

The Quality of Psychotherapy Case Formulations: A Comparison of Expert, Experienced, and Novice Cognitive–Behavioral and Psychodynamic Therapists

Tracy D. Eells, Kenneth G. Lombart, Edward M. Kendjelic, L. Carolyn Turner, and Cynthia P. Lucas
University of Louisville

Sixty-five expert, experienced, and novice cognitive–behavioral and psychodynamic psychotherapists provided “think aloud” case formulations in response to 6 standardized patient vignettes varying in disorder and prototypicality. The 390 formulations were reliably transcribed, segmented into idea units, content coded, and rated on multiple dimensions of quality. As hypothesized, the formulations of experts were more comprehensive, elaborated, complex, and systematic. Judges did not rate them as more coherent or precise in the use of language. In addition, the treatment plans of experts were more elaborated and linked better to the formulations. Effect sizes for overall ratings of quality ranged from medium to large. Few differences based on therapy orientation were observed. Results are discussed in terms of therapist training.

Keywords: case formulation, therapist expertise, cognitive–behavioral therapy, psychodynamic therapy

Therapist expertise is a relatively neglected, but potentially powerful, variable in psychotherapy process and outcome research (Beutler, Machado, & Neufeldt, 1994; Wampold, 2001). Crits-Christoph and colleagues’ (Crits-Christoph et al., 1991) meta-analysis estimated therapist effects to account for about 9% of variance in outcome among 15 comparative psychotherapy studies analyzed, which equates to an effect size of about .60. In comparison, Wampold and colleagues (Wampold et al., 1997) estimated an upper bound effect size for treatments of about .20. In their review of four psychotherapy outcome studies, Luborsky et al. (1986) found considerable differences among therapists in their average success rates, as well as variability in outcome within therapists. Similarly, Blatt, Sanislow, Zuroff, and Pilkonis (1996) found differences in effectiveness among therapists in the National Institute of Mental Health Treatment of Depression Collaborative Research Program, a study in which considerable effort was devoted to reducing variability in outcome caused by therapist differences. Najavits and Weiss (1994) analyzed therapist effects in seven outcome studies investigating psychosocial treatments of

substance abuse. They concluded that therapists showed diverse rates of effectiveness and that such differences appear to be independent of therapists’ professional background and of patient factors at the start of therapy. Other studies also support the hypothesis that therapist expertise is an important therapy outcome variable (Luborsky, McLellan, Woody, O’Brien, & Auerbach, 1985; Ricks, 1974). Similar findings have been observed for both cognitive–behavioral (Kingdon, Tyrer, Seivewright, Ferguson, & Murphy, 1996), and psychodynamic (Barber, Crits-Christoph, & Luborsky, 1996) therapy. However, some studies show that training, which presumably correlates with expertise, may not predict treatment outcome (e.g., Jacobson, 1995). Some psychotherapy process research suggests that expertise may rival or even exceed theoretical orientation in significance. In two recent studies explicitly focused on examining differences between therapists, the researchers found that peer-nominated expert cognitive–behavioral and psychodynamic therapists were more similar than different in exploring emotionally significant events in therapy (Goldfried, Raue, & Castonguay, 1998; Wiser & Goldfried, 1998).

One reason that therapist expertise has not been studied more is that psychotherapy researchers have focused more on seeking treatment differences and, from the statistical standpoint, have viewed the therapist as a within-treatment error variable. Such studies have attempted to minimize therapist variability by training therapists to adhere to a treatment manual, rather than exploring therapist behavior as a variable of interest. Even among these studies, however, some therapists consistently perform better than others (Blatt et al., 1996; Shaw et al., 1999). Psychotherapy researchers have been cautioned about the significant risk of Type I error rates when nonrandom therapist differences are not controlled for (Crits-Christoph & Mintz, 1991; Martindale, 1978; Wampold & Serlin, 2000). When they are not, researchers risk misattributing to treatment those effects that are actually due to therapists.

Tracy D. Eells, Department of Psychiatry and Behavioral Sciences, University of Louisville; Kenneth G. Lombart, Edward M. Kendjelic, L. Carolyn Turner, and Cynthia P. Lucas, Department of Psychological and Brain Sciences, University of Louisville.

This study was supported in part by National Institute of Mental Health Grant R03MH53996-01, a University of Louisville Research Incentive Grant, and a University of Louisville Medical School Grant. We thank the therapists who served as subjects in this study and the following individuals who provided research assistance: Andrew Brothers, Camille Frey, Katherine Hoover, Greg Lynch, Maryrose Manshadi, Lisa Milliner, Richard Morris, Diane Taylor, Troy Raffield, and Julia Ward.

Correspondence concerning this article should be addressed to Tracy D. Eells, Department of Psychiatry and Behavioral Sciences, University of Louisville, 550 South Jackson Street, Louisville, KY 40202. E-mail: eells@louisville.edu

The goal of this study was to test the hypothesis that therapist expertise is an important variable in psychotherapy, with a specific focus on case formulation. A *case formulation* can be defined as a set of “hypotheses about the causes, precipitants, and maintaining influences of a person’s psychological, interpersonal, and behavioral problems” (Eells, 2002, p. 815). It is a succinct, flexible, and appropriately comprehensive account of a patient’s psychological problems that includes a treatment plan closely linked to that account. Although traditionally more emphasized in psychodynamic therapies (Luborsky et al., 1993), case formulation has been increasingly emphasized among cognitive therapists (e.g., Persons, 1989), behavioral therapists (e.g., Hersen & Porzelius, 2002), and those practicing integrationist approaches to psychotherapy (e.g., Caspar, 1995). *Psychotherapy technique*, as opposed to case formulation, refers to the therapist’s interventions and tactics during treatment that are directed at helping the patient.

Case formulation is a core psychotherapy skill (Eells, 1997). According to Shaw and Dobson (1988, p. 667), case formulation skills include

complex events such as the extent to which the therapist elicits relevant data and integrates the information into a conceptualization on the patient’s main problem . . . the extent to which the therapist enacts a coherent plan for change that follows logically from the conceptualization of the problem . . . [and] the extent to which the interventions are in the sequence or context specified by the theory.

Skill in case formulation appears related to the “common factors” of psychotherapy, which Wampold (2001) estimates collectively explain at least 70% of total variance in therapy outcome. Common factors are shared aspects of virtually all therapies and are considered by some to be the active therapeutic ingredients rather than specific components of various forms of treatment. According to Frank and Frank (1991), one ingredient common to all effective therapies is that the therapist provides a “rationale, conceptual scheme, or myth that provides a plausible explanation for the patient’s symptoms and prescribes a ritual or procedure for resolving them” (p. 42). The functions of this conceptual scheme are to instill hope in the patient, provide new learning experiences, arouse emotions, enhance self-efficacy, and provide opportunities for practice. A related shared characteristic between the common factors model of psychotherapy and case formulation skill is that both focus on enhancing the therapeutic alliance. Eells (1997) asserts that one function of a good case formulation is to increase therapist’s empathy for a client through the understanding offered by the formulation. A formulation may also increase therapist confidence, thus contributing further as a common factor.

Among the few investigators of expertise in case formulation, Cummings, Hallberg, Martin, Slemon, and Hiebert (1990) examined the conceptualizations of two novice and two experienced counselors and found greater consistency of concepts over time and between clients for the experienced counselors, which they interpret as indicating well-developed underlying schemas for explaining patient change and for identifying individual client problems. Hillerbrand and Claiborn (1990) studied a larger sample of novice and expert therapists who generated diagnoses for hypothetical cases and provided a rationale for their choices. As compared with novices, experts felt more knowledgeable and confident and saw the cases as being clearer. Kivlighan and Quigley (1991) found that experienced therapists identified more complex pro-

cesses in group therapy and made more refined distinctions among group members. Martin, Slemon, Hiebert, Hallberg, and Cummings (1989), however, found inconsistent results in the case conceptualizations of experienced and novice therapists.

What Is Expertise?

Up to this point, we have used the term *expertise* without defining it or differentiating it from related terms such as *competence*, *experience*, *proficiency*, *effectiveness*, or *mastery*. In the present study, we have adopted the concept of expertise as it has been investigated in cognitive science. Research in cognitive science has shown that experts in a variety of knowledge domains process information differently than novices (Chi, 1988). The classic studies of chess players by Chase and Simon (1973), which followed those of de Groot (1965), demonstrated that master chess players are much better and quicker than weaker players in perceiving meaningful patterns among chess pieces and in recalling the arrangement of pieces on a board after a brief exposure to them. Whereas these studies focused on superior pattern recognition skills, Glaser and Chi (1988) identified several other characteristics of experts. These include the use of problem-solving strategies emphasizing comprehension rather than rote performance and the representation of a problem in more abstract and meaningful conceptual categories than is done by novices, who tend to categorize problems according to more superficial properties.

The organization of expert knowledge in psychotherapy may be characterized as ill-structured, as opposed to the well-structured knowledge domains that are often studied by cognitive psychologists. *Well-structured* knowledge, such as that involved in solving chess or physics problems, has clearly defined initial states and goal states, permitting a systematic way to decide when a proposed solution is acceptable (Simon, 1973). *Ill-structured* knowledge domains are less well defined, and agreement on the “correctness” of a proposed solution is more difficult. The process of constructing a psychotherapy case formulation meets criteria for ill-structured problems. The client’s initial states and goal states are both unknown and changeable. Further, highly trained individuals within and across theoretical orientations may disagree on the correctness of the content of a formulation.

A standard paradigm for studying expertise in a knowledge domain is to present problems to experts and novices, then compare their performances. Protocol analysis, as developed by Ericsson and Simon (1993) from an information-processing theoretical framework, is typically used to evaluate performances. A subject is presented with a problem and instructed to “think aloud” as he or she solves it. Ericsson and Simon showed that verbal protocols can be reliably scored and that verbalizing information affects cognitive processes only if the instructions require verbalization of information that would not otherwise be attended to. When applied to ill-structured knowledge domains, the think-aloud paradigm is often augmented with experimenter-provided verbal probes designed to elicit additional information regarding the nature of the knowledge being used. A modified Ericsson–Simon paradigm was used in the present research.

It is possible that expert therapists view patients in systematically different ways than do novice therapists. Similarly, expert therapists may view patients in different ways than do experienced

therapists; that is, expertise and experience may not be analogous concepts. One may not be able to assume that a psychotherapist is an expert just because he or she has been practicing psychotherapy for many years. The relationship between expertise and experience was also examined in this study.

Hypothesis

Our hypothesis was that the formulations of expert therapists would be of higher quality than those of both the experienced and novice therapists. Specifically, we predicted greater comprehensiveness, elaboration of ideas within formulations, elaboration of ideas within treatment plans, use of a systematic reasoning process across vignettes, precision of language, complexity, coherence, and goodness-of-fit between the formulation and the treatment plan. The experienced therapists were expected to perform at a level between that of the experts and novices. We included both psychodynamic and cognitive-behavioral therapists because these are two of the most common forms of psychotherapy practiced. We did not make any specific predictions with regard to treatment orientation reported by the therapists.

Method

Therapists

Of the 65 psychotherapists participating in the study, 24 were novices, defined as clinical psychology graduate students with less than 1,500 hr of supervised psychotherapy experience. Eleven of the novices identified themselves as psychodynamically oriented (PD), and 13 as cognitive-behaviorally oriented (CB). Nineteen therapists were categorized as experienced because they had 10 or more years of experience practicing either as a CB therapist ($n = 8$) or a PD therapist ($n = 11$). None of these 19 met our criteria for case formulation experts. Twenty-two therapists were experts, defined as experienced psychiatrists or clinical psychologists meeting one or more of three criteria: (a) developed a method of psychotherapy case formulation; (b) led one or more workshops for professionals on how to construct case formulations; or (c) published one or more scientific articles, books, or book chapters on the topic of psychotherapy case formulation. We sought individuals who were recognized as national experts on case formulation. Eleven experts were CB and 11 were PD.

Table 1 shows descriptive information about the therapists. As shown, the PD therapists were significantly older than their CB counterparts, $F(1, 57) = 20.1, p < .001$. The novices, on average, were in their 3rd year of

graduate school. The PD experts had practiced longer than their CB counterparts, $t(19) = 2.21, p < .05$, although a corresponding comparison between the PD and CB experienced therapists fell just short of statistical significance, $t(16) = 1.87, p = .08$, because of greater variability within these groups. Both the PD experienced, $t(17) = 2.21, p < .05$, and expert, $t(19) = 2.33, p < .05$, therapists had more years of supervisory experience than their CB counterparts. Finally, as required for the study, the experts of both treatment modalities had given more workshops and had published more on the topic of case formulation than their experienced and novice counterparts.

Vignettes

Six vignettes were constructed to describe patients with one of three common psychiatric disorders: generalized anxiety disorder, major depressive disorder, and borderline personality disorder. The patients described were either highly prototypical of the condition or had a relatively low number of prototypic features. The purpose of including a variety of vignettes is that previous research has shown that experts' performance varies depending on specifics of problems presented in vignettes (Elstein, Shuyman, & Sprafka, 1978). The mean length of the vignettes was 405 words (range = 368–424). Each vignette contained details regarding identifying information, presenting condition, history of presenting condition, past history of mental health care, developmental history, social history, and mental status.

In order to conduct a manipulation check of the vignettes, we asked each therapist to rate how prototypical each vignette patient was of the target disorder. On a 9-point scale (1 = *minimally prototypical*, 9 = *extremely prototypical*), the intended high prototypical anxiety ($M = 7.56$ vs. 4.63), major depressive disorder ($M = 8.04$ vs. 4.42), and borderline personality disorder ($M = 8.37$ vs. 4.09) cases were rated as more prototypical than their intended low prototypical counterparts, $t(56) = 8.93, 13.14$, and 15.31, respectively; $ps < .001$.

Procedures

The experienced and expert therapists were recruited through direct contacts by Tracy D. Eells, by telephone or at professional conferences. Some were contacted after word-of-mouth referrals from professional associates. The students were recruited through the cooperation of two American Psychological Association-accredited graduate programs in clinical psychology. All therapists were offered \$40 for their participation. Tracy D. Eells or an advanced clinical psychology graduate assistant conducted the interviews either in person or by telephone.

Table 1
Characteristics of Therapists

Therapist characteristic	Cognitive-behavioral therapists			Psychodynamic therapists		
	NOV	EXP	EXT	NOV	EXP	EXT
<i>n</i>	13	8	11	11	11	11
Sex (male/female)	7/6	7/1	7/4	3/8	9/2	9/2
Age (years)	27.1	49.3	46.3	37.3	57.8	56.1
Year in school	3.1			3.5		
Years of experience as therapist	1.6	21.1	21.2	2.4	29.4	28.4
Years of experience as supervisor	0.0	13.4	15.8	0.1	23.7	24.3
Number of therapists supervised	0.0	37.8	90.7	0.3	55.7	147.7
Number of case formulation publications	0.0	0.0	8.1	0.0	1.4	13.3
Number of case formulation workshops taught	0.0	0.3	29.5	0.0	2.5	23.4

Note. NOV = novice; EXP = experienced; EXT = expert.

After obtaining informed consent, the interviewer read a set of instructions to the therapist explaining the purpose of the study, the time required for participation, and the procedure to be followed. We provided each therapist with a written copy of the vignettes and gave them permission to take notes while listening to the vignettes. We then presented the vignettes in a fixed randomized order, with the constraint that vignettes describing the same disorder or presenting a disorder at the same level of prototypicality were not presented consecutively. The audio recording was in the voice of Tracy D. Eells and was about 2 min in length per vignette. After listening to each vignette, the therapist was given 5 min to "think aloud about your conceptualization of the patient. . . [to] construct a case formulation . . . as best you can, addressing whatever you feel is important." After 5 min, the participant was interrupted and given 2 min to "think aloud about how you would treat the patient in psychotherapy." After the therapist completed the think-aloud formulations and treatment plans for all six vignettes, he or she was given a set of six postinterview questionnaires, one for each vignette. A self-addressed stamped envelope was included for return of the questionnaires.

Transcribing and Content Coding the Formulations

The formulations were transcribed in accordance with standards proposed by Mergenthaler and Stinson (1992). These standards include multiple checks on the accuracy of the transcription. The transcribed formulations were then segmented into "idea units," following procedures described by Stinson, Milbrath, and Reidbord (1994). The method involves three judges reading text and segmenting it into relatively small portions (usually a sentence or less) according to the judge's decision of what represents a complete idea. The method is based on the notion that human judges are reliably able to identify complete ideas even if they are not able to articulate the rules used to do so. Stinson et al. report mean reliability estimates of .87 using their method.

Four graduate students and Tracy D. Eells served as text segmenters, working in teams of three. All were trained on practice vignettes until they achieved consistent percentage-of-agreement levels of .85. We estimated reliability by following two formulas proposed by Scott and Hatfield (1985) for data of this type. The more conservative of the two formulas yielded a mean agreement of 87% across all six vignettes, with a range of 86% to 88%. The less conservative method produced mean agreements ranging from 94% to 95%.

Case Formulation Content Coding Method (CFCCM)

Eells, Kendjelic, and Lucas (1998) developed the CFCCM to provide a tool for reliably and comprehensively categorizing the information that a clinician uses in conceptualizing a patient, and for rating the quality of the formulation. A major goal in developing the CFCCM was to make it applicable across multiple approaches to psychotherapy. Toward this end, its developers reviewed several writings on case formulation and case formulation construction methods and identified four broad categories of information that are contained in most methods: (a) symptoms and problems, (b) precipitating stressors or events, (c) predisposing life events or stressors, and (d) a mechanism that links the preceding categories together and offers an explanation of the precipitants and maintaining influences in the individual's problems. Eells et al. demonstrated excellent reliability estimates across both content and quality categories (mean $\kappa = .86$).

The CFCCM was revised and expanded for the current study (Eells, Kendjelic, Lucas, & Lombart, 1998). As with the original system, a key goal was to develop a system for organizing and integrating the range of information clinicians working from various theoretical orientations use in developing formulations. The revised method was also adapted so that it could be applied to each idea unit in a formulation. As shown in Table 2, the revised system is hierarchically organized into four general sections: descriptive information, diagnostic information, inferential information, and treatment planning, each of which contains subcategories.

The goal of the content coding step was to achieve a set of reliable consensus codes for each idea unit in each formulation. The primary task was to determine the presence or absence of a coding system element within an idea unit. Coding rules permitted each idea unit to be coded in any or all of the four major categories of descriptive, diagnostic, inferential, or treatment information. The primary task of the coders was to read the transcript and mark whether or not a particular element from each of these four categories was present in an idea unit. Each idea unit could therefore receive a maximum of one code for each of these major content areas, and a minimum of no codes. For example, the statement "Since [the vignette patient] describes his parents as quite critical, he might have a core belief that others are critical" would be coded for both descriptive and inferential information, but not for diagnosis or treatment. Six clinical or counseling psychology graduate students, working in teams of three, content coded the transcripts. Coder training entailed reading the coding manual, attending a series of training sessions led by Tracy D. Eells, and practicing on several formulations developed for training purposes. After practicing on each formulation, the coders met as a group with Tracy D. Eells and reviewed the codes. These meetings also led to revisions in the coding manual that clarified unanticipated ambiguities. Once it appeared that the coders understood the system, they began coding the transcripts collected for this study. The coders were blind to the treatment orientation and experience of the therapist. At the outset, consensus meetings were held after 5 transcripts were independently coded. At these meetings, each code was reviewed and disagreements were resolved through discussion. A general rule was that a code would be accepted if two of the three coders applied it; however, this rule could be broken if the third coder made a persuasive case for another code. As coding continued, consensus meetings were held after 10–20 transcripts were independently coded.

To estimate the reliability of the codes, we randomly selected 10% of the 390 transcripts and compared the codes of the three-member team. The resulting sample yielded 1,390 idea units, or 16,680 separate coding decisions (1,390 idea units \times 4 coding categories \times 3 coders). We calculated multirater kappa coefficients using the formula discussed in Siegel and Castellan (1988), obtaining values of .61, .81, .62, and .69, respectively, for the descriptive, diagnosis, inferential, and treatment categories. According to Fleiss (1981), these values reflect good to excellent agreement beyond chance. Next, we recoded the sample of codes into dummy codes so they denoted only the presence or absence of any code within a major category rather than a specific code within a major category. Resulting kappas were .64, .82, .70, and .81 for the descriptive, diagnosis, inferential, and treatment categories. Comparing the two sets of kappa suggests that the coders agreed more about the presence or absence of a major coding category than about the specific code within the major category that should be applied. For example, they agreed more that a piece of inferential information was offered than about whether it indicated, say, a psychological mechanism, a predisposing experience, or a problem in global functioning.

Measures of Case Formulation Quality

Expanding on the methods of Strupp (1955, 1958), we measured case formulation quality using eight criteria: comprehensiveness, formulation elaboration, precision of language, complexity, coherence, goodness-of-fit of the formulation to the treatment plan, treatment plan elaboration, and the extent to which the therapist appeared to follow a systematic formulation process across all six vignettes. These measures were developed to be conceptually independent.

Comprehensiveness. *Comprehensiveness* refers to the range of inference categories the therapist discussed in each formulation. It was operationally defined as the number of the 10 inferential domains (diagnosis plus nine inferred information categories) the therapist used in formulating the vignette, as derived from the consensual content codes. If an inferred information category was coded in at least one idea unit, it added to the

Table 2
Description of Content Coding System

Major category	Subcategory
Descriptive information	(a) Identifying information; (b) symptom identification (information given in vignette); (c) history of present or previous episode of mental health problems or care (in self or family); (d) medical history; (e) developmental history; (f) adult life history; (g) mental status information; (h) other descriptive information; (i) more specific descriptive information needed.
Diagnosis	(a) <i>DSM-III-R</i> or <i>DSM-IV</i> Axis I diagnosis; (b) <i>DSM-III-R</i> or <i>DSM-IV</i> Axis II diagnosis; (c) Axis I and Axis II diagnosis in same idea unit; (d) alcohol-substance abuse or dependency.
Inferential information	(a) Problems in global psychological, social, or occupational functioning; (b) inferred symptoms or problems; (c) predisposing experiences, events, traumas, or stressors inferred as explanatory; (d) precipitating or current stressors and/or events; (e) inferred psychological mechanisms (including problematic aspects-traits of the self, problematic aspects of relatedness to others, dysfunctional thoughts and/or core beliefs, affect regulation or dysregulation, defense mechanisms-problematic coping style, skills or social learning deficit); (f) inferred biological mechanisms; (g) inferred social or cultural mechanisms (including absence of or poor psychosocial support; demographic-cultural factors as source of a problem; role conflict; role strain, role transition, role dispute); (h) strengths in global psychological, social, or occupational functioning (including strengths-adaptive skills, aspects of traits of self, adaptive perceptions of or beliefs about others, positive motivation for treatment, adaptive wishes, hopes or goals, good psychosocial support); (i) identification of potential therapy-interfering events.
Treatment planning	(a) Type of treatment; (b) evaluation-assessment; (c) specific techniques; (d) possible "red flag" issues; (e) treatment contract-expectations; (f) therapist-patient relationship; (g) signs and symptoms; (h) predisposing experiences, events, or traumas; (i) psychological mechanisms; (j) social and/or cultural factors; (k) biological factors-psychopharmacology; (l) strengths in global psychological, social, or occupational functioning.

Note. *DSM-III-R* = *Diagnostic and Statistical Manual of Mental Disorders, Third Edition—Revised*; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*.

comprehensiveness score, regardless of the extent to which the idea was elaborated.

Formulation elaboration. The goal of the elaboration ratings step was to assess the extent to which the clinician articulated and developed an inferred information category that the coders had consensually agreed was present in the formulation. Two coders independently assigned a score of 0 to 4 to each of the nine inferred information categories and to the diagnosis category (0 = *element not present*, 1 = *present but not elaborated*, 2 = *present and slightly elaborated*, 3 = *present and moderately elaborated*, and 4 = *present and highly elaborated*). Discrepancies were resolved through consensus decisions made by the two coders. Excluding the categories in which a content code was absent, we calculated a two-way random effects intraclass correlation coefficient (ICC) for these elaboration scores, yielding the following results: all categories, .89; diagnosis, .86; problems in global functioning, .81; inferred symptoms or problems, .69; predisposing experiences, .87; precipitating stressors, .80; psychological mechanisms, .85; biological mechanisms, .75; social or cultural factors, .74; strengths, .82; and identification of potential therapy-interfering events, .84. According to Shrout and Fleiss (1979), the overall ICC indicates excellent reliability, and the ICCs for specific categories range from good to excellent.

Precision of language. We defined *precision of language* as the extent to which the therapist used language to describe a unique individual in the formulation. *Precision* refers only to the degree of articulation and specificity in the use of language, not to the quality or the amount of information covered. We assumed that highly precise language facilitates construction of a formulation that is tailored to a unique individual. A 5-point, verbally

anchored Likert scale was used to describe precision of language (0 = *insufficient information*, 1 = *very little precision*, 2 = *little precision*, 3 = *moderate precision*, 4 = *high precision*). Coders achieved a two-way random effects ICC of .77 on precision of language, which is excellent agreement.

Complexity. *Complexity* was defined as the extent to which therapists integrated several facets of the person's problems into a meaningful presentation. Highly complex formulations either integrated several aspects of the person's problems and functioning or developed one or two themes extensively. Coders achieved a two-way random effects ICC of .83 using a 5-point scale (0 = *insufficient information*, 1 = *very little complexity*, 2 = *little complexity*, 3 = *moderate complexity*, 4 = *high complexity*), which may be considered excellent.

Coherence. *Coherence* was the extent to which the formulation provided an internally consistent account of the individual's problems. Coders were trained to judge coherence by determining the extent to which the formulation could be summarized in a short and highly meaningful sentence. Coders achieved a two-way random effects ICC of .75 using a 5-point scale (0 = *insufficient information*, 1 = *very little coherence*, 2 = *little coherence*, 3 = *moderate coherence*, 4 = *high coherence*). This is also considered excellent reliability.

Treatment plan elaboration. For this variable, coders were instructed to rate how well the clinician explained or developed the treatment plan. The following scale was used: 0 = *insufficient information*, 1 = *very little elaboration*, 2 = *little elaboration*, 3 = *moderate elaboration*, 4 = *high elaboration*. An ICC of .86 was achieved, which shows excellent reliability.

Goodness-of-fit. *Goodness-of-fit* was the extent to which the treatment plan was viewed as consistent with the formulation. Coders rated the extent to which the treatment plan addressed the issues raised in the formulation on the basis of the following scale: 0 = *insufficient information*, 1 = *very little consistency*, 2 = *little consistency*, 3 = *moderate consistency*, 4 = *high consistency*. The ICC for this variable was .73, which is considered good.

Systematic process. We defined *systematic process* as the degree of evidence that the clinician followed an a priori scheme, that is, a predetermined, structured method that is independent of specific patient information, in organizing clinical information into a formulation. Each therapist's set of six formulations were independently examined by two advanced clinical psychology graduate students using a 5-point scale: 1 = *no evidence or nearly no evidence*, 2 = *little evidence*, 3 = *moderate degree of evidence*, 4 = *clear and convincing evidence*, 5 = *evidence beyond a reasonable doubt*. The raters held consensus meetings to compare ratings and resolve discrepancies, producing a single systematic process score for each therapist. Reliability was estimated by means of a two-way random effects ICC, yielding a value of .77, which is considered excellent.

Procedures for Quality Coding

The procedure for quality coding was first to randomly sequence the formulations, then, in one pass, code complexity, precision of language, coherence, goodness-of-fit of formulation to treatment plan, and elaboration of treatment plan. Formulation elaboration and systematic process were coded in separate passes. As noted, we calculated comprehensiveness on the basis of the content codes.

Results

A key initial question is whether the therapists found the information contained in the vignettes adequate to develop a formulation. We addressed this question by asking the therapists to rate the adequacy of each vignette on a 9-point scale (1 = *completely inadequate*; 9 = *perfectly adequate*). The mean response in the sample was 5.37 ($SD = 1.91$), indicating that the vignettes could be considered moderately adequate on average. We observed statistically significant main effects for experience level, $F(2, 341) = 11.38, p < .01$; therapy mode, $F(1, 341) = 16.90, p < .01$; and vignette, $F(5, 311) = 12.78, p < .01$. (Note: Error degrees of freedom differ from stated sample size because of missing data.) CB therapists ($M = 5.74, SD = 1.92$) found the information in the vignettes to be more adequate for developing a formulation than did PD therapists ($M = 5.00, SD = 1.84$). Experts ($M = 4.75, SD = 2.17$) found the information less adequate than did either the

novices ($M = 5.54, SD = 1.87$), $t(237) = -3.01, p < .01$, or the experienced therapists ($M = 5.81, SD = 1.47$), $t(219) = -4.24, p < .01$. When each vignette was compared with the mean of all six, Vignettes 2, 3, and 5 were significantly different. Vignette 2 presented the low-prototypical personality disordered patient; it was rated as lowest in adequacy and was commented on by many of the therapists as the most difficult to formulate. Its mean rating of 4.25 ($SD = 1.70$) may still be considered in the moderately adequate range. Vignettes 3 ($M = 6.14, SD = 1.73$) and 5 ($M = 6.46, SD = 1.79$) were rated relatively high in adequacy. These are both high-prototypical vignettes, the first for anxiety and the second for a personality disorder. In sum, the vignettes may be considered moderately adequate as a basis for developing a formulation.

Next, we examined the formulations descriptively. A total of 14,499 codes were applied to the idea units, an average of 44.9 codes per each of the 390 formulations. A mean of 13.2 (29.5%) of the codes pertained to descriptive information, 2.8 (6.1%) were diagnostic in nature, 16.5 (36.8%) dealt with inferential information, and 12.4 (27.6%) were about treatment.

Quality of Formulations

Prior to investigating differences among the groups on the quality measures, we examined intercorrelations among the eight measures of formulation quality. We wanted assurance that the measures were not so highly intercorrelated that they essentially served as proxies for each other rather than as measures of unique aspects of case formulation quality. As Table 3 shows, these measures appear to tap into separate aspects of quality. The two most highly correlated measures, treatment elaboration and goodness-of-fit, account for .46 of the variation measured by each. This appears a reasonable degree of overlap because both measure treatment aspects of the formulation. Also, as designed, comprehensiveness and elaboration of formulation are uncorrelated, thus measuring independent aspects of the formulation process. Note that the systematic process measure correlates most highly with both elaboration measures and with complexity, suggesting that following a systematic formulation process leads to more detailed and multilayered formulations. Similarly, the correlation between complexity and elaboration of formulation shows that some development of the formulation may be required to achieve complexity.

Table 4 shows means and standard deviations for each of the formulation quality measures for each of the six groups. We began

Table 3
Pearson Product-Moment Correlations Among Quality Measures

Quality measure	1	2	3	4	5	6	7	8
1. Comprehensiveness	—							
2. Elaboration of information	-.09	—						
3. Complexity	.40**	.67**	—					
4. Precision	.24	.37**	.57**	—				
5. Coherence	.04	.51**	.61**	.32*	—			
6. Goodness-of-fit	.33**	.21	.38**	.23	.45**	—		
7. Treatment elaboration	.35**	.40**	.57**	.16	.32**	.68**	—	
8. Systematic process	.17	.42**	.40**	.01	.18	.15	.37**	—

Note. $N = 65$.

* $p < .05$. ** $p < .01$.

Table 4
Formulation Quality Means and (Standard Deviations)

Formulation quality measure	Cognitive-behavioral therapists			Psychodynamic therapists		
	NOV	EXP	EXT	NOV	EXP	EXT
Comprehensiveness	5.31 (1.37)	4.81 (1.05)	5.86 (1.41)	4.89 (1.53)	5.20 (1.29)	5.30 (1.41)
Formulation elaboration	2.23 (0.39)	2.16 (0.40)	2.31 (0.42)	2.27 (0.43)	2.09 (0.36)	2.37 (0.42)
Diagnosis	1.47 (1.19)	1.53 (1.15)	1.55 (1.36)	1.79 (1.31)	1.48 (1.15)	2.33 (1.09)
Problems in global functioning	0.48 (0.83)	0.68 (0.84)	0.70 (0.91)	0.42 (0.71)	0.46 (0.74)	0.77 (0.97)
Inferred symptoms or problems	1.27 (0.81)	1.03 (0.86)	1.42 (0.88)	1.27 (0.83)	1.21 (.97)	1.74 (0.88)
Predisposing stressors	2.65 (1.00)	2.48 (1.14)	2.94 (1.07)	2.62 (0.97)	2.40 (.92)	2.39 (1.15)
Precipitating stressors	0.65 (0.96)	0.70 (0.93)	0.73 (0.99)	0.74 (0.96)	0.54 (0.85)	0.77 (0.96)
Psychological mechanisms	2.95 (0.90)	2.45 (1.03)	2.97 (1.01)	2.86 (0.98)	2.75 (1.00)	3.14 (0.80)
Biological mechanisms	0.00 (0.00)	0.15 (0.40)	0.14 (0.46)	0.17 (0.47)	0.08 (0.28)	0.21 (0.60)
Social or cultural factors	0.38 (0.78)	0.36 (0.73)	0.53 (1.00)	0.62 (0.96)	0.56 (0.92)	0.68 (0.98)
Strengths	0.65 (0.87)	0.74 (0.95)	0.97 (1.05)	.85 (1.02)	0.50 (0.77)	0.77 (0.94)
Potential therapy-interfering events	0.30 (0.68)	0.58 (1.01)	0.53 (1.07)	0.28 (0.66)	0.23 (0.63)	0.50 (1.01)
Precision of language	3.69 (0.36)	3.38 (0.61)	3.49 (0.63)	3.63 (0.42)	3.44 (0.60)	3.59 (0.53)
Complexity	3.02 (0.67)	2.83 (0.72)	3.28 (0.67)	2.94 (0.65)	2.95 (0.70)	3.20 (0.60)
Coherence	2.90 (0.62)	2.67 (0.62)	2.80 (0.59)	2.82 (0.62)	2.79 (0.61)	2.99 (0.64)
Goodness-of-fit of formulation to treatment plan	3.11 (0.58)	3.01 (0.61)	3.41 (0.52)	3.18 (0.59)	3.00 (0.60)	3.22 (0.51)
Treatment plan elaboration	2.99 (0.75)	2.82 (0.74)	3.43 (0.63)	2.78 (0.72)	2.59 (0.69)	2.99 (0.70)
Systematic formulation process	1.71 (0.75)	1.38 (0.52)	2.91 (1.04)	1.91 (0.83)	1.45 (0.69)	2.36 (1.28)

Note. NOV = novice, EXP = experienced, EXT = expert.

our analysis of the quality of the formulations by conducting a three-way (Experience Level \times Therapy Mode \times Vignette) multivariate analysis of variance, entering all quality measures as dependent variables, except systematic process because its unit of analysis was the therapist, not the vignette. Unless stated otherwise, all analyses used a Type I error rate of .05. Using the Wilks's lambda criterion for significance, results showed a significant Experience Level \times Therapy Mode interaction, $F(14, 696) = 1.79$, $p < .05$, and main effects for experience level, $F(14, 696) = 6.67$, $p < .01$, and therapy mode, $F(7, 348) = 3.70$, $p < .01$. Univariate analyses, including those for systematic process, are summarized in Table 5. As shown, we found an experience level main effect for all quality measures except coherence, and a therapy mode main effect for treatment plan elaboration. In addition, we observed an Experience Level \times Therapy Mode interaction for comprehensiveness. In order to produce a single measure of formulation quality, we summed all eight quality measures into a total quality measure (giving systematic process the same weight as the other variables) and subjected it to an analysis of variance, which produced a main effect for experience level and an Experience Level \times Therapy Mode interaction. We also conducted univariate analyses of each elaboration inferential category. These showed experience level main effects for elaboration of diagnosis, elaboration of problems in global functioning, elaboration of inferred symptoms or problems, and elaboration of psychological mechanisms.

In order to understand these main effects and interactions, we conducted pairwise comparisons for the three experience level groups. We used a Bonferroni correction for each set of comparisons because these tests are not independent, as well as to control for experiment-wise Type I error. The results (see Table 5) show that the formulations of experts were rated as more comprehensive, elaborated, and complex compared with those of both the novices and the experienced therapists. The treatment plans of the experts were more elaborated and were rated as better fitting to the

formulations compared with those of the novices and the experienced therapists. In addition, compared with those of the novices and experienced therapists, the sets of six formulations of the experts showed more evidence of a consistent, structured process being followed. Test results also show that the experts elaborated more than either the novices or experienced therapists on possible diagnoses, problems in global functioning, symptoms or problems that were inferred, and psychological mechanisms. Finally, as shown in Figure 1, the total quality rating showed the experts to be superior to the novice and experienced therapists. To our surprise, the formulations of the novices were rated as higher in overall quality than those of the experienced therapists.

In order to gain an additional perspective of the comparative performances of the therapists, we conducted a median split of all 390 vignettes on the basis of the total quality rating. Figure 2 shows that the experts' formulations were far more likely than those of the other group to be above the 50th percentile in overall quality. We also examined the 20 formulations rated highest in total quality and found that 18 were produced by experts, 1 by an experienced therapist, and 1 by a novice.

Effect Sizes

In order to determine the magnitude of the observed differences, we examined effect sizes of each quality measure for which statistically significant differences were observed, as shown in Table 5. Following Cohen (1988), we chose the product-moment correlation coefficient as the measure of effect size, correlating each dependent variable with index values (i.e., dummy codes) of the independent variables. Cohen suggested that an r of .10 be considered small, an r of .30 be considered medium, and an $r = .50$ be considered large. On the basis of these criteria, the observed effect sizes range widely, although the effect size for total quality is in the medium to large range. Systematic process and treatment

Table 5
Summary of Analysis of Variance

Formulation quality measure	Effect or pairwise comparison	Statistical test results	<i>r</i>
Comprehensiveness	Experience level	$F(2, 384) = 6.56, p < .01$	N/A
	Expert vs. novice	$t(274) = 2.67, p < .01$.16
	Expert vs. experienced	$t(244) = 3.22, p < .01$.20
Overall formulation elaboration	Experience Level \times Therapy Mode	$F(2, 384) = 4.10, p < .05$	N/A
	Experience Level	$F(2, 384) = 8.57, p < .01$	N/A
	Expert vs. experienced	$t(244) = 4.31, p < .01$.27
Diagnosis	Experienced vs. novice	$t(256) = -2.62, p < .01$	-.16
	Experience level	$F(2, 384) = 4.19, p < .05$	N/A
	Expert vs. experienced	$t(244) = 2.75, p < .01$.17
Problems in global functioning	Therapy mode (CB vs. PD)	$F(1, 384) = 8.05, p < .01$	-.16
	Experience Level \times Therapy Mode	$F(2, 384) = 3.62, p < .05$	N/A
	Expert vs. novice	$F(2, 384) = 3.85, p < .05$	N/A
Symptoms or problems	Expert vs. novice	$t(274) = 2.76, p < .01$.16
	Experience level	$F(2, 384) = 9.23, p < .05$	N/A
	Expert vs. novice	$t(274) = 3.03, p < .01$.18
Predisposing experiences	Expert vs. experienced	$t(244) = 4.16, p < .01$.26
	Therapy mode (CB vs. PD)	$F(1, 384) = 4.35, p < .05$.10
	Precipitating stressors	All effects and interaction	<i>ns</i>
Psychological mechanisms	Experience level	$F(2, 384) = 6.91, p < .01$	N/A
	Expert vs. experienced	$t(244) = 3.85, p < .01$.24
	Experienced vs. novice	$t(256) = -2.64, p < .01$	-.16
Biological mechanisms	All effects and interaction	<i>ns</i>	
	Therapy mode (CB vs. PD)	$F(1, 384) = 4.50, p < .05$	-.11
	Social or cultural factors	All effects and interaction	<i>ns</i>
Strengths	Potential therapy-interfering events	All effects and interaction	<i>ns</i>
	Experience level	$F(2, 384) = 7.20, p < .01$	N/A
	Experienced vs. novice	$t(256) = -4.03, p < .01$	-.24
Precision of language	Experience level	$F(2, 384) = 9.50, p < .01$	N/A
	Expert vs. novice	$t(274) = 3.36, p < .01$.20
	Expert vs. experienced	$t(244) = 4.01, p < .01$.25
Complexity	All effects and interaction	<i>ns</i>	N/A
	Experience level	$F(2, 384) = 8.91, p < .01$	N/A
	Expert vs. novice	$t(274) = 2.58, p < .01$.15
Coherence	Expert vs. experienced	$t(244) = 4.28, p < .01$.26
	Experience level	$F(2, 384) = 16.44, p < .01$	N/A
	Expert vs. novice	$t(274) = 3.69, p < .01$.22
Goodness-of-fit of formulation to treatment plan	Expert vs. experienced	$t(244) = 5.86, p < .01$.35
	Therapy mode (CB vs. PD)	$F(1, 384) = 16.59, p < .01$	-.21
	Experience level	$F(2, 59) = 9.99, p < .01$	N/A
Treatment plan elaboration	Expert vs. novice	$t(44) = 2.90, p < .01$.40
	Expert vs. experienced	$t(39) = 4.06, p < .01$.55
	Experience level	$F(2, 384) = 42.22, p < .01$	N/A
Systematic formulation process	Expert vs. novice	$t(274) = 5.82, p < .01$.33
	Expert vs. experienced	$t(244) = 8.82, p < .01$.49
	Experienced vs. novice	$t(256) = -3.62, p < .01$	-.22
Overall quality	Experience Level \times Therapy Mode	$F(2, 384) = 3.10, p < .05$	N/A

Note. The index values used for the independent variables to calculate effect sizes were 3 for expert, 2 for experienced, and 1 for novice. For therapy mode, index values were 1 for cognitive-behavioral (CB) and 2 for psychodynamic (PD). Empty cells signify lack of statistical significance in analysis of variance.

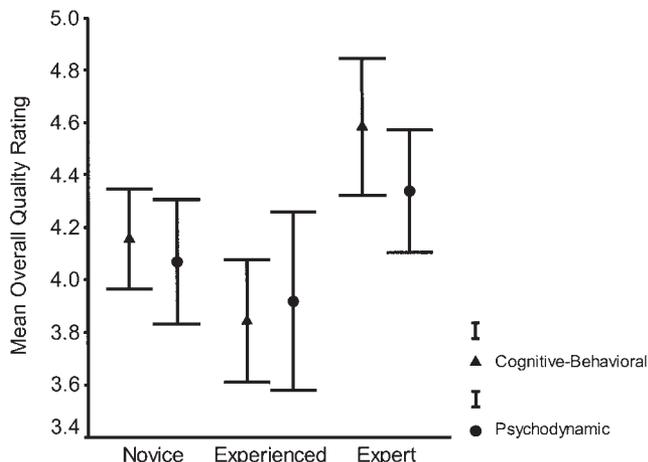
plan elaboration produced the largest effect sizes among the eight quality measures.

Discussion

These results demonstrate the overall superiority of the case formulation skills of the expert therapists, compared with the novice and experienced therapists, across multiple types of patient vignettes, regardless of whether theoretical orientation was cognitive-behavioral or psychodynamic. When separate components of case formulation quality were considered, effects were seen for comprehensiveness; formulation elaboration, including elaboration of diagnosis, problems in global functioning, inferred

symptoms or problems, and explanatory psychological mechanisms; complexity; goodness-of-fit to the treatment plan; treatment plan elaboration; and use of a systematic formulation process. Effect sizes falling in the medium to large range are those most likely to be clinically meaningful. These include those for overall quality, systematic formulation process, and treatment plan elaboration.

One question these results leave unanswered is "What cognitive processes enabled the experts to produce better formulations?" A partial answer is suggested by Hayes (1985), who found that experts in various scientific and artistic fields required at least 10 years of practice before acquiring a "genius" level of performance,



Note. Confidence intervals are 95%

Figure 1. Mean overall case formulation quality

although as our results suggest, practice in itself may not ensure achieving an expert level of performance. In their summary of research on expertise Glaser and Chi (1988) identify several characteristics of expertise. One is that experts excel mainly in their own domain. Thus, one would not necessarily expect the experts in the present study to excel in other aspects of the practice of their profession, including perhaps applying the formulation in therapeutic interventions. Neither would one predict greater levels of intelligence in the experts as compared with the other subjects.

A second characteristic identified by Glaser and Chi (1988) is that experts perceive large, meaningful patterns in their domain, a skill that is thought to reflect a difference in the organization of the knowledge base rather than general perceptual superiority. It is possible that the experts in this study recognized a greater number of patterns in the descriptive data of the vignettes and drew upon that knowledge to construct more complex, elaborated, and nu-

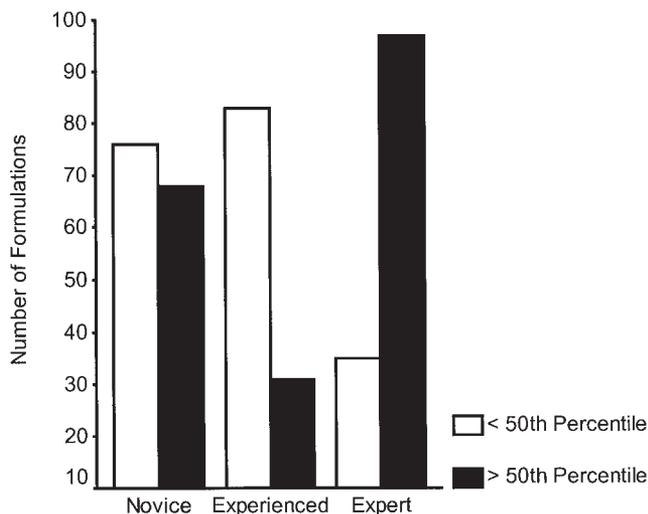


Figure 2. Median split of overall quality rating

anced formulations. We are presently analyzing the reasoning processes of the therapists, as a reflection of the organization of knowledge, to determine whether differences can be observed.

A third characteristic of experts is that they see and represent a problem in their domain at a more principled or deeper level than novices, who are more likely to represent a problem by its surface features. For example, Weiser and Shertz (1983) found that expert computer programmers organized programming problems according to solution algorithms, whereas novices sorted them according to areas of application. Analogously, the expert case formulators may have organized formulation “solutions” around a set of a priori formulation schemas, such as “anxiety predicts a schema of fear of loss of control, which is related to a history of unpredictable caretaking and ambivalent withdrawal as a coping response.” The finding that experts were more likely to use a consistent, systematic formulation process across their six vignettes also provides evidence of a priori cognitive structures guiding their formulation process.

A fourth characteristic of experts is that they have strong self-monitoring skills. They are more aware of when they make errors, why they fail to comprehend, and when they need to recheck their solutions. This characteristic may explain why the experts found the vignettes less adequate than the other therapists.

We were surprised that novices performed better than experienced therapists in some comparisons, particularly total formulation quality. One explanation may be that the learning of the novices, as compared with that of the experienced therapists, more closely matches the teaching of the experts. If so, one might expect greater similarity between the performance of the experts and that of the novices. Experts may use cognitive tools to keep themselves “calibrated” to a high standard of case formulation skill, whereas the novices were more recently calibrated through their studies, as compared with the experienced therapists who have been out of formal training for several years. A consequent risk to experienced therapists may be overconfidence, which does not alert one to the need for calibration (Dougherty, Gettys, & Ogden, 1999; Oskamp, 1965).

These results have implications for the training of therapists in case formulation skills. It may be possible to accelerate the training curve if one can identify specific skills differentiating experts from other therapists and train for these specific skills. Because it showed our largest effect size ($r = .55$ when comparing experts and experienced therapists), the learning of a systematic formulation method may be one such skill. Along these lines, Kendjelic and Eells (2004) tested a generic case formulation training model that emphasized categories to include in a formulation (symptoms and problems, precipitating stressors, predisposing events, and inferential information) as well as the importance of qualitative measures such as comprehensiveness, precision of language, complexity, and an appropriate degree of inference. They found that therapists undergoing a 2-hr training produced higher quality formulations than a matched group of therapists who did not receive the training. Another skill that may merit more training focus is elaborating the treatment plan; in the present study, it showed a medium effect size in comparisons between expert and experienced therapists.

These results should be viewed in light of certain limitations. First, the sample is relatively small, limiting generalizability, and differed in ways other than theoretical orientation and degree of

experience/expertise, introducing possible confounds. One could object to our decision to combine the therapists into two groups, PD and CB, when each of these groups has many diverse theoretical variations within it. We were able to explore this possibility because we asked the therapists to identify a more narrow orientation within the broader one. Almost without exception, the CB therapists identified with the approach of Aaron T. Beck (e.g., Beck & Emory, 1979; Beck & Freeman, 1990; Beck, Rush, Shaw, & Emory, 1979), whereas the PD therapists identified themselves within an interpersonal-relational tradition. Second, these results relate only to case formulation skills, not to the association between case formulations and psychotherapy process or outcome variables. Third, although the kappa values of the CFCCM coding system fell into the good to excellent range, we would have preferred that they had been uniformly excellent. Fourth, some may object to the use of vignettes, claiming that they are too distant from the actual clinical situation. On the other hand, the use of vignettes provided several benefits. We were better able to control the information provided to the therapist, for example, by ensuring that similar categories of information were contained in each. Vignettes also permitted us to systematically vary disorder and prototypicality, which made possible the presentation of three different disorders at two distinct levels of prototypicality. Although engagement was not quantitatively measured, we were struck by how engaged the therapists became in producing formulations. Several commented on how realistic the vignettes were and how similar they were to patients they had seen; or they asked whether the vignettes represented actual patients. Finally, one could object that the vignettes were not sufficiently adequate, with a mean rating of only 5.37 on a 9-point scale. We were not surprised at this finding, however. When designing the study, constraints regarding how much time we could realistically ask of therapists forced a choice between constructing six sufficiently adequate vignettes or fewer, more adequate vignettes. Our choice to construct six vignettes varying by disorder and prototypicality provides firmer ground on which to generalize our findings.

A final observation is that these results indicate that expertise and experience may not be synonymous. As one clinician put it, "Twenty years of experience is not the same as one year of experience repeated 20 times." These results clearly argue for continued effort to develop one's skills.

References

- Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology, 64*, 619–622.
- Beck, A. T., & Emery, G. (1979). *Cognitive therapy of anxiety and phobic disorders*. Philadelphia: Center for Cognitive Therapy.
- Beck, A. T., & Freeman, A. (1990). *Cognitive therapy of personality disorders*. New York: Guilford Press.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- Beutler, L. E., Machado, P. P. P., & Neufeldt, S. A. (1994). Therapist variables. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 229–269). New York: Wiley.
- Blatt, S. J., Sanislow, C. A., Zuroff, D. C., & Pilkonis, P. A. (1996). Characteristics of effective therapists: Further analysis of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*, 1276–1284.
- Caspar, F. (1995). *Plan analysis: Toward optimizing psychotherapy*. Seattle, WA: Hogrefe & Huber.
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.
- Chi, M. T. H. (Ed.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., et al. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1*, 81–91.
- Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59*, 20–26.
- Cummings, A. L., Hallberg, E. T., Martin, J., Slemmon, A., & Heibert, B. (1990). Implications of counselor conceptualizations for counselor education. *Counselor Education and Supervision, 30*, 120–134.
- de Groot, A. (1965). *Thought and choice in chess*. The Hague, the Netherlands: Nouton.
- Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*, 180–209.
- Eells, T. D. (1997). Psychotherapy case formulation: History and current status. In T. D. Eells (Ed.), *Handbook of psychotherapy case formulation* (pp. 1–25). New York: Guilford Press.
- Eells, T. D. (2002). Formulation. In M. Hersen & W. Sledge (Eds.), *The encyclopedia of psychotherapy* (pp. 815–822). New York: Academic Press.
- Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation? Development and use of a content coding manual. *Journal of Psychotherapy Practice and Research, 7*, 144–153.
- Eells, T. D., Kendjelic, E. M., Lucas, C. P., & Lombart, K. G. (1998). *Manual for case formulation and treatment plan coding*. Unpublished manuscript, University of Louisville.
- Elstein, A. S., Shuyman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.
- Frank, J. D., & Frank, J. B. (1991). *Persuasion and healing: A comparative study of psychotherapy* (3rd ed.). Baltimore: Johns Hopkins University Press.
- Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. XV–XXVIII). Hillsdale, NJ: Erlbaum.
- Goldfried, M. R., Raue, P. J., & Castonguay, L. G. (1998). The therapeutic focus in significant sessions of master therapists: A comparison of cognitive-behavioral and psychodynamic-interpersonal interventions. *Journal of Consulting and Clinical Psychology, 66*, 803–810.
- Hayes, J. R. (1985). Three problems in teaching general skills. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions* (Vol. 2, pp. 391–405). Hillsdale, NJ: Erlbaum.
- Hersen, M., & Porzelius, L. K. (Eds.). (2002). *Diagnosis, conceptualization, and treatment planning for adults: A step-by-step guide*. Mahwah, NJ: Erlbaum.
- Hillerbrand, E. T., & Claiborn, C. D. (1990). Examining reasoning skill differences between expert and novice counselors. *Journal of Counseling and Development, 68*, 684–691.

- Jacobson, N. (1995). The overselling of therapy. *Family Therapy Networker*, 19(March/April), 41–47.
- Kendjelic, E. M., & Eells, T. D. (2004). *Generic psychotherapy case formulation training improves formulation quality*. Manuscript submitted for publication.
- Kingdon, D., Tyrer, P., Seivewright, N., Ferguson, B., & Murphy, S. (1996). The Nottingham study of neurotic disorder: Influence of cognitive therapists on outcome. *British Journal of Psychiatry*, 69, 93–97.
- Kivlighan, D. M., & Quigley, S. T. (1991). Dimensions used by experienced and novice group therapists to conceptualize group process. *Journal of Counseling Psychology*, 38, 415–423.
- Luborsky, L., Barber, J. P., Binder, J., Curtis, J., Dahl, H., Horowitz, L., et al. (1993). Transference-based measures: A new class based on psychotherapy sessions. In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 326–341). New York: Basic Books.
- Luborsky, L., Crits-Christoph, P., McLellan, T., Woody, G., Piper, W. E., Liberman, B., et al. (1986). Do therapists vary much in their success? Findings from four outcome studies. *American Journal of Orthopsychiatry*, 56, 501–512.
- Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, 42, 602–611.
- Martin, J., Slemmon, A. G., Hiebert, B., Hallberg, E. T., & Cummings, A. L. (1989). Conceptualizations of novice and experienced counselors. *Journal of Counseling Psychology*, 36, 395–400.
- Martindale, C. (1978). The therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 46, 1526–1530.
- Mergenthaler, E., & Stinson, C. H. (1992). Psychotherapy transcription standards. *Psychotherapy Research*, 2, 125–142.
- Najavits, L. M., & Weiss, R. D. (1994). Variations in therapist effectiveness in the treatment of patients with substance use disorders: An empirical review. *Addiction*, 89, 679–688.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29, 261–265.
- Persons, J. B. (1989). *Cognitive therapy in practice: A case formulation approach*. New York: Norton.
- Ricks, D. F. (1974). Supershrink: Methods of a therapist judged successful on the basis of adult outcomes of adolescent patients. In D. F. Ricks, M. Roff, & A. Thomas (Eds.), *Life history research in psychopathology* (pp. 275–297). Minneapolis: University of Minnesota Press.
- Scott, M. M., & Hatfield, J. C. (1985). Problems of analyst and observer agreement in naturalistic narrative data. *Journal of Educational Measurement*, 22, 207–218.
- Shaw, B. F., & Dobson, K. S. (1988). Competency judgments in the training and evaluation of psychotherapists. *Journal of Consulting and Clinical Psychology*, 56, 666–672.
- Shaw, B. F., Olmsted, M., Dobson, K. S., Sotsky, S. M., Elkin, I., Yamaguchi, J., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*, 67, 837–846.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4, 181–201.
- Stinson, C. H., Milbrath, C., & Reidbord, S. P. (1994). Thematic segmentation of psychotherapy transcripts for convergent analysis. *Psychotherapy*, 31, 36–48.
- Strupp, H. H. (1955). Psychotherapeutic technique, professional affiliation, and experience level. *Journal of Consulting Psychology*, 19, 97–102.
- Strupp, H. H. (1958). The psychotherapist's contribution to the treatment process. *Behavioral Science*, 3, 43–67.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin*, 122, 203–215.
- Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, 5, 425–433.
- Weiser, M., & Shertz, J. (1983). Programming problem representation in novice and expert programmers. *Instructional Journal of Man-Machine Studies*, 19, 391–398.
- Wiser, S., & Goldfried, M. R. (1998). Therapist interventions and client emotional experiencing in expert psychodynamic–interpersonal and cognitive–behavioral therapies. *Journal of Consulting and Clinical Psychology*, 66, 634–640.

Received January 7, 2004

Revision received April 15, 2004

Accepted May 20, 2004 ■