Psychotherapy: Theory, Research, Practice, Training
2006, Vol. 43, No. 2, 154–172

Copyright 2006 by the American Psychological Association
0033-3204/06/$12.00 DOI: 10.1037/0033-3204.43.2.154

A NARRATIVE AND META-ANALYTIC REVIEW OF HELPING SKILLS TRAINING: TIME TO REVIVE A DORMANT AREA OF INQUIRY

CLARA E. HILL AND ROBERT W. LENT
University of Maryland

The authors review previous narrative and meta-analytic reviews on the effectiveness of overall helping skills training programs. The authors then review narrative reviews and conduct a new meta-analysis of specific methods used to teach helping skills within these programs. Our meta-analysis found that, in the aggregate, training methods substantially outperformed no training conditions, and that effect sizes did not vary as a function of trainee educational level (graduate vs. undergraduate students) or the type of criterion measure (interview-based vs. analogue-based empathy measures). Direct comparison of the training methods revealed that modeling outperformed instruction and feedback, and multimethod outperformed single-method training. The authors critique the literature and suggest that the studies in the helping skills literature generally fail to meet contemporary methodological standards, thereby limiting the conclusions that can be drawn. The authors appeal for better research on helping skills training, especially as it is currently practiced.

Keywords: helping skills training, training of psychotherapists

The late 1960s to early 1970s was an exciting time for helping skills training. Numerous training programs (e.g., Carkhuff, 1969; Ivey, 1971) were developed to teach discrete helping skills (e.g., reflection of feelings) to beginning therapist trainees, and training was extended to many new populations as well (e.g., peer helpers, parents, teachers, and children). In addition, research on helping skills training was prolific at that time, helping to establish the empirical basis for the training programs (see reviews by Alberts & Edelstein, 1990; Baker & Daniels, 1989; Baker, Daniels, & Greeley, 1990; Ford, 1979; Kurtz, Marshall, & Banspach, 1985; Matarazzo, 1971, 1978; Matarazzo & Patterson, 1986; Matarazzo, Wiens, & Saslow, 1966; Russell, Crimmings, & Lent, 1984).

Although helping skills training has become quite common as the first step in therapist training programs, research on helping skills training has slowed to a virtual standstill. In recent years, attention has shifted to supervision and the practicum level of therapist development (e.g., Ellis, Ladany, Krengel, & Shult, 1996), which most often occurs after helping skills training. This trend is reflected in the major reviews of therapist training and supervision over the past few decades. For example, the 1st edition of the *Handbook of Counseling Psychology* included a chapter covering both training and supervision (Russell et al., 1984), but the 2nd and 3rd editions focused almost exclusively on supervision (Goodyear & Guzzardo, 2000; Holloway, 1992). Likewise, the *Handbook of Psychotherapy and Behavior Change* included chapters on training in

Clara E. Hill, Department of Psychology, University of Maryland, and Robert W. Lent, Department of Counseling and Personnel Services, University of Maryland.

Portions of this paper were presented at the annual meeting of the Society for Psychotherapy Research, Rome, Italy, June 2004, and at the American Psychological Association, August, 2005. We are grateful to Bruce E. Wampold for consultation on the meta-analysis, and to Charles J. Gelso and Sarah Knox for commenting on a draft of the manuscript.

Correspondence regarding this article should be addressed to Clara E. Hill, PhD, Department of Psychology, University of Maryland, College Park, MD 20742. E-mail: hill@psyc.umd.edu

its first 3 editions (Matarazzo, 1971, 1978; Matarazzo & Patterson, 1986), but this focus was absent from the 4th and 5th editions.

This deceleration of inquiry into helping skills training deserves closer scrutiny. Perhaps trainers and researchers have assumed that the effectiveness of helping skills training programs has been well-established (see Baker et al., 1990). A closer look reveals, however, that numerous methodological and conceptual problems, as well as uneven study of key issues, limit the conclusions that can be drawn from the early research (see Gormally & Hill, 1974; Lambert, DeJulio, & Stein, 1978; Russell et al., 1984). Neither has there been much effort to discern which specific methods of training are most valuable and for whom training is most helpful.

The purpose of this paper is to revisit the helping skills literature, evaluating what is known about helping skills training and what remains unknown. Given that helping skills training is an integral part of training novice therapists, we need to know more about its effectiveness and the best methods for training. Our goal is to stimulate new research on helping skills training by approaching the literature from a fresh perspective, thereby encouraging a deeper level of understanding about how to train novice therapists. We thus review the narrative and meta-analytic reviews related to the effectiveness of overall helping skills training programs. We then review the narrative reviews of the effectiveness of specific training methods and conduct a meta-analysis of these studies since no meta-analysis has yet been done on this literature. In addition, we review the few studies on moderating variables that may affect the process and outcome of helping skills training. We then offer a critique of this body of research and make recommendations for future research.

Prior to beginning our review, we should clarify some terms that we use throughout the manuscript. By *helping skills*, we refer to specific verbal skills such as open questions, reflections of feelings, interpretations, and direct guidance (these skills are also referred to as the *targets of training*). By *trainees*, we generally refer to undergraduate or graduate students who are training to become mental health professionals. By *training programs*, we refer to multimethod programs designed to teach a set of helping skills. *Specific methods* of training refer to particular techniques (e.g., instruction, modeling) used to teach helping skills. *Moderators* refer to factors that modify the effectiveness of training (e.g., trainee variables).

We should also clarify that helping skills training differs from the broader area of therapist training or clinical training in that it is more delimited. Helping skills training involves teaching specific verbal skills to novice therapists rather than more broadly training students in the whole gamut of issues related to becoming a therapist (e.g., assessment, core psychology, specific theoretical orientations, and empirically supported manuals). Helping skills training is typically implemented prior to individualized supervision and practicum training, and is sometimes also referred to as prepracticum training.

The Effectiveness of Helping Skills Training Programs

The helping skills training programs developed more recently (e.g., Brammer, 1973; Brammer & MacDonald, 2003; Cormier & Cormier, 1998; Danish & Hauer, 1973; Egan, 1975, 1994; Evans, Hearn, Uhlemann, & Ivey, 1998; Gazda, Asbury, Balzer, Childers, & Walters, 1977; Goodman & Dooley, 1976; Hill, 2004; Hill & O'Brien, 1999; Meier & Davis, 2001; Okun, 2002) were all variations of training programs that were developed in the late 1960s and 1970s (e.g., Carkhuff, 1969; Ivey, 1971). These early training programs were all based loosely on the foundation of Rogers' (1942, 1951, 1957) client-centered therapy, which focused on therapist facilitative conditions as being the necessary and sufficient conditions for effective therapy. Although Rogers (1942) first considered the facilitative conditions as skills (e.g., restatements, reflections of feelings) that could be taught, he later came to the conclusion that they were attitudes that could not be taught (Rogers, 1957). Many of his followers (e.g., Carkhuff, 1969; Truax & Carkhuff, 1967), however, believed that the facilitative attitudes were skills that could be taught and hence developed systematic training programs to teach them.

The helping skills training programs that have received the most empirical attention are those developed by Carkhuff (1969); Ivey (1971), and Kagan (1984). Because most current training programs are based on these 3 foundational programs, we briefly describe them before reviewing the narrative and meta-analytic reviews of the effectiveness of these programs.

Carkhuff's (1972) Human Relations Training (HRT), also called Integrated Didactic Experiential Training (IDET), became one of the most popular training programs. In HRT, therapists are taught to progress through 3 stages with clients: (a) self-exploration (facilitated by therapist empathy, as communicated by nonverbal attending and reflection of feelings); (b) understanding (facilitated by advanced empathy, using skills of interpretation, self-disclosure, immediacy, confrontation, genuineness, and concreteness); and (c) action (facilitated by direct guidance skills such as problem-solving, decision-making, and behavioral techniques). Closed questions and sympathy are considered to be unproductive and even detrimental.

Ivey's (1971) microcounseling (MC) emerged as another popular training program. Although Ivey did not propose a stage model for counseling, he focused on skills similar to those in HRT and arranged them on a pyramid ranging from the easiest and most fundamental skills (e.g., attending behavior), to increasingly more complicated skills (e.g., reflection of feelings), to integration of the skills into a personal style and theory. MC involves teaching helping skills through (a) having trainees perform a baseline interview; (b) providing instruction and modeling on a specific skill; (c) having the trainee practice; and (d) providing feedback on how well the skill was implemented.

A third major training approach was Kagan's (1984) Interpersonal Process Recall (IPR). In IPR, trainees conduct an interview, followed by a recall session with an "inquirer," who asks the trainee to reflect on in-session thoughts and feelings, including those that might have interfered with the trainee's ability to help the client. IPR was based on the assumption that trainees typically already possess the requisite helping skills but get blocked from using them because of performance anxiety and the need for impression management.

Narrative Reviews of Training Programs

The earliest reviews of the helping skills training literature were narrative (i.e., nonmeta-analytic) reviews of the effects of training. For example, Matarazzo (1971) concluded that warmth and empathy can be taught to both lay and professional personnel, although she noted a number of methodological problems with the research (e.g., the skills were not operationally defined, rating scales used to measure attainment of skills were crude and subjective). In 1978, Matarazzo dismissed outcome studies of HRT because of the multitude of methodological problems (e.g., aspects of the training were unspecified, controls conditions were not adequate, the same rating scales were used both for training and to assess outcome, outcome was assessed through analogue methods rather than actual interview behavior, and judges of outcome measures were inadequately trained). At the same time, she concluded that there was sufficient evidence that MC could be taught effectively in a relatively short period of time to appropriately selected students. She was particularly impressed that MC involved well-defined behavioral skills, carefully programmed teaching methods, and videotape feedback and ratings. In 1986, Matarazzo and Patterson (1986) added that the optimal sequencing of training methods had received little empirical attention.

In their review of MC studies, Kasdorf and Gustafson (1978) concluded that MC was effective but that skill acquisition varied among skills and individuals. Furthermore, they noted that little was known about the maintenance of skills and how skills are used in naturalistic interview settings, but concluded that skills might deteriorate if not practiced. They also concluded that MC compared favorably with other training formats but that its combination with IPR was even more effective. Ford (1979) concluded that MC and HRT were both effective training programs. Similarly, Russell et al. (1984) concluded that MC "represents perhaps the most clearly articulated and systematic program for training in basic counseling skills" (p. 650).

Meta-Analytic Reviews of Training Programs

In more recent years, meta-analytic techniques have been developed that offer the ability to aggregate findings across studies quantitatively, producing more precise estimates of the effects of a given training program. Using these meta-analytic techniques, Baker et al. (1990) reported overall large (1.07), medium (.63), and small (.20) effect sizes for HRT, MC, and IPR, respectively, when used with graduate-level therapists. These effect sizes, which can be interpreted as the difference in standard deviation units between training and control groups, must be considered

cautiously because they were generated from a relatively small set of studies (HRT = 8, MC = 23, IPR = 10). Moreover, there may have been a confound between type of training and length of training, given that the average length was 37 hours for HRT compared to 19 hours for IPR and 9.5 hours for MC.

On balance, the meta-analytic findings suggest that trainees profited more from HRT than MC or IPR, and more from MC than IPR. Although the reasons for the effect size differences are not entirely clear, they may reflect the utility of a more structured, instructor-directed focus on learning specific helping skills, particularly at earlier stages of training when trainees are less sure about what skills to use within particular therapy scenarios, such as establishing a relationship or facilitating client exploration. Indeed, findings in the larger educational literature suggest the utility of instructor guidance and curricular focus over less structured, discovery-oriented learning methods (Mayer, 2004). Being shown by an expert what is likely (and not likely) to work may simply be more efficient than trying to discover this largely on one's own, albeit with the aid of an inquirer.

These comments also suggest that an important difference across training programs involves the role of the trainer: In HRT and MC, the trainer is essentially a didactic teacher who also employs experiential tools (e.g., modeling, practice); the IPR inquirer, by contrast, functions more like a therapy supervisor who relies to a greater degree on experiential methods (e.g., expanding the therapist's range of awareness). It is likely, therefore, that IPR is more effective after basic helping skills have been mastered (or at least acquired at minimally effective levels) than as an initial training method. When trainees are struggling with basic issues of what to do or say next to keep the therapy interview going, they may be less likely to have the cognitive and emotional wherewithal to achieve more abstract insights about themselves or their clients. Thus, the style of instruction in IPR may be best matched to the practicum stage of training that typically occurs after helping skills training. Unfortunately, as noted by Matarazzo and Patterson (1986), the optimal sequencing of training methods has received little empirical attention.

Baker et al.'s (1990) data also raised interesting questions about the cost-effectiveness of the different training programs (i.e., the effect sizes of each program in relation to the amount of time required to produce the training effect). HRT produced a larger ES but involved nearly 4 times more training time to produce this effect than MC. One interpretation of these results is that MC is an acceptable and less expensive alternative to HRT. We caution against such premature interpretations, however, given the different goals of the training programs, the fact that even 40 hours of training is modest compared to how training is currently implemented in undergraduate and graduate training (usually one to two semesters), and the methodological problems with this literature.

Finally, an earlier meta-analysis of MC (Baker & Daniels, 1989) found a large effect size for undergraduate trainees (1.18) but only a moderate effect size for graduate trainees (.66). Baker et al. (1990) suggested that it may be more difficult to teach MC to graduate than undergraduate students. However, as Goodyear and Guzzardo (2000) observed, such differences may have arisen from ceiling effects on the skill measures. That is, graduate students may have begun training with higher skill levels, on average, than did the undergraduates, and so would have had less room to show improvement. Given that students who apply to graduate training programs in therapy naturally tend to have good communication skills, Goodyear and Guzzardo's interpretation has merit. Regardless of the interpretation, however, Baker and colleagues' findings highlight the importance of accounting for level of training or professional status in training research. That a certain technique or program shows a large training effect with one group does not guarantee that its effects will be equally dramatic with another.

Investigations of Currently Used Helping Skills Training Programs

Although a fair amount of research has been conducted on HRT, MC, and IPR, which were the primary training programs used in the 1970s, other training programs currently in use (e.g., Brammer & MacDonald, 2003; Cormier & Cormier, 1998; Egan, 1994; Evans, Hearn, Uhlemann, & Ivey, 1998; Meier & Davis, 2001; Okun, 2002) have received little, if any, validation through outcome studies. One could argue that their effectiveness is suggested by generalizability from the more established programs they resemble (e.g., HRT and MC), but it is important

to test the effectiveness of each program relative to other programs and no-training.

One exception is recent research on Hill's (Hill, 2004; Hill & O'Brien, 1999) training program. This program integrates aspects of HRT, MC, IPR, and is based on extensive research on the effects of the skills in the therapy process. In this program, students learn about the skills using a 3-stage framework: Exploration (involves attending skills, open questions, restatements, and reflections of feelings), insight (involves challenges, interpretations, self-disclosures of insight, and immediacy), and action (involves information and direct guidance). Hill and Kellems (2002) found increases across the course of training in trainees' use of the helping skills, ability to establish a therapeutic relationship, and ability to conduct a good session, as assessed by volunteer clients. Although these initial findings are promising, more research is needed on the effectiveness of this training program in comparison to control conditions and alternative training programs.

Summary

In sum, narrative and meta-analytic reviews of the research on helping skills training programs indicate that HRT and MC are effective in comparison to no training control conditions, although less support has been found for IPR. In addition, 1 study has shown promising results for Hill's training model, although this study needs to be replicated. Unfortunately, conclusions drawn from the reviews are very global and tell us little about the specifics of why, how, and with whom these training programs are effective. Furthermore, as we will describe later, significant methodological problems limit our confidence in such results. More research is needed before we can make definitive statements about the effects of helping skills training programs.

The Effectiveness of Specific Methods of Helping Skills Training

Having reached, however guardedly, the conclusion that at least some training programs are effective at teaching helping skills, we still need to know more about what makes them effective. Hence, in this section, we consider what is known about the effectiveness of specific methods of training (e.g., instruction, modeling, feedback).

Narrative Reviews of the Effectiveness of Specific Methods

The only reviews that have been done thus far on this topic are narrative reviews. For example, in their review of MC, Kasdorf and Gustafson (1978) concluded that better results were obtained when more of the methods of MC (instruction, modeling, supervision, feedback, and self-observation) were used. They also noted that skill complexity influenced training, such that instructions were adequate with attending behaviors, whereas more complex skills such as reflection of feelings required more methods of training. In his review, Ford (1979) concluded that feedback, modeling, programmed texts, instructions, cocounseling, videotape self-confrontation, and behavioral rehearsal were all effective training methods. Matarazzo's (1978) review suggested that the most helpful methods were modeling, explicit definition of behaviors to be learned, practice, feedback, and deconditioning of trainee anxiety. Russell et al. (1984) suggested that modeling was the most effective method for teaching basic counseling skills. Matarazzo and Patterson (1986) found evidence for the effectiveness of supervision, modeling, and practice. Alberts and Edelstein (1990) concluded that instructions, modeling, feedback, and/or rehearsal lead to trainees' acquisition and demonstration of basic therapy responses. In sum, modeling and rehearsal/practice were each cited as effective in 4 reviews; instruction and feedback/supervision were each cited in 2 reviews; and self-observation/confrontation, cocounseling, and deconditioning of anxiety were each cited in 1 review.

Meta-Analysis of the Effectiveness of Specific Methods

Given the limitations of narrative reviews (e.g., impressionistic weighting of different studies' findings), we conducted our own meta-analysis of this literature. We gathered studies through a manual search of the *Journal of Counseling Psychology* and *Counselor Education and Supervision* (where most of the studies were published) from 1967 (when the first studies appeared) to the present; a review of the reference sections of identified articles; an examination of studies listed in a compendium of MC research (Daniels, 2001); a computerized search of PsycInfo (using

key words such as helping skills training, counseling training, therapist training, prepracticum training). We restricted our meta-analysis to studies focusing on exploration skills broadly operationalized as empathy, restatement, reflection of feelings, or "tacting" (defined vaguely by authors as following the client), and so excluded the few studies on decision-making. In addition, we excluded dissertations or unpublished research (because of the difficulty of finding complete data given how long ago these studies were completed), studies focused on training for participants who were not students or mental health professionals (e.g., peer counselors, teachers, parents, children), and studies that did not provide sufficient data to determine effect sizes. In studies that involved variations of a method (e.g., high vs. low empathy models), we used the higher quality variation in our analysis to provide the best chance for demonstrating the effectiveness of the method. Studies included in our meta-analysis are noted with asterisks in the Reference section; Table 1 provides a description of each study. Because we focused on methods whereas Baker et al. (1990) focused on entire training programs, there was minimal overlap between the studies included in 2 meta-analyses.

Because most of the research on methods is based on MC, which was influenced by Bandura (1969), the training methods that have been discussed most often are instruction, modeling, practice, and feedback. Instruction typically involved brief (5 to 10 minute) written and/or taped didactic information about the target skill (e.g., definition and rationale for using the skill). Modeling involved demonstrations of appropriate ways to perform the target skill. Modeling stimuli were typically presented via brief (less than 30 min) videotapes in which expert therapists demonstrated the skill in response to brief client statements. Feedback involved immediate reinforcement delivered via lights to signal that a given trainee response was appropriate or inappropriate, verbally via earphones or speakers (e.g., "excellent response"), or through extended (20–30 min) interpersonal interaction (which often involved feedback about the quality of the trainee's responses, along with modeling, didactic/instructional supervision, and experiential supervision). For the purpose of this meta-analysis, the various feedback methods were aggregated (a subsequent test of homogeneity did not reveal differences among the effect sizes of studies of different types of feedback). Although practice (also called rehearsal or role-playing) is another distinct training method (e.g., Fuqua & Gade, 1982; Schwebel, 1953; Spivack, 1973), we found only 1 study that compared practice to a no-training control (Hazler & Hipple, 1981) and so did not include practice in our review.

We first examined findings from studies that isolated and compared training methods to no-training controls. Within this group of studies, we further examined whether type of method, type of outcome measure, and level of trainee made a difference. We then examined studies that compared methods within the same study. We calculated g (i.e., the difference between the means on the outcome measure for the 2 conditions divided by the pooled standard deviation). To correct for the bias in g, unbiased estimates of the population effect size (d) were calculated (Hedges & Olkin, 1985). The variance of d was also estimated. Multiple measures were aggregated within each study by calculating a mean effect size and standard error, assuming correlations among the measures (Hedges & Olkin, 1985; Wampold, 1997), so that each study contributed only one effect size to any single meta-analysis. In these analyses, effect size estimates reflect the difference, expressed in standard deviation units, between 2 conditions (e.g., if d for the comparison between Conditions A and B is .50, the difference between the means of A and B is one half of a standard deviation). Using Cohen's (1988) criteria, ds of .2, .5, and .8 reflect small, medium, and large effect sizes, respectively.

We then aggregated ds across a set of studies, producing a meta-analytic estimate of population effect sizes (d_+) that gives more weight to studies with smaller variances (and larger sample sizes). A 95% confidence interval was used to determine whether d_+ differed reliably from 0. Tests of homogeneity (Q) were used to assess whether the studies within a particular set of comparisons produced relatively homogeneous or heterogeneous effect sizes (Hedges & Olkin, 1985). A significant Q statistic (when compared with a χ^2 distribution with $df = k - 1$, where k = the number of effect sizes) indicates substantial heterogeneity among the effect sizes; it is standard practice to search for moderators that may account for substantial hetereogeneity. Finally, a between-groups test (Q_B) was used to ascertain whether type of method, type of skill measure,

TABLE 1. Characteristics of Studies in the Meta-Analyses of the Effectiveness of Specific Training Methods

Study	Training and control conditions	Target of training	Participants	Dependent variables (all involved judgments from trained judges)
Baum & Gray (1992)	Instruction, modeling	Minimal encouragers, open-ended questions, reflections of feelings	24 undergraduates (12 per condition)	% facilitative responses (reflection, minimal encourager, open-ended question) in 5-min interviews
Berg & Stone (1980)	Feedback, no-training control	Reflections of feelings	40 undergraduates (20 per condition)	# and quality of reflections, empathy in response to audiotaped client statements
Canada (1973)	Feedback, no-training control	Open-ended leads	20 employment interviewers (10 per condition)	# open-ended leads, % client talk time, average length of client responses in 20-min interview
Chasnoff (1976)	Modeling, no-training control	Unspecified interview behaviors	52 masters students (26 per condition)	16 counselor interview behaviors (not defined) in 15-min interviews
Dalton et al. (1973)	Instruction, no-training control	Empathy	64 advanced undergraduates (32 per condition)	Empathy in response to written client statements
Eisenberg & Delaney (1970)	Modeling, feedback, no-training control	Tacting response leads	23 masters students (8 in each training condition; 7 in control)	Tacting responses in a 30-min interview
Kuna (1975)	Instruction (lecture, lecture + reading), instruction (lecture + reading) + modeling, no-training control	Restatement	92 advanced undergrads (50 in instruction conditions, 32 in instruction + modeling, 10 in control condition)	% restatements in response to audiotaped client statements
Perry (1975)	Instruction, modeling, modeling + instruction, no-training control	Empathy	44 clergymen (11 per condition)	Empathy in response to written client statements in 15-min interview
Quartaro & Rennie (1983)	Instruction, no-training control	Nondirective skills	31 advanced undergrads (17 in instruction, 14 in control)	% open invitations, closed inquiry, interpretations/advice in 10-min interview
Reddy (1968)	Feedback, no-training control	Empathy	24 introductory psychology students (12 per condition)	Empathy in response to videotaped client statements
Robinson et al (1979)	Modeling, no-training control	Tacting response leads	65 masters students (52 in modeling conditions, 13 in control)	Quantity, content fidelity, type appropriateness, and delivery of tacting responses to written client statements
Ronnestad (1977)	Modeling, feedback, no-training control	Empathic communication	30 masters students (10 per condition)	Empathy in response to videotaped client statements
Stone & Stein (1978)	Instruction, modeling, no-training control	Reflections of feelings	30 introductory psychology students (10 per condition)	Empathy in response to written and videotaped client statements; % reflections in 15-min interview and to written client statements
Stone & Vance (1976)	Instruction, modeling, feedback, instruction + modeling, no-training control	Empathic responses	30 introductory psychology students (6 per condition)	Empathy in response to written client statements and in 6-min interview
Uhlemann et al. (1976)	Instruction, modeling, instruction + modeling, no-training control	Facilitative responding	40 introductory psychology students (10 per condition)	Empathy response to written client statements; % reflections/empathy in 1st 10 of 15-min interview

and type of trainee produced different effect sizes (Hedges & Olkin, 1985).

Investigations of Aggregated Methods in Relation to No-Training Controls

The aggregated effect size (d_+) over 14 studies involving 526 participants was .89 (confidence interval of .71 to 1.07), suggesting that, in the aggregate, the training methods produced a large effect relative to no training. In practical terms, the average student receiving at least 1 of the methods performed better on the outcome measures than about 80% of those in the no-training control conditions (cf. Wampold, 2001). The Q statistic, however, was significant (55.49, $df = 13$, $p < .001$), indicating substantial variability among the effect sizes from the individual studies. One obvious source of variability was the study by Kuna (1975) that had an effect size much larger than other studies ($d = 3.90$). Examination of this study suggested that its control group performed much more poorly on the outcome measure than was typical of control conditions in other studies. After removing this study, Q was reduced to 17.54 ($df = 12$, $p > .05$), and d_+ was still relatively large (.79, confidence interval of .60 to .97), so we did not look for additional moderators. Given its potential to disproportionately influence effect sizes, especially in a relatively small set of studies, the outlier study was omitted in our subsequent analyses.

Type of method versus no training. Because aggregating across methods could have obscured differences among the individual methods relative to no treatment, we examined effect sizes separately for training methods. Relative to no-training, the different methods produced effect sizes varying in magnitude from medium (instruction = .63, $k = 6$) to large (modeling = .90, $k = 8$; feedback = .89, $k = 6$). In the between-groups test, effect size differences among the 3 methods were not statistically significant, Q_B (2) = 2.22, $p > .05$.

Type of skill measure. We were also interested in examining the extent to which effect sizes might vary as a function of how skill acquisition was assessed, especially because aggregation of measures within studies in the first meta-analysis could have masked variability of effects. Across the set of studies, 26 different measures were used. The 2 most frequent types were judges' ratings of trainee empathy based on behavior in an interview with a coached or practice client ($k = 5$) and judges' ratings of trainee empathy based on behavior in an analogue situation, typically written/taped responses to written/taped stimuli; $k = 9$). When aggregated across training methods, effect sizes for interview-based (.75) and analogue-based measures (.62) did not differ significantly, Q_B (1) = 2.94, $p > .05$. Similar findings were observed when we restricted the comparison to those studies that included both interview-based (.75) and analogue-based (.57) empathy measures in the same study (Q_B [1]) = 1.66, $p > .05$), which provides a better control for extraneous variables. Hence, we can conclude that the type of skill measure used did not substantially affect the findings.

Type of trainee. Finally, we explored the possibility that effect sizes might vary as a function of trainees' educational level. Specifically, we examined studies that compared training methods versus no training for graduate students ($k = 4$) or undergraduates ($k = 6$) as trainees. Both graduate (.88) and undergraduate (.77) students showed relatively large skill improvements, with no significant differences between them, Q_B (1) = .29, $p > .05$. Hence, unlike Baker and Daniels' (1989) finding in their meta-analysis of training programs that undergraduates outperformed graduate students (1.18 vs. .66), trainees' level of education did not substantially affect outcome in our meta-analysis of more specific training methods.

Studies Directly Comparing Methods of Training

In this section, we compare studies that isolated and contrasted the training methods directly with one another, offering evidence of their relative effectiveness in fostering skill gains. Seven studies, containing 140 students, were located in which modeling was compared with either instruction or feedback in the same study. This set of studies produced a relatively small aggregate effect size ($d_+ = .43$; confidence interval = .13 to .74) favoring modeling over the other 2 methods. The within-category test of homogeneity, however, revealed significant variability among the studies, Q (6) = 19.90, $p < .01$. Closer examination of this set of studies revealed that only 1 study yielded a negative effect ($d = -.68$; Uhlemann Lea, & Stone, 1976), and its authors

noted possible problems in administering the modeling condition. Recalculation of the aggregate effect size, excluding this outlier, produced a medium-sized effect favoring modeling (d_+ = .67; confidence interval = .33 to 1.00). The recalculated Q statistic (9.36, $df = 5$, $p > .05$) was not significant, reflecting homogeneity of effects within the set of studies. Hence, modeling appeared to produce moderately stronger effects than either feedback or instruction.

Studies of the Additive Effects of Training Components

We next examined studies that compared a combination of training methods (e.g., modeling plus instruction) with single method conditions (e.g., either modeling or instruction alone). Though small in number ($k = 5$; $n = 201$), these comparisons are worth examining because they reveal the additive effects of the training methods and, perhaps, better reflect the realities of actual training wherein trainers rely on multiple methods to foster skill development. The comparison yielded a relatively large effect size of .76 (confidence interval = .49 to 1.03) favoring the multimethod over the single-method conditions. However, the test of homogeneity was significant ($Q = 11.57$, $df = 4$, $p < .05$), suggesting substantial variability of effects among the set of studies. Removal of 1 apparent outlier (which produced a d of 1.58; Dalton, Sundblad, & Hylbert, 1973) reduced the aggregate effect size to .51 (confidence interval = .20 to .82), and Q was no longer significant. The average participant in combined method training outperformed about two thirds of the trainees in the single method conditions, suggesting that the additive advantage of the training methods was not trivial.

Summary

The results of our meta-analysis confirmed the conclusions reached in the narrative reviews about the effectiveness of instruction, modeling, and feedback for teaching trainees to use exploration skills. These 3 training methods, in the aggregate, produced a large effect relative to no training, with modeling outperforming instruction and feedback, and multimethod training outperforming single-method training. We found no differences between interview-based and analogue-based judgments of empathy (the 2 most typical types of measures) used to assess the effects of the methods, or between undergraduate and graduate trainees. Although it is initially encouraging to find evidence for the effectiveness of these 3 training methods, we caution that the meta-analysis was based on only 14 studies and that there were serious methodological flaws in the set of studies (see forthcoming section). Higher-quality research is needed before conclusions can be drawn about the best methods for training students.

Variables that Moderate the Effects of Helping Skills Training

Although many trainee, trainer, and situational variables could influence the process and outcome of helping skills training, only a few trainee variables have received any attention. For example, as noted earlier, Baker and Daniels (1989) found a larger effect size for undergraduate than graduate trainees in the outcomes of training programs, but we found no differences between undergraduate and graduate trainees in our meta-analysis of methods.

Furthermore, a few studies were found that investigated characteristics of trainees who profit most from training. These studies investigated dominance (Chasnoff, 1976), sex (Ronnestad, 1977), conceptual level (Berg & Stone, 1980), positive attitudes toward the target skill (Hirsch & Stone, 1982), and pretraining expectations for nondirective versus directive therapy style (Quartaro & Rennie, 1983). Unfortunately, no firm conclusions can be drawn about which trainee characteristics influence training outcome because these studies involved different trainee and target variables, and none of their findings have been replicated.

Kasdorf and Gustafson (1978) reported that 2 (apparently unpublished) MC studies found no sex differences in outcomes. Given the conflicting evidence regarding demographic variables in the psychotherapy literature (Sue & Lam, 2002), we speculate that psychosocial processes such as gender role socialization values or interaction styles associated with particular cultures would have more of a bearing on the perceived relevance of, or experience with, helping skills training than would demographic variables such as sex or race. For example, because helping is often seen as a gender-typed skill in Western culture, women, on average, may initially feel more com-

fortable with the helping skills than do men. Likewise, certain skills may feel more familiar or natural to students from some cultures than others (e.g., because challenge can appear somewhat confrontational, it may be less valued and used in collectivistic cultures, where greater emphasis is placed on relational harmony).

We acknowledge that helping skills training, and the verbal therapies with which it is associated, were developed in the United States and thus do not represent the only (or most) appropriate avenue for assisting all help-seekers. For example, Atkinson, Kim, and Caldwell (1998) identified a variety of helping roles that, indeed, may be preferred by help-seekers in different cultural contexts. Although a full consideration of the cultural parameters of psychotherapy and training is beyond the scope of this paper, it seems important to consider the role of culture with regard to the perceived relevance, credibility, and comfort level of specific helping skills for particular trainees.

In sum, a few trainee variables have been investigated, but the findings have not been replicated, limiting confidence in the stability of the findings. We were particularly surprised and dismayed to find no investigations of trainer effects and structural variables (e.g., the context in which training is offered, the length and spacing of training) on the outcomes of training.

Critique of Studies on Helping Skills Training

In general, the quality of the studies on helping skills training programs and training methods is not up to the current standards for process and outcome research. Indeed, the methodological limitations of these studies make it difficult to draw firm conclusions about the effects of training. In this section, we summarize the major methodological problems, drawing from previous critiques of this literature (Gormally & Hill, 1974; Lambert, DeJulio, & Stein, 1978; Matarazzo, 1971, 1978; Russell et al., 1984). We cover problems involved in investigating training programs and in studying specific training methods.

Overall Training Programs

Very little is known about the content of training programs, given that, in general, specific manuals were not used for the training. Furthermore, none of the studies required adherence to manuals, so we do not know what trainers actually did during training. Since much of the training was done within the context of educational programs, it is likely that there was a lot of variation because of different educational settings and objectives.

A further problem is that most studies involved only 1 trainer. Results thus confounded the training and the trainer. Given the findings (albeit somewhat conflicting) within the psychotherapy literature about the potential influence of therapists, particularly when therapists are allowed to exercise judgment and do not have to adhere closely to preexisting treatment protocols (Crits-Christoph, 2006; Crits-Cristoph et al., 1991; Kim, Wampold, & Bolt, 2006; Wampold, 2001), we would expect that trainers could have a major influence on the process and outcome of training. For example, trainers can be empathic and encourage students to develop their own style, or they can rigidly discourage deviations from the training program being taught.

Another serious concern is the lack of random assignment to conditions in some studies. If 1 class of students who voluntarily sign up for helping skills training is compared to another class of students who voluntarily sign up for something else (e.g., theories of therapy), differences in motivation to learn the skills could influence the results. Of course, we acknowledge that most training takes place within academic settings, where it is very difficult to randomly assign students to conditions.

Other concerns relate to the structure of the training. Many of the training programs were far briefer (e.g., 10 hours) than those currently employed in actual training (1 to 2 semesters). Furthermore, the size of the training group may influence the process and outcome of training. It may be easier to learn skills in smaller groups where trainees can receive more individual attention. In addition, the spacing of training has not been studied. Training spread over several weeks may have a different impact than training condensed into a shorter time span.

Training outcomes. Outcome has been assessed through coding the appropriate use of interview behaviors (e.g., pauses, length of responses; Matarazzo, Phillips, Wiens, & Saslow, 1965; Matarazzo & Wiens, 1967; Saslow & Matarazzo, 1959); judging how well trainees discrim-

inate the helping skills (e.g., ability to distinguish reflections from restatements or interpretations; Lee, Zingle, Patterson, Ivey, & Haase, 1976); rating how well trainees demonstrate helping skills under relatively simple, analogue (i.e., nonclinical) conditions (e.g., ability to write an empathic response to a videotaped client; Smit & van der Molen, 1995, 1996a, 1996b, 1997, 1996a, 1996b, 1997); and determining how well trainees perform the helping role with clients (either coached, volunteer, or actual clients). Many authors have critiqued how outcome has been assessed (e.g., Chinsky & Rappaport, 1970; Gormally & Hill, 1974; Rappaport & Chinsky, 1972). Most assert that assessing change through written or oral responses to analogue clients is not an adequate test of training. A common suggestion is that skills be tested in situations as closely representative of actual clinical interviews as possible, with volunteer clients presenting real, unscripted problems. Further, it has been argued that researchers should test whether trainees use the skills at adequate levels of competence in their interactions with clients, and whether clients improve as a result of their interaction with the trainee. Just using more of a given skill is not equivalent to using the skills appropriately and empathically in a therapy session. In addition, assessment of skill-related behaviors (e.g., empathy) often has involved the use of trained judges, but inadequate attention was paid to the judgment process, with many studies employing vague definitions of behaviors to be judged and using only 1 judge for the majority of the judgments (raising questions about rater bias; see Hill & Lambert, 2004).

Finally, a major problem is that many of the training programs "taught to the test." The focus in many studies was on teaching specific skills (e.g., reflection of feelings), with posttests being used to assess the acquisition of these specific skills. Hence, when experimental group participants were tested after training, they knew what was expected of them (e.g., that they were supposed to reflect feelings). In contrast, participants in the control groups typically were unaware of what behavior was desired on the posttest. If they had simply been told to reflect feelings (and been given a definition of this term), for example, they may have been able to do so. This "teaching to the test" may have artificially accentuated training effects when compared with no-training and may have confused training with the simple "cueing" of participants to produce predefined "correct" responses (cf. Quartaro & Rennie, 1983; Resnikoff, 1972).

Specific Training Methods

In this section, we discuss several methodological issues related to studies investigating training methods. We do not repeat criticisms discussed earlier under the critique of the training programs, even though some of these also apply here (e.g., assessment of outcome).

Targets of training. All of the studies reviewed in our meta-analysis focused on empathy or exploration skills (e.g., reflections, open questions, tacting, and restatements) rather than other skills that may be more difficult to teach (e.g., interpretations). In addition, skills were taught in isolation (e.g., a study might have only focused on reflections of feelings). In contrast, most training programs teach a full range of skills and teach them sequentially. For example, it may be easier to learn reflection after having mastered open questions and restatements. Similarly, it may be easier to learn challenges and interpretations after having mastered exploration skills.

Furthermore, definitions of some of the skills (e.g., tacting) were vague and not operationalized behaviorally. As another example, authors seemed to assume that readers knew what was meant by empathy, but reviews of the literature suggest that empathy can mean many different things (see Duan & Hill, 1996). These definitional problems often make it difficult to ascertain whether researchers were truly targeting the same skills despite using the same labels.

Finally, a number of other important targets of training have not been emphasized or investigated systematically in conjunction with helping skills. Specifically, we believe that, in practice, helping skills training is often accompanied by efforts to teach students empathy, case conceptualization and case management skills, a theoretical framework, a facilitative attitude, an awareness of trainee intentions and client reactions, self-awareness, and professional ethics. Without attention to these other foci of training, skills training could be perceived as mechanistic and decontextualized, and trainees might not be encouraged to become optimally empathic and responsible in their roles. These topics have received varying degrees of attention in the

literature, but rarely within the context of helping skills training.

Definition and operationalization of methods. Authors were often vague and imprecise about how they defined and implemented the individual training methods. The most obvious example of this problem is the overlap between instructions and modeling. Instruction usually seemed to involve at least some modeling in that examples were provided to illustrate the skill. Similarly, modeling typically involved at least some instruction when researchers described what was being modeled. Another problem is that practice has rarely been studied as a distinct training method, even though it seems essential to the success of training. For example, Bandura (1997) suggested that personal successes (and failures), brought about through practice, constitute a particularly compelling source of self-efficacy and skill development.

The methods were implemented in a variety of ways across studies. For example, instruction and modeling were sometimes implemented through written materials, audiotapes, or videotapes, making it difficult to compare results across studies. Furthermore, live presentations by expert instructors were not investigated, even though students typically learn about the skills from instructors in actual training applications. Finally, as noted earlier, feedback was operationalized in a variety of ways, including immediate, focused reinforcement and less structured supervision interactions.

An additional problem is that the most of the studies on methods involved extraordinarily brief interventions, typically ranging between 5 minutes to 1 hour. These researchers seemed to believe (and often stated so explicitly in the introductions to the studies) that training should ideally be done as briefly and efficiently as possible through the use of technology (e.g., videotape modeling). Our experience is that for most trainees to truly learn how to use skills effectively with clients (rather than reproduce a simple, modeled response in an analogue situation), they need considerable training and practice. Therefore, the studies may, in general, poorly reflect the conditions and requirements of actual training.

Sample characteristics. We were struck with the small sample size in many of the studies (often as small as 6–12 participants per condition, see Table 1), with attendant restriction in statistical power. Furthermore, there was a wide range of participants (introductory psychology students who participated for extra credit, upper-level undergraduates in counseling-related courses, master's-level students in beginning practicum courses, and mental health professionals) in studies. The use of different samples is not a problem if the intent is to examine the impact of training for each of these groups, but it appeared that, at least in certain cases, undergraduates were employed as proxies for mental health trainees. Given that introductory psychology students may have participated for extra credit rather than because they were motivated to learn the skills, their adequacy as proxies is uncertain. Finally, as with studies of the effects of overall training programs, those studies that involved trainers typically used only 1 trainer (if indeed the trainer or supervisor was even described). Hence, type of training was often confounded with trainer, making it difficult to ascertain the degree to which findings can be generalized across trainers or training contexts.

Recommendations for Future Research

Our review of this literature leads us to conclude tentatively that at least 2 helping skills training programs (HRT and MC) and several specific methods (modeling, instruction, and feedback) are useful in promoting helping skill acquisition. However, these broad conclusions need to be qualified in light of the methodological problems cited above. It is perhaps not surprising that these studies do not meet the methodological standards required today for research given that most of them were conducted over 30 years ago. Hence, we need more and better research assessing the effectiveness of the training programs and of the specific methods of training to better support current training practices. We also need to answer questions about the who, what, and how of training to improve training. In this section, we offer recommendation for future research on training programs, methods, and moderating variables.

Investigating the Effectiveness of Overall Training Programs

There is a need to test whether helping skills training as it is currently conducted is effective (a) when all trainees know what skills they should be using and (b) when training effects are assessed in relation to volunteer clients who

present real, nonscripted problems. In addition, studies comparing the relative effectiveness of the various training programs are needed, as is research on the ideal length and spacing of training, and the consolidation and maintenance of skills over time. To pursue these important questions, researchers might profit from using multisite designs which would allow for a number of instructors each teaching several classes to control for or investigate such variables as trainer effects and classroom climate. Furthermore, it will be important to develop training manuals and to assess adherence to them, which would allow a clearer linkage to be drawn between training content and outcome.

Skills training conditions could also be compared to alternative training conditions (such as students taking a didactic course on theories of counseling), so that trainees have equivalent expectations that they are learning to become therapists but without exposure to the experiential practices that are the hallmark of helping skills training. Issues of random assignment are of course key, and researchers will need to think of innovative ways to deal with this problem.

Effects of training might best be measured with multidimensional assessment programs, tapping changes in trainee knowledge, proficiency of trainee skill use in actual counseling sessions, and client reactions. Where possible, it may be best to use at least 2 sessions with different clients at each assessment point because of variability associated with different clients. For example, if 1 client at the beginning of the semester is relatively cooperative, verbal, and psychologically sophisticated, whereas a second client at the end of the semester is much more difficult, this might lead to an artificially pessimistic appraisal of the trainee's progress. In addition, multiple perspectives (e.g., trainee, supervisor, external judge, client) should be used to assess the trainee's performance during the session, given empirical data indicating the lack of congruence among perspectives (see Hill & Lambert, 2004).

A number of measures could be used. Clients, trainees, and supervisors could use measures such as the Helping Skills Measure (Hill & Kellems, 2002) to assess how much the trainee used exploration, insight, and action skills. External judges could code the helping skills used in sessions, using measures such as those developed by Hill (2004) or Stiles (1979, 1992). Additional measures of training effects could include anxiety, trainees' self-efficacy (e.g., Lent, Hill, & Hoffman, 2003) and hindering self-awareness (Williams, Hurley, O'Brien, & DeGregorio, 2003). Furthermore, the ability to form a good relationship with the client could be assessed with measures of the therapeutic relationship or working alliance (e.g., Hill & Kellems, 2002; Hatcher, 2006); the ability to conduct sessions could be assessed with measures of the depth and quality of sessions (Hill & Kellems, 2002; Stiles & Snow, 1984); and client outcome could be assessed through change on measures such as the Outcome Questionnaire (Lambert et al., 1996).

Although we agree with other reviewers (e.g., Stein & Lambert, 1995) that client change or improvement is a critical consideration, a caveat is needed given that therapist interventions are only one of many influences on client change. The change process is a complex phenomenon that is linked to multiple client, therapist, relational, and extratherapy determinants (Beutler et al., 2004; Clarkin & Levy, 2004; Hill & Williams, 2000; Orlinsky, Ronnestad, & Willutzki, 2004). The link from helping skills to client outcomes may, therefore, be less strong and direct than one might expect, calling for study of mediating paths. For example, use of helping skills is probably linked to client satisfaction and other outcomes through intervening paths, such as the quality of the therapeutic relationship, rather than directly (Hill & Kellems, 2002). The linking of all of these variables merits empirical attention.

The timing of the evaluations is also an important matter. Training outcomes should be assessed not only at the beginning and end of training but also during training to assess the learning curve. We have noticed in training that sometimes trainees' self-efficacy declines as they realize that the skills they have used successfully in friendships are not the same as those used in helping situations (e.g., challenge, interpretation, disclosure, and immediacy are used much more often in helping situations whereas self-disclosure and advice are used only at particular times for particular reasons). The effects of training also need to be measured at follow-up points to determine how well trainees maintain the skills. In a study examining the persistence of skills six to nine months following 40 hours of HRT training, Gormally, Hill, Gulanick, & McGovern (1975) found that, although both graduate and undergraduate students increased in empathy over the course of training, graduate stu-

dents continued to increase in empathy whereas undergraduates decreased in empathy at follow-up. Gormally et al. speculated that graduate students were more empathic at follow-up because they had opportunities to practice their skills, whereas undergraduates did not. Skill maintenance thus may depend on subsequent skill use and feedback experiences (e.g., enacting the therapy role in practicum, serving as a trainer for other students who are learning helping skills).

Researchers could also assess not only whether trainees are able to use skills in interviews with clients at differing follow-up periods, but also whether trainees' basic skills generalize to new situations, such as handling client anger directed at them personally (e.g., Hess, Knox, & Hill, 2006) or feelings of sexual attraction to clients (e.g., Ladany et al., 1997). Trainees might be able to use skills competently with relatively easy client problems or situations, but might have a more difficult time using them under more challenging scenarios. Additional practice and more sophisticated training strategies might be needed to help trainees apply the skills, and manage their anxiety, effectively in more difficult situations.

In addition, there is an urgent need to test the effects of training programs as they are currently implemented. We need to be assured, using good methodology, that our training is effective. Although no formal survey has been conducted, our sense is that trainers currently tend to combine elements from the various programs, the training is longer (at least 1 to 2 semesters) and followed closely by supervised practicum training, and training focuses on teaching skills within the context of teaching students about empathy. In addition, rather than just teaching skills/facilitative attitudes, trainers often teach the larger art of becoming a therapist, which includes training in theoretical orientations, client dynamics, the need for self-awareness, managing sessions, ethics, and professional behavior (see Hill & Lent, 2006). These additional components need to be investigated for their role in fostering professional growth.

Finally, structural aspects of training need investigation. Matarazzo and Patterson (1986) mentioned the need for studying the optimal sequencing of training methods, but we are not aware of any studies on the sequencing of methods. It seems logical to teach exploration skills before insight and action skills because exploration theoretically provides a foundation for insight and action, (Hill, 2004), but this sequence needs to be tested. In addition, the sequence of teaching self-awareness and helping skills needs to be tested. Anecdotally, it seems that self-awareness of one's intentions and motivations is necessary for trainees to successfully implement the skills, but we do not know empirically if this is true. If so, it may be better to train in facilitative self-awareness before training in helping skills.

Investigating Training Methods

First, the methods (e.g., instruction, modeling, practice, feedback) need to be defined more clearly and distinguished from each other. Second, we need to know more about practice as a method of training since it has received minimal empirical attention. Third, other methods should be identified and studied (e.g., self-supervision through watching tapes or writing process notes, coaching, cocounseling, reducing anxiety). Fourth, we need to study the methods in relation to targets other than empathy and exploration skills. Different methods may be effective, for example, for teaching insight and action skills than for teaching exploration skills.

Finally, there is a need for new theoretical bases for studying training methods. Most of the investigations of the methods have been based at least loosely on Bandura's (1969) social learning theory and have been broadly concerned with the effects of instruction, modeling, practice, and/or feedback. Although important and heuristic, this theoretical base has been supplanted in more recent years by Bandura's (1986, 1997) more comprehensive social–cognitive theory, which offers a more detailed rendering of the intervening processes through which behavior change and skill acquisition occur. This general framework has given rise to notable theoretical efforts to account for certain aspects of counselor development (Larson, 1998; Lent, Hackett, & Brown, 1998), but there is need for more empirical tests of this position specifically within the context of teaching and learning helping skills.

Other theories could also be developed or adapted to study helping skills training. For example, the literature in cognitive psychology (e.g., Etringer & Hillerbrand, 1995; Lichtenberg, 1997; Reisberg, 1997) on the development of expertise could prove helpful for studying the effects of training (e.g., how persons progress

from attending to the surface structure of the problem to gaining the deep structure or, in the case of helping skills, moving from parroting the skills to using them in a flexible, client-centered manner). Binder (2004) presents a good example of the application of cognitive theory to therapist training.

Recommendations for Investigating Moderating Variables in Training

We speculate that trainee and trainer variables moderate the outcome of training. Hence, these variables need to receive more research attention than they have in the past.

Trainee variables. One variable that merits investigation is the natural helping ability of the trainee (Stahl, Hill, & Kivlighan, in preparation). Many students who enter training programs have had considerable helping experiences throughout their lives. Their friends and family members often turn to them for help because they have shown a natural ability to offer support or guidance. People with such natural helping tendencies may be easier to train given that training essentially involves shaping their already existing skills. On the other hand, training may not have as great an impact on their skill levels, compared to those with lesser natural ability, because they are starting out at a higher baseline (a similar issue was raised earlier in relation to graduate vs. undergraduate trainees; Baker et al., 1990; Goodyear & Guzzardo, 2000).

Parenthetically, this consideration of natural helping ability recalls the "input-output" controversy in higher education; namely, one would ideally want to know that superior universities truly do a better job of teaching their students, and that they do not merely capitalize on the preexisting capabilities of the students they select (cf. Magoon & Holland, 1984). If student input factors alone were sufficient to explain therapist effectiveness, the magic of training would essentially be reduced to admitting the "best" students and polishing the skills they already have via supervision, rather than troubling to teach them new skills. Then again, selecting the "best" students is more difficult than it sounds given the absence of established measures of therapy aptitude (Matarazzo & Patterson, 1986.)

Another trainee variable that might be worth investigating is reactance level, which refers to the ease with which a person is provoked to resist external demands (Beutler, Moleiro, & Talebi, 2002) or an individual's defense when he or she feels his or her freedom is threatened (Brehm & Sensenig, 1966). In the psychotherapy literature, reactant clients did better with nondirective treatment methods, whereas nonreactant clients did better with structured interventions (Beutler et al., 2002). Extrapolating to the training context, it is possible that reactant trainees would prefer nondirective training methods because of the greater self-direction such methods allow. In contrast, nonreactant trainees might profit from either structured or unstructured training because they may be less sensitive to the issue of external constraint.

Similarly, conceptual level may also affect trainee preference for, or skill gains from, structured versus less structured training methods. In this regard, Berg and Stone (1980) found that more structured (didactic) supervision produced higher ratings of trainee satisfaction, perceived helpfulness, and perceived learning among trainees with low conceptual level, whereas less structured (experiential) supervision yielded higher ratings among those with high conceptual level. Hence, trainees with high conceptual levels, as opposed to those with low conceptual levels, might profit more from less structured training methods.

Trainer variables. One variable that likely influences the outcome of training is the trainer's competence as both a trainer and a therapist. Carkhuff (1969) posited that a trainee could only rise to the competence level of the trainer. We do not agree with this unconditional stance because trainer competence level is not the only factor influencing trainee development. We do suggest, however, that it is probably more difficult for trainees to learn the skills when the trainer cannot competently model the skills or when the trainer gives them information that is discrepant with their text. Other potential variables, like trainers' instructional self-efficacy or their ability to promote counseling self-efficacy in their students, could be identified and explored in terms of their potential to moderate the effects of structured training programs.

Conclusions

Our challenge in this paper has been to consider what is still vital and relevant about helping skills training; what is known and still unknown about how to conceptualize, assess, and modify

helping skills; and what new research and theoretical directions may revive inquiry into, and advance understanding of, helping skills. Our review of the existing literature indicates that at least two helping skills training programs (HRT and MC) and several individual methods of training (e.g., instruction, modeling, feedback) promote skill acquisition. Although encouraging and relevant to the actual practice of skill training, we also noted limitations in methodology that have prevented this literature from making greater contributions to the understanding and practice of helping skills training. For example, there has been surprisingly little study of whether and how helping skills training influences the process and outcome of therapy.

We also highlighted methodological challenges in assessing the effects of training. Unlike some other career performance contexts (e.g., sales figures in business, win-loss records in sports), the outcomes of helper training and subsequent trainee performance may not be captured sufficiently with a single, consensually accepted measure. Although most trainers would agree that helping skills training should ultimately aid clients in some palpable way, we also need to document that the skills are, in fact, learned appropriately and maintained over time. The multiple targets and outcomes of training, and the variety of perspectives on these outcomes (e.g., trainer, client, external judges) suggest the value of using multidimensional assessments of trainee growth and performance.

We hope that our review, methodological critique, and suggested directions for research will stimulate new inquiry on helping skills training as it is currently conducted. We encourage researchers to rediscover this topic and rekindle research programs on helping skills training, given its presumed centrality to the growth and development of therapists. It would be especially exciting if theorists used the results of investigations to modify existing theories of helping skills, which then could provide a solid theoretical base for additional research.

References

References marked with an asterisk indicate studies included in the meta-analysis of specific methods of training.

Alberts, G., & Edelstein, B. (1990). Therapist training: A critical review of skill training studies. *Clinical Psychology Review, 10,* 497–511.

Atkinson, D. R., Kim, B. S. K., & Caldwell, R. (1998). Ratings of helper roles by multicultural psychologists and Asian American students: Initial support for the three-dimensional model of multicultural counseling. *Journal of Counseling Psychology, 45,* 414–423.

Baker, S. B., & Daniels, T. G. (1989). Integrating research on the microcounseling program: A meta-analysis. *Journal of Counseling Psychology, 36,* 213–222.

Baker, S. B., Daniels, T. G., & Greeley, A. T. (1990). Systematic training of graduate-level therapists: Narrative and meta-analytic reviews of three major programs. *The Counseling Psychologist, 18,* 355–421.

Bandura, A. (1969). *Principles of behavior modification.* New York: Holt, Rinehart, & Winston.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

*Baum, B. E., & Gray, J. J. (1992). Expert modeling, self-observation using videotape, and acquisition of basic therapy skills. *Professional Psychology: Research and Practice, 23,* 220–225.

Benjamin, A. (1981). *The helping interview* (3rd ed.). Boston: Houghton Mifflin.

*Berg, K. S., & Stone, G. L. (1980). Effects of conceptual level and supervision structure on therapist skill development. *Journal of Counseling Psychology, 27,* 500–509.

Beutler, L., Moleiro, C. M., & Talebi, H. (2002). Resistance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 129–144). New York: Oxford University Press.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., et al. (2004). Therapist variables. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.

Binder, J. L. (2004). Key competencies in brief dynamic psychotherapy: Clinical practice beyond the manual. New York: Guilford Press.

Brammer, L. M. (1973). *The helping relationship: Process and skills.* Oxford, England: Prentice Hall.

Brammer, L. M., & MacDonald, G. (2003). *The helping relationship: Process and skills* (8th ed.). Needham Heights, MA: Allyn & Bacon.

Brehm, J. W., & Sensenig, J. (1966). Social influence as a function of attempted and implied usurpation of choice. *Journal of Personality and Social Psychology, 4,* 703–707.

*Canada, R. M. (1973). Immediate reinforcement versus delayed reinforcement in teaching a basic interview technique. *Journal of Counseling Psychology, 20,* 395–398.

Carkhuff, R. R. (1969). *Human and helping relations* (Vols. 1 & 2). New York: Holt, Rinehart, & Winston.

Carkhuff, R. R. (1972). *The art of helping.* Amherst, MA: Human Resource Development Press.

*Chasnoff, S. S. (1976). The effects of modeling and ambiguity tolerance on interview behavior. *Counselor Education and Supervision, 16,* 46–51.

Chinsky, J. M., & Rappaport, J. (1970). Brief critique of

the meaning and reliability of "Accurate Empathy" ratings. *Psychological Bulletin, 73,* 379–382.

CLARKIN, J. F., & LEVY, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

CORMIER, W., & CORMIER, S. (1998). *Interviewing strategies for helpers* (4th ed.). Monterey, CA: Brooks/Cole.

CRITS-CRISTOPH, P. (2006). Therapist effects in the TDCRP and other psychotherapy studies. *Psychotherapy Research, 16,* 178–178.

CRITS-CRISTOPH, P., BARABACKIE, K., KURCIAS, J. S., CARROLL, K., LUBORSKY, L., MCLELLAN, T., ET AL. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91.

*DALTON, R. F., SUNDBLAD, L. M., & HYLBERT, K. W. (1973). An application of principles of social learning to training in communication of empathy. *Journal of Counseling Psychology, 20,* 378–383.

DANIELS, T. G. (2001, August). Microcounseling, microskills, and medical education: Empirical Investigations. Paper presented at the annual meeting of the American Psychological Association, San Francisco.

DANISH, S. J., & HAUER, A. (1973). *Helping skills.* New York: Behavioral Publications.

DUAN, C., & HILL, C. E. (1996). Theoretical confusions in the construct of empathy: A review of the literature. *Journal of Counseling Psychology, 43,* 261–274.

EGAN, G. (1975). *The skilled helper: Model, skills, and methods for effective helping.* Monterey, CA: Brooks/Cole.

EGAN, G. (1994). *The skilled helper: A problem-management approach to helping (5th ed.*). Belmont, CA: Brooks/Cole.

*EISENBERG, S., & DELANEY, D. (1970). Using video stimulation of counseling for training counselors. *Journal of Counseling Psychology, 17,* 15–19.

ELLIS, M. V., LADANY, N., KRENGEL, M., & SHULT, D. (1996). Clinical supervision research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology, 43,* 35–50.

ETRINGER, B. D., & HILLERBRAND, E. (1995). The transition from novice to expert counselor. *Counselor Education and Supervision, 35,* 4–18.

EVANS, D. R., HEARN, M. T., UHLEMANN, M. R., & IVEY, A. E. (1998). Essential interviewing: A programmed approach to effective communication (5th ed.). Belmont, CA: Brooks/Cole.

FORD, J. D. (1979). Research on training therapists and clinicians. *Review of Educational Research, 49,* 87–130.

FUQUA, D. R., & GADE, E. M. (1982). A critical reexamination of the practice method in counselor education. *Counselor Education and Supervision, 21,* 282–294.

GAZDA, G., ASBURY, F., BALZER, F., CHILDERS, W., & WALTERS, R. (1977). *Human relations development.* Boston: Allyn & Bacon.

GOODMAN, G., & DOOLEY, D. A. (1976). A framework for help-intended communication. *Psychotherapy: Theory, Research, and Practice, 13,* 106–117.

GOODYEAR, R. K., & GUZZARDO, C. R. (2000). Psychotherapy supervision and training. In S. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed., pp. 83–108). New York: Wiley.

GORMALLY, J., & HILL, C. E. (1974). Guidelines for research in Carkhuff's training model. *Journal of Counseling Psychology, 21,* 539–547.

GORMALLY, J., HILL, C. E., GULANICK, N., & MCGOVERN, T. (1975). The persistence of communication skills for undergraduate and graduate trainees. *Journal of Clinical Psychology, 31,* 369–372.

HATCHER, R. (2006). Development and validation of a revised short version of the Working Alliance Inventory (WAI-SR). *Psychotherapy Research, 16,* 12–25.

HAZLER, R. J., & HIPPLE, T. E. (1981). The effects of mental practice on counseling behaviors. *Counselor Education and Supervision, 20,* 211–218.

HEDGES, L. V., & OLKIN, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

HESS, S., KNOX, S., & HILL, C. E. (2006). Teaching graduate student trainees how to manage client anger: A comparison of three types of training. *Psychotherapy Research, 16,* 282–292.

HILL, C. E. (2004). *Helping skills: Facilitating exploration, insight, and action (*2nd ed.). Washington DC: American Psychological Association.

HILL, C. E., & KELLEMS, I. S. (2002). Development and use of the Helping Skills Measure to assess client perceptions of the effects of training and of helping skills in sessions. *Journal of Counseling Psychology, 49,* 264–272.

HILL, C. E., & LAMBERT, M. J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed., pp. 84–136). New York: Wiley.

HILL, C. E., & LENT, R. W. (2006). Training for novice therapists: Skills plus. *Psychotherapy Bulletin, 41,* 11–16.

HILL, C. E., & O'BRIEN, K. (1999). *Helping skills: Facilitating exploration, insight, and action.* Washington DC: American Psychological Association.

HILL, C. E., & WILLIAMS, E. N. (2000). The process of individual therapy. In R. W. Lent & S. D. Brown (Eds.), *Handbook of counseling psychology* (pp. 670–710). New York: Wiley.

HIRSCH, P. A., & STONE, G. L. (1982). Attitudes and behavior in counseling skill development. *Journal of Counseling Psychology, 29,* 516–522.

HOLLOWAY, E. L. (1992). Supervision: A way of teaching and learning. In S. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (2nd ed., pp. 177–214). New York: Wiley.

IVEY, A. E. (1971). *Microcounseling: Innovations in interviewing training.* Springfield, IL: Charles C Thomas.

KAGAN, N. (1984). Interpersonal process recall: Basic methods and recent research. In D. Larson (Ed.), *Teaching psychological skills: Models for giving psychology away* (pp. 229–244). Monterey, CA: Brooks/Cole.

KASDORF, J., & GUSTAFSON, K. (1978). Research related to microtraining. In A. E. Ivey & J. Authier (Eds.), *Microcounseling: Innovations in interviewing, counseling, psychotherapy, and psychoeducation* (pp. 323–376). Springfield, IL: Thomas.

KIM, D., WAMPOLD, B. E., & BOLT, D. M. (2006). Ther-

apist effects in therapy: A random effects modeling of the NIMH TDCRP data. *Psychotherapy Research, 16,* 161–172.

*Kuna, D. J. (1975). Lecturing, reading, and modeling in therapist restatement training. *Journal of Counseling Psychology, 22,* 542–546.

Kurtz, P. D., Marshall, E. K., & Banspach, S. W. (1985). Interpersonal skill-training research: A 12-year review and analysis. *Counselor Education and Supervision, 24,* 249–263.

Ladany, N., O'Brien, K., Hill, C. E., Melinkoff, D., Knox, S., & Peterson, D. (1997). Sexual attraction toward clients: A qualitative study of psychotherapy predoctoral interns. *Journal of Counseling Psychology, 44,* 413–424.

Lambert, M. J., DeJulio, S. S., & Stein, D. M. (1978). Therapist interpersonal skills: Process, outcome, methodological considerations, and recommendations for future research. *Psychological Bulletin, 85,* 467–489.

Lambert, M. J., Hansen, N. B., Umphress, N. J., Lunnen, K., Okiishi, J., Burlingame, G. M., et al. (1996). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.2*). Wilmington, DE: American Professional Credentialing Services.

Larson, L. M. (1998). The social cognitive model of counselor training. *The Counseling Psychologist, 26,* 219–273.

Lee, D. Y., Zingle, H. W., Patterson, J. G., Ivey, A. E., & Haase, R. F. (1976). Development and validation of a Microcounseling Skill Discrimination Scale. *Journal of Counseling Psychology, 23,* 468–472.

Lent, R. W., Hackett, G., & Brown, S. D. (1998). Extending social cognitive theory to counselor training: Problems and prospects. *The Counseling Psychologist, 26,* 295–306.

Lent, R. W., Hill, C. E., & Hoffman, M. A. (2003). Development and validation of the Therapist Activity Self-Efficacy Scales. *Journal of Counseling Psychology, 50,* 97–108.

Lichtenberg, J. W. (1997). Expertise in counseling psychology: A concept in search of support. *Educational Psychology Review, 9,* 221–238.

Magoon, T. M., & Holland, J. L. (1984). Research training and supervision. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (pp. 682–715). New York: Wiley.

Matarazzo, J. D., & Wiens, A. N. (1967). Interviewer influence on durations of interviewee silence. *Journal of Experimental Research in Personality, 2,* 56–59.

Matarazzo, R. G. (1971). The systematic study of learning psychotherapeutic skills. In A. E. Bergin & S. L. Garfield, *Handbook of psychotherapy and behavior change: An empirical analysis* (pp. 895–924). New York: Wiley.

Matarazzo, R. G. (1978). Research on the teaching and learning of psychotherapeutic skills. In S. L. Garfield & A. E. Bergin, *Handbook of psychotherapy and behavior change: An empirical analysis* (2nd ed., pp. 941–966). New York: Wiley.

Matarazzo, R. G., & Patterson, D. R. (1986). Methods of teaching therapeutic skill. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed.). New York: Wiley.

Matarazzo, R. G., Phillips, J. S., Wiens, A. N., & Saslow, G. (1965). Learning the art of interviewing: A study of what beginning students do and their pattern of change. *Psychotherapy: Theory, Research, and Practice, 2,* 49–60.

Matarazzo, R. G., Wiens, A. N., & Saslow, G. (1966). Experimentation in the teaching and learning of psychotherapy skills. In L. A. Gottschalk & A. H. Auerbach (Eds.), *Methods of research in psychotherapy* (pp. 597–635). New York: Appleton-Century-Crofts.

Mayer, R. E. (2004). Should there be a three-strikes rule against discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14–19.

Meier, S. T., & Davis, S. R. (2001). The elements of counseling (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Okun, B. F. (2002). *Effective helping: Interviewing and counseling techniques* (6th ed.). Belmont, CA: Brooks/Cole Publishing Co.

Orlinsky, D., E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.

*Perry, M. (1975). Modeling and instructions in training for therapist empathy. *Journal of Counseling Psychology, 22,* 173–179.

*Quartaro, G. K., & Rennie, D. L. (1983). Effects of trainee expectancies and specific instructions on counseling skill acquisition. *Canadian Journal of Behavioral Science, 15,* 174–185.

Rappaport, J., & Chinsky, J. M. (1972). Accurate empathy: Confusion of a construct. *Psychological Bulletin, 77,* 400–404.

*Reddy, W. B. (1969). The effects of immediate and delayed feedback on the learning of empathy. *Journal of Counseling Psychology, 16,* 59–62.

Reisberg, D. (1997). *Cognition: Exploring the science of the mind.* New York: Norton.

Resnikoff, A. (1972). Critique of the human resource development model from the viewpoint of rigor. *The Counseling Psychologist, 3,* 46–55.

*Robinson, S. E., Froehle, T. C., & Kurpius, D. J. (1979). Effects of sex of model and media of model presentation on skill development of therapist trainees. *Journal of Counseling Psychology, 26,* 74–80.

Rogers, C. R. (1942). *Counseling and psychotherapy.* Boston: Houghton Mifflin.

Rogers, C. R. (1951). *Client-centered therapy: Its current practice, implications, and theory.* Boston: Houghton Mifflin.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21,* 95–103.

*Ronnestad, M. H. (1977). Effect of modeling, feedback, and experimental methods on counselor empathy. *Counselor Education and Supervision, 16,* 194–201.

Russell, R. R., Crimmings, A. M., & Lent, R. W. (1984). Therapist training and supervision: Theory and research. In S. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (pp. 625–281). New York: Wiley.

Saslow, G., & Matarazzo, J. D. (1959). A technique for studying changes in interview behavior. In E. A. Rubenstein & M. B. Parloff (Eds.). *Research and psy-*

chotherapy. Washington DC: American Psychological Association.

SCHWEBEL, M. (1953). Role-playing in therapist training. *Personnel and Guidance Journal, 32,* 196–201.

SMIT, G. N., & VAN DER MOLEN, H. T. (1995). Development and evaluation of a video test for the assessment of interviewing skills. *Journal of Cancer Education, 10,* 195–199.

SMIT, G. N., & VAN DER MOLEN, H. T. (1996a). Three methods for the assessment of communication skills. *British Journal of Educational Psychology, 66,* 543–555.

SMIT, G. N., & VAN DER MOLEN, H. T. (1996b). Simulations for the assessment of counselling skills. *Assessment and Evaluation in Higher Education, 21,* 334–346.

SMIT, G. N., & VAN DER MOLEN, H. T. (1997). The construction and evaluation of simulations to assess professional interviewing skills. *Assessment in Education, 4,* 353–364.

SPIVACK, J. D. (1973). Critical incidents in counseling: Simulated video experiences for training counselors. *Counselor Education and Supervision, 12,* 263–270.

STAHL, J., HILL, C. E., & KIVLIGHAN, D. M. (in preparation). Identifying natural helpers.

STEIN, D. M., & LAMBERT, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63,* 182–196.

STILES, W. B. (1979). Verbal response modes and psychotherapeutic technique. *Psychiatry, 42,* 49–62.

STILES, W. B. (1992). *Describing talk: A taxonomy of verbal response modes.* Newbury Park, CA: Sage.

STILES, W. B., & SNOW, J. S. (1984). Counseling session impact as viewed by novice counselors and their clients. *Journal of Counseling Psychology, 31,* 3–12.

*STONE, G. L., & STEIN, M. L. (1978). Effects of modeling and instructions as a function of time, task, and order. *Journal of Counseling Psychology, 25,* 150–156.

*STONE, G. L., & VANCE, A. (1976). Instructions, modeling, and rehearsal: Implications for training. *Journal of Counseling Psychology, 23,* 272–279.

SUE, S., & LAM, A. G. (2002). Cultural and demographic diversity. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 401–422). New York: Oxford University Press.

TRUAX, C. B., & CARKHUFF, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice.* Chicago: Aldine.

*UHLEMANN, M. R., LEA, G. W., & STONE, G. L. (1976). Effect of instructions and modeling on trainees low in interpersonal-communication skills. *Journal of Counseling Psychology, 23,* 509–513.

WAMPOLD, B. E. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes." *Psychological Bulletin, 122,* 203–215.

WAMPOLD, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Mahwah, NJ: Erlbaum.

WILLIAMS, E. N., HURLEY, K., O'BRIEN, K., & DEGREGORIO, A. (2003). Development and validation of the Self-Awareness and Management Strategies (SAMS) Scales for therapists. *Psychotherapy, 40,* 278–288.