THE DEVELOPMENT AND VALIDATION OF THE FACES-IV-SF

Jacob B. Priest
University of Iowa

Elizabeth O. Parker
Seattle Children's Hospital

Angela Hiefner
UT Southwestern Medical

Sarah B. Woods
University of Texas Southwestern Medical Center at Dallas

Patricia N. E. Roberson
University of Tennessee

The Family Adaptability and Cohesion Scale IV (FACES-IV) was developed to capture the balanced and unbalanced levels of cohesion and flexibility in families. Although this measure has been shown to be valid and reliable, its length at 62 items limits utility and uptake in clinical and research settings. This paper details the development of a shorter form of the FACES-IV (the FACES-IV Short Form) using two studies. In the first study, three item-level analyses were used to identify 24 items that provided the best measurement of each of the scales of the FACES-IV. In the second study, the reliability, and convergent and divergent validity of the short form was tested. Results suggest that the FACES-IV-SF is a valid and reliable measure that adheres to the theory underlying the original FACES-IV but may be better utilized in clinical and research settings due to its brevity.

Specific theoretical models for conceptualizing family functioning are critical for the development of evidence-based, family-oriented clinical approaches. Providing clear pathways by which unique constructs within family process impact one another to create specific patterns of family functioning is necessary to guide therapeutic intervention and scientific inquiry. One such theoretical model which has garnered a great deal of evidence is Olson's (2010; 2011) Circumplex Model. The Circumplex Model assumes three key dimensions in conceptualizing family functioning and interactions: cohesion, flexibility, and communication. This approach hypothesizes that healthy functioning families have balanced levels of cohesion (i.e., separated or connected, rather than disengaged or enmeshed) and flexibility (i.e., flexible or structured, rather than chaotic or rigid), while problematically functioning families report unbalanced levels of both. The third dimension of the model, communication, operates to facilitate the relative levels of cohesion and flexibility within families. In other words, communication in families allows family members to vacillate in their flexibility and cohesion. When communication is healthy, it supports family members' engagement with and responsiveness to one another.

Measurement is a critical next step in conveying family functioning theory to clinical practice. Accordingly, alongside the development of the Circumplex Model, Olson (2010) developed the

Jacob B. Priest, PhD, University of Iowa, Iowa city, IA; Elizabeth O. Parker, PhD, Seattle Children's Hospital, Seattle, WA; Angela Hiefner, PhD, UT Southwestern Medical, Dallas, TX; Sarah B. Woods, PhD, University of Texas Southwestern Medical Center at Dallas, Dallas, TX; Patricia N. E. Roberson, PhD, University of Tennessee, Knoxville, TN.

Address correspondence to Jacob B. Priest, Department of Psychological and Quantitative Foundations, University of Iowa, Iowa City, Iowa; E-mail: jacob-b-priest@uiowa.edu

Family Adaptability and Cohesion Scale IV (FACES-IV) to assess family functioning as specified by his theoretical approach. Specifically, the FACES-IV was developed to capture the balanced and unbalanced levels of cohesion and flexibility (Olson, 2011). The FACES-IV has three cohesion scales, assessing enmeshment and disengagement (i.e., unbalanced levels of cohesion), as well as cohesion balance. The FACES-IV also has three scales assessing flexibility, including rigidity and chaos (i.e., unbalanced levels of flexibility), as well as balanced levels of flexibility. In addition, FACES-IV also includes communication and satisfaction scales. In total, the FACES-IV includes eight scales to fully assess the range of cohesion and flexibility in families (Olson, 2011).

Not only is the FACES-IV a reliable and valid measure of family functioning (Olson, 2011; Priest, Parker, & Woods, 2018), it also provides clinical utility. Unlike many measures of family functioning, the FACES-IV (and the Circumplex Model) allows clinicians to assess the complexity of family functioning by capturing the family processes that are linked to outcomes. For example, instead of measuring whether families are highly satisfied, or whether families have high levels of conflict, the FACES-IV can help clinicians gather an in-depth assessment of what patterns of interactions are occurring in families (Olson, 2000), which can be more meaningfully addressed in systemic therapy. The FACES-IV also reflects systems-based assumptions that fit many couple and family therapy approaches (e.g., Structural Family Therapy, Bowen's Family Systems Theory). These strengths of the FACES-IV make it an important assessment for clinicians aiming to assess meaningful change in relational therapy. In addition, the FACES-IV is an effective measure for researchers working to understand the links between families and wellbeing outcomes through a systemic lens (Olson, Waldovegel, & Schlieff, 2019).

Limitations of the FACES-IV

Though the FACES-IV is a theoretically-based, psychometrically-sound, practice-relevant measure, the length of the assessment, at 62 items, limits its utility and uptake in research and clinical settings. For example, survey-based research is increasingly being conducted online. Research on survey length suggest that the longer an online survey is, the less likely participants are to begin or complete the survey (Galesic & Bosnjak, 2009). In addition, questions that appear later in surveys are answered more quickly and answers are more likely to be uniform (Galesic & Bosnjak, 2009). This type of responding can result in misclassification and poor measurement (Rolstad, Adler, & Ryden, 2011). As a result, researchers conducting web-based studies are less likely to receive accurate information from lengthy questionnaires.

The length of the FACES-IV also limits its use in large data collection projects. Many nationally representative datasets include multiple measures assessing for a wide variety of variables, including family functioning (e.g., Midlife in the United States; Ryff et al., 2013–2014), Because these surveys hope to gain information on many constructs, briefer assessments are preferred to limit response fatigue (Whisman, 2007). However, unlike the FACES-IV, these brief measures are infrequently theory-driven, nor clinically applicable, and rarely go beyond asking about family states of conflict and relationship satisfaction. This is problematic as a substantial amount of research (e.g., Kutschke, Bengston, Seeman, & Harris, 2018; Priest et al., 2018) using secondary datasets, and their less complex measures of family functioning, have begun to demonstrate associations between family relationships and major public health issues (e.g., mental health, obesity, and chronic health conditions). In other words, though the findings of these studies are critically important, the generalizability and clinical utility of their results may be incomplete due to the limitations of the measures included. Few psychometrically-validated family measures also limit research examining the impacts of family involvement in chronic disease management, a critically important public health initiative (Gillis, Pan, & Davis, 2019; Torenholt, Schwennesen, & Willaing, 2014). Therefore, to ensure a full, accurate understanding of how family functioning impacts public health issues broadly, it is necessary to use rich, conceptually complex, empirically-supported, theory-based measures that can be included in research projects due to their brevity.

Lastly, the length of the FACES-IV also limits its use in clinical settings. One clinical setting where the FACES-IV may be especially useful is in primary care, and integrated behavioral health. As noted previously, family relationship issues are prevalent in primary care (Woods, Priest, Fish, Rodriguez, & Denton 2014), and are consistently linked to many health outcomes, as well as disease management (Gillis et al., 2019; Woods & Denton, 2014). Given the impact of families on health,

patients accessing primary care may be best served by integrated, systemic behavioral health providers, such as medical family therapists (Woods, Bridges, & Carpenter, 2019). However, routine and accurate screening of patients to determine the need for integrated family-based services is challenging in the current volume-based healthcare environment. In the current system, healthcare providers are increasingly burdened by growing productivity expectations and more time-consuming administrative demands (Rothenberger, 2017). As such, barriers to the routine use of assessment tools in primary care include length, proprietary nature, limited usefulness for assessment and treatment planning, and requirement of clinician administration rather than patient self-report (Spitzer et al., 2006). As such, many primary care practices do not routinely assess for family functioning despite the importance of screening for family issues (Woods, Priest, & Denton, 2015). This limits the ability to effectively triage patients in need of family-based services, as well as limiting broader needs assessments that could support the uptake of integrated family-based behavioral healthcare. For clinical assessment tools to be effectively adopted as routine practice in primary care, as one example, they must fit within the time demands and clinical workflows of today's healthcare systems.

For the FACES-IV to be most effectively utilized by behavioral health providers in primary care settings, the tool (administration and scoring) must fit within medical visit time limitations, such that it does not overburden either patient or staff, while providing high quality family relationship information that can inform an effective approach to each patient's unique health issues and family context. Though research recognizes the family's influence on patient health, the length, administration, and scoring of the FACES-IV (as with alternate family functioning measures) prohibits its feasible use in primary care.

Present Study

The goal of this study was to develop a short form of the FACES-IV, known as the Family Adaptability and Cohesion Evaluation Scale Short Form (FACES-IV-SF). We aimed to dramatically reduce the assessment to approximately one-third of its length. This reduction would increase the likelihood of research and clinical implementation, yet still retain the richness and theoretical complexity of the measure. Therefore, we reduced the 62-item assessment to a 24-item assessment while maintaining the reliability and validity of the FACES-IV. This was accomplished with two studies. The first study was used to identify which questions from the FACES-IV scales could be removed but still allow the assessment to adhere to the Circumplex Model (Olson, 2011). The second study used the 24 items from the first study to examine the convergent and divergent validity of the FACES-IV-SF.

METHOD

Study 1

Procedure. Respondents for this study were asked to complete an online survey. Three strategies were used to recruit respondents to complete the online survey. First, participants were recruited from a university research listserv. An email was distributed on this listserv that had information about the survey and a link to access it. Second, respondents were recruited through a Google AdWords advertisement. If a person searched for key terms, (e.g. family functioning, family therapy, parenting, couple therapy, etc.) would be shown an advertisement which would link them to the survey. Finally, a similar advertisement was posted on social media websites where those who interacted with the advertisement could access the survey. To participate, respondents were required to be 18 years old or older and speak English.

Participants. Our sample included $n = 522$ respondents. These respondents were majority female (83%) and had an average age of 39.94. The majority were Caucasian (86.3%), and 2.9% were Hispanic/Latinx, 2.2% were Asian, 1.4% were African American, and 5.1% reported "other" as their race/ethnicity. Approximately one-third of the sample had a college degree (32.7%), 20.4% had a master's degree, 10.6% had a professional or doctoral degree, 20.9% reported attending at least some college, 9.8% had an associate degree, and 4.7% reported having complete high school. Nearly half the sample (47%) were married and had children, while 11.4% were married with no children; 23.3% were single, 8.2% were living with a partner, 8% were divorced, and 1.4% were widowed.

Measures

Family adaptability and cohesion evaluation scale (FACES). The FACES-IV (Olson, 2011) is also a self-report assessment of family functioning. It contains 62 items and has eight scales. These scales assess for levels of adaptability, cohesion, and satisfaction of family members with the levels of cohesion and flexibility. This assessment has demonstrated validity and reliability (Olson, 2011). Questions on the scales were asked on a five-point Likert scale ranging from 1 "Strongly Disagree" to 2 "Strongly Agree." FACES-IV has a copyright, and permission from the original author was granted to use this assessment in this study. In this study one question from the Chaotic scale, "Our family has a hard time keeping track of who does various household tasks" was inadvertently left out of the online survey. As such, analyses conducted on the Chaotic scale only includes 6 items.

Data Analysis

In accordance with prior literature using multiple types of analyses, and item response theory (IRT) to develop brief, meaningful measures of psychosocial constructs (e.g., Funk & Rogge, 2007), we used an iterative, deductive data analysis process to determine the highest quality, most informative FACES-IV items to create a brief version of the measure. Thus, data analysis was conducted in two steps. The goal of the first step was to identify the items that had the largest effect on the measurement of constructs of the FACES-IV to create the FACES-IV-SF. To identify these items, three item-level analyses were conducted. The second step consisted of comparisons between short-form versions of the FACES-IV and the original form. These comparisons were made to test whether the short-form scales provided similar, better, or worse measurement that the original form scales.

Item-level analyses. The first item-level analysis was total item correlation. Total item correlation assesses the correlation between a singular item and the overall scale score. The second item-level analysis was a factor analysis. The goal of this step was to identify factor loadings of each item onto the latent trait. Items with higher factor loadings have a larger effect on the measurement of the latent trait than those with lower factor loadings.

The third item-level analysis used an IRT approach. IRT uses model-based measurement; this measurement depends on both a person's responses and the properties of the questions that were administered (Embretson & Reise, 2013). IRT establishes a model that identifies the probability of observing each response option to a question as a function of a latent trait (De Ayala, 2013; Reckase, 2009). This is done by estimating response curves for each question and by estimating latent trait scores for the respondents (Funk & Rogge, 2007). By doing this, it is possible to estimate how much information each question offers to the latent trait being measured. By using an IRT approach, it is thus possible to examine the precision with which each question can be used to estimate the latent trait. This level of precision is known as information. Questions that provide large amounts of information are better able to distinguish those with high levels of a latent trait from those with lower levels of a latent trait. In other words, these questions are more precise or accurate. Items that provide small amounts of information do a poor job at distinguishing levels of a latent trait, and therefore only contribute to the error variance of the total measure (De Ayala, 2013; Reckase, 2009).

As item response curves are calculated uniquely for each item, and within any measure for which IRT analysis is used, discriminating among curves, and the information each item provides, is a visualization task. Therefore, the authors visually evaluated each item's response curve, and identifies the three items which provided the greatest amount of information for each FACES-IV subscale. For subscales where the best three curves (i.e., most precise items providing the greatest amount of information specific to the latent trait measured) were unclear (e.g., five response curves appeared to be equally indicative of highest quality items, as with the Satisfaction subscale), the authors then determined the best three by accounting for the items with the highest total item correlation, and highest factor loading. Lastly, for subscales where it remained difficult to discern the highest quality items (i.e., after considering items with most informative response curves, highest item correlations, and highest factor loadings), the authors made a decision on which three items to retain based on which items were most theoretically consistent with the subscale, based on how Olson (2010) specified the subscale's construct (e.g., an item assessing whether family members

get along may score high in each area, however, is less reflective of the Communication subscale construct than an item that assesses whether family members can calmly discuss problems with each other).

After the item level analyses were run, and the items that provided the best measurement for each construct of the FACES-IV were chosen, these items were used to create the subscales of the FACES-IV-SF. Next, the second step of the analysis was conducted, to compare measure iterations.

Measure comparisons.   The first comparison evaluated the model fit of the constructs for both the FACES-IV and the FACES-IV-SF using confirmatory factor analysis (CFA). Since model fit is affected by complexity (Kline, 2011), we randomly selected 3 items from each of the subscales the original FACES-IV to compare to the items chosen through the item level analysis. Specifically, six CFAs were conducted. The first CFA assessed the fit of the cohesion dimension scales of the Circumplex Model Cohesion (the Cohesion, Enmeshment, and Disengaged scales) for the FACES-IV-SF. The second CFA assed the fit of the cohesion dimension scales for the randomly selected items of FACES-IV. The next two CFA replicated this process but with the flexibility dimension scales (the Flexibility, Chaotic, and Disengaged scales) – separate CFAs were run for both the FACES-IV_SF flexibility dimension scales and the randomly selected flexibility dimension items from the FACES-IV. The final two CFAs assessed the Satisfaction and Communication Scales for both the randomly selected items of FACES-IV and FACES-IV-SF. Fit statistics (root mean squared approximation, comparative fit index, Tucker-Lewis index, and standardized root mean square residual) were used to compare the fit of the Cohesion, Flexibility, and the Communication and Satisfaction CFAs between the randomly selected items and the items of the FACES-IV-SF.

The second comparison examined the differences between the Total Circumplex Ratio Scores for the two scales. The ratio score is used to assess the degree to which a family is healthy/balanced or unhealthy/unbalanced (Olson, 2011). Though not used in clinical practice, Olson (2011) suggests this ratio score is critical for testing Circumplex Model hypotheses in research studies. The Total Circumplex Ratio is calculated by dividing the average of the Flexibility and Cohesion (i.e., balanced) scales by the average of the Rigid, Chaotic, Enmeshed, and Disengaged (i.e., unbalanced) scales. Ratio scores that are 1 or greater represent balanced (i.e., healthy) families. Ratio scores that are less than 1 represent unbalanced (i.e., unhealthy) families. The goal in calculating total ratio scores was to determine whether the FACES-IV-SF would classify respondents similarly to the FACES-IV. Once ratio scores were calculated, scores were dichotomized as 0 or 1–0 representing ratio scores less than 1 (families with problematic functioning) and 1 representing scores greater than or equal to 1 (families with balanced, healthy functioning). A chi-square difference test was conducted to test whether there was a significant difference between the measures on their classifications.

Results

Total item correlations and factor loadings for all 62 items on the FACES-IV are available in the supplemental information available online. Item information curves for all 62 items are reported in Figure 1. For each scale, three items that had the highest total item correlations, factor loadings, provided the greatest amount of information, and most clearly aligned with hypotheses of the Circumplex Model were used to create the scales of the FACE-IV-SF. The items chosen for each scale along with their total item correlations and factor loadings are reported in Table 1. The item information curves for each question of the FACES-IV are reported in Figure 1.

Cronbach's alpha for the FACES-IV-SF and the original FACES-IV are reported in Table 2. For the short form, all scales' Cronbach's alphas were greater than .75 except for the Enmeshed scale. For the original FACES-IV, all the scales Cronbach's alphas were greater than .70 with the exception of the Rigid scale.

Fit statistics of the CFAs are reported in Table 3. Overall, the short form version of the scales provided better fit for the data than the original form scales; however, the fit of the randomly selected items from the original FACES-IV also provided good fit.

Using ratio scores, the original version of the FACES-IV classified 9.9% of families as unhealthy and 90.1% of families as healthy, in accordance with the nosology indicated by Olson's
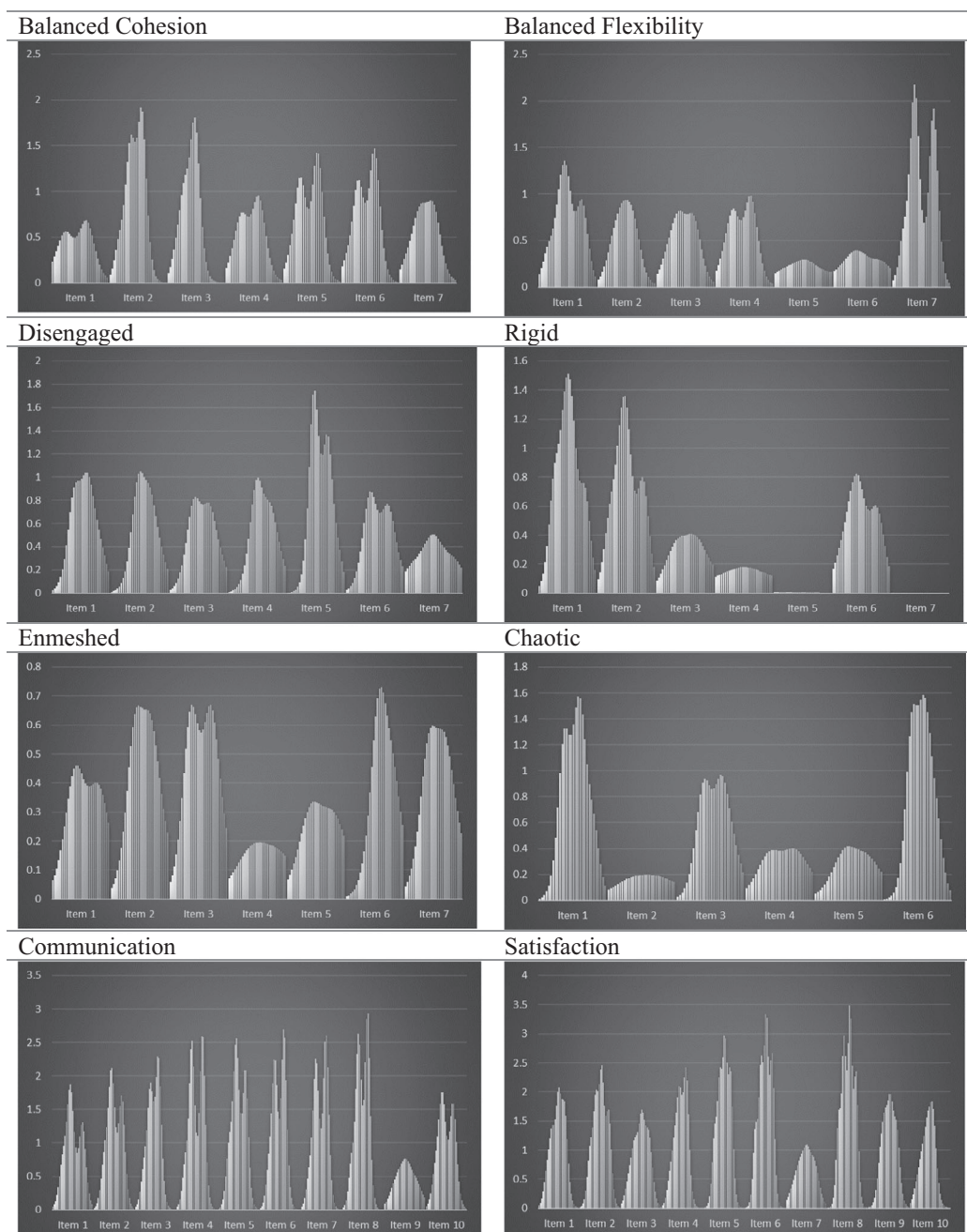
Figure 1. Information curves for all 62 items of the FACES-IV.

(2011) Circumplex Model; the short form version of the FACES-IV classified 11.4% of families as unhealthy and 88.6% of families as healthy. The results of the χ^2 test between these two distributions showed no significant difference ($\chi^2 = 0.45$, $p > .05$).

Overall, the results from Study 1, suggest that the short form of the FACES has similar reliability and slightly better measurement fit than the original FACES-IV.

Study 2

Like the previous study, this study asked participants to complete an online survey. The purpose of this survey was to test the convergent and divergent validity of the FACES-IV-SF. Subjects

Table 1
Total Item Correlation and Factor Loading of FACES-IV-SF Scales for Study 1

Scale	Item	Total item correlation	Factor loading
Balanced cohesion	FACES-IV Item 7	.86	.76
	FACES-IV Item 13	.84	.72
	FACES-IV Item 25	.82	.73
Balanced flexibility	FACES-IV Item 2	.80	.69
	FACES-IV Item 20	.76	.62
	FACES-IV Item 38	.87	.75
Disengaged	FACES-IV Item 3	.76	.68
	FACES-IV Item 9	.76	.63
	FACES-IV Item 27	.84	.72
Enmeshed	FACES-IV Item 10	.68	.51
	FACES-IV Item 16	.69	.53
	FACES-IV Item 40	.66	.50
Rigid	FACES-IV Item 5	.82	.57
	FACES-IV Item 11	.80	.46
	FACES-IV Item 35	.71	.53
Chaotic	FACES-IV Item 6	.83	.64
	FACES-IV Item 18	.76	.57
	FACES-IV Item 42	.84	.64
Communication	FACES-IV Item 46	.90	.82
	FACES-IV Item 47	.89	.84
	FACES-IV Item 50	.91	.85
Satisfaction	FACES-IV Item 57	.92	.86
	FACES-IV Item 58	.93	.87
	FACES-IV Item 60	.93	.87

Table 2
Cronbach's Alpha Coefficients for Scales for Study 1

	Cronbach's alpha FACES-IV-SF	Cronbach's alpha FACES-IV
Cohesion	.83	.86
Flexibility	.82	.84
Disengaged	.79	.86
Enmeshed	.63	.73
Rigid	.76	.61
Chaotic	.80	.78
Communication	.89	.95
Satisfaction	.93	.96

were recruited using Qualtrics Panels. When using this method, Qualtrics recruits subjects based on specific criteria. For the purposes of this study, a general population census match was used, so that participants would reflect the general population of the United States. Specifically, participants were matched to census data based on gender, age, and race/ethnicity. Respondents who completed the study were given $5 as compensation.

Table 3
Fit Statistics for the FACE-IV-SF and a Random Selection of Items From the FACES-IV

	FACES-IV-SF	FACES-IV
Cohesion scales	$\chi^2(24) = 43.74, p = .01$, RMSEA $= 0.04$, CFI $= 0.98$, TLI $= 0.97$, SRMR $= 0.03$	$\chi^2(24) = 69.87, p < .001$, RMSEA $= 0.06$, CFI $= 0.95$, TLI $= 0.93$, SRMR $= 0.05$
Flexibility scales	$\chi^2(24) = 38.41, p = .03$, RMSEA $= 0.03$, CFI $= 0.99$, TLI $= 0.98$, SRMR $= 0.03$	$\chi^2(166) = 50.15, p < .001$, RMSEA $= 0.05$, CFI $= 0.97$, TLI $= 0.96$, SRMR $= 0.04$
Communication and satisfaction scales	$\chi^2(8) = 12.89, p = .11$, RMSEA $= 0.03$, CFI $= 0.99$, TLI $= 0.99$, SRMR $= 0.01$	$\chi^2(168) = 32.96, p < .001$, RMSEA $= 0.08$, CFI $= 0.98$, TLI $= 0.96$, SRMR $= 0.02$

Sample. A total of $n = 260$ individuals completed the questionnaire. The average duration was 7.3 min. Respondents were 51.2% female, 12.3% were between the age of 18–24, 17.7% were 25–34, 16.9% were 35–44, 17.3% were 45–54, 16.5% were 55–64, and 19.2% were 65 or older. Of the sample 63.1% reported being Caucasian, 17.3% were Hispanic/Latinx, 13.1% reported being African American/Black, 5.8% were Asian, and 3.9% reported "other" as their race/ethnicity. More than half (53.8%) were married, 5% were cohabiting with a partner, 7.3% were divorced, 1.2% were separated, 6.9% were partnered, 3.5% were dating, 3.1% were widowed. The vast majority of the sample (82.8%) identified their sexual orientation as straight; 3% identified as gay or lesbian, 6.2% identified as bisexual, 1.1% reported that they had another sexual orientation.

Measures

In addition to the FACES-IV-SF (permission was also obtained from the FACES-IV author to use these items in Study 2) and demographic items, measures of family support and strain, depression, and anxiety were collected.

Family support was assessed with 4 items. This scale asked the respondents to indicate on a scale from 1 (*a lot*) to 4 (*not at all*) how much their family members: (1) care about them; (2) understand the way they feel; (3) can be relied on if they have a serious problem; and (4) can be opened up to. Scores were summed and averaged to create a composite score. Cronbach's alpha for the sample was $\alpha = .90$. Family strain was also assessed with four times. On a scale from 1 (*a lot*) to 4 (*not at all*), these items asked respondent to indicate how often their family members: (1) criticize them; (2) make too many demands on them; (3) let them down; and (4) get on their nerves. Cronbach's alpha for the sample was $\alpha = 78$. Both scales have been used in large representative surveys previously, and have also demonstrated reliability (e.g., Alegria, Jackson, Kessler, & Takeuchi, 2003; Ryff, Almeida, Ayanian, Binkley, Carr, Coe, et al, 2014).

Depression was measured using the 9 item Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001). The nine items of this measure are aligned with DSM-IV criteria for depression. Specifically, the items of PHQ-9 ask the respondent to indicate if the have experienced: (1) little interest in doing things; (2) feeling depressed or hopeless; (3) trouble sleeping; (4) having little energy; (5) changes in appetite; (6) feeling bad about yourself; (7) trouble concentrating; (8) moving or speaking slowly; and (9) thoughts of hurting yourself. Responses ranged from 1 (*not at all*) to 5 (*always*). Cronbach's alpha for the sample was $\alpha = .96$.

Anxiety was measured with the 7 item Generalized Anxiety Disorder Scale (GAD-7; Spitzer, Kroenke, Williams, & Lowe, 2006). The 7 items of the GAD-7 ask the respondents if in the past 7 days they have: (1) found it hard to focus on anything other than anxiety; (2) been overwhelmed by worries; (3) felt uneasy; (4) felt nervous; (5) felt anxious; (6) felt tense; and (7) felt like they needed help for their anxiety. Responses ranged from 1 (*not at all*) to 5 (*always*). Cronbach's alpha for the sample was $\alpha = .96$.

Data analysis

Replicating the previous study, Cronbach's alphas for each of the scales of the FACES-IV-SF were computed, and a CFA of the cohesion dimension scales (Cohesion, Enmeshment, and Disengaged scales), a CFA of the flexibility dimension scales (Flexibility, Chaotic, and Disengaged scales), and a CFA of the the Communication and Satisfaction scales were run, and fit statistics were evaluated. Then, to determine the convergent and divergent validity of the short form, correlations between the scales, the Circumplex Ratio, and the family support, family strain, depression, and anxiety measures were conducted. The measures of family support and strain were used to establish convergent validity and the depression and anxiety measures were used to establish divergent validity. Additionally, using the Total Circumplex Ratio, the percentage of healthy/balanced and unhealthy/unbalanced families were computed (Olson, 2011). This classification was then used to compare the means for Family Support, Family Strain, Depression, and Anxiety between healthy and unhealthy families.

Results

Cronbach's alphas for the FACES-IV-SF are reported in Table 4. Each of the subscales had alphas near or above .70, typically considered an acceptable level of reliability. The results of the CFAs for the cohesion dimension, flexibility dimension, and Communication and Satisfaction scales are reported in Table 5. The fit statistics suggested that each model provided good fit for the data with the possible exception of flexibility dimension scales (Flexibility, Chaotic and Disengaged scales) which has a mediocre fit. Specifically, the TLI and the SRMR were slightly outside of acceptable range indicating that correlations among the items may not be high enough to be classified as "good fit"; however, the CFI does fall within acceptable rage despite being less stringent assessment of fit compared to the TLI (Kenny, 2015). The significant chi-square for the Flexibility CFA may indicate that the model (e.g., subscale of items) proposed does not fit the data. However, chi-square is only thought to be a reasonable measure of fit for samples 75–200 and our sample is just over this threshold (Kenny, 2015). Nevertheless, the findings with the Flexibility CFA should be interpreted within the shortcoming of the fit indices. Unlike the other CFAs, items in the flexibility dimension scales may have less internal reliability across samples.

The results of the correlations are reported in Table 6. Most of scales of the FACES-IV-SF were correlated with the family support and strain scales in directions that suggest convergent validity and divergent. Most scales were correlated with the measures of Family Support, Family Strain, Depression and Anxiety as predicted. For example, higher levels of Cohesion were associate with greater family support, lower strain, and lower levels of depression and anxiety; whereas, a higher level of Chaos was associated with less support, more strain, and greater anxiety and depression. The only non-significant association was between the Rigid scale and family support.

Table 4
Cronbach's Alpha Coefficients for Scales for Study 2

	Cronbach's alpha FACES-IV-SF
Cohesion	.83
Flexibility	.80
Disengaged	.68
Enmeshed	.69
Rigid	.80
Chaotic	.80
Communication	.89
Satisfaction	.94

Using the Total Circumplex Ratio, 16.2% of respondents' families were classified as unhealthy and 83.8% were classified as healthy, in accordance with Olson's (2011) original calculation. Respondents whose families were classified as unhealthy had significantly lower levels of family support (MD = 4.06, $p < .01$) and the effect size for this difference was large ($d = 1.41$). Those in the unhealthy family group also had significantly higher levels of family strain (MD = 2.45, $p < .01$) and the effect size for this difference was also large ($d = 0.97$). Those in the unhealthy family group had significantly higher levels of depression (MD = 8.33, $p < .01$) and anxiety (MD = 5.04, $p < .01$) and the effect size for both differences were large ($d = 0.94$ and $d = 0.77$ respectively).

DISCUSSION

Though the FACES-IV is a theoretically derived and clinically applicable assessment, the length of the assessment limits its utility in many research and clinical settings. By reducing the number of items on the assessment, the goal was to create a short form of the FACES-IV that continues to provide a multi-faceted measurement of family functioning but reduces the burden on respondents.

The results presented here suggest that the FACES-IV-SF provides reliable measurement of the flexibility and cohesion dimensions of the initial scale (and the Circumplex Model; Olson, 2011). In both samples, the scales of the FACES-IV-SF showed adequate reliability and good fit to the data. In addition, the short form demonstrated adequate convergent and divergent validity in its ability to classify healthy and unhealthy families. Given these results, it may be that the FACES-IV-SF could provide better measurement, especially when used in online survey research. When surveys are long, participants are less likely to begin or complete surveys and answers at the end of long surveys tend to be more uniform (Galesic & Bosnjak, 2009); this can lead to poorer measurement accuracy (Rolstad et al., 2011). By reducing the number of items of the original FACES-IV, it is likely that when used in online research, participants will be more likely to complete the assessment and have less uniformity in their responses. This may allow for better classification of respondents.

The FACES-IV-SF could also provide clinically relevant measurement in large, population-based surveys. Given the associations between family relationships and health, it is important to have comprehensive family functioning assessments that are clinically relevant. Many assessments used in current population-based surveys ask solely about satisfaction or conflict. The FACES-IV-SF is a theoretically driven and clinically relevant assessment. If used in population-based surveys, it could help improve knowledge regarding associations between family function and health. Specifically, by looking at processes in families and/or different types of classifications of families, research could shed more light on how families effect health across the life course.

Table 5
Fit Statistics for Cohesion, Flexibility, and Communication and Satisfaction Scales for Study 2

	FACES-IV-SF
Cohesion scales	$\chi^2(24) = 38.94$, $p = .03$, RMSEA = 0.05, CFI = 0.97, TLI = 0.96, SRMR = 0.04
Flexibility scales	$\chi^2(24) = 73.043$ $p < .01$, RMSEA = 0.09, CFI = 0.92, TLI = 0.88, SRMR = 0.08
Communication and satisfaction scales	$\chi^2(8) = 11.28$, $p = .19$, RMSEA = 0.04, CFI = 0.99, TLI = 0.99, SRMR = 0.02

Table 6

Correlations Between FACES-IV-SF and Measure of Family Support and Strain, Depression and Anxiety

	Support	Strain	Depression	Anxiety
Cohesion	.67**	−.34**	−.34**	−.25**
Flexibility	.55**	−.31**	−.26**	−.18**
Disengaged	−.52**	.47**	.35**	.32**
Enmeshed	−.15*	.34**	.28**	.24**
Rigid	−.01	.21*	.17**	.15*
Chaotic	−.48**	.51**	.460**	.36**
Communication	.68**	−.51**	−.40**	−.29**
Satisfaction	.69**	−.53**	−.43**	−.36**
Circumplex ratio	.58**	−.52**	−.42**	−.33**

*$p < .05$.

**$p < .01$.

In addition to its utility in supporting scientific inquiry, the FACES-IV-SF can enable better translation of family-based intervention research into clinical practice, particularly in primary care. While the length of the original FACES-IV prohibits its feasible use in the current volume-based healthcare reimbursement system, the FACES-IV-SF can more suitably fit within primary care time constraints without overburdening patients or integrated behavioral health providers. A brief but rich family functioning assessment can provide useful information to assist therapists embedded in primary care, such that patients reporting unbalanced family systems may be more accurately referred for family-based interventions that aim to support positive disease management. The FACES-IV-SF may also be useful for family therapists to communicate characteristics about a patient's family system to primary care physicians, thus promoting collaborative and patient-centered care. A family-oriented approach to healthcare, involving the regular use of family functioning screening tools, may be especially relevant in family medicine, which may be most adept at this type of approach (McDaniel, Campbell, Hepworth, & Lorenz, 2005).

Lastly, for family therapists broadly, and other practitioners seeking to meaningfully assess family functioning with quick results and without unduly burdening clients or staff, the FACES-IV-SF may provide an important alternative to currently available assessment tools. Each of the eight FACES-IV-SF subscales provides valuable information about systemic family processes that may be targeted in family therapy, including, for example rigidity, communication, enmeshment, and flexibility. The present findings suggest a potentially feasible option for therapists seeking to enhance their systemic practice by assessing an array of family relationship characteristics at intake, and beyond.

Limitations & Future Research

The results of the studies summarized above lend support for the use of the FACES-IV-SF in cases where the original, longer version may impose a burden on respondents. However, the support found for the FACES-IV-SF should be viewed with restraint given some of the limitations of the studies and the evidence provided. For example, the reduction process used in the first study was done on a self-selected, non-representative sample. It could be that had the sample been more representative, the items selected may have been different. The items chosen from study 1 did perform well when examined for a more representative sample in study 2; however, future research needs to test the FACES-SF-IV with diverse samples to further validate the results here.

In addition, the Circumplex Model and the FACES-IV were designed to measure family functioning. In both studies reported here, only one family member was reporting on their family

functioning. Future test of the FACES-IV-SF would benefit from including reports from multiple people from the same family. This would allow for a better understanding of the validity of the items chosen, and the overall measure. Similarly, using the Circumplex Model clinical rating scales have been developed (Thomas & Olson, 1994). On way to further validate the results of the findings here would to use these clinical rating systems to see if clinicians would score and classify families similarly to how individual respondents would score and classify themselves.

Finally, previous tests of the FACES-IV have used measurement invariance testing to assess how measurement of the constructs varies depending on which members of the family a respondent was thinking about. The results of test have shown that measurement may not be consistent across different family types (Priest et al., 2018). It would be important to replicate similar test for the for the FACES-IV-SF to see if the same limitations are present. It would also be important to use measurement invariance testing to assess whether measurement is consistent across multiple groups. In other words, it may be that measurement of the constructs of the FACES-IV-SF vary across gender, race/ethnicity/, sexual orientation, gender identity or age. More psychometric work needs to be completed on the FACES-IV-SF to evaluate its utility for all families.

REFERENCES

Alegria, M., Jackson, J. S., Kessler, R. C., & Takeuchi, D. (2003). *Collaborative Psychiatric Epidemiology Surveys (CPES), 2001–2003 [United States]*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016–03-23.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. New York, NY: Guilford Publications.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Mahwah, NJ: Psychology Press.

Funk, J. L., & Rogge, R. D. (2007). Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology*, *21*(4), 572–583.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360.

Gillis, C. L., Pan, W., & Davis, L. L. (2019). Family involvement in adult chronic disease care: Reviewing the systematic reviews. *Journal of Family Nursing*, *25*(1), 3–27.

Kenny, D. A. (2015).Measuring model fit. Retrieved from June 15, 2019, http://www.davidakenny.net/cm/fit.htm

Kline, R. B. (2011). Convergence of structural equation modeling and multilevel modeling. In M. Williams & W. P. Vogt (Eds.), *Handbook of methodological innovation* (pp. 562–589). London: Sage.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613.

Kutschke, J., Bengtson, M. B., Seeman, T. E., & Harris, J. R. (2018). Social support and strain across close relationships: A twin study. *Behavior genetics*, *48*(3), 173–186.

McDaniel, S. H., Doherty, W. J., Hepworth, J., & Lorenz, A. (2005). *Family-oriented primary care*. New York, NY: Springer-Verlag.

Olson, D. H. (2000). Circumplex model of marital and family systems. *Journal of Family Therapy*, *22*(2), 144–167.

Olson, D. H. (2010). *FACES IV manual*. Minneapolis, MN: Life Innovations.

Olson, D. (2011). FACES IV and the Circumplex Model: Validation study. *Journal of Marital and Family Therapy*, *37*(1), 64–80.

Olson, D. H., & Gorall, D. M. (2006). *FACES IV and the Circumplex Model*. Minneapolis, MN: Life Innovations.

Olson, D. H., Waldvogel, L., & Schlieff, M. (2019). The Circumplex Model of Marital and Family Systems: An update. *Journal of Family Theory and Review*, *11*, 199–211.

Priest, J. B., Parker, E. O., & Woods, S. B. (2018). Do the constructs of the FACES IV change based on definitions of "Family?" A measurement invariance test. *Journal of Marital and Family Therapy*, *44*(2), 336–352.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer-Verlag.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health*, *14*(8), 1101–1108.

Rothenberger, D. A. (2017). Physician burnout and well-being: A systematic review and framework for action. *Diseases of the Colon & Rectum*, *60*(6), 567–576.

Ryff, C., Almeida, D., Ayanian, J., Binkley, N., Carr, D. S., Coe, C., et al. (2014). *Midlife in the United States (MIDUS 3), 2013–2014*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019–04-30.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Thomas, V., & Olson, D.H. (1994). Circumplex model: curvilinearity using clinical rating scale (CRS) and Faces III. *The Family Journal*, *2*(1), 36–44.

Torenholt, R., Schwennesen, N., & Willaing, I. (2014). Lost in translation – the role of family in interventions among adults with diabetes: A systematic review. *Diabetic Medicine*, *31*, 15–23.

Whisman, M. A. (2007). Marital distress and DSM-IV psychiatric disorders in a population-based national survey. *Journal of Abnormal Psychology*, *116*(3), 638.

Woods, S. B., Bridges, K., & Carpenter, E. N. (2019). The critical need to recognize that families matter for adult health: A systematic review of the literature. *Family Process*. https://doi.org/10.1111/famp.12505

Woods, S. B., & Denton, W. H. (2014). The biobehavioral family model as a framework for examining the connections between family relationships, mental, and physical health for adult primary care patients. *Families, Systems, & Health*, *32*(2), 235–240.

Woods, S. B., Priest, J. B., & Denton, W. H. (2015). Tell me where it hurts: Assessing mental and relational health in primary care using a biopsychosocial assessment intervention. *The Family Journal*, *23*(2), 109–119.

Woods, S. B., Priest, J. B., Fish, J. N., Rodriguez, J. E., & Denton, W. H. (2014). Relational discord in urban primary care: Prevalence rates and psychiatric comorbidities. *Journal of Health Care for the Poor and Underserved*, *25*(2), 913–929.