This resource succinctly describes and critiques the most common assessment instruments used in systemic supervision to evaluate supervisees’ competencies. The first section focuses on measures of individual counseling skills that have been used in creative ways to assess systemic therapist development. The second section covers instruments developed to specifically assess systemic supervisees trained as couple, marriage, and family therapists. Some of the instruments were developed specifically for use by supervisors. Together, they offer systemic supervisors a catalogue of instruments to select from when methods for evaluation are within their discretion and important background information about an instrument when an educational or training program may be requiring a particular instrument in supervision. (See Chapter 19 in the accompanying book for the role of instruments within the evaluation process.) Many of these instruments have been in used in research to help supervisors identify key aspects to focus on in supervision and in therapy. Chapter 14 in the accompanying book reviews this research.

—Editors’ introduction

This article extends the work of Todahl and Perosa (2006) by focusing only on measures of trainee clinical skills typically utilized either in individual counseling
role-play simulations or of the second section on AAMFT core competencies sessions with actual clients, or in family therapy role-play simulations and sessions with client families, or in research studies on trainee development and effectiveness. Measures will be described in terms of their historical development, their psychometric properties, their intended use in clinical training or research, and the level of skills performance (e.g., single behavior or complex competency) assessed by their items and scales. They will be critiqued according to their ability to fulfill their stated purpose (e.g., within-program evaluation vs. research), their psychometric soundness, and their relation to the core competencies and outcome-based education in marriage and family therapy (MFT).

**Measures of Individual Counseling Skills**

**Supervisor rating scales**

The Skilled Counseling Scale (Urbani et al., 2002). Skills-based training models for use with psychology and school and community counseling trainees originally were developed by Carkhuff (1987) and Ivey (1971) in the 1970s and 1980s. More recently, these models have been merged and expanded upon by Smaby, Maddux, Torres-Rivera, and Zimmick (1999) to form the Skilled Group Counselor Training Model (SGCTM) to be utilized in group counseling sessions and the Skilled Counselor Training Model (SCTM) to be used with individual counseling sessions (Urbani et al., 2002). As part of this project, two measures were developed: the Skilled Group Counseling Scale (SGCS; Smaby, Maddux, Torres, & Zimmick, 1997) and a modification of that measure to assess individual skills called the Skilled Counseling Scale (SCS). The skills assessed by both measures are very similar; minor changes were made to descriptions of the skill described on the SGCS to the skill’s description on the SCS to reflect the use of the skill with an individual rather than with a group. For example, the focus of the skill of eye contact was changed from looking at the entire group to looking at the individual. The scales were designed to be used by trained raters evaluating counselors in experiential courses and also for research.

The SCS contains 18 items, each describing a single skill in terms of observable behavior (e.g., “eye contact, direct gaze with occasional breaks, the latter usually intentional”; “immediacy, recognizes immediate feelings expressed between the client and counselor when discussing the problem”). A total score and subscale scores including Attending, Questions and Reflecting, Interchangeable Empathy, Decision-Making, and Contracting can be computed. Table 6E.1 presents a description of the item format used for rating responses by respondents, reliability, and validity information for all measures reviewed in this article. A reading of Table 6E.1 shows that the SCS uses a 5-point Likert-type scale for responses and has demonstrated interrater and internal consistency reliability and content and predictive validity.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Item format</th>
<th>Focus</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled Counseling Scale (SCS)</td>
<td>5-pt. Likert: 1 = not at all; 5 = always</td>
<td>Supervisor rate</td>
<td>Interrater = 0.79–0.90; Cronbach’s alpha = 0.81–0.89 (Schaefer, Smaby, Maddux, &amp; Cates, 2005; Urbani et al., 2002)</td>
<td>Content; predictive (Eriksen &amp; McAuliffe, 2003)</td>
</tr>
<tr>
<td>Counseling Skills Scale (CSS)</td>
<td>5-pt. Likert: 1 = highly developed; −5 = major adjustment needed</td>
<td>Supervisor rate</td>
<td>Cronbach’s alpha = 0.91 (total score) (Eriksen &amp; McAuliffe, 2003)</td>
<td>Construct (Eriksen &amp; McAuliffe, 2003)</td>
</tr>
<tr>
<td>Supervisor or Peer Rating System (SOPRS)</td>
<td>5-pt. Likert: 1 = used inappropriately; 5 = used appropriately</td>
<td>Supervisor rate self-report</td>
<td>Internal consistency; test–retest internal consistency; test–retest</td>
<td>Construct; concurrent; predictive (Hill, 2001; Hill &amp; Kellems, 2002)</td>
</tr>
<tr>
<td>Counseling Activity Self-Efficacy Scale (CASES)</td>
<td>9-pt. Likert: 0 = no confidence; 9 = complete confidence</td>
<td>Observer rate</td>
<td>Interrater = 0.66 (Kung, 2000)</td>
<td>Face; constructive; predictive (Kung, 2000)</td>
</tr>
<tr>
<td>Rating Scale of Therapists’ Systemic Responses in an Individual Treatment Context (RSTSRITC)</td>
<td>5-pt. Likert: 1 = not systemic; 5 = strongly systemic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Therapist Rating Scale (FRS)</td>
<td>7-pt. Likert: 0 = ineffective; 6 = maximally effective</td>
<td>Supervisor rate</td>
<td>Interrater (Piercy, Laird, &amp; Mohammed, 1983)</td>
<td>Face; criterion (Piercy et al., 1983)</td>
</tr>
<tr>
<td>MFT Internship Evaluation Instrument (MFTIE)</td>
<td>No numerical score</td>
<td>Supervisor rate</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Postgraduate Competency Document (PGCD)</td>
<td>No numerical score</td>
<td>Supervisor rate</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Scoring Rubric Counselor-Trainee Clinical Work (SRCTCW)</td>
<td>Rating of 5, 3, or 1</td>
<td>Supervisor rate</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Competency Evaluation Inventory (CEI)</td>
<td>5-pt. Likert: 1 = weakness; 5 = very strong</td>
<td>Supervisor rate self-report</td>
<td>Internal consistency</td>
<td>Concurrent (Davenport, Northey, Ratliff, Todahl, &amp; Perosa, 2007)</td>
</tr>
<tr>
<td>Family Therapy Skills Observer Rating Form (FTSORF)</td>
<td>5-pt. Likert: 1 = weakness; 5 = very strong</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Critique. The SCS is designed to measure clear, behaviorally defined skills related to a skills-based training model implemented by trainers/supervisors. The instrument has been useful in determining the effects of the training model on the development of student competence. According to the instrument developers, the SCS includes lower-level skills (e.g., summarizing), middle-level skills (e.g., asking for concrete and specific expressions), and higher-level skills (e.g., deciding). However, the meaning of the terms lower level versus middle or higher level for skills is not clarified. The rating system to evaluate the performance of each skill appears to be based on the frequency with which the skill is performed rather than on the appropriateness or effectiveness of the application of the skill by the counselor trainee or the type of mental processing involved in demonstrating the skill. In comparison, the marriage and family therapy (MFT) core competencies (CC) involve more complex cognitive processes. Thus, this measure appears to be more relevant to assess some basic skills such as attending, questioning, and summarizing that may be useful for MFT students in beginning role-play simulations at the beginning of their training.

The Counseling Skills Scale (Eriksen & McAuliffe, 2003). In 2003, Eriksen and McAuliffe revised the SCS to develop the Counseling Skills Scale (CSS). The 22 items comprising the CSS form six subscales. The first four subscales—Shows Interest and Appreciation (which includes items such as “using minimal encouragers helpfully” and “expressing a caring vocal tone”), Encouraging Exploration (which includes “requesting specific examples” and “reflection of content”), Deepens the Session (including “using immediacy” and “observing themes and patterns”), and Encourages Change (including “determining goals” and “using strategies for change”)—represent possible stages of a helping interview, according to the authors. The last two, Develops the Therapeutic Relationship and Manages the Session, represent more global conditions for effective helping that cut across sessions. The scoring system used in the SCS was revised by Eriksen and McAuliffe to require raters to make judgments about the appropriateness and helpfulness of the skills used by the trainee. Scale scores are obtained by summing the ratings for items and averaging them. Therefore, skills items on a subscale that are less frequently applied but are applied successfully to help accomplish the purpose of the scale do not lower the subscale score (which could happen with the SCS scoring). A total score for counseling competence is obtained by summing the subscale scores.

Critique. The revisions to the SCS incorporated in the CSS appear to have made ratings of counselor performance by supervisors more reflective of counselor judgments involved in applying the skill. The CSS appears to contain both less complex skills based on single behaviors (e.g., “using minimal encouragers helpfully”) and more complex skills (e.g., “observing themes and patterns”). Internal consistency reliability and construct validity have been found in two studies (see Table 6E.1). One difficulty with the measure is that item-to-scale correlations for the subscales showed that some items correlated higher with other scales than
their own scale, which may indicate that the scores do not represent true factors (Eriksen & McAuliffe, 2003). In addition, interrater reliability studies are needed. The skills assessed by the CSS appear to be limited to skills learned early in MFT skills courses involving role-play simulations.

Integrated sets of instruments

**Hill’s Helping Skills System** (Hill, 2001, 2009). The Hill’s Helping Skills System (HHSS) model includes a set of reliable and valid measures to be utilized by trainers/supervisors and students/trainees that link the counseling competency and outcome-based education (OBE) philosophies. They are intended to be used in training and research. The set of instruments includes observer ratings of therapist verbal responses (i.e., the Supervisor or Peer Rating System, SOPRS) and self-report ratings by counselor trainees of those same verbal responses (i.e., the Counseling Activity Self-Efficacy Scale, CASES) that will be discussed later in this resource. Other instruments clearly define the skills that are to be practiced and assessed by the supervisor and trainee (i.e., the HHSS sheet), intentions that guide counselor trainees’ choices of skills employed throughout the counseling process (i.e., the Helper Intentions List, HIL), client reactions to the counselor’s responses (i.e., the Client Reactions System, CRS), and client ratings of the therapist’s skills, the therapeutic relationship, and session evaluation (i.e., the Session Process and Outcome Measure, SPOM; Hill, 1992).

The HHSS sheets include definitions of counseling verbal skills, such as “challenging” (e.g., “points out discrepancies, contradictions, defenses, and irrational beliefs of which the client is unaware, unable to deal with, or unwilling to change”). These skills and definitions are followed by tape script examples such as the following:

Client: I don’t have any problems. Everything in my life is going really well.
Helper: You say everything is going well, but you keep getting sick. I wonder if it is difficult for you to look at your situation.

These skills, their definitions, and examples in verbal exchanges help to enhance clarity when supervisors rate the performance of students they observe using the SOPRS. They also help clarify skills students are expected to master when they rate their confidence in performing each skill using the CASES.

Hill’s HIL includes goals such as setting limits, offering support, encouraging catharsis, and identifying maladaptive cognitions. Each intention is clearly defined. An example of a counselor intention and its definition is as follows.

<table>
<thead>
<tr>
<th>Intention</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus</td>
<td>To help the client get back on track, change the subject, or channel or structure the discussion if the client is unable to begin or has been diffuse or rambling</td>
</tr>
</tbody>
</table>
Hill also has a CRS sheet, which clients use to rate their reactions to helper statements such as the following.

<table>
<thead>
<tr>
<th>Client reaction</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstuck</td>
<td>I overcame a block and felt freed up and more involved in what I have to do in helping</td>
</tr>
</tbody>
</table>

These forms help students note their intentions for using each skill and get feedback on the effectiveness of their interventions by having clients rate their reactions to the counselor’s responses using the CRS forms. These forms are useful for training.

The SOPRS. On this instrument, the supervisor/trainer rates how appropriately the counselor trainee performs each skill listed on a 5-point Likert-type scale (see Table 6E.1). A total score is calculated by summing scores for each item. The rater also can list examples for each skill as well as strengths and areas that need improvement for the counselor trainee.

Critique. Because the SOPRS is associated with a training program that includes forms to help the trainees see the relation between their intentions and skills and also instruments that help them receive feedback from clients about the impact of the skills applied, the SOPRS appears to assess both behavior (applying skills) and higher cognitive processes involving planning and evaluating. The written feedback by the rater makes the SOPRS even more valuable as a training tool. The SOPRS has been used in many research studies that provide evidence for its reliability and validity.

The CASES. The CASES allows the student/trainee to rate his or her degree of confidence using each skill on a 9-point Likert-type scale. It consists of three sections: the first contains 15 items pertaining to skills used at each stage of the counseling process, the second has 10 items that involve more complex activities such as building a clear conceptualization of the client and his or her counseling issues, and the third has 16 items focusing on diversity and difficult client behaviors. Scores can be obtained for each section by summing scores for each item.

Critique. The CASES has demonstrated reliability and validity in several studies (see Table 6E.1). It can be used in both teaching, as a way of obtaining feedback on student levels of confidence, and in research.

Overall, Hill’s package of assessments is important for both training and research. The combination of forms allows students to consider their intentions and use of skills and to receive feedback on their skills performance from the supervisor and client. The instruments have been adapted for beginning courses in MFT programs involving student role-plays of individual skills that can be used in doing a genogram or beginning an initial session with a family to help family members tell their story. They also have served as a template for developing an integrated set of measures assessing therapist use of family skills by the supervisor, the trainee, and the clients (see the following on how this is being done at the...
University of Akron by the authors). The measures can be used by researchers to study trainee growth in achieving competence from the perspectives of the supervisor, counselor, and client. The impact of the use of skills by the counselor on client change and session progress also can be tracked from the perspective of the counselor and the client.

Summary of Research on Counselor Training Based on These Measures

Each of the instruments described in the preceding text has demonstrated acceptable levels of reliability and validity for research purposes. As a result, researchers are able to use them in research to monitor student growth over time, and evidence is beginning to accrue that training does make a difference in student performance and self-efficacy beliefs in techniques classes, in practica, and in later work with clients. For example, Schaeffe et al. (2005) found that the scores of counseling students improved significantly on the SCS before training, immediately after training, and at the end of the master’s degree program on the total scale score and on 14 out of 18 skills. Effect sizes ranged from medium to high. Eriksen and McAuliffe (2003) found that scores for students in a theories and techniques course changed significantly in a positive direction on the overall score on the SCS and for all subscales except Develops the Therapeutic Relationship. The effect size was 0.80, indicating that the change was meaningful. Less is known about the impact of internships and postdegree experience on counselor mastery, although Melchert, Hays, Wiljanen, and Kolocek (1996) found that increased training in doctoral applied psychology programs augments counseling self-efficacy more than post master’s-level clinical experience. Student ratings of their counseling self-efficacy increased around one standard deviation from pre- to posttest in a practicum class and, in general, were higher for students who had more years of counseling experience and more semesters of supervision in a study by Larson et al. (1992) as well.

Initial research by Urbani et al. (2002) also found that differences in ratings of counseling skills competency exist between faculty, students in training, and a control group not receiving skills training, with trainees overrating themselves before training in a skills class but underrating themselves after training compared to ratings by supervisors. Students receiving training also rated themselves significantly higher on counseling self-efficacy at the end of the skills course than the control group. Research by Crews et al. (2005) indicates that personality traits such as low or high degrees of self-monitoring did not have a significant impact on the skills performance of counselors in training. In a recent study with counselors and clients by Hill and Kellem (2002), client perceptions of therapist skills predicted session outcome evaluation over and above client ratings of the relationship. Changes across sessions were found for counselors’ use of skills with exploration skills scores increasing initially and then insight scores and finally action skills scores in later sessions.
Thus, the advantage of developing psychometrically sound measures related to clinical competence is that systematic research on the effectiveness of training can be conducted. This kind of research can provide valuable feedback to students and teachers/supervisors to help improve training, which is the cornerstone of evidence-based education or OBE.

**Measures of Relational Skills Used With Individual Clients**

Although the instruments described earlier have been developed to assess the use of skills associated with individual counseling sessions, it is important to note that none of these measures include a rating of the counselor’s use of a systems focus when working with an individual client. Doherty and Simmons’s (1996) national survey of marital and family therapists revealed that half of the treatment provided by these practitioners consisted of individual therapy. Yet what distinguishes the systems approach with individual clients is the attention given to individuals or problems in context—the tracking of interaction sequences between individuals, subsystems, and whole systems from the perspective of the client by the therapist and the focus on circular processes and patterns that connect over time and across generations that impact the client in the current situation.

In short, when individual counseling occurs by a family-trained therapist, the focus is on relational and systems issues; the “I” in individual sessions for a family therapist is always the “self” embedded in relationships. In contrast, the “I” for the individually trained therapist is the internal psychological state and internal self of the client, and the focus of therapy is on helping the client explore internally generated thoughts, feelings, and behaviors from a linear perspective.

**The Rating Scale of Therapists’ Systemic Responses in an Individual Treatment Context** (Kung, 2000). Kung’s development of the Rating Scale of Therapists’ Systemic Responses in an Individual Treatment Context (RSTSRITC) is an initial attempt at assessing a systemic focus by a therapist in individual treatment. The RSTSRITC consists of two scales: (1) Interactional Sequences and (2) Consideration of Context. The rating manual describing Interactional Sequences (i.e., from the individual client’s perspective) gives examples of how a therapist can help a client see how his or her attempted solution to a problem may escalate, reinforce, or reduce behaviors of his or her partner in a symmetrical or complementary way. A sample tape script demonstrating symmetrical escalation reads:

T: What were you feeling when you had this exchange with your husband?
P: I didn’t really care if he went to the concert or not.
T: How did you think your husband was feeling?
P: I don’t know. I … He knew that I was going whether he went or not.
T: Yeah, but I wonder if you didn’t convey to him a sense of “It doesn’t matter to me whether you go or not. I don’t care.”
P: Probably.
T: And how do you think he felt about that?
P: He probably felt bad.
T: So how do you think that would have affected his decision about whether to go or not?
P: He’d have the same attitude that I did.
T: Exactly. So what …
P: So what we’re doing is fighting against one another.
T: And this battle just goes on and on, an undeclared war.

Sample examples of Consideration of Context items include descriptions of therapist responses that highlight life cycle developmental stresses, or the additional stresses on couples due to unemployment, or family-of-origin legacies straining interactions between partners.

Five different types of measures were developed to assess the constructs: (1) the number of interactional responses, (2) the number of contextual responses, (3) a global interactional rating, (4) a global contextual rating, and (5) the overall global systemic rating. The first two scores are based on counting the number of interactional responses and counting the number of contextual ones. The global ratings for interactional and contextual responses involve judgments on whether the responses were interpretive (e.g., the therapist interprets the circular causality of an interactional sequence, higher score) or exploratory (e.g., tracking only, lower score). The overall global systemic rating is used to reflect the raters’ overall impression of the therapists’ systemic responses. The global ratings are based on 5-point Likert-type scales (see Table 6E.1).

Critique. The measure is developing evidence for reliability and validity, so it is useful for future research studies (see Table 6E.1). The manual gives sample tape scripts and rating criteria that may require practice by the rater before the scales can be used in research. The global scales can be helpful to trainers to rate the quality of students’ systemic responses with individual clients and provide feedback on their performance.

The addition of a scale that assesses systemic factors utilized in individual treatment is important because it delineates unique skills MFT students employ when working with individuals.

The Usefulness of Individual Measures for MFT Training

Marriage and family therapy programs have developed a variety of ways to design curricula to train competent practitioners. At the University of Akron, we have incorporated Hill’s model and instruments to teach basic microcounseling skills as part of an MFT techniques/skills course that students take before prepracticum and practicum. Individual counseling skills, such as attending, using open-ended questions, restating content, reflecting feelings, and summarizing, are practiced by students in dyads and rated using Hill’s measures. These individual skills then are practiced along with circular questioning in family role-play simulations. In later sessions, students also practice some of these individual skills as they carry out a genogram with a couple from the simulated family they are counseling (see Perosa & Perosa, 2007; Perosa, Perosa, Heberling, & Williams, 2006). During the couple and family role-plays, these individual skills are integrated into a systems perspective.

Hill’s questionnaires and rating sheets also have provided a template for us to develop an integrated set of family skills measures that are used by students and trainers throughout the
simulated family therapy sessions. The set of family measures are described in the
next section of this article.

We believe that the availability of these individual counseling skills inventories
and the comprehensive training models that were the foundation for their
development provide a bridge to link individual and family training that follows
an OBE philosophy. We have presented a description of all of the measures to
provide a menu of different instruments for MFT educators to select from as they
implement OBE.

**Measures of Family Therapy Skills**

In spite of the relative lateness of the MFT field to adopt OBE, MFT faculty have
a rich tradition of developing instruments for evaluating student performance
during clinical experiences. Although the majority of instruments focus on
providing feedback to students within programs, there have been efforts to develop
measures that can be used in research across programs.

**Supervisor rating scales**

**The Family Therapist Rating Scale** (Piercy et al., 1983). The Family Therapist
Rating Scale (FTRS) consists of five categories; each contains ten skills. Two of
the categories, Structuring and Relationship, include general skills that cut
across theoretical models. A sample item on Structuring includes “asks
open-ended questions.” A Relationship item is “engenders hope.” The other
categories are Historical, Structural/Process, and Experiential and are based on
Levant’s (1980) classification of various family therapy theoretical models. An
item in the Historical category is “avoids becoming triangulated by the family.”
Structural/Process items include “checks out pronouns to see who did what to
whom.” Experiential behaviors include “asks for current feelings.” Each cate-
gory is rated on a 7-point Likert-type scale ranging from 0, not present, to 6,
maximally effective. A profile of skills for each category is obtained by rating
only actual behaviors observed in a session for each category and then calcul-
lating the mean for the category. Another scoring option is to add the total
points within each category (raw scores) and place them on the profile. The
profile provides a concise way of presenting information on student strengths
and skills areas that need more practice.

**Critique.** The measure has demonstrated reliability and validity (see Table 6E.1).
Consequently, it is useful for training and for research. Each skill is presented in a
single item, which clarifies the rating process. The measure appears to cover some
of the CC, such as eliciting family conflicts, alliances, and coalitions. However,
many CC are not evaluated on the FTRS, so the scale is more useful in the early
part of training when faculty are focusing on having students understand and
practice basic skills. Additional skills from other theories that have become popular since the 1980s (e.g., noting exceptions, externalizing problems) could easily be added to the measure.

The Basic Skills Evaluation Device (Nelson & Johnson, 1999). The Basic Skills Evaluation Device (BSED) was developed using empirically derived skills from the Basic Family Therapy Skills Project (see Figley & Nelson, 1989, 1990; Nelson & Figley, 1990; Nelson, Heilbrun, & Figley, 1993) to assess a wide range of trainee skills and professional development. The BSED is comprised of five core dimensions—Conceptual, Perceptual, Executive Skills, Professional Skills, and Evaluative Skills—and an optional dimension called Theory-Specific Skills. Each core dimension is subdivided into related subdimensions. For instance, Executive Skills include joining, assessing, hypothesizing, interventions, communication skills, personal skills, and session management. Descriptions for each underlying dimension comprising a core dimension are provided in the form of profiles describing what the trainee is able to do under each category. Under Executive Skills, the subcategory of “Interventions” includes the following profile.

The trainee demonstrates an understanding of intervention techniques by structuring interventions that defuse violent or chaotic situations, deflect scapegoating and blaming, and interrupt negative patterns and destructive communication cycles. The ability to intervene also includes appropriately challenging clients on their position, explicitly structuring or directing interactions among family members, and helping families establish boundaries. The trainee is able to elicit family/client strengths and utilize them in both session discussions and homework assignments.

Other interventions that illustrate skill include normalizing the problem when appropriate, helping clients develop their own solutions to problems, giving credit for positive changes, reframing, and appropriately using self-disclosure. The trainee uses theory-specific interventions appropriately and is able to articulate a rationale for these interventions (Nelson & Johnson, 1999, p. 26).

The anchors applied by raters to evaluate students on these subcategories range from “inadequate information” to “deficient,” to “below expectation,” to “meets expectation,” and to “exceptional skills.” Raters are told that the phrase meets expectations means “in your experience compared with other trainees with this level of experience and training.” The authors recognize that this method of rating does not allow supervisors and researchers to compare students-in-training across programs (although the instrument may be useful within programs) since each rater is using his or her own subjective standards. Therefore, users are told they can change the anchors for items from the general statement “compared to peers with similar levels of experience” to a scaling rating based on anchors ranging from “no demonstrated skill” to “skills level of a program graduate” or “skills level of a very experienced clinician.”

In order to help raters have some common standard for expected trainee behaviors that might apply across programs as they rate students, the BSED does provide
general guidelines regarding developmental levels. The profile of a beginning trainee reads as follows.

**Beginner**: First 50–75 hr of experience, less, perhaps, if under intensive live supervision. The beginner will need more direction and structure, clearer session plans, and more freedom to go in a direction that may seem less productive but which follows the trainee’s plan for the session and the supervisor’s plan for what the trainee is currently working on. For example, the supervisor may see an opportunity for a paradoxical or solution-oriented approach, but the trainee may be working on structuring the session with parents and children. The trainee can discuss case material based on one theoretical perspective but may get confused if trying to use more than one. The trainee is eager for supervision and may feel confused or anxious in new situations.

These profiles are helpful because descriptions of trainee skills are tied to hours of clinical experience and practice with families and are meant to enhance consistency in using the BSED.

**Critique.** Strengths of the BSED are that the authors have provided evidence of reliability and content validity (see Table 6E.1). In addition, a mapping of the CC onto the BSED items reveals that several CC appear to be tapped into by the BSED. These competencies include understanding theories, recognizing contextual and systemic dynamics, understanding models of assessment, and hypothesizing. The BSED clearly fits the conceptualization of the CC and is unique in its use of profiles to capture the complexity of the skills that need to be assessed.

However, because several skills are embedded together within each description of core dimensions and subdimensions and trainee-level profiles, evaluators are unable to rate specific deficiencies or single skills as weaknesses if the student is stronger on other skills within a profile; only overall ratings for the entire profile are possible. Perhaps, having a place for comments after each profile may allow the rater to indicate specific skills that need to be strengthened. Concurrent research studies involving the BSED and other measures using specific skills also could increase our understanding of the relationship between the profiles and ratings of more simple skills assessed by other measures.

The ambiguity of anchors adds to the lack of clarity in rating student performance because the anchors are not standardized or grounded in findings from research determining the skills of program graduates or very experienced clinicians. Future research in the MFT field may provide this information.

No description of skills, cognitive abilities, and behaviors is provided for advanced-level trainees with more clinical experience. However, according to Nelson et al. (2007), “Nelson is working on a broader instrument, similar to the BSED, that may be sufficiently flexible for both MFT graduate programs and for advanced training” (p. 427). The instrument will be a welcome addition to the field.

Because of its inclusion of more complex higher-order skills, the BSED may be more suitable for use in practicum and internship courses than for beginning role-play simulations in theories or skills/techniques classes. The addition of the measure currently being developed by Nelson may be even more sensitive to assessing advanced
complexities inherent in the CC. The BSED is important because of the breadth of skills and complex cognitive processing related to the CC that are captured in the instrument. It has set a pattern for other measures that are designed for advanced training, such as the Postgraduate Competency Document (PGCD). The BSED also is one of the few measures that can be used in research across programs.

The MFT Internship Evaluation Instrument (AAMFT, n.d.). The MFT Internship Evaluation Instrument (MFTIE), a document in the AAMFT’s Practice Management Forms, consists of two scales: Professional Development and Counseling Process. Professional Development is assessed by 10 items such as “maintains professional and ethical behavior,” “completes case reports and records punctually,” and “assesses and diagnoses clients in accordance with DSM criteria.” The Counseling Process scale contains 20 items related to counseling, such as “demonstrates joining skills”; “assists family definition of problem and goals”; “identifies covert family conflicts, alliances, and coalitions”; and “uses interventions related to process rather than content.” Each item is rated as “excellent,” “acceptable,” or “low.” Space is given for supervisors to comment on interns’ strengths and areas that need improvement.

Critique. P. Parr (personal communication, April 2007) has mapped the CC onto the MFTIE, and it appears that the MFTIE scale items target several CC related to family therapy, assessment, and diagnosis, as well as ethical and professional behaviors. The instrument was intended for rating trainees and providing feedback to students and program faculty at the internship stage. Although each item is rated as “excellent,” “acceptable,” or “low,” no numerical scores are provided and no guidelines are given as to what behaviors exemplify low, acceptable, or excellent degrees of competency. Consequently, no reliability or validity information is provided.

The Postgraduate Competency Document (Storm, York, Vincent, McDowell, & Lewis, 1997). The PGCD is intended to be utilized with postgraduate therapists. There are seven competency areas on the PGCD: General Case Management, Therapeutic Relationship, Perceptual, Conceptual, Structuring, Intervention, and Professional Development competencies. Each competency area is defined and contains three to six specific competencies, skills, or behaviors to be rated. For example, the area of Conceptual Competencies is defined as “the abilities to integrate observations with theory, resulting in appropriate interventions and decisions about treatment goals.” A specific competency under that category is “therapist bases hypotheses and goals upon theory.” Each of these competency items is rated on a continuum with nine markers and five anchors ranging from “unacceptable” to “below,” to “expected,” to “above,” and to “advanced” levels of competency. Definitions for anchors are provided at the “expected” and at the “advanced” levels for each specific competency. The description for the “expected” anchor for the competency item on theory reads, “can view cases from more than one theory.” The description for the “advanced” level anchor reads, “can view cases from multiple theories including preferred one.” Instructions at the top of the instrument
indicate that “a rating below competency level signifies additional ability is required in order to obtain a competent rating.” The PGCD adds an Overall Competency rating for the supervisee based on this point in his or her career development plus space for Overall Comments and for Goals for Future Development.

**Critique.** A comparison of this measure with the CC indicates that several competencies are addressed on the PGCD. Each competency item is linked to a single higher-order (more complex) skill; therefore, the measure provides clear indicators of trainees’ behaviors at the postgraduate level of experience. The PGCD was intended for use within programs for feedback and evaluation and does fulfill that purpose. Although there are no numbers associated with the ratings of “unacceptable” to “advanced,” the instrument could easily be updated so that it could be used in research. Adding scoring procedures would allow programs to collect reliability information and to conduct research on the development of their trainees.

**The Scoring Rubric Counselor-Trainee Clinical Work** (Hanna, 1997). Another approach taken to assess student clinical performance is the scoring rubric. The Scoring Rubric Counselor-Trainee Clinical Work (SRCTCW) was developed for supervisors to assess students’ performance levels in individual and family counseling programs related to specific criteria. The 20 items comprising the SRCTCW are broken down into two sections: I. Professional Responsibility (subdivided into Individual Supervision, Group Supervision, Completion of Paperwork, Professional Behavior, and Ethical Behavior) and II. Theoretical and Procedural Knowledge (subdivided into Initiates Intakes, Diagnostic and Case Conceptualization Skills, Marriage/Family Theories and Techniques, Establishing the Counseling Relationship, Listening Skills, Knowledge of Systems, Understanding Couple/Family Reality, Therapist Judgment, Case Management, Clinic Policy and Procedures, Self Evaluation of Therapy Skills, Personal Growth, Therapy Style Development, Interactions with Colleagues, and Willingness to Work on Personal Issues).

Rating alternatives on the SRCTCW are defined by exemplars (i.e., specific descriptions of behavior that would exemplify a person receiving a particular rating). One exemplar is given at each end point (for a rating of 5 or 1) and at midpoint (for a rating of 3) for each of the 20 items on the instrument. A sample of the scoring rubric for “Knowledge of Systems” is as follows:

5. Demonstrates understanding of the roles of internal/external systems, the self as counselor, the role of self as impacts on couples/families, and the interactive nature of all systemic roles

4. (between 5 and 3) comment:

3. Has moderate understanding of the roles of the clients, the self as counselor, the self as impact on clients, and the interactive nature of all roles but has some difficulty translating knowledge into action

2. (between 3 and 1) comment:

1. Has significant difficulty in understanding or recognizing the multiple systems of roles and their interaction when working with clients

A total score is calculated by summing scores on the 20 items.
Critique. A mapping of the CC onto this instrument reveals that two or more competencies are assessed by each of the 20 items. Thus, although many competencies are assessed on the instrument, each item assesses more than one competency. Consequently, the instrument does not provide a clear rating of specific competencies; instead, ratings are blurred by covering more than one behavior in each exemplar. (Remember that the trainee is compared to the whole description of behaviors provided in the exemplar.) In addition, no reliability or validity information is described.

Integrative Sets of Family Skills Rating Forms

The Competency Evaluation Inventory (Davenport et al., 2007). A promising new measure being developed for use with MFT faculty and students is the Competency Evaluation Inventory (CEI). The CEI was initially formulated by faculty from St. Mary's University, and empirical research to revise it was supported by other members of the Beta Test Group. The revised CEI is a 20-item questionnaire developed in two stages to assess the CC. Two parallel forms are available: one for trainees and another for supervisors. The newest version of the instrument emphasizes the CC facets of attending to the therapeutic relationship and designing and conducting treatment. Sample items include the following: (the trainee) “attends to clients’ emotional climate in order to build trust,” “engages clients in the therapeutic process,” and “assists the client in developing effective problem-solving abilities.” Items are rated on a 5-point Likert-type scale ranging from 1 (this is a weakness for the trainee; she or he should continue to work on this skill) to 3 (the trainee is generally competent in this area, but she or he should continue to work on this skill) to 5 (the trainee is very strong in this area with all clients; this comes naturally for him or her now). There also is a column for NA (I am not able to assess the trainee on this competency at this time). A total score is derived by summing scores for items. A qualitative component was added to the CEI to allow supervisors to describe the trainees’ strengths, areas for future growth, specific concerns the supervisor might have regarding the trainees’ skills, and an explanation for any “NA” responses.

Critique. Both the original 50-item version and the revised version have demonstrated good internal consistency reliability and have been able to significantly discriminate between self-ratings of competence by trainees with less than 100 hr of experience, those between 101 and 300 hr, and those between 301 and 500 hr. Thus, the CEI appears to be sensitive to growth demonstrated over time by students as reported by their self-evaluations from a cross-sectional (but not from a longitudinal) perspective. As a result, the CEI is beginning to show some evidence of construct validity (see Table 6E.1). However, supervisor ratings of trainees’ skills did not strongly correlate with the students’ self-reported evaluations of their competence. This finding parallels the results of research in the individual counseling field.
The revised forms of the CEI are valuable tools that allow for self-efficacy ratings of skills performance by students and ratings of trainee competence by supervisors. The CEI items focus on global complex skills rather than basic simple skills and on generic competencies rather than on skills and strategies associated with a specific family therapy model. Each competency is assessed discretely and clearly in a single item. Although the CEI yields a total score, factor analytic studies are needed to discern whether there is only one or several factors underlying the structure of the CEI. Because the CEI has demonstrated evidence of reliability and validity, it is now possible for large-scale studies to be conducted comparing student growth across programs rather than just within one particular program in order to gain a broader picture of student development throughout the discipline.

**The Family Therapy Skills Rating Forms** (Perry & Perosa, 2007). The set of Family Therapy Skills Rating Forms developed by Perosa and Perosa emulate the work of Hill in individual counseling referred to earlier that focuses on skills, intentions, feedback, and outcomes. The set of family therapy instruments includes a rating form to be completed by trainees listing their intentions and the skills they applied to carry out their intentions in family therapy sessions; it also includes measures for a supervisor to rate trainee competence and for the trainee to rate his or her self-efficacy using family therapy skills for use in role-plays in family counseling techniques/skills classes and for sessions with clients in practicum.

**The Family Skills: Helper’s Intentions Form.** The Family Skills: Helper’s Intentions Form (FSHIF) was created to help trainees understand the relationship between their intentions and the choice of skills they select to accomplish those personal goals. An example of an item on the FSHIF includes an intention and its definition, such as the following.

<table>
<thead>
<tr>
<th>Intention</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>To understand the influence of family-of-origin experiences on current family functioning</td>
<td>To help family members understand how rules of interacting learned while growing up influence behaviors in the current nuclear family. List skills you used: What did you do well? What else could you have said to carry out your intention?</td>
</tr>
</tbody>
</table>

**The Family Therapy Skills Supervisor Rating Form.** The Family Therapy Skills Supervisor Rating Form (FTSSRF) contains a section listing skills for use at each stage of the family therapy process plus a section that assesses more complex conceptual, perceptual, executive, and professional competencies and tasks. One skill to be rated reads, “[The trainee] uses probes to gain information, and identifies family-of-origin patterns that influence current family functioning when conducting a genogram.” An example of a more complex executive task is, “[The
trainee] can promote changes in family members’ cognitions, perceptions, and narratives.” Items are rated on a 5-point Likert-type scale ranging from 1 (this is a weakness for the trainee) to 3 (the trainee is generally competent in this area, but she or he should continue to work on this skill) to 5 (the trainee is very strong in this area with all clients). There also is a column for NA (I am not able to assess the trainee on this competency at this time). A total score for each section is derived by summing scores for items and then dividing by the number of items answered. There also is a profile sheet to list scores on each section and items the trainee is below 3 (below competency) on. Supervisors also can describe the trainees’ strengths, areas for future growth, and an explanation for any “NA” responses. These are similar to the CEI ratings.

**The Family Therapy Skills Observer Rating Form.** The Family Therapy Skills Observer Rating Form is used for rating the skills level of a student, demonstrating the skill that can be completed by the student after observing one or more of his or her sessions on tape. It is similar to the FTSSRF.

**Critique.** The Family Therapy Skills Rating Forms and other training sheets are tied to a training model that has been successful with students in individual counseling programs. The set of instruments can be used in techniques classes as well as in practicum and internship classes to monitor student growth utilizing specific skills, complex conceptualizations and tasks at each stage of the counseling process, and professional behaviors from the perspective of the supervisor and trainee. The measures have demonstrated reliability and validity, and a longitudinal research study is being planned incorporating the instruments to study trainee growth over the duration of training.

**Summary of the Usefulness of Current Measures of Family Therapy Competencies**

The CC and OBE are intertwined, with the CC representing a final step in the OBE process. Therefore, a variety of measures and assessment tools are needed that span the stages of training leading to competency. Over the years, the field of family therapy has collected a treasure chest of measures that can be used throughout the training process to provide feedback to students and evaluate their performance.

A comparison of the family rating scales described earlier indicates that each one emphasizes different aspects of the therapeutic process that are covered by the CC. Some inventories include a broader range of topics and more categories from the 128 CC than other measures do. For example, the FTRS and the CEI focus on intervention skills, whereas the BSED, the MFTIE, the PGCD, and the SRCTCW assess not only intervention skills but also other professional behaviors. Some measures target beginning, less complex skills (e.g., “asks open-ended questions” on the FTRS), whereas others emphasize more complex sets of skills
that form competencies (e.g., “establishes a sequence of treatment processes in a treatment plan” on the CEI). Some instruments include both simple and more complex skills and tasks (e.g., the MFTIE, the FTSSRF). Some measures are designed for techniques and skills classes (e.g., the FTSSRF), others for practicum and internship (e.g., the BSED, the CEI, the MFTIE, the SRCTCW, and the FTSSRF), and others for postgraduate training (e.g., the PGCD). A variety of scoring procedures are utilized on the measures, ranging from complex profiles on the BSED to single scales forming profiles on the FTRS and the FTSSRF to one global score on the CEI. The measures differ on their purpose, with some aimed at within-program training and others intended for research.

As a result, MFT educators can choose from a variety of instruments to evaluate students in their clinical training and obtain feedback on the effectiveness of faculty teaching and supervision strategies so that curriculum can be revised. The current measures also serve as models for developing assessment instruments that can be used in research. For example, future research needs to examine whether there is a rating of trainee and practitioner competency that is consistent across programs or whether there are differences within and between programs as to what constitutes competence at various levels of training. Are there differences between the supervisor and trainee on ratings of trainee competence that remain throughout training?

References


Perosa, S., & Perosa, L. (2007). *A semester course for training students in marriage and family counseling skills*. Paper presented at the Association for Counselor Education and Supervision Convention, Columbus, OH.


Todahl, J., & Perosa, L. (2006). *The evaluation of core competencies: Constructs, models and resources for marriage and family therapy*. Core Competency Beta Test Group Website, American Association for MFT.