain	Error Score								
	-4	-3	-2	-1	0	1	2	3	4
Income		0.8%	1.3%	1.3%	92.5%	3.3%	0.4%	0.4%	
Employment		0.4%	0.4%	7.6%	88.2%	3.4%			
Housing	0.4%		2.9%	8.3%	81.3%	1.3%	5.8%		
Transportation	0.4%		1.7%	5.4%	70.7%	18.0%	1.7%	2.1%	
Food Security		0.4%	0.4%	1.7%	87.5%	4.2%	2.1%	3.8%	
Child Care		0.5%	3.9%	0.5%	82.8%	3.0%	9.4%		
Child Educ.		0.7%		2.9%	89.9%	5.8%		0.7%	
Adult Educ.		1.3%	1.3%	9.6%	85.8%	0.8%	0.4%	0.8%	
Cash Savings				3.8%	89.6%	5.8%	0.4%	0.4%	
Debt Manage.		0.4%	0.8%	7.1%	90.4%	1.3%			
Health Cover.			0.4%	0.8%	93.3%	4.6%	0.4%	0.4%	
Phys. Health		0.4%		9.6%	77.1%	10.4%	1.3%	0.8%	0.4%
Mental Health	0.4%		2.1%	17.5%	69.6%	10.0%		0.4%	
Sub. Abuse	0.4%		3.3%	6.3%	85.4%	4.2%	0.4%		

Table 5

Bivariate Correlations and Standardized Multiple Regression Coefficients of Sociodemographic Characteristics with Mean Absolute Value of Rater Error

Sociodemographic Variable	N	r	beta
Age	24	.03	.41
Highest Degree Obtained	24	.09	-.42*
Length of FRC Employment	22	-.30	
Length of Time Working with Families	24	-.18	-.39
Count CFSAs in Last Month	22	.64*	.68*
Strengthening Families Certificate	23	.11	-.02
Family Development Credential (FDC)	23	-.41*	-.43*

* $p < .05$