

Marquette University

e-Publications@Marquette

---

Psychology Faculty Research and Publications

Psychology, Department of

---

2009

## Laying the Foundation for Progress Research in Family, Couple, and Individual Therapy: The Development and Psychometric Features of the Initial Systemic Therapy Inventory of Change

William M. Pinsof

*The Family Institute at Northwestern University*

Richard E. Zinbarg

*The Family Institute at Northwestern University*

Jay L. Lebow

*The Family Institute at Northwestern University*

Lynne M. Knobloch-Fedders

*Marquette University*, [lynne.knobloch-fedders@marquette.edu](mailto:lynne.knobloch-fedders@marquette.edu)

C. Emily Durbin

*The Family Institute at Northwestern University*

*See next page for additional authors*

Follow this and additional works at: [https://epublications.marquette.edu/psych\\_fac](https://epublications.marquette.edu/psych_fac)



Part of the [Psychology Commons](#)

---

### Recommended Citation

Pinsof, William M.; Zinbarg, Richard E.; Lebow, Jay L.; Knobloch-Fedders, Lynne M.; Durbin, C. Emily; Chambers, Anthony L.; Latta, Tara A.; Karam, Eli; Goldsmith, Jacob Z.; and Friedman, Greg, "Laying the Foundation for Progress Research in Family, Couple, and Individual Therapy: The Development and Psychometric Features of the Initial Systemic Therapy Inventory of Change" (2009). *Psychology Faculty Research and Publications*. 369.

[https://epublications.marquette.edu/psych\\_fac/369](https://epublications.marquette.edu/psych_fac/369)

---

**Authors**

William M. Pinsof, Richard E. Zinbarg, Jay L. Lebow, Lynne M. Knobloch-Fedders, C. Emily Durbin, Anthony L. Chambers, Tara A. Latta, Eli Karam, Jacob Z. Goldsmith, and Greg Friedman

Marquette University

**e-Publications@Marquette**

***Psychology Faculty Research and Publications/College of Psychology***

***This paper is NOT THE PUBLISHED VERSION.***

Access the published version via the link in the citation below.

*Psychotherapy Research*, Vol. 19, No. 2 (March 2009): 143-156. [DOI](#). This article is © Taylor & Francis (Routledge) and permission has been granted for this version to appear in [e-Publications@Marquette](#). Taylor & Francis (Routledge) does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Taylor & Francis (Routledge).

# Laying The Foundation for Progress Research in Family, Couple, And Individual Therapy: The Development and Psychometric Features of The Initial Systemic Therapy Inventory of Change

**William M. Pinsof**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

**Richard E. Zinbarg**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

**Jay L. Lebow**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

**Lynne Marie Knobloch-Fedders**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

### **Emily Durbin**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

### **Anthony Chambers**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

### **Tara Latta**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

### **Eli Karam**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

### **Jacob Goldsmith**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

### **Greg Friedman**

Center for Applied Psychological and Family Studies, The Family Institute at Northwestern University, Evanston, Illinois & Department of Psychology, Northwestern University, Evanston, Illinois

## **Abstract**

This article details the development and methodological characteristics of the Systemic Therapy Inventory of Change (STIC), the first measurement system designed to assess change in family, couple, and individual therapy from a multisystemic and multidimensional perspective. The article focuses specifically on the developmental process that resulted in the five valid and reliable scales that comprise the core measure of the system, the INITIAL STIC, which is administered to clients just before beginning therapy. The scales focus on five systemic domains: individual adult, family of origin, couple, family, and individual child. This article describes the five system scales, the results of the factor analytic process that created them, as well as data on their convergent and discriminant validity.

## **Keywords**

couples and family systems therapy, integrative treatment models, outcome research, process research, alliance

Client-focused progress research (Howard, Moras, Brill, Martinovich, & Lutz, [16]; Lambert, Hansen, & Finch [21]) quantitatively tracks client change over the course of therapy and feeds these change data back to therapists and other clinical stakeholders during therapy. The progress research paradigm was developed by Ken Howard and colleagues at Northwestern University (Lyons, Howard, O'Mahoney, & Lish, [22]). They created the COMPASS (Grissom & Howard, [13]; Howard, Martinovich, & Black, [15]) to measure outpatients' well-being, symptoms, role performance, and experience of the therapeutic alliance. The paradigm was extended by Michael Lambert and colleagues at Brigham Young University (Lambert, Gregersen, & Burlingame, [19]; Mueller, Lambert, & Burlingame, [26]; Umphress, Lambert, Smart, Barlow, & Clouse, [33]). Their Outcome Questionnaire-

45 (OQ-45) assesses outpatient well-being, symptoms, and current relationships. In Chicago, Scott Miller and Barry Duncan (Duncan, Sparks, & Miller, [10]; Miller, Duncan, & Hubble, [24]; Miller, Duncan, Sorrell, & Brown, [25]) created a simple measurement system with two items for assessing outpatient well-being and the alliance.

Outside the United States, Hans Kordy (Kordy & Bauer, [18]; Lambert, Hannover, Nisslmuller, Richard, & Kordy, [20]; Percevic, Lambert, & Kordy, [27]), in collaboration with Lambert, developed the Heidelberg–Stuttgart system for tracking inpatient change. At the University of Leeds, Michael Barkham and colleagues (Barkham, Gilbert, Connell, Marshall, & Twigg, [4]; Cahill et al., [9]; Evans et al., [12]) created the 34-item CORE to track outpatients' global distress, well-being, problems/symptoms, and life/social functioning.

Progress research has produced methodologically sound instruments as well as findings that suggest that providing therapists with feedback about client change improves efficacy and efficiency. However, this research has been limited to individual clients in individual outpatient or inpatient therapy. Additionally, assessment in these studies has primarily been limited to individual variables, ignoring other interpersonal systems, which emerging research suggests are associated with outcome in individual therapy (e.g., Pinsof, Zinbarg, & Knobloch-Fedders, [30]; Zinbarg, Lee, & Yoon, [35]). None of the measures has been developed to study change in family or couple therapy or from a multisystemic perspective in individual therapy. Whether the promising findings from individual therapy will generalize to family and couple therapies remains to be seen. Future systemic research on psychotherapy, particularly couple and family therapy, requires a new client change measurement instrument.

This article presents the first measurement system designed to track change in family and couple therapy as well as the first measurement system that brings a multisystemic perspective to the study of client change in individual therapy: the Systemic Therapy Inventory of Change (STIC). The STIC measurement system consists of three distinct instruments. The first and the focus of this article is the INITIAL STIC, which is administered to clients before therapy begins. The second instrument, derived from the INITIAL STIC, is the INTERSESSION STIC. It is substantially shorter than the INITIAL STIC and is designed to be administered to clients before every therapy session after the first. The third instrument—the Short Form Integrative Psychotherapy Alliance Scales (Pinsof et al., [30])—is administered to clients along with the INTERSESSION STIC.

This article presents two studies. Study I delineates the development of the INITIAL STIC instrument, focusing primarily on the factor analytic process through which it was created. Study II preliminarily explores the convergent and discriminant validity of the INITIAL STIC. Last, we explore the scientific and clinical implications of the STIC and the measurement system of which it is a part. Because of space limitations, we do not present data on the INTERSESSION STIC, the alliance scales, or the Internet-based STIC feedback system.

## Instrument Development Guidelines

Five methodological/theoretical guidelines informed the development of the STIC measurement system. The first was that the STIC would use client self-report as opposed to observational methodology. Because the STIC was developed to build a large and growing data base eventually containing thousands of cases, client self-report was the most efficient and cost-effective data collection method. In addition, this method allows therapists to be immersed in client feedback in real time. Such data immersion is difficult on a large scale with observational methods. Also, if there is an outcome "bottom line," it is client self-report.

The second guideline—that change would be measured on a session-by-session basis—was important for two reasons. First, to provide a comprehensive picture of the process and shape of change in psychotherapy, change must be measured on a session-by-session basis. Second, it would be more routine for clients to learn to complete a questionnaire before every session than before sessions on irregular or intermittent schedules. This guideline necessitated creating two versions of the STIC. It seemed reasonable to ask clients to spend about an

hour before their first therapy session (either at home or in the office) completing an extensive battery of measures, including demographic data. This battery is the INITIAL STIC. It also seemed necessary to create a shorter version that would not take more than 5 to 7 min to complete before every subsequent session after the first. This became the INTERSESSION STIC.

The third guideline was to bring a multisystemic perspective to the change process in individual, couple, and family therapies. Thus, the STIC aims to capture change in the primary systemic domains of a person's life: individual adult, individual child, couple/marital, and family. This allows the study of how changes in one domain relate to changes in others. For instance, how does individual therapy affect a client's marriage, family, and children? Similarly, to what extent does couple therapy affect individual adult, child, and family functioning.

As part of this guideline, we also took an integrative perspective on client change and the psychotherapeutic process. We did this in two ways. The first was that within each system domain we focused on behavior, cognition, and emotion. We also wanted to add a historical perspective to our primarily here-and-now view of a client system. To accomplish this, we decided to add a fifth system scale to the INITIAL STIC to focus on the reporting clients' family of origin when they were growing up. We hoped this integrative perspective would make the STIC relevant to therapists from different intellectual traditions (e.g., behavioral, cognitive, affective, psychodynamic), encouraging movement toward a more universal or generic understanding of change.

The fourth guideline was to provide a rich, empirically based, and clinically relevant "picture" of a case and its change process. Within each system scale (individual adult, family of origin, couple/marital, current family, and child), the INITIAL STIC aims to give therapists a detailed, multidimensional, and clinically rich picture of a client's initial status, and the INTERSESSION STIC gives a comparable picture of the case's change process on key variables that can differ from case to case. Ultimately, the goal was to create a multidimensional assessment framework that would be broad enough for therapists to test idiographic hypotheses about change for each case. We believe that multidimensional assessment better aids clinician efforts to delineate the "target" problems that define a case and represent the change foci for therapy.

The fifth guideline was, in addition to providing information about initial status and client change on multiple dimensions in different systemic domains, to provide information about the therapeutic alliance over the course of therapy. The alliance is the process variable most consistently linked to positive outcomes in treatment research (Horvath & Greenberg, [14]). Thus, the INTERSESSION STIC is administered along with the Short Form Integrative Psychotherapy Alliance Scales (Pinsof, [28]; Pinsof & Catherall, [29]; Pinsof et al., [30]), which measure the alliance from an integrative and systemic perspective in family, couple, and individual therapy.

## Study I: Creating the Initial stic

The five system scales that comprise the INITIAL STIC were created in the four-step process detailed next. However, before delineating that process, it is worth noting that what we report was not our initial scale development strategy. We began this program of research by selecting what we thought were the very best extant measures in each of the five domains. We administered them as a package twice (before therapy and just before the eighth session) to more than 200 clients over a 3-year period. We then conducted a variety of analyses and selected a smaller sample of items from each measure that correlated highly with the original dimension on which they loaded and that were sensitive to change. We organized these items into a comprehensive measure and sought copyright release for the items we had selected from the copyright holders of the measures we had selected. All of the copyright holders declined to release the subsets of items we had selected, and some even threatened legal action if we used them. Under legal advice, we discarded that derived measure and started from scratch with the development process outlined next.

## Step I: Domain, Dimension and Initial Item Identification

Having decided to develop a multisystemic and multidimensional measure of client change from scratch and after delineating the criteria to guide its creation, our research team decided to focus the instrument on five domains or interpersonal systems: Individual Problems and Strengths (IPS), Family of Origin (FOO), Relationship with Partner (RWP), Family/Household (FH), and Child Problems and Strengths (CPS). We chose these domains because we viewed them as the five most clinically relevant systems that could be consistently investigated in family, couple, and individual therapy.

We considered adding two additional domains that would focus on work/school and social relations/friends but did not because increasing the STIC (INITIAL and INTERSESSION) to seven domains would have made it too long. However, we included items and dimensions from these domains in the IPS and the CPS Scales.

We wanted the FOO Scale to focus on adults' recollections of their family of origin when they were growing up. The rationale for including it as a systemic domain in the INITIAL STIC was not that the family-of-origin recollections of adults would constitute a primary outcome domain but rather that such recollections might predict client change. In line with the growing interest in attachment theory in family psychology (Wood, [34]), we believe that recollected information about family-of-origin relations may moderate client change, particularly in couple and family therapy. It may also moderate alliance development in individual, couple, and family therapy.

On the basis of our knowledge of the literature and clinical experience, we generated a list of the most clinically relevant dimensions (10–15) within each domain. Typical dimensions included depression and anxiety symptoms in the IPS and CPS domains; sexual dissatisfaction, intimacy, and conflict resolution in the RWP domain; and communication, roles, and problem solving in the FH domain. After generating the dimension list in each domain, we generated a brief preliminary list of generic items to exemplify the questions that we wanted for each domain as well as our typical 5-point Likert scale response format (e.g., *never, rarely, sometimes, often, all of the time*). Typical items in each domain were "We argue with each other too much" (FH), "I love my partner" (RWP), "My child does not get along with other children" (CPS), and "During the past month, how well have you been doing in the following areas: ( 1) performing work/school/household tasks or ( 2) managing day-to-day life?" (IPS).

We viewed the generation of the set of dimensions for each domain as the articulation of a low-level or first-order factorial model within that domain. The lists of original dimensions generated by our research team within each domain are not presented here as a result of space limitations but are available from William M. Pinosof on request. Because our research team included a variety of clinician-researchers from different intellectual traditions (integrative, cognitive-behavioral, family systems, psychodynamic, experiential), our specification of the first-order set of dimensions within each domain did not derive from one distinct theoretical model but rather from an integrative, multisystemic perspective on family systems and therapeutic change.

## Step II: Clinical Expert Item Generation

For the second step, we approached four renowned clinicians in the Chicago area who each had more than 20 years of clinical and teaching experience. We invited the adult psychiatrist (Jesse Viner, MD) in the group to generate the items for the IPS domain; the clinical psychologist (Jean Goldsmith, PhD) to generate the items for the RWP domain; the clinical psychologist (JG) and the clinical social worker (Douglas Breunlin, MSW) to generate the items for the FH domain; and the child psychiatrist (Neil Fialkow, MD) to generate the items for the CPS domain. We asked them to generate three to five items for each dimension, for a total of approximately 60 to 100 items within their domain.

This group was chosen for two reasons. First, they were highly experienced and thoughtful clinicians. Second, they could work within what intellectual property law calls a "clean room procedure." By and large, they were not active researchers and were not familiar with the major research instruments that had been developed to study their particular domain. In generating items, we asked them to exclusively use their own creativity and clinical knowledge. Items were to be likely to detect change over the course of therapy and had to be in simple, common language. Any items they created that resembled items from other extant research measures were the product of independent creation within the "clean room" and, therefore, did not infringe on the copyright or patent protection of those other measures.

For the FOO Scale, we used the item pool that our expert psychologist and social worker generated for the current FH scale. We changed the instructions to ask clients to focus on their family of origin when they were growing up and changed the items to the past tense. This made the FH and the FOO scales as similar as possible to facilitate comparison.

Each expert prepared a list of items that our research team organized into five Long Form INITIAL STIC scales. The IPS Long Form Scale had 39 items divided into three parts: personal well-being (six items), current life functioning (21 items), and attitudes about myself (12 items). The FOO Long Form Scale had 104 items and the RWP Long Form Scale 90 items. The FH Long Form Scale had 104 items plus an additional six items asking about current relationships with specific family members (e.g., for children with their mother, father; for parents with their children) and with the family overall. The CPS Long Form Scale had 64 items. In the long-form questionnaires, the items in each scale were scrambled so that items from the same dimension were not spatially associated with each other.

### Step III: Administration of the INITIAL STIC Scales—Long Form: Data Collection

#### *Participant Recruitment*

For each STIC system scale, we administered a long-form scale to clients seeking individual, couple, or family therapy at our Family and Child Clinic. Clients were recruited over the phone by their therapists when the therapist called to arrange the first appointment. Therapists were practicum students in an accredited master's program in marital and family therapy or an accredited doctoral program in clinical psychology at Northwestern University. Seventy-four therapists recruited all of the client-participants.

If clients consented to participate, they were mailed and asked to complete a packet of long-form response scales and the research participation consent form before the first session. If they were partnered adults living with their children, they filled out the RWP, FH, and CPS Long-Form Scales (the couple/family packet). If they were partnered adults without children they completed the IPS, FOO, and RWP Long-Form Scales (couple packet). If they were single, unpartnered adults without children, they completed the IPS and FOO Long-Form Scales (single-adult packet). If they were unpartnered adults living with children at least some of the time, they filled out the IPS, FH and CPS Long-Form Scales (single-adult/family packet). Clients received the demographically appropriate scale packet, regardless of the type of therapy they sought. Approximately 50% of the clients sought individual therapy, 30% couple therapy, and 20% family and/or child therapy.

To eliminate data dependency in our final samples, we used one respondent from a couple or family. This cut our final sample sizes for couples and families almost in half (e.g., from approximately 250 to 134 for the RWP Scale). When we had multiple respondents from the same family on the RWP, FH, or CPS Long-Form Scales, a research assistant randomly selected one adult's form on each scale from that couple or family.



## Participants

### *Individual problems and strengths*

The IPS Long-Form Scale was completed by 188 clients (71 men [38%], 117 women [62%]). Their average age was 31 years. Most were single, although 53 (28%) were married, cohabiting, or engaged. More than 70% (132) were European Americans, with African Americans and Latinos comprising another 12% (23) and 8% (15), respectively. Eighteen (10%) did not specify ethnicity or race. More than 23% (44) had bachelor's degrees, over 14% (27) had advanced degrees (master's or doctorate), and over 94% (177) had graduated high school. Median income was just over \$30,000.

### *Family of origin*

The average age of the 169 clients (69 men [41%], 100 women [59%]) who completed the FOO Long-Form Scale was 39 years. Of these clients, 105 (62%) were married, cohabiting, or engaged; 27 (16%) were divorced or separated; and 37 (22%) were single and never married. In terms of ethnicity, (85%) were European American, 7 (4%) were African American, and 9 (5%) were Asian American. Nine (5%) did not specify ethnicity or race. Over half of the sample (85) had completed college, and almost 40% (67) had advanced degrees. Median annual income was just over \$50,000.

### *Relationship with partner*

The average age of the 134 clients (64 men [48%], 70 women [52%]) completing the RWP Long-Form Scale was 41 years. Over 80% (108) had completed college, and more than half (55) of this group had advanced degrees. Almost 70% (93) were married, 14% (19) were cohabiting, and 16% (22) were either engaged or in a committed relationship. Eighty-eight percent (118) were European American and 4% (5) were African American. Eight percent (11) did not specify ethnicity or race. Four percent (5) of the respondents were in gay or lesbian relationships. Median household income was approximately \$85,000.

### *Family household*

Ninety-nine clients (44 men [44%], 55 women [56%]) with at least one child between the ages of 4 and 18 years completed the FH Long-Form Scale. Average age was 45 years. Eighty-four (83) percent had completed college, and of this group 52% (51) had advanced degrees. Twenty-one (21) percent were divorced or separated. Eighty-four (83) percent were European American, 5% (5) were African American, and the rest did not specify ethnicity or race. Almost 66% (65) of the sample reported total household incomes greater than \$100,000.

### *Child problems and strengths*

Ninety-seven parent-clients (35 men [36%], 62 women [64%]) with at least one child between the ages of 4 and 18 years completed the CPS Long-Form Scale. Their average age was 44 years. Over 75% (73) had completed college, and of this group, almost 60% (43) had advanced degrees. Median household income was just over \$100,000. More than 25% (25) of the parents were divorced or separated. More than 90% (88) of the reporting parents were European American and just over 3% (3) were African American. The rest did not specify ethnicity or race.

## Step IV: Factor Analysis and the Creation of the Five INITIAL STIC System Scales

Our factor analytic approach represents an intermediate point in the progression from exploratory factor analysis (EFA) to confirmatory factor analysis (CFA), as described by Anderson and Gerbing ([3]). In pure EFAs, there is no a priori model specification. In creating the STIC, we began with a low-level a priori measurement model that specified the number of first-order factors (dimensions) and the items that loaded on each factor for each scale. However, as Anderson and Gerbing noted is often the case, our initial model failed to provide acceptable fit. We engaged in what they called "the necessary respecification and reestimation using the same data," which ultimately resulted in a model that met CFA criteria but was "not exclusively confirmatory" (Anderson & Gerbing, [3], p. 412).

We used a CFA statistical program in contrast to an EFA program for two reasons. First, we had a hypothesized model of the first-order factor structure of each of the five STIC system scales (based on the dimensions used to generate the original long-form items). CFA programs, in contrast to EFA programs, generate indexes of fit that specify how well the observed covariance matrix corresponds to the hypothesized model. Second, CFA programs deal better with different factor relationships (e.g., orthogonal, oblique, hierarchical). With EFA programs, all factors are constrained to be orthogonal or the correlations among them are freely estimated. With CFA programs, a given factor can be constrained to be orthogonal with a second factor, but its correlation with a third factor can be freely estimated (Bollen, [7]).

CFA involves three steps. The first specifies the factors and the items that should load on each of them. This step implies a covariance structure in which items that load on a factor covary more with each other than with items not hypothesized to load on that factor. The second step compares the covariance structure implied by the model with the observed covariance structure. Subtracting the hypothesized covariance structure from the observed structure produces a matrix of residual covariances: the portions of the observed covariances among the items not accounted for by the model. Finally, CFA computes model fit indexes from the residuals that specify how well the implied structure fits the observed.

We used three fit indexes. For the first, the comparative fit index (CFI), we set a criterion of at least .9 as an acceptable level. For the second, the root mean square error of approximation (RMSEA), and the third, the standardized root mean square residual (SRMR), we followed the recommendations of Hu and Bentler ([17]) and set criteria of .06 or less and .08 or less, respectively. Although the optimal scale derived from this procedure should have a factor structure that met all three criteria, we followed the recommendations of Marsh, Hau, and Wen ([23]) against using cutoff scores rigidly.

The factor analytic procedure that created the STIC scales was a process of successive approximation, eventually producing the set of items and factors that best met the three CFA indexes of fit. Initially, we took the dimensions we gave our experts and the items they generated for them as our a priori model. None of the a priori models converged on the first try. We then looked at the correlations between our original factors as well as the correlated error terms for the items. Using these data, as well as modification indexes estimating what item loadings would have been on factors other than the ones we had prespecified, we moved items to different dimensions and collapsed highly correlated dimensions. Typically, the first respecification produced better results but still did not converge. We used data from that analysis to organize another factor structure that we tested in a second respecification. Eventually, this successive respecification process resulted in five scales with factor structures that converged and met the fit criteria. The limitations of this procedure are discussed later.

The CFAs for the IPS, FOO, and RWP Scales were based on the long forms that had been completed in full by respondents. In contrast, the CFAs for the FH and CPS Scales, which had lower sample sizes, included participants with missing data for some items based on full information maximum likelihood estimation.

## Results: The STIC Scales and their Dimensions

The factor analytic procedure described previously resulted in the five INITIAL STIC scales. The fit criteria findings of the final models are presented in Table I. Three of the five INITIAL STIC scales emerged with nine factors and two with eight. We viewed each of these factors as measuring a distinguishable dimension of the system domain tapped by the scale. In the following discussion of the measures, we refer to the set of items comprising each dimensional factor as a subscale. The factor structure for each of the five scales that emerged in the factor analytic procedure, the number of items that comprised each factorial subscale, the mean score, standard deviation and range of scores, as well as the alpha reliabilities for each subscale on the initial samples are presented in Table II, Table III, Table IV, Table V, Table VI. The tables also present representative items from

each subscale. Each scale and factor structure presented represents the system scale models that best met the fit criteria.

Table I. STIC Scales and Confirmatory Factor Analysis Indexes of Fit

STIC scale	N	CFI $\geq$ .90	RMSEA $\leq$ .06	SRMR $\leq$ .08
Individual Problems and Strengths	188	.94	.06	.05
Family of Origin	169	.93	.08	.07
Relationship with Partner	134	.95	.06	.05
Family/Household	99	.95	.05	.06
Child Problem and Strengths	97	.90	.06	.08

Note. STIC = Systemic Therapy Inventory of Change; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual.

Table II. STIC Individual Problems and Strengths Scale: CFA Factors/Dimensions and Alpha Reliabilities

Factor/dimension	No. items	M	SD	Range	$\alpha$
Negative Affect (e.g., Felt sad most of the day; Felt tense or anxious; How well have you been getting along emotionally these days?)	8	2.46	0.76	1.00–4.75	.89
Disinhibition (e.g., Had urges or impulses you could not control)	3	1.37	0.55	1.00–3.67	.59
Life Functioning (e.g., Managing day-to-day life)	2	3.83	0.91	1.00–5.00	.84
Open Expression (e.g., I can speak up for myself when the situation calls for it)	2	3.69	0.84	1.50–5.00	.62
Flexibility/Resilience (e.g., When I get upset, I find healthy ways to make myself feel better)	3	3.37	0.78	1.33–5.00	.74
Self-Misunderstanding (e.g., I don't understand why I do the things I do)	2	2.26	0.89	1.00–5.00	.56
Substance Abuse (e.g., Drank too much alcohol)	2	1.40	0.66	1.00–5.00	.54
Self-Acceptance (e.g., I am comfortable with who I am)	2	3.34	0.94	1.00–5.00	.69

Note. N=188. STIC = Systemic Therapy Inventory of Change; CFA = confirmatory factor analysis.

Table III. STIC Family of Origin Scale: CFA Factors/Dimensions and Alpha Reliabilities

Factor/dimension	No. items	M	SD	Range	$\alpha$
Positivity (e.g., I knew I was loved in my family)	5	3.51	0.95	1.00–5.00	.90
Negativity (e.g., I felt like nobody really understood me)	5	2.69	1.00	1.00–5.00	.90
Mutuality of Expectations (Clear Expectations) (e.g., I knew the right thing to do in my family)	2	3.62	1.05	1.00–5.00	.80
Family Pride (e.g., I was proud of my family)	2	3.82	1.04	1.00–5.00	.86
Intrusiveness (e.g., You could be pretty sure that if you needed to be alone, someone in my family was going to bother you)	2	2.49	0.96	1.00–5.00	.70
Physical Abuse (e.g., Someone in my family pushed people around physically to get his or her way)	2	1.75	1.08	1.00–5.00	.86
Sexual Abuse (e.g., There was inappropriate sexual behavior between some of the members of my family)	2	1.20	0.72	1.00–5.00	.97
Abusive Climate (e.g., I was afraid of someone in my family)	2	2.12	1.22	1.00–5.00	.83

Substance Use (e.g., Someone in my family thought that I drank alcohol too much)	3	1.59	0.88	1.00–5.00	.77
--	---	------	------	-----------	-----

Note. N=169. STIC = Systemic Therapy Inventory of Change; CFA = confirmatory factor analysis.

Table IV. STIC Relationship with Partner Scale: CFA Factors/Dimensions and Alpha Reliabilities

Factor/dimension	No. items	M	SD	Range	$\alpha$
Positivity (e.g., We enjoy doing things together; I love my partner; My partner really listens to me when we discuss things; We are each other's best friend)	8	3.84	0.78	1.62–5.00	.92
Trust/Betrayal (e.g., I feel betrayed by my partner)	3	4.02	0.95	1.00–5.00	.87
Commitment (e.g., I am sure we will make it as a couple)	2	4.04	0.96	1.00–5.00	.81
Inequity (e.g., I am expected to do too much)	2	2.21	0.88	1.00–5.00	.36
Anger/Contempt (e.g., I am filled with anger toward my partner)	2	2.19	0.87	1.00–4.50	.74
Sexual Dissatisfaction (e.g., I am sexually frustrated in this relationship)	2	2.31	0.95	1.00–5.00	.84
Physical Abuse (e.g., We get into shoving or hitting each other when we fight)	2	1.22	0.43	1.00–3.50	.60
Substance Abuse (e.g., My partner uses drugs or alcohol too much)	2	1.56	0.72	1.00–5.00	.22

Note. N=134. STIC = Systemic Therapy Inventory of Change; CFA = confirmatory factor analysis.

Table V. STIC Family/Household Scale: CFA Factors/Dimensions and Alpha Reliabilities

Factor/dimension	No. items	M	SD	Range	$\alpha$
Physical Abuse (e.g., Someone in my family is physically abusive to other family members)	2	1.38	0.73	1.00–4.00	.78
Abusive Climate (e.g., I feel abused by someone in my family)	2	1.44	0.70	1.00–4.00	.67
Sexual Abuse (e.g., There is someone in my family who is sexually abusive to other family members)	2	1.03	0.11	1.00–1.50	.03
Positivity (e.g., People in my family respect each other's feelings and thoughts; We feel loved and supported by each other; We know how to have fun together; I know my family will be there for me; People in my family are honest with each other)	10	4.06	0.66	2.20–5.00	.93
Decision Making (e.g., Everyone has a say in my family)	2	3.68	0.66	2.00–5.00	.61
Negativity (e.g., Talking together as a family is a nightmare; If people in my family know what you are really feeling, they use it against you; I feel like a prisoner in my family; My family is too much in my business)	8	2.11	0.78	1.00–4.88	.89
Boundary Clarity (e.g., I know the right thing to do in my family)	2	3.89	0.69	2.00–5.00	.53
Feeling Misunderstood (e.g., I feel like nobody in my family really understands me)	2	2.45	1.06	1.00–5.00	.87
Family Pride (e.g., I'm proud of my family)	2	4.24	0.79	2.00–5.00	.65

Note. N=99. STIC = Systemic Therapy Inventory of Change; CFA = confirmatory factor analysis.

Table VI. STIC Child Problems and Strengths Scale: CFA Factors/Dimensions and Alpha Reliabilities

Factor/dimension	No. items	M	SD	Range	$\alpha$
Depression (e.g., My child is sad)	3	2.30	0.61	1.00–4.33	.55
Anxiety (e.g., My child worries)	3	2.18	0.66	1.00–4.00	.56
Antisocial (e.g., My child starts physical fights; My child blames others for his/her problems; My child lies)	6	2.52	0.82	1.17–4.67	.69
Self-Control/Impulsivity (e.g., My child has difficulty controlling his/her reactions)	4	2.84	0.88	1.00–5.00	.82
Substance Abuse (e.g., My child uses marijuana or other illegal drugs)	2	1.15	0.62	1.00–5.00	.64
Parent–Child Alliance (e.g., My child enjoys spending time with his/her parents)	2	3.57	0.89	1.00–5.00	.85
Prosocial (e.g., My child is thoughtful and caring toward others)	3	3.17	0.76	1.67–5.00	.79
Food/Weight (e.g., I am concerned about my child's attitude about food)	2	2.08	1.02	1.00–5.00	.60
Social/Academic (e.g., My child does well at school)	4	3.52	0.65	1.75–5.00	.70

Note. N=97. STIC = Systemic Therapy Inventory of Change; CFA = confirmatory factor analysis.

### *Individual Problems and Strengths*

As shown in Table I, the IPS factor model met all three fit criteria. This best fitting model, described in Table II, had eight factors (subscales) with 24 items. The factor with the largest number of items was Negative Affect, which encompasses depression, anxiety, and well-being. IPS subscale reliabilities ranged from .54 to .89.

### *Family of Origin*

As indicated in Table I, the FOO factor model met two of the three fit criteria (CFI and SRMR) and came close to the third (RMSEA). Table III shows that the best fitting FOO scale resulted in nine factors covering 22 items. Alpha reliabilities for the nine subscales ranged from .70 to .90.

### *Relationship with Partner*

Table I shows that the RWP Scale factor model met all three fit criteria. As Table IV illustrates, the best fitting model had eight factors with 25 items. The factor with the largest number of items, Positivity, covered aspects of couple life, such as fun, love, communication, and intimacy. Two subscales had alphas below .50: Inequity (.36) and Substance Abuse (.22). Despite these poor alphas, we retained these subscales for their clinical relevance and because the final model did not converge without them. The rest of the alphas ranged from .60 to .92.

### *Family/Household*

Table I reveals that the best fitting FH factor model met all three fit criteria. As Table V shows, this model included nine factors covering 32 items. The two factors with the largest number of items, Positivity (10) and Negativity (8) tapped, respectively, aspects of family life like respect, nurturance, fun, loyalty, and good communication on the one hand and bad conflict, nastiness, feeling trapped, and intrusiveness on the other. The alpha estimate for Sexual Abuse was probably quite low (.03) in large part because the items were rarely endorsed within our sample. However, we retained it because of its clinical relevance and the model did not converge without it. Other FH subscale alphas ranged from .53 to .93.

### *Child Problems and Strengths*

Table I shows that final CPS factor model met all three fit criteria. Table VI reveals that this model had nine factors with 29 parent-report items. The largest factor, Antisocial (six items), taps aspects of child functioning like aggression, defiance, and delinquency. CPS subscale alphas ranged from .55 to .85.

### Discussion of Factor Analytic Results

The factor analyses of the INITIAL STIC scales resulted in five scales with 43 empirically supported subscales, which had, with some exceptions, adequate to good internal consistency reliabilities. The factor analytic procedure that we used resulted in a number of different dimensions with different numbers of items for each of the five system scales. The final set of factors (subscales) for each scale was virtually the only solution of the many that we tried that worked. For instance, once the final model for a scale emerged, if we deleted an item from a factor or an entire factor, the model would not work (i.e., meet the fit criteria). Thus, the final version of each scale's factors was the only version of our low-level theoretical model that was supported by the data. Of course, we did not try every possible permutation of each scale's items and potential factors, but the final model was the product of an active sequential process that examined many factorial permutations before finding one that worked.

In the factor analytic progression discussed by Anderson and Gerbing ([3]), ideally a factor structure that emerges out of a procedure such as ours that is intermediate between EFA and CFA should be replicated on a sample that differs from the one on which it was derived. Unfortunately, our sample was too small to divide into a derivation subsample and a cross-validation subsample. We intend to test the factor structure reported in this article in subsequent studies with a nonclinical sample as well as other clinical samples. Until that replication, we cannot say definitively that we have confirmed the structures that emerged from our intermediate factor analytic process.

That the alpha reliabilities of 18 subscales were low (below .70) is a major methodological concern. That the existence of the factors was supported by the factor analysis procedure and that their reliabilities were low suggest that they "exist" but that our subscales were not measuring them adequately. A problem with a number of the low-reliability subscales is the low number (two) of items that they contain. In future iterations of the STIC, we plan to add specific items to increase the reliability of the subscales with low alphas.

Another possible explanation for some of the low alpha estimates is that there was restricted variability in our sample. We believe this to be the case for our physical and sexual abuse, suicide (depression), substance abuse, and food/weight items in our sample. As is the case with many outpatient programs, the population from which our sample was drawn was not extremely disordered, and the problems targeted by the just-mentioned subscales tended to have relatively low frequencies and, hence, restricted variabilities within our sample. More definitive assessment of the reliabilities of these dimensions must await future studies with more disordered populations or samples with specific disorders (e.g., abuse, addiction).

On the basis of the construct and predictive validity alliance scale findings from Pinsof et al.'s ([30]) study that presented the factor analysis and validity data for the alliance scales in the STIC measurement system, we knew not only that two of the three alliance (individual and couple) scales contained empirically validated factors but also that the subscales measuring those factors were significantly associated with progress in individual and couple therapy. Building on that research, the next question that emerged for us was whether the five factorially valid system scales (and their subscales) that we found for the INITIAL STIC would have convergent validity: Would their subscales correlate as expected with other well-validated and widely used research instruments? We also aimed to provide a preliminary test of their discriminant validity.

## Study 2: Construct Validity of the Initial Stic System Scales

To test the convergent validity and, in some cases, the discriminant validity of the five empirically derived INITIAL STIC scales, we conducted a preliminary evaluation with a small sample of clients (none of whom were involved in Study 1) who completed the INITIAL STIC along with one or more well-validated instruments for assessing individual adult, individual child, couple, or family functioning. We primarily examined the relationship between specific subscales of each STIC system scale and specific subscales of the criterion measures.

This section actually presents five ministudies, one for each of the STIC system scales. Each ministudy is presented independently and, with one exception, involved a different sample of client participants. All of the data were collected just before the first session of therapy. For the criterion measures, we tried to select the most widely used and well-validated measures of a similar phenomenon. In these studies, we focused primarily on convergent validity but, when possible, we addressed discriminant validity as well. We report the convergent validity findings in regard to the significant expected correlations between specific dimensions of the STIC scale and specific dimensions of the criterion measure (as appropriate) in each ministudy. We report the discriminant validity findings in regard to the specific dimensions of the STIC scales that were expected to be largely uncorrelated with specific dimensions of the criterion measure. We viewed these five ministudies as an effort to collect preliminary and initial data on the construct validity of the INITIAL STIC scales. We also hoped that the results of these studies could be used to guide further scale development.

The data on the participants in each of these studies are presented separately for each study in Table VII. The sample sizes ranged from a low of 27 for the CPS-Child Behavior Checklist (CBCL) study to a high of 87 for the FOO–Marital Satisfaction Inventory–Revised (MSI-R) and RWP-MSI-R studies. Both of the latter studies used the same sample. As evident in Table VII, in general the samples were well educated, middle to upper-middle class in income, and primarily, although not exclusively, European American.

Table VII. INITIAL STIC Sample Characteristics

	Study					
Characteristic	2a	2b	2c	2c	2d	2e
N	31	87	48	87	32	27
Men	8	33	23	33	12	9
Women	23	54	25	54	20	18
Age (years; mdn)	38	35	42	35	39	45
Race/ethnicity						
European American	25	61	41	61	24	21
African American	2	7	2	7	1	2
Latino/a	1	11	3	11	4	4
DNS	4	8	2	8	3	0
Education						
Completed college	21	65	41	65	24	20
Advanced degree	7	39	25	39	8	10
Median income	\$50,000	\$55,000	\$60,000	\$55,000	\$60,000	\$100,000
Married or cohabitating	12	35	41	41	24	22

Note. STIC = Systemic Therapy Inventory of Change; DNS = did not specify; Study 2a: Individual Problems and Strengths with Beck Depression Inventory/Beck Anxiety Inventory; Study 2b: Family of Origin with Marital Satisfaction Inventory–Revised; Study 2c: Relationship with Partner with Revised Dyadic Adjustment Scale; Study 2c: Relationship with Partner with Marital Satisfaction Inventory–Revised; Study 2d: Family Household with Family Assessment Device; Study 2e: Child Problems and Strengths with Child Behavior Checklist.

## Study 2a: INITIAL STIC IPS Scale

### *Method and Measures*

To assess the concurrent validity of the INITIAL STIC IPS Scale, which measures individual functioning, symptoms, and well-being, we administered it along with the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, [6]) and the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, [5]) to 31 clients presenting for individual therapy. These are two of the most reliable, valid, and widely used measures for assessing depression and anxiety in clinical research. Depression and anxiety are two of the major variables targeted specifically by the Negative Affect factor of IPS Scale.

### *Results*

Seven of the eight empirically derived IPS subscales correlated significantly (at least  $p < .05$ ) with the BDI. All of the significant correlations were in the expected direction (negative IPS dimensions correlated positively with the BDI and positive IPS dimensions correlated negatively). Significant correlations ranged from .81 (Negative Affect) to .38 (Self-Misunderstanding). Substance Abuse was the only subscale predicted to be largely uncorrelated with the BDI, and it indeed was the only subscale that did not correlate significantly with the BDI. Six of the eight IPS subscales correlated with the BAI. All of the significant correlations were in the expected direction. Correlations ranged from .66 (Negative Affect) to .36 (Disinhibition). Substance Abuse was again the only subscale predicted to be largely uncorrelated with the BAI and was one of only two subscales that did not correlate significantly with the BAI. The results largely support the convergent and discriminant validity of the IPS with the BDI and BAI.

## Study 2b: INITIAL STIC FOO Scale

### *Method and Measures*

We administered the STIC FOO Scale with the MSI-R (Snyder, [31]) History of Distress subscale to 87 clients seeking couple therapy. The MSI-R is a well-validated and widely used comprehensive and multidimensional measure for assessing marital functioning (described later). The History of Distress subscale is the only MSI subscale tapping family of origin functioning and one of the few extant measures addressing adults' reported experience of their family of origin.

### *Results*

Seven of nine FOO subscales correlated significantly with MSI History of Distress. The FOO Intrusiveness and Sexual Abuse subscales did not correlate with MSI History of Distress. All of the significant correlations were in the expected direction and ranged from  $-.66$  (Positivity) to .27 (Substance Use). These results support the convergent validity of the STIC INITIAL FOO scale with the MSI History of Distress subscale.

## Study 2c: INITIAL STIC RWP Scale

### *Method and Measures*

We evaluated the RWP Scale in relation to the Revised Dyadic Adjustment Scale (RDAS; Busby, Christensen, Crane, & Larson, [8]) and the MSI-R (Snyder, [31]). The RDAS is a short form of the original Dyadic Adjustment Scale (DAS; Spanier, [32]), the most widely used measure of marital satisfaction. We used the RDAS because of its improved psychometric properties relative to the DAS. The current study examined the RDAS total score, which has excellent internal consistency ( $\alpha = .88$ ).

The MSI-R (Snyder, [31]) provides a detailed assessment of marital satisfaction. The measure has 150 true–false items comprising 10 subscales, a global distress scale, and a validity scale. The MSI-R is the most comprehensive and multidimensional extant marital assessment device. It is widely used in the marital field, and studies repeatedly find good reliabilities.



As a result of several factors—adequately sized sample, men and women frequently reporting different experiences of their marriages, and the problem of data dependency in couple data—we analyzed the RWP data separately for men and women. As shown in Table VII, the sample of clients who completed the RWP and the RDAS was not the same sample who completed the RWP and the MSI. Forty-eight clients seeking couple therapy completed the RWP and RDAS. The 87 clients who completed the RWP and MSI-R were the same clients who completed the FOO and the MSI-R described previously.

### *Results*

For men, six (of eight) RWP subscales correlated with the RDAS, with correlations ranging from  $-.86$  (Positivity) to  $.53$  (Sexual Dissatisfaction). Six (of eight) RWP subscales correlated with at least one of eight (of nine) MSI subscales; there were 33 significant correlations between the MSI and the RWP in all. Correlations ranged from  $.80$  (Sexual Dissatisfaction subscales from both measures) to  $-.37$  (MSI Role Orientation with RWP Betrayal). Twenty-four of the 25 correlations predicted to be convergent validity coefficients were significant and in the expected direction. In contrast, only one of 22 correlations predicted to be discriminant validity coefficients was significant.

For women, six of the eight RWP correlated in the predicted direction with the RDAS, with correlations ranging from  $-.72$  (Positivity) to  $.43$  (Sexual Dissatisfaction). Seven of the eight RWP subscales correlated with at least one of nine MSI subscales; there were 33 significant correlations in all. Correlations ranged from  $.75$  (Sexual Dissatisfaction from both measures) to  $-.75$  (RWP Positivity with MSI Global Distress) to  $.29$  (RWP Inequity with MSI Global Distress). Twenty-two of the 26 correlations predicted to be convergent validity coefficients were significant, and they were all in the expected direction. In contrast, only seven of 21 correlations predicted to be discriminant validity coefficients were significant. For men and women, the findings from this analysis support the convergent and discriminant validity of the INITIAL STIC RWP Scale in regard to the two most widely used and well-validated measures of couple functioning.

## Study 2d: STIC INITIAL FH Scale

### *Method and Measures*

We correlated the STIC INITIAL FH scale with the Family Assessment Device (FAD; Epstein, Baldwin, & Bishop, [11]) Healthy and Unhealthy subscales for 32 adult parents seeking family therapy. The FAD is a widely used and well-validated self-report measure of family functioning. It contains 60 items on seven different scales. Participants rate each item on a 4-point Likert scale (*strongly agree, agree, disagree, strongly disagree*). Unfortunately, the sample was not large enough to conduct separate gender analyses.

### *Results*

Of the nine FH subscales, three correlated significantly with FAD Healthy. Five FH subscales correlated significantly with FAD Unhealthy. All significant correlations were in the expected direction. Correlations ranged from  $-.75$  (FH Feeling Misunderstood with FAD Healthy) to  $-.39$  (FH Decision Making with FAD Unhealthy). The results of the FH validity study were weaker than those of the studies with the other four INITIAL STIC system scales but lent partial support to the convergent validity of the FH scale.

## Study 2e: INITIAL STIC CPS Scale

### *Method and Measures*

With 27 parent-clients seeking child or family therapy, we correlated the STIC INITIAL CPS Scale with the subscales of the Child Behavior Check List (CBCL, Achenbach & Edelbrock, [1]), the most widely used child functioning measure in developmental and child psychology. The CBCL (Achenbach & Rescorla, [2]) is a 118-item standardized measure assessing internalizing and externalizing behavior problems in children between the ages of 6 and 18 years. Parents rate their child's behavior on a 3-point scale (0 = *not true*, 1 = *somewhat true*, 2 = *very*

*true or often true*). Achenbach and Rescorla ([2]) found that the CBCL had high internal consistency. Again and unfortunately, because of the small number of clients in this sample, we were unable to conduct separate gender analyses.

### *Results*

Seven of the nine CPS subscales correlated with at least one of the eight CBCL subscales. There were 37 correlations between the seven CPS and the eight CBCL subscales. Correlations ranged from .86 (CPS Antisocial with CBCL Aggressive Behavior) to  $-.41$  (CPS Prosocial with CBCL Rule Breaking Behavior). Eighteen of the 21 correlations predicted to be convergent validity coefficients were significant, and they were all in the expected direction. In contrast, only four of 17 correlations predicted to be discriminant validity coefficients were significant. The results support the convergent and discriminant validity of the INITIAL STIC CPS Scale in regard to the most widely used measure of child functioning.

### *Discussion of Construct Validity Findings*

The convergent and discriminant validity findings, on a set of largely European American, highly educated middle- to upper-middle-class client samples, provide preliminary support for the construct validity of the five INITIAL STIC system scales. Each of the scales had at least three subscales that correlated significantly with its criterion measures. Some of the strongest preliminary findings were that the Negative Affect subscale of the STIC IPS shared two thirds of its variance with the BDI and 44% of its variance with the BAI. The STIC FOO Positivity subscale shared 44% of its variance with the MSI-R Family History of Distress Scale. For men, STIC RWP Positivity shared 74% of its variance with the RDAS and 50% of its variance with MSI-R Global Distress. Similarly, RWP Anger/Contempt shared 56% of its variance with MSI-R Global Distress and 48% of its variance with the RDAS. For women, RWP Positivity shared 56% of the variance with MSI-R Global Distress and 52% with the RDAS. For men and women, RWP Sexual Dissatisfaction shared 64% and 56%, respectively, of its variance with MSI-R Sexual Dissatisfaction. The STIC CPS Depression and Anxiety subscales, respectively, shared 46% and 38% of their variance with CBCL Anxious/Depressed. CPS Anxiety also shared 49% of its variance with CBCL Attention Problems. CPS Antisocial shared 48% of its variance with CBCL Rule Breaking Behavior and 74% of its variance with CBCL Aggressive Behavior. These findings, together with the many nonsignificant discriminant validity coefficients for the RWP and the CPS, suggest that the STIC IPS, FOO, RWP, and CPS Scales are, in fact, measuring what they purport to measure.

The findings for the STIC FH Scale were not as strong as those for the other four scales. Only three of the eight FH subscales correlated with a FAD subscale beyond .50. The one strong finding for the STIC FH Scale was that Feeling Misunderstood shared, respectively, 56% and 53% of the score variance with FAD Healthy and Unhealthy.

Across the ministudies, the findings for the abuse (substance, physical, sexual, and climate) subscales on the STIC scales were also not as strong as the findings from the other scales. Only four of the seven correlations involving the abuse scales that we predicted to be convergent validity coefficients were significant. Obviously, this may reveal something about their validity. A rival hypothesis derives from the low frequency of endorsement of the items that comprise these dimensions in the outpatient population sampled in this research, which would result in decreased variability in the abuse scores. Clearly, any definitive insight into the validity of the STIC abuse dimensions must await evaluation of the STIC in samples with higher frequencies of abuse.

A limitation of these findings concerns the data from the STIC FH and CPS scales. The data were not strictly independent in both samples, with a number of the participants (almost half in each sample) coming from the same families. For instance, the 32 clients making up the FH sample came from 16 families. This fact violates a core assumption (independence) of the statistical analyses and compromises the validity and increases the

preliminary nature of the findings for the FH and CPS scales. Despite the low sample sizes, we decided to conduct these analyses in order to collect some preliminary data. However, future research needs to evaluate the concurrent validity of the STIC system scales with larger samples that do not violate the independence assumption. Additionally, we did not test the discriminant validity of the STIC FOO or FH scales, and the discriminant validity tests of the STIC IPS were very limited. Last, future research needs to examine the relationship between the STIC system scales and other widely used and well-validated criterion measures. For more information about specific subscale correlations with criterion dimensions contact William Pinsof.

## Conclusion

This article presented the developmental process and methodological characteristics of the INITIAL STIC, the core measure of a new measurement system for the multisystemic and multidimensional assessment of change in family, couple, and individual therapy. The developmental process culminated in the delineation of five scales for assessing five client systems. Four of the five INITIAL STIC scales had strong convergent validity with well-validated and widely used criterion measures. The evidence for the convergent validity of the STIC FH Scale was not as strong as that for the other four scales and remains to be determined. Some evidence of discriminant validity also emerged for the STIC RWP and CPS Scales. With several exceptions, the internal consistency reliabilities of the STIC subscales were adequate to good.

This development process created the foundational instrument of the first measurement system designed explicitly to study change in family, couple, and individual therapy from a multisystemic and multidimensional perspective. The initial results of this research suggest that the STIC has great promise as a methodologically sound and productive measure for systemic assessment and potentially for the study of change in psychotherapy.

The INITIAL STIC scales and subscales, along with the INTERSESSION STIC, have the potential to test an enormous variety of hypotheses about psychotherapy process and outcome from a multisystemic perspective. For instance, what is the relationship between change in different systems? In couple therapy with a distressed couple in which one of the spouses is depressed, does the alleviation of depression typically precede the reduction in marital distress or vice versa? Also, what is the sequential change relationship between adult depression, marital distress, and amelioration of symptoms in one of their children? The STIC measurement system also has the potential to test hypotheses about the relationship between specific dimensions of the therapeutic alliance and change in different systems. It also could be used to compare and contrast the change process in different modalities of therapy (e.g., individual, couple) and well as in different types of therapy (e.g., cognitive-behavioral, experiential, psychodynamic). These represent a fraction of the hypotheses about psychotherapy that the STIC system could be used to test and explore from an integrative, multisystemic, and multidimensional perspective.

The promising findings reported in this article need to be replicated with other (and in certain cases larger) samples. A number of studies are currently underway within our research group. One with the INITIAL STIC involves a comparison of a new clinical sample with a nonclinical, random sample of participants that is representative of the U.S. population. This study allows for the confirmation of the factor structure of the INITIAL STIC presented in this article on a new clinical sample as well as a nonclinical sample. Additionally, it will hopefully permit the determination of clinical cutoff scores on the INITIAL STIC. A second study involves evaluation of the factor structure of the INTERSESSION STIC and its capacity to detect clinically and statistically significant change. Several studies are looking at the development of the alliance in individual and couple therapy over the first 10 therapy sessions. Future research with the various components of the STIC measurement system needs to be conducted with other research teams on other more diverse and disturbed populations.

## Acknowledgements

This research was supported by the Kovler Family and A. C. Nielsen Family Foundations.

## References

- Achenbach, T. M. and Edelbrock, C.1991. *Child Behavior Checklist (CBCL)*, Burlington: University of Vermont, Department of Psychiatry and Psychology.
- Achenbach, T. M. and Rescorla, L. A.2001. *Manual for the ASEBA school-age forms & profiles*, Burlington: University of Vermont, Research Center for Children, Youth, & Families.
- Anderson, J. C. and Gerbing, D. W.1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103: 411–423.
- Barkham, M., Gilbert, N., Connell, J., Marshall, C. and Twigg, E.2005. Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry*, 186: 239–246.
- Beck, A. T., Epstein, N., Brown, G. and Steer, R. A.1988. An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56: 893–897.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. and Erbaugh, J.1961. An inventory for measuring depression. *Archives of General Psychiatry*, 4: 561–571.
- Bollen, K. A.1989. *Structural equations with latent variables*, New York: Wiley.
- Busby, D. M., Christensen, C., Crane, D. R. and Larson, J. H.1995. A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: Construct hierarchy and multi-dimensional scales. *Journal of Marital and Family Therapy*, 33: 392–405.
- Cahill, J., Barkham, M., Stiles, W. B., Twigg, E., Hardy, G. E.Rees, A.2006. Convergent validity of the CORE measures with measures of depression for clients in cognitive therapy for depression. *Journal of Counseling Psychology*, 180: 51–60.
- Duncan, B. L., Sparks, J. A. and Miller, S. D.2006. "Client, not theory, directed: Integrating approaches one client at a time". In *A casebook of psychotherapy integration*, Edited by: Stricker, G. and Gold, J.225–240. Washington, DC: American Psychological Association.
- Epstein, N. B., Baldwin, L. M. and Bishop, D. S.1983. McMaster Family Assessment Device (FAD). *Journal of Marital and Family Therapy*, 9: 171–180.
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G.Mellor-Clark, J.2002. Towards a standardized brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, 180: 51–60.
- Grissom, G. R. and Howard, K. I.2002. "Directions and COMPASS-PC". In *Handbook of psychological assessment in primary care settings*, Edited by: Maruish, M. E.255–275. Mahwah, NJ: Erlbaum.
- Horvath, A. O., & Greenberg, L.(1994). *The working alliance: Theory, research and practice*. New York: Wiley.
- Howard, K. I., Martinovich, Z. and Black, M.1997. Outpatient outcomes. *Psychiatric Annals*, 27: 108–112.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z. and Lutz, W.1996. Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51: 1059–1064.
- Hu, L. and Bentler, P. M.1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6: 1–55.
- Kordy, H. and Bauer, S.2003. The Stuttgart-Heidelberg model of active feedback driven quality management: Means for the optimization of psychotherapy provision. *International Journal of Clinical and Health Psychology*, 3: 615–631.
- Lambert, M. J., Gregerson, A. and Burlingame, G. M.2004. "The Outcome Questionnaire-45". In *The use of psychological testing for treatment planning and outcome assessment*, 3rd ed, Edited by: Maruish, M.191–234. Mahwah, NJ: Erlbaum.
- Lambert, M. J., Hannover, W., Nisslmuller, W., Richard, M. and Kordy, H.2002. Questionnaire on the results of psychotherapy: Reliability and validity of the German translation of the Outcome Questionnaire 45.2 (OQ-45.2). *Zeitschrift fur Klinische Psychologie und Psychotherapie: Forschung und Praxis*, 31: 40–46.

- Lambert, M. J., Hansen, N. B. and Finch, A. E.2001. Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69: 159–172.
- Lyons, J. S., Howard, K. I., O'Mahoney, M. T. and Lish, J. D.1997. *The measurement & management of clinical outcomes in mental health*, New York: Wiley.
- Marsh, H. W., Hau, K.-T. and Wen, Z.2004. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11: 320–341.
- Miller, S. D., Duncan, B. L. and Hubble, M. A.2005. "Outcome-informed clinical work". In *Handbook of psychotherapy integration*, 2nd ed, Edited by: Norcross, J. C. and Goldfried, M. R.84–102. London: Oxford University Press.
- Miller, S. D., Duncan, B. L., Sorrell, R. and Brown, G. S.2005. The partners for change outcome management system. *Journal of Clinical Psychology*, 61(2): 199–208.
- Mueller, R. M., Lambert, M. J. and Burlingame, G. M.1998. Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, 70: 248–262.
- Percevic, R., Lambert, M. and Kordy, H.2004. Computer-supported monitoring of patient treatment response. *Journal of Clinical Psychology*, 60: 285–299.
- Pinsof, W. M.1994. "An integrative systems perspective on the therapeutic alliance: Theoretical, clinical and research implications". In *The working alliance: Theory, research and practice*, Edited by: Horvath, A. and Greenberg, L.173–195. New York: Wiley.
- Pinsof, W. M. and Catherall, D.1986. *The Integrative Psychotherapy Alliance: Family, Couple and Individual scales*, Evanston, IL: The Family Institute at Northwestern University.
- Pinsof, W. M., Zinbarg, R. E. and Knobloch-Fedders, L.2008. Factorial and construct validity of the revised short form Integrative Psychotherapy Alliance Scales for Family, Couple, and Individual Therapy. *Family Process*, 47(3): 281–301.
- Snyder, D. K.1997. *Manual for the Marital Satisfaction Inventory–Revised*, Los Angeles: Western Psychological Services.
- Spanier, G. B.1976. Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, 38: 15–25.
- Umphress, V. J., Lambert, M., Smart, D. W., Barlow, S. H. and Clouse, G.1997. Concurrent and construct validity of the outcome questionnaire. *Journal of Psychoeducational Assessment*, 15: 40–55.
- Wood, B.(2002). Attachment theory and family therapy [Special issue]. *Family Process*, 41(3).
- Zinbarg, R., Lee, J. E. and Yoon, L.2007. Dyadic predictors of outcome in a cognitive-behavioral program for patients with generalized anxiety disorder in committed relationships: A "spoonful of sugar" and a dose of non-hostile criticism may help. *Behaviour Research and Therapy*, 45: 699–713.