

## Evaluations of supervisees: Brief commentary and research report

By: [L. DiAnne Borders](#) and Margaret L. Fong

Borders, L. D., & Fong, M. L. (1992). Evaluations of supervisees: Brief commentary and research report. *The Clinical Supervisor*, 9(2), 43-51. [https://doi.org/10.1300/J001v09n02\\_05](https://doi.org/10.1300/J001v09n02_05)

This is an Accepted Manuscript of an article published by Taylor & Francis in *The Clinical Supervisor* on 13 April 1992, available online:

[http://www.tandfonline.com/10.1300/J001v09n02\\_05](http://www.tandfonline.com/10.1300/J001v09n02_05).

\*\*\*© 1992 The Haworth Press. Reprinted with permission. No further reproduction is authorized without written permission from Taylor & Francis. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. \*\*\*

### **Abstract:**

A fairly substantial body of literature has investigated sources of variability in counselor performance ratings. No comparisons of supervisors' and external judges' evaluations, however, have been reported. In this study, individual supervisors (n = 27) evaluated their counseling students (n = 43), who were at three training levels (first practicum, second practicum, and internship). External judges rate audiotapes of students' actual counseling sessions. The correlation between the two ratings was not significant. Follow up analyses of variance revealed that, while external judges' ratings increased across training levels, supervisors' evaluations of second practicum students were significantly lower than those of their less or more experienced counterparts. Results seemed to support Holloway's (1984) assertion that supervisors' evaluations of counseling performance are influenced by their interactions with students during supervision sessions.

**Keywords:** counseling | clinical supervision | student counselors | counselor supervision | training

### **Article:**

Accurate evaluations are "an ethical responsibility the supervisor has to the counselor, the profession, and to future clients of the supervisee" (Bernard, 1987, p. 54). Supervisors are the primary source of evaluations of counseling performance (Galassi & Trent, 1987), and rating scales are the predominant method of evaluation (Fuqua, Newman, Scott, & Gade, 1986). It is becoming increasingly clear, however, that evaluating counselor performance is a complex and inexact task, and that even "objective" rating scales are clearly fallible (Fuqua et al., 1986).

Past research has identified several sources of variability in ratings of counseling performance. One primary source of variance is the rater. Results of a number of studies (e.g., Bishop, 1971; Borman & Ramirez, 1975; Bozarth & Grace, 1970; Brown & Cannaday, 1969; Burck, Jacobs, Sauber, Stone, & Thomson, 1973; Fish, 1970; Fuqua, Johnson, Newman, Anderson, & Gade,

1984; Hansen, Moore, & Carkhuff, 1968; Kurtz & Grummon, 1972) indicated that evaluations often differ depending on whether supervisors, supervisees, peers, or external judges complete the scale. Other researchers reported that ratings vary based on which scale evaluators complete (Fuqua et al., 1986), what type of training raters receive (Newman & Scott, 1988), and whether the client improves or deteriorates (Ward, Friedlander, Schoen, & Klein, 1985).

Rating scales themselves also have been criticized. First, reliability and validity data are lacking for most measures (Ford, 1979; Newman & Scott, 1988; Ponterotto & Furlong, 1985; Scofield & Yoxtheimer, 1983). Second, counseling performance is considered a multidimensional construct, yet most rating scales attempt to measure a single, global factor (Newman & Scott, 1988; Scott, Fuqua, & Newman, 1986). In addition, items on the rating scales typically assess only in-session counseling behavior. Holloway (1984) asserted, however, that supervisors' ratings also are influenced by counselors' behaviors during *supervision* sessions.

To explore Holloway's (1984) concern about supervisors' ratings, we utilized a subset of rating data that had been collected as part of a larger study on supervision (Borders, 1985). Our focus was the relationship between supervisors' evaluations and independent judges' ratings of supervisees, a comparison that has not been investigated in previous studies.

## **METHOD**

### **Participants**

Participants were counseling students enrolled in a field-based practicum or internship and their supervisors at a major southeastern university. Out of a potential pool of 71, 43 (29 females, 14 males) students and their 27 assigned supervisors agreed to participate. The students represented all programs in the counselor education department (e.g., school counseling, community mental health, student personnel, counseling psychology). All had successfully completed introductory skills and theories courses; second practicum and internship students had completed varied additional coursework. Ages ranged from 22 to 51 ( $M = 30.4$ ,  $SD = 7.1$ ). Of the 43 participants, 23 were enrolled in first practicum, 7 in second practicum, and 13 in internship.

Supervisors included 10 counselor education faculty members, 4 adjunct faculty, 8 approved on-site supervisors, and 5 counselor education doctoral students. Supervisors met individually with their supervisees for one hour weekly, and critiqued a minimum of 11 audiotaped counseling sessions during the semester.

The two external judges were an advanced doctoral counseling student and a Ph.D. counselor who were trained to use the rating scale according to suggested procedures (Strupp, 1981).

### **Measures**

Our review of previous research indicated measures of facilitative skills were used almost exclusively, even when more experienced counselors were evaluated. In this study, both instruments included beginning (facilitative) skills and more advanced counseling behaviors.

*Supervisors' evaluations.* Supervisors completed the Counselor Evaluation Rating Scale (CERS; Myrick & Kelly, 1971) for each of their supervisees. The CERS consists of 27 Likert-type items; 13 comprise the counseling performance subscale; another 13, the supervision subscale. Ratings on the two subscales and a final global item ("Can be recommended for a counseling position without reservation") are summed for a total rating of counseling effectiveness.

Myrick and Kelly (1971) reported a split-half reliability coefficient of .95, a 4-week test-retest reliability coefficient of .94, and a correlation of .86 for the two subscales; Jones (1974) reported a split-half reliability of .87. Loesch and Rucker (1977) found some support for the counseling and supervision subscales in a factor analysis, but concluded the total CERS score was more valid.

Total CERS scores were used in this analysis, as recommended by Loesch and Rucker (1977). Scores ranged from 71 to 189 ( $M = 162.28$ ,  $SD = 24.99$ ). Three students received perfect scores (189) and 11 had scores of 180 or higher.

*Vanderbilt Psychotherapy Process Scales.* External judges used the Vanderbilt Psychotherapy Process Scales (VPPS; O'Malley, Suh, & Strupp, 1983; Strupp, 1981; Suh, Strupp, & O'Malley, 1986) to rate supervisees' audiotapes. The VPPS includes positive and negative behaviors and attitudes the authors assume either facilitate or impede progress in counseling. The 80 Likert-type items are scored on a scale ranging from 1 (not at all) to 5 (a great deal). Factor analyses of items on the current form yielded eight subscales, five addressing client behavior and attitudes and three addressing therapist behavior and attitudes. Internal consistency of the subscales ranged from .82 to .96. Subscale scores have been used in past research rather than a total VPPS score.

For the purposes of this study, only the three subscales assessing therapist behavior and attitude (items 44-80) were used: (1) Therapist Warmth and Friendliness; (2) Negative Therapist Attitude; and (3) Therapist Exploration. These subscales include items reflecting facilitative skills and more advanced counseling behaviors.

The external judges rated the tapes using the systematic sampling procedure (first, middle, and closing 5 minute segments) advocated by Suh et al. (1986). Their interrater reliability ( $r = .86$ ) fell within the acceptable range (Suh et al., 1986). The sum of the two ratings was used as the final score for each audiotape. VPPS scores ranged from 154 to 267 ( $M = 211.63$ ,  $SD = 31.42$ ).

## **Procedure**

Volunteer participants were asked to submit an audiotape of a "working" counseling session (at least third session, excluding termination) that they believed was representative of their work. Each participant received an audiotape that was marked only with an identification number. Consent forms were obtained from supervisees, audiotaped clients, and supervisors.

During the tenth week of the semester, individual supervisors were mailed a copy of the CERS (one per supervisee), a letter of instructions for completing the scale, and a stamped return envelope. The letter paired each student with his/her identification number; the CERS was

marked only with the student's identification number. Supervisors were asked to return the CERS within one week; all supervisors responded within two weeks.

## DATA ANALYSIS AND RESULTS

Means and standard deviations of CERS and VPPS scores by training levels are presented in Table 1. Both measures had large standard deviations, indicating substantial variance among individual ratings. To answer the research question, we computed a Pearson product-moment correlation between the CERS and VPPS ratings. The correlation ( $r = .1231$ ) was not significant ( $p = .4317$ ), indicating there was no significant relationship between supervisors' global evaluations and external judges' ratings of counseling performance during a particular session.

**Table 1.** Means and Standard Deviations of CERS and VPPS Scores by Training Level

	Training Level			
	First Practicum ( $n=23$ )	Second Practicum ( $n=7$ )	Internship ( $n=13$ )	Total ( $n=43$ )
CERS				
<i>M</i>	164.74	142.14	168.77	162.28
<i>SD</i>	17.34	41.98	21.42	24.99
VPPS				
<i>M</i>	204.65	214.43	222.46	211.63
<i>SD</i>	28.88	16.60	39.53	31.42

It can be noted that while means on the VPPS increased as training level advanced, mean scores on the CERS declined at second practicum. To investigate this unexpected pattern of mean ratings, two follow up analyses of variance were conducted. There were significant quadratic (but not linear) differences,  $F(1,40) = 6.17, p = .0173$ , in the mean CERS ratings across the three training levels. Supervisors' ratings of supervisees in second practicum were significantly lower than their ratings of supervisees in first practicum and internship. There was a linear (but not quadratic) trend,  $F(1,40) = 2.72, p = .1070$ , in the mean VPPS ratings across the three training levels. External judges tended to give higher ratings to supervisees with more experience.

## DISCUSSION

This study found no relationship between supervisors' global ratings of their supervisees' counseling performance and external judges' ratings of the supervisees' counseling behaviors and attitudes during an actual counseling session. These results suggest that, though both sets of raters evaluated counseling behavior, the two sets of raters may have based their evaluations on different criteria. These results, along with the relatively large variances among individual supervisors' ratings (see Table 1), lend support to Holloway's (1984) assertion that supervisors' evaluations may be influenced by their interactions with supervisees during weekly supervision sessions.

We might expect ratings of counseling performance to increase with experience, and interns did receive the highest ratings on both measures. But while external judges' ratings increased across training levels, supervisors' ratings of second practicum students were significantly lower than either their less or more experienced counterparts.

One possible explanation for the lower evaluations of supervisees in second practicum is supervisors' higher expectations for counseling performance. For example, while empathic responses and evidence of facilitative conditions may have been sufficient for high ratings at first practicum, the same skill level may have been rated lower at second practicum, with supervisors expecting more advanced skills and sophisticated interventions.

A theoretical explanation for our findings, and one that is related to Holloway's (1984) assertion, can be found in developmental models of supervision (e.g., Loganbill, Hardy, & Delworth, 1982; Stoltenberg, 1981; Stoltenberg & Delworth, 1987). These models suggest that supervisees in the second developmental stage experience confusion, conflict, instability, and feelings of incompetence about their counseling. Because these supervisees now begin to perceive the complexities of client dynamics and have more choices of skills to use, they may appear less effective than some supervisees at the first stage (Stoltenberg & Delworth, 1987). They also vacillate between dependence and autonomy in the supervisory relationship. In contrast to the assumptions they held during first practicum, supervisees now realize their supervisors are not omnipotent and all-knowing experts. As a result, they may express disappointment and anger either directly or indirectly to their supervisors (Loganbill et al., 1982).

If the second-practicum students in this study were in the second developmental stage, supervisors may have been influenced by students' "uncooperativeness" or by tension in the supervisory relationship when completing the evaluations. In contrast, first practicum supervisees may have been more dependent, obedient, and grateful, and interns may have approached their supervisors in a respectful, collegial manner (cf., Loganbill et al., 1982; Stoltenberg, 1981; Stoltenberg & Delworth, 1987).

The results of this study are subject to several limitations. Two different "work samples" were evaluated, although supervisors had critiqued audiotaped counseling sessions and, we assume, used that data in making their overall evaluations on the CERS. External judges, however, rated only one self-selected counseling session; it is quite possible that supervisees submitted an audiotape of their "best" work. In contrast, supervisors most likely reviewed audiotapes representing the range of supervisees' counseling abilities.

In addition, two different ratings scales were compared. Although both the VPPS and CERS address similar behaviors and attitudes of supervisees during counseling, the CERS also includes items for assessing behaviors and attitudes during supervision. Based on the results, it is unknown whether the lack of a relationship between the two ratings was due to this difference in the content of the measures. There also was variation in field supervisors, field settings, and clients, unavoidable by-products when one uses actual counseling experiences in research. Finally, the differential number of students at each training level also may have influenced the results.

## CONCLUSION

The results of this study suggest some directions for future investigations of supervisors' evaluations. Needed follow up studies include an investigation of the *criteria* supervisors use in evaluating supervisees at different training levels, and comparisons of supervisors' and external

judges' ratings of the same work sample on the same rating scale. In addition, it appears researchers should consider more variables, such as supervisees' training levels and developmental stages, and the interactive dynamics of the supervisory relationship. Such studies are needed to further clarify sources of variability in ratings of counselor performance.

## REFERENCES

Bernard, J.M. (1987). Ethical and legal considerations for supervisors. In L. D. Borders & G. R. Leddick, *Handbook of counseling supervision* (pp. 52-57). Alexandria, VA: American Association for Counseling and Development.

Bishop, J.B. (1971). Another look at counselor, client, and supervisor ratings of counselor effectiveness. *Counselor Education and Supervision, 10*, 319-323.

Borders, L. D. (1985). An exploration of the relationship of ego development theory to counselor development (Doctoral dissertation, University of Florida, 1984). *Dissertation Abstracts International, 45*, 2762A.

Borman, C. A., & Ramirez, C. (1975). Evaluating counseling practicum students. *Counselor Education and Supervision, 15*, 48-54.

Bozarth, J. D., & Grace, D. P. (1970). Objective ratings and client perceptions of therapeutic conditions with university center clients. *Journal of Clinical Psychology, 26*, 117-118.

Brown, D., & Cannaday, M. (1969). Counselor, counselee, and supervisor ratings of counselor effectiveness. *Counselor Education and Supervision, 8*, 113-118.

Burck, H., Jacobs, E., Sauber, R., Stone, W., & Thomson, W. (1973). Teaching a course in counseling theories and practice: Description and evaluation. *Counselor Education and Supervision, 12*, 162-171.

Fish, J. M. (1970). Empathy and the reported emotional experiences of beginning psychotherapists. *Journal of Consulting and Clinical Psychology, 35*, 64-69.

Ford, J. D. (1979). Research on training counselors and clinicians. *Review of Educational Research, 49*, 87-130.

Fuqua, D.R., Johnson, A. W., Newman, J. L., Anderson, M. W., & Gade, E. M. (1984). Variability across sources of performance ratings. *Journal of Counseling Psychology, 31*, 249-252.

Fuqua, D. R., Newman, J. L., Scott, T. B., & Gade, E. M. (1986). Variability across sources of performance ratings: Further evidence. *Journal of Counseling Psychology, 33*, 353-356.

Galassi, J. P., & Trent, P. J. (1987). A conceptual framework for evaluating supervisory effectiveness. *Counselor Education and Supervision, 26*, 260-269.

Hansen, J. C., Moore, G. D., & Carkhuff, R. R. (1968). The differential relationships of objective and client perceptions of counseling. *Journal of Clinical Psychology, 24*, 244-246.

Holloway, E. L. (1984). Outcome evaluation in supervision research. *The Counseling Psychologist, 12*(4), 167-174.

Jones, L. K. (1974). The Counselor Evaluation Rating Scale: A valid criterion of counselor effectiveness? *Counselor Education and Supervision, 14*, 112-116.

Kurtz, R. R., & Grumman, D. L. (1972). Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *Journal of Consulting and Clinical Psychology, 39*, 106-115.

Loesch, L. C., & Rucker, B. B. (1977). A factor analysis of the Counselor Evaluation Rating Scale. *Counselor Education and Supervision, 16*, 209-216.

Loganbill, C., Hardy, E., & Delworth, U. (1982). Supervision: A conceptual model. *The Counseling Psychologist, 10*(1), 3-42.

Myrick, R. D., & Kelly, F. D., Jr. (1971). A scale for evaluating practicum students in counseling and supervision. *Counselor Education and Supervision, 10*, 330-336.

Newman, J. L., & Scott, T. B. (1988). The construct problem in measuring counseling performance. *Counselor Education and Supervision, 28*, 71-79.

O'Malley, S. S., Suh, C. S., & Strupp, H. H. (1983). The Vanderbilt Psychotherapy Process Scale: A report on the scale development and a process-outcome study. *Journal of Consulting and Clinical Psychology, 51*, 581-586.

Ponterotto, J. G., & Furlong, M. J. (1985). Evaluating counselor effectiveness: A critical review of rating scale instruments. *Journal of Counseling Psychology, 32*, 597-616.

Scofield, M. E., & Yoxtheimer, L. L. (1983). Psychometric issues in the assessment of clinical competencies. *Journal of Counseling Psychology, 30*, 413-420.

Scott, T. B., Fuqua, D. R., & Newman, J. L. (1986). Further evidence for the unidimensionality of measures of counseling performance. *Psychological Reports, 58*, 983-990.

Stoltenberg, C. (1981). Approaching supervision from a developmental perspective: The counselor complexity model. *Journal of Counseling Psychology, 28*, 59-65.

Stoltenberg, C. D., & Delworth, U. (1987). *Supervising counselors and therapists: A developmental approach*. San Francisco: Jossey-Bass.

Strupp, H. H. (1981). *Vanderbilt Psychotherapy Process Scales (VPPS): Rater manual* (rev. ed.). Unpublished manuscript, Vanderbilt University, Nashville.

Suh, C. S., Strupp, H. H., & O'Malley, S. S. (1986). The Vanderbilt process measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 285-323). New York: Guilford.

Ward, L. G., Friedlander, M. L., Schoen, L. G., & Klein, J. G. (1985). Strategic self-presentation in supervision. *Journal of Counseling Psychology, 32*, 111-118.