

A Measure of Counselor Competency

Karen Eriksen & Garrett McAuliffe

Counselor educators need to be able to demonstrate their effectiveness in training new counselors; however, currently few valid or reliable measures exist for assessing educators' impact. The authors describe the development of such an instrument, the Counseling Skills Scale. They began by revising an existing scale and then they solicited feedback from experts and a focus group. They used the instrument to compare beginning counselors-in-training with those who had completed a counseling skills course. Finally, they conducted an item analysis. A paired *t* test showed significant improvements in counseling skills ($t = 4.51, p < .000$) from pretest to posttest. Cronbach's alpha showed internal consistency to be .90.

The counseling profession must be able to demonstrate its effectiveness if it is to gain the confidence of its public and if its members are to practice ethically. In parallel fashion, individuals in the counselor education field itself must also measure their ability to produce effective and ethical counselors.

However, many unanswered questions exist about the best way to assess counseling students' competency. Although efforts have been made to develop instruments that might measure counseling effectiveness, the reliability and validity of these instruments are generally unsatisfactory (Alberts & Edelstein, 1990; McLennan, 1994; Ponterotto & Furlong, 1985; Shaw & Dobson, 1988; Strupp, 1986). In light of this need and these limitations, we describe in this article the development of a valid and reliable measure of counselor competency, one that is particularly targeted at assessing the skills of beginning counseling students.

Literature Review: Measuring Skills Competence

The so-called skills training approach (e.g., Ivey, 1971; Truax & Carkhuff, 1967) has dominated counselor training for the past 30

Karen Eriksen, Department of Counselor Education, Radford University; Garrett McAuliffe, Educational Leadership and Counseling, Old Dominion University. Karen Eriksen is now at the Department of Counseling, Argosy University. The authors thank members of the focus group, Heather Cutler, Alan Forrest, and Sherry Lynch, and others who offered input on the instrument—Lois Benishek, Dale Pietrzak, Carolyn Woodbury, Robert Urofsky, Glenda Reynolds, Wally Scott, Avis Quinn, Donald Anderson, Russell Sabella, Joy Burnham, Sandra Wolf Protulipa, David Duys, Jo Hayslip, Jennifer Baggerly, Susan Adams, Larry Kontosh, Montse Casado, Sandy Magnuson, and Heather Cutler. They also thank the following statistics consultants: Marty Aylesworth, Steve Zerwas, Dale Pietrzak, Kristin Bellando, and Christin Talbot. Correspondence concerning this article should be addressed to Karen Eriksen, 13341 Eton Place, Santa Ana, CA 92705 (e-mail: keriksen@argosyu.edu).

years (Young, 1998). The counseling skills method of teaching breaks down the moment-by-moment utterances of the counselor into categories such as *reflection of feeling*, *paraphrase*, and *open questions* (e.g., Ivey & Ivey, 2003). The skills are then modeled for students and practiced separately by them. Often they are taught in a sequence of *read*, *discuss/lecture*, *see a demonstration*, *try out*, *critique*, and *try again*. This sequence of skills, linear though it may seem, spirals upward as students integrate understanding and skills into their repertoire through new practice. In the words of Kolb's (1984) cycle of learning model, students typically move from *concrete experience*, through *reflective observation*, *abstract conceptualization*, and *active experimentation* as they learn counseling skills. The method aims at having students integrate these skills and attitudes into an *intentional* interview (Ivey & Ivey, 2003). In the best pedagogical models, students learn to demonstrate empathy, regard, and congruence and to engage in ongoing reflections about clients, reflections that inform counseling choices. The teaching of counseling skills has thus developed into a relatively high art during these decades.

Measuring the attainment of these skills has been much more problematic, however. Because of the need in the field of counselor education for a research and feedback tool for measuring counseling skills, we reviewed the most recent attempts to assess counseling skills. We gathered extant instruments from an international literature search, a CESNET inquiry, and relevant conference presentations. The result was information on instruments from the 1960s to the present. We found, similar to the findings of Shaw and Dobson (1988), that overall methods of assessing counselor competence varied in the unit of analysis (i.e., the data that were assessed), the raters that were used, the validity and reliability of the measure, the method of rating skills, the refinement of the scales, and the nature of the analysis itself. We grouped these variations into the five categories, as follows.

Global Versus Specific Measurement

Regarding the unit of analysis, many instruments have relied on case notes and global supervisor impressions to assess counseling competency, rather than on specific, moment-by-moment critical analysis of skill performance. For example, one of the classic, often-cited instruments, the Counselor Evaluation Rating Scale (Myrick & Kelly, 1971), asks for global impressions of counselor functioning. Questions assess general attitudes, ranging from "Is open to self-examination during supervision" to "Verbal behavior . . . is appropriately flexible and varied" (p. 336).

Clients Versus Experts as Raters

The expertise and roles of raters also vary. Some instruments ask clients themselves for reports on their counseling session experiences, whereas others rely on expert judgment about in-session counselor behavior. Client report of counselor competency is problematic (Shaw & Dobson, 1988; Strupp, 1986): Internal variations in clients and factors that are external to counseling contribute to successful or unsuccessful client outcomes. Clients are also generally unable to consistently distinguish effective counselor behavior and often vary in their perception of the same counselor (Strupp, 1986). Expert judgment seems to have the greatest potential as a source of accurate rating of counselor behavior. Such expertise is best defined as being possessed by those (i.e., instructors) who teach counseling skills, those who are the gatekeepers for the profession.

Absence of Validity Checks

Even when specific counseling sessions are the focus of analysis and when experts judge counselor behavior, few valid and reliable measures of actual counseling skills exist. Virtually none of the measures that were analyzed in literature reviews and meta-analyses by Alberts and Edelstein (1990), McLennan (1994), Ponterotto and Furlong (1985), and Shaw and Dobson (1988) had been adequately tested for construct validity. A full 88% of the rating scales reviewed by Ponterotto and Furlong relied only on face validity, and only 43% reported any reliability check at all. The others added only content validity. Counselor educators might therefore erroneously influence counseling students, their programs of study, and their future behavior by relying on nonvalidated methods.

Counting Versus Making Judgments About Skills

Another problem with many skills measures involves the unrefined nature of the rating scales. Many scales are designed merely to count the number of times that a skill is exhibited, without reference to context. Such counting is based on, in Alberts and Edelstein's (1990) words, "an assumption that the more these responses are displayed in an interview, the more efficacious the therapist's performance" (p. 508). The counting approach cannot be considered a valid assessment of counseling ability because the quality of student utterances is not evaluated, nor are counselor responses considered in context. Effective counseling requires weighing of evidence and considered action. Alberts and Edelstein proposed that counselors' responses be judged on whether the skills are "selectively and strategically applied in order to show that [counselors] can modulate or orchestrate the various skills in a therapeutic manner" (p. 508).

Dichotomous Scales

A related problem in analysis of counselor competency lies in the existence of merely dichotomous (good/bad, pass/fail, or yes/no) scales. In the place of these bipolar scales, Ponterotto and Furlong (1985) have recommended that more refined ordinal categories be used, such as acceptable, good, and exemplary. The use of these more refined discriminations can provide both more specific feedback to counselor trainees and greater gradation of their skillfulness for research purposes.

Emergence of the Counseling Skills Scale

This analysis of the literature on measures of counselor competency led us to name five desirable criteria for the measurement of counseling skills. Measures must (a) be valid and reliable; (b) rely on observations of actual in-session performance of counseling skills; (c) be accessible—that is, have face validity, be easy to use, and be relevant for students and instructors as a feedback device; (d) rely on ratings by expert judges, rather than on ones by students, clients, or peers; and (e) require qualitative judgments as to the contextual appropriateness of the use of particular skills. Three extant instruments met at least some of these criteria: the Counselor Interaction Analysis scale (Altekruse & Brown, 1999), the Skilled Counseling Scale (SCS; Urbani et al., 2002), and the Counseling Strategies Checklist (Hackney & Cormier, 1994). All of these measures were attempts to assess in-session performance of counseling skills, and all could be used by expert judges. However, for each measure, the validity data were weak or nonexistent, and the scales relied on counting, not evaluating, the use of counseling skills.

We believed that Urbani et al.'s (2002) SCS held the most promise because, even though the authors indicated that they had not carried out validity studies, they had achieved face validity from our perspective in the use of 5-point Likert scales, the division of individual skills into subscales, and the content. Furthermore, they had achieved content validity by basing their scale on a review of current counseling techniques textbooks. For instance, they included both subscales and related items, such as Attending (items: eye contact, body language, verbal tracking), Questions and Reflecting (items: questions, paraphrases, summarization), Interchangeable Empathy (items: reflection of feelings, self-disclosure, concrete and specific examples), Additive Empathy (items: immediacy, identifying patterns, confronting caringly), Decision Making (items: deciding, choosing, consequences), and Contracting (items: agreements, deadlines, reviewing needed actions and potential outcomes). Despite its assets, the limitations of the SCS warranted revisions. Thus we began work

on a revised scale, the Counseling Skills Scale (CSS). It emerged through three iterations, as described in the following sections.

First Iteration of the CSS: Revisions of the SCS

In addition to improving verbal clarity, shifting to a clearer format, and creating clearer descriptions to operationalize the items, we changed the SCS in the following ways.

Step 1: More precise scoring. The SCS method of counting how often counseling skills were used during counseling sessions as a means of determining counselor competence gave way to asking raters to make judgments about the appropriateness and helpfulness of the skills used. In the SCS, Urbani et al. (2002) had measured student performance during a 10-minute segment of a videotaped counseling session and had merely indicated on a 1 to 5 Likert scale whether a student performed the skills *not at all* (1), *a little* (2), *somewhat* (3), *a great deal* (4), or *always* (5). Such an accumulative rendering failed to assess the intentional and strategic use of skills in the context of the interview. For instance, if a student spent 10 minutes interrogating the client, she or he would get a high score on the SCS for the "questioning" item, even if the questions were off target or overused. In contrast, we thought that overuse of questioning by the student should result in a reduced score.

Another limit of the SCS involved its absoluteness: If a student did not use a particular skill at all during the observed 10-minute segment, she or he would be given 1 point on the SCS for *not at all*, despite the fact that she or he may have used that skill quite appropriately at an unobserved time during the session. Furthermore, if the student failed to use a skill that was needed during the observed segment, she or he would get that 1 point. In fact, harmful or ineffective performance of any of the skills on the part of the student still garnered points on the SCS because there was no way to deduct or withhold points.

To correct these limitations of the SCS, we created a 5-point Likert scale in consultation with other counselor educators. This new scale ranged from +2 to -2, extending from *highly developed* (+2) to *well developed* (+1), *developing skills* (0), *continue practice* (-1), and *major adjustment needed* (-2). We also created definitions for each of the five scores to help raters decide among them. The items were grouped into subscales, and after rating individual items, evaluators were to average the items belonging to a subscale into a subscale score.

Raters were further directed to use NN (not needed) for any skills that had been appropriately left out of the session. These unnecessary skills would then not affect subscale averages. Thus, student

counselors' appropriate choices *not* to use particular skills would not lower their subscale scores. For instance, under the subscale of Deepening the Session, only immediacy might be used in a given session. If only immediacy was relevant to that particular 10-minute section of the counseling session, then it seemed that the student should receive full credit for the subscale instead of being penalized for failing to use the other skills that were listed under Deepening the Session. If that student thus used immediacy to deepen the session in a helpful manner, used the skill expertly, and refrained properly from using any other skills in that subscale, she or he would receive +2 points for the subscale Deepening the Session. We hoped that such a change would make the instrument useful for many research and evaluation situations—for example, for researchers who observe 10 minutes of a session, or 10 minutes at the beginning and the end of the session, or for instructors who only require students to submit 10-minute transcripts.

Step 2: Adding items. Important dimensions of the counseling interview had not been incorporated into items of the SCS. We used the literature (e.g., Ivey & Ivey, 2003), consultations with counselor educators, and our own experiences to add one new subscale and several specific skills to the SCS. The new subscale was called Develops Therapeutic Relationship. It assesses the counselor's ability to develop a high-quality relationship with the client based on the core conditions. Under other subscales, we added the items "using a helpful vocal tone," "reflecting meaning and values," and "determining goals," based on a review of Ivey and Ivey's (2003) microskills. Finally, we introduced an item called "using strategies for creating change" to assess active interventions such as guided imagery, behavioral checklists, or the solution-focused consideration of exceptions.

Step 3: Eliminating and adjusting existing items. We omitted some counseling skills from the SCS because they were not necessary indicators of good counseling, especially if overused. Other skills overrepresented a particular theoretical perspective. For example, we eliminated self-disclosure as a separate category and, after focus group input, included it as an item under "using strategies for creating change." We also thought that too much attention was paid in the SCS to the decision-making and action dimension of counseling, given the short segments that would be viewed, particularly for counselors operating outside of the behavioral paradigm, so we combined all of the SCS's six action-oriented microskills into four skills under one subscale, Encourages Change.

Second Iteration: Feedback on the CSS From the Experts

Our next goals were to ensure face and content validity for the emerging instrument, which was now renamed the Counseling Skills

Scale (CSS). Therefore, we sought feedback from a convenience sample of CESNET electronic mailing list members who teach counseling skills (prepracticum), practicum, or internship courses. We sent a draft of the CSS with a request for suggestions for improvement to each participant. Nineteen counselor educators with varying backgrounds, gender, and race responded.

We made several changes based on the experts' feedback. First, we changed the rating scale by making its descriptions clearer and more explicit, as noted previously in Step 1: More Precise Scoring. Next, we made the following two changes to items under the subscale Encourages Exploration: We added "reflection of content" as a synonym for the item "paraphrasing," and we added the qualifier "limited use of questions when theoretically consistent" under the item "questioning" to allow for a strongly person-centered approach. To the subscale Encourages Change, we added "intentionality" of interventions, "theoretically consistent," and "uses reframing" as qualifiers to the item "using strategies for creating change." Finally, we added a section for comments by the instructor, and we changed the second subscale name from the earlier name Builds Trust to Shows Interest and Appreciation.

Third Iteration: Reliability and Validity for the CSS

Next, we sought to further increase reliability and validity for the CSS and, in the process, created a third iteration of the instrument. We began by pilot testing the instrument in a focus group to solicit feedback and to obtain initial measures of interrater reliability. We then used the CSS to rate student counseling skills (a) as students began their counseling program (pretest) and (b) as students finished their first-semester counseling skills course (posttest). We compared their pretest and posttest scores to obtain a measure of construct validity. We used their posttest scores to compute the internal consistency of the instrument.

Members of the focus group were five counselor educators (including the two authors), from two universities, who had varying degrees of teaching experience. All members of the group were currently teaching counseling skills, practicum, and/or internship courses. These five individuals participated in a pilot test of the CSS by watching student videotapes as a group and then independently rating student skills using the CSS. The focus group was partially intended to be a training session for instructors who would be using the CSS during the upcoming semester as an evaluative tool in the Counseling Techniques course. Therefore, the participants were a convenience sample of all the faculty at one university who would be teaching the course (one of whom was the first author) and one faculty member at the other university who was to be teaching the course (the second author). The group consisted of

two men and three women, ages 28 to 53, all Caucasian. They watched 10 minutes of videotape from the beginning of the session and 10 minutes from the end of the session.

After the initial ratings, the group members were asked what their ratings were and the reasons for giving those ratings. If the other member of the group had a different understanding of the items than we did, the first author explained the intention of the items. During this discussion, group members made suggestions about how to make the items clearer. Each member was then given a chance to re-rate the items, if she or he desired, based on the discussion. Computations of interrater correspondences overall increased from 56.84% to 76.8% as a result of discussions held among focus group members. The increase implies that training in the use of the instrument will be necessary in the future to ensure acceptable levels of interrater reliability. Although the goal of the discussions was to move the focus group members toward consensus so that on future ratings they would be more likely to use the instrument in a manner similar to other raters, greater consensus may also indicate that group members were unduly influenced by the discussion to change their ratings in ways that they will not continue to do on their own. This affects the ability to depend on the interrater reliability results. However, because the instrument has been altered based on focus group suggestions, there is already a need to reassess interrater reliability from ratings of actual research participants in this study or future studies.

The discussion among focus group members was helpful in other ways. The rich, thoughtful ideas about how to alter the scale to improve interrater reliability contributed a great deal to the newest version of the CSS. Accordingly, we made the following changes. We reordered some of the subscales and added *Manages the Session* as a subscale. We revised two items of the subscale *Shows Interest and Appreciation* as follows: added the words "professional appearance" to the item "body language," which now reads "keeping open body language and professional appearance"; changed "verbal tracking" to "minimal encouragers," which now reads "using minimal encouragers helpfully." In the subscale *Deepens the Session*, we changed "confronting" to "challenging" in the item "challenging and pointing out discrepancies." In the subscale *Encourages Change*, we added more options under the item "using strategies for creating change." Finally, we placed the rating scale definitions at the top of each page, as a reminder, and placed NN (not needed) at the end of most of the Likert scales, also as a reminder.

The "Final" Instrument

In its final iteration, the CSS is a 22-item measure of the effectiveness of counseling skills performance. The performance of skills is rated on a 5-point Likert-type scale by experts (e.g., instructional

faculty) who have viewed a full counseling session, either live or videotaped. Faculty rate each of the 22 skills as *highly developed* (+2), *well developed* (+1), *developing skills* (0), *continue practice* (-1), or *major adjustment needed* (-2). Each marker is defined further on the instrument itself in order to help the rater to determine a score. The 22 items are subsumed into six subscales, which convey essential conditions for helping clients. The first four subscales represent possible stages of a helping interview, from showing interest through encouraging change. The last two subscales represent more global conditions for effective helping, namely Develops Therapeutic Relationship and Manages the Session. The first subscale is Shows Interest and Appreciation, which includes the items "maintaining eye contact," "keeping an open body language and professional appearance," "using minimal encouragers helpfully," "expressing a caring vocal tone," and "evoking and punctuating client strengths." The second subscale is called Encourages Exploration and includes the separate items "questioning," "requesting concrete and specific examples," "paraphrasing (reflection of content)," and "summarizing." Deepens the Session is the third subscale, under which the items "reflecting feeling," "using immediacy," "observing themes and patterns," "challenging and pointing out discrepancies," and "reflecting meaning and values" are included. The fourth subscale is Encourages Change, which consists of the items "determining goals and desired outcomes," "using strategies for creating change," "considering alternatives and their consequences," and "planning action and anticipating possible obstacles." Within these first four subscales, which incorporate 18 of the 22 items, are the sequenced steps of a solution-oriented counseling interview. The fifth major subscale is Develops Therapeutic Relationship, which consists of a single global item that refers to consistently engaging in a caring, genuine, respectful relationship with the client. It is most obviously representative of the core conditions (Rogers, 1957) for helping. The final subscale, Manages the Session, consists of three items by which the rater evaluates, after watching the entire session, the counselor's success at managing the session: "opening the session," "directing the session," and "closing the session." Each of the 22 items is explicitly defined on the instrument in a short paragraph that follows its label.

For scoring purposes, the rater averages the scores for the items under each of the six major subscales, leading to six subscale scores, namely Shows Interest and Appreciation, Encourages Exploration, Deepens the Session, Encourages Change, Develops Therapeutic Relationship, and Manages the Session. Raters may also indicate NN, or not needed, for any of the individual items when the counselor uses other skills in that subscale to successfully accomplish the purpose of that subscale. In such a case, only the items receiving a numerical score would be averaged into the subscale score.

A total score for counseling competence is created by adding together the six subscale scores.

Determining Construct Validity and Internal Consistency of the CSS

Once we had developed the "final" version of the CSS, we sought to examine construct validity and internal consistency.

Method. Two instructors (trained during the focus group, but not the authors) used the CSS as a pretest and posttest measure in three sections of the course Theories and Techniques of Counseling. We thereby tested the ability of the CSS to measure the construct "counseling skills performance" by comparing student ratings before and after a course in which they were taught counseling skills. During the first week of class, a baseline tape was made: Students paired up with a classmate to do a 10-minute, videotaped session in which they were to demonstrate their natural ways of helping (Ivey & Ivey, 2003). These students' pretest videotapes were then evaluated by the course instructor using the CSS. At the end of the semester, the same students' final (i.e., posttest) videotapes were evaluated by the instructors. The instructors did not have access to the initial ratings after they had rated the students. Means and ranges for all variables were computed. Cronbach's alpha was computed on the posttests. The means of the pretest and posttest total CSS scores were compared using a paired sample *t* test. The means of the pretest and posttest subscale scores were also compared using paired sample *t* tests.

Participants. The entire population ($N = 29$) of all three sections of the Theories and Techniques of Counseling course (a first-semester course) that were offered in a small, rural, teaching university in the mid-Atlantic region of the United States participated in the research. Participants' ages ranged from 22 to 42 years, with a mean of 26.38 years. Sixty-two percent ($n = 18$) were female, and 38% ($n = 11$) were male. Seventy-six percent ($n = 22$) were White and non-Hispanic, 17% ($n = 5$) were African American, 3% ($n = 1$) were Asian American, and 1 participant did not identify a race. Seventy-six percent ($n = 22$) of those who chose to answer ($n = 24$) were full-time students, and 7% ($n = 2$) were part-time students. Thirty-one percent ($n = 9$) of those who chose to answer ($n = 26$) were in the school counseling track, 24% ($n = 7$) in community counseling, 24% ($n = 7$) in student affairs administration, 7% ($n = 2$) in student affairs counseling, and 3% ($n = 1$) in other university programs. (Percentages have been rounded.)

Results. Cronbach's alpha conducted on the posttest was .90, a high level of internal consistency, particularly considering the small sample size. Statistics indicated that removing Items 1 and 20 would improve the internal consistency. These were items related to maintaining eye contact and opening the session smoothly.

When these two items were removed from calculations, Cronbach's alpha was .91. These two items were later integrated into other items because (a) they were skills that everyone used and therefore did not help to distinguish between counselors who were more skilled and those who were less skilled and (b) it made sense to leave them in the instrument somewhere because students would be expected to use them.

Construct validity was assessed by examining pre- to postcourse changes in student performance and by conducting an item analysis. Table 1 shows descriptive statistics and the results of paired sample *t* tests conducted on the pretest and posttest total and subscale CSS scores. Significant changes in student performances were found for overall scores ($t = 4.51, p < .000$) and for five of the six subscales. The mean of the posttest scores was 2.33 points higher than the mean of the pretest scores. The effect size of this change was .80, indicating that the change can be considered meaningful. The only subscale that did not reflect significant improvement despite its increase from pretest to posttest was Develops Therapeutic Relationship, a measure of how well students practiced the core conditions. This result may reflect a selection factor; that is, perhaps most students who choose counseling as a profession find it easy to demonstrate the core conditions even without training.

Item analysis examines the degree to which each item and each subscale contribute to the total score. Validity of items depends on each item's contribution to overall counselor competency. Cronbach's alpha and item total statistics were computed for the subscale scores and for the items within each subscale. All items and subscale scores correlated positively with the total score. Corrected item correlations with the total score ranged from .18 to .71, with 68% above .41. Corrected subscale correlations with the total score ranged from .44 to .73, with an alpha of .78. Ten items correlated no higher with any other subscale than with their own subscale, but 11 items correlated higher with at least one other subscale than with their own. Therefore, although the scales may have face validity for counselor educators and content validity with counseling skills texts, the items within the scales do not represent true factors. With more participants and the ability to perform a factor analysis, we will be able to determine what the factors are and which items fit within those factors.

Discussion

The CSS builds on the work of previous attempts to determine counselor competence. Those attempts were especially challenging because of the complexity of the counseling act; that is, client variability, context, and external situations can all influence the

TABLE 1

Descriptive Statistics and Paired Sample *t* Tests on Pretest and Posttest Total and Subscale CSS Scores

Pair	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% CI of Difference		<i>t</i>	<i>df</i>	Sig. (2-tailed)
				Lower	Upper			
1 Total before	2.42	2.22	0.41					
Total after	4.75	1.75	0.33					
Change in total before to after	2.33	2.79	0.52	3.39	1.27	4.51	28	.000
2 Interest before	0.86	0.41	0.08					
Interest after	1.16	0.35	0.07					
Change in Interest before to after	0.29	0.54	0.10	0.50	0.09	2.90	28	.007
3 Exploration before	0.14	0.43	0.08					
Exploration after	0.73	0.46	0.09					
Change in Exploration before to after	0.60	0.72	0.13	0.87	0.32	4.46	28	.000
4 Deepens before	0.06	0.56	0.10					
Deepens after	0.41	0.35	0.06					
Change in Deepens before to after	0.48	0.68	0.13	0.74	0.22	3.82	28	.001
5 Change before	0.16	0.64	0.12					
Change after	0.56	0.42	0.08					
Change in Change before to after	0.40	0.69	0.13	0.66	0.14	3.14	28	.004
6 Relationship before	1.00	0.46	0.09					
Relationship after	1.14	0.58	0.11					
Change in Relationship before to after	0.14	0.69	0.13	0.40	0.13	1.07	28	.293
7 Management before	0.33	0.48	0.09					
Management after	0.75	0.33	0.06					
Change in Management before to after	0.41	0.63	0.12	0.65	0.17	3.55	28	.001

Note. CSS = Counseling Skills Scale. CI = confidence interval. Total = overall scores. CSS subscale names: Interest = Shows Interest and Appreciation; Exploration = Encourages Exploration; Deepens = Deepens the Session; Change = Encourages Change; Relationship = Develops Therapeutic Relationship; Management = Manages the Session.

counseling process. The CSS is an attempt to incorporate those subtleties into an instrument, while maintaining the precision of a skills-based evaluation. To do so, the CSS asks raters to note the context appropriateness of the use of 22 skills grouped into

six subcategories based on a moment-by-moment viewing of a counseling session. The items for the CSS were derived from the current canon of counseling skills and based on the research on effective counseling (e.g., Wampold, 2001), international usage in teaching counseling skills, and wide consultation with counselor educators. The precision of the CSS items and the context sensitivity of its rating scales are advances over the more global and merely enumerative assessment of counselor competence that characterized previous measures. In our attempts to increase face and content validity, we drew from the content and critique derived from the years of clinical and teaching experience of 24 counselor educators who participated in either a survey or a focus group and from counseling techniques texts. The data reported in this study provide an early indication of support for the validity of the CSS, because it measured significant positive changes occurring after a course in counseling skills. Furthermore, all items and subscale scores varied in a positive direction with the total score, and all items within subscales varied in a positive direction with their subscale scores. Finally, the CSS's overall internal consistency was also well established.

Much future research is needed on the instrument because this initial study was limited by its use with a small group of students, two raters, and one university. It is further limited by the lack of interrater reliability information on the data from the study. In addition, although it successfully measured change in the level of the students' counseling skills from the beginning to the end of the skills course, more important data might emerge from measuring such change from the beginning to the end of the counseling program. Finally, it is possible that posttest ratings were affected by the instructors' memories of the students' pretest ratings.

We will, therefore, continue to study the CSS performance in measuring counselor competence of students in techniques, practicum, and internship courses. We will train raters in order to encourage use of the CSS and to increase our sample size. Finally, we will explore criterion validity by examining the relationships between the CSS and other established measures of characteristics that are related to counseling skills, for instance, measures of empathy.

Conclusion

We believe that we have begun to develop an instrument that moves the measurement of counseling skills performance to higher levels of reliability and validity. We hope that the CSS will serve as a vehicle to assist in the training of counseling students, as an accountability measure for accreditation purposes, as a means to

determine how to improve teaching and programming to ensure greater counselor competence, and as a research tool in continuing attempts to determine how to improve counselor training.

Note. Copies of the Counseling Skills Scale may be obtained from the first author.

References

- Alberts, G., & Edelstein, B. (1990). Therapist training: A critical review of skill training studies. *Clinical Psychology Review, 10*, 497-511.
- Altekruse, M. K., & Brown, C. (1999, October). *An instrument for analyzing counselor responses*. Paper presented at the meeting of the Association for Counselor Education and Supervision, New Orleans, LA.
- Eysenck, H. J. (1966). *The effects of psychotherapy*. New York: International Science Press.
- Hackney, H., & Cormier, S. (1994). *Counseling strategies and interventions*. Boston: Allyn & Bacon.
- Ivey, A. E. (1971). *Microcounseling: Innovations in interview training*. Springfield, IL: Thomas.
- Ivey, A. E., & Ivey, M. B. (2003). *Intentional interviewing and counseling*. Pacific Grove, CA: Brooks/Cole.
- Kolb, D. (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice Hall.
- McLennan, J. (1994). The skills-based model of counseling training: A review of the evidence. *Australian Psychologist, 29*, 79-88.
- Myrick, R. D., & Kelly, D. F. (1971). A scale for evaluating practicum students in counseling and supervision. *Counselor Education and Supervision, 12*, 330-336.
- Ponterotto, J. G., & Furlong, M. J. (1985). Evaluating counselor effectiveness: A critical review of rating scale instruments. *Journal of Counseling Psychology, 32*, 597-616.
- Rogers, C. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95-103.
- Shaw, B. F., & Dobson, K. S. (1988). Competency judgments in the training and evaluation of psychotherapists. *Journal of Consulting and Clinical Psychology, 56*, 666-672.
- Strupp, H. H. (1986). Psychotherapy: Research, practice, and public policy (how to avoid dead ends). *American Psychologist, 41*, 120-130.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy*. Chicago: Aldine.
- Urbani, S., Smith, M. R., Maddux, C. D., Smaby, M. H., Torres-Rivera, E., & Crews, J. (2002). Skills-based training and counseling self-efficacy. *Counselor Education and Supervision, 42*, 92-106.
- Wampold, B. E. (2001). *The great psychotherapy debate*. Mahwah, NJ: Erlbaum.
- Young, M. E. (1998). Skills-based training for counselors—Micro-skills or mega-skills? *Counseling and Human Development, 31*, 1-12.

Copyright of Counselor Education & Supervision is the property of American Counseling Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.