HELPFUL AND HINDERING EVENTS IN PSYCHOTHERAPY: A PRACTICE RESEARCH NETWORK STUDY

LOUIS G. CASTONGUAY AND
JAMES F. BOSWELL
Penn State University

SANNO E. ZACK
Stanford University

SALLY BAKER,
MARY A. BOUTSELIS, AND
NANCY R. CHISWICK
Independent Practice, State College, PA

DIANA D. DAMER
University of Texas at Austin

NEAL A. HEMMELSTEIN,
JEFFREY S. JACKSON,
MAROLYN MORFORD,
STEPHEN A. RAGUSEA, AND
J. GOWEN ROPER
Independent Practice, State College, PA

CATHERINE SPAYD
Independent Practice, Duncansville, PA

TARA WEISZER
University of Georgia

THOMAS D. BORKOVEC
Penn State University

MARTIN GROSSE HOLTFORTH
Friedrich Schiller University

This paper presents the findings of a psychotherapy process study conducted within the Pennsylvania Psychological Association Practice Research Network (PPA-PRN). The investigation was the product of a long-term collaborative effort, both in terms of the study design and implementation, between experienced clinicians of various theoretical orientations and full-time psychotherapy researchers. Based on a relatively large sample of clients seen in independent practice settings, close to 1,500 therapeutic events (described by clients and therapists as being particularly helpful or hindering) were collected. These events were

Louis G. Castonguay, James F. Boswell, and Thomas D. Borkovec, Department of Psychology, Penn State University; Sanno E. Zack, Department of Psychiatry and Behavioral Science, Stanford University; Sally Baker, Mary A. Boutselis, Nancy R. Chiswick, Neal A. Hemmelstein, Jeffrey S. Jackson, Marolyn Morford, Stephen A. Ragusea, and J. Gowen Roper, Independent Practice, State College, PA; Diana D. Damer, Counseling and Mental Health Center, University of Texas at Austin; Catherine Spayd, Independent Practice, Duncansville, PA; Tara Weiszer, Counseling and Psychological Services, University of Georgia; and Martin Grosse Holtforth, Department of Psychology, Friedrich Schiller University.

Correspondence concerning this article should be addressed to Louis G. Castonguay, PhD, 308 Moore Building, University Park, PA 16802. E-mail: lgc3@psu.edu

coded by three independent observers using a therapy content analysis system. Among the findings, both clients and therapists perceived the fostering of self-awareness as being particularly helpful. The results also point to the importance of paying careful attention to the therapeutic alliance and other significant interpersonal relationships. The merits and difficulties of conducting scientifically rigorous and clinically relevant studies in naturalistic contexts are also discussed.

Although applied psychology is based on a scientist-practitioner model (Raimy, 1950), the state of the relationship between psychotherapy research and clinical practice may at best be described as tenuous. In the past several years, a number of strategies have been proposed to build stronger connections between researchers and clinicians, such as the conduct of effectiveness research (aimed at testing the external validity of randomized controlled trials [RCTs]) and the publication of treatment manuals that have received support for their efficacy in RCTs.

Although these strategies have contributed to the enhancement of evidence-based practice, they also seem to reflect a "top-down" approach to the accumulation and application of scientifically based information. In both cases, the knowledge about how to practice psychotherapy is generated from researchers and is then theoretically disseminated to the practitioners, hopefully generalizing to clinical practice. Inadvertently, such strategies may have contributed to what has been described elsewhere as "empirical imperialism" (see Castonguay in Lampropoulos et al., 2002) where scientists who often treat very few patients inform clinicians who rarely participate in research on what should be studied to understand and improve psychotherapy.

A more fruitful strategy to facilitate the integration of science and practice may lie in the formation of Practice Research Networks (PRNs). These PRNs are based on an active collaboration between researchers and clinicians in the development of clinically relevant studies that are at the same time scientifically rigorous. To our knowledge, the Pennsylvania Psychological Association Practice Research Network (PPA-PRN) is the first PRN to be specifically devoted to this type of collaborative research on psychotherapy. Established under the leadership of a full-time academician (Tom Borkovec) and a full-time clinician (Steve Ragusea), the PPA-PRN launched its first study in the mid-1990s, with the goal of testing the feasibility of conducting scientifically sound research within the practice setting using a core assessment battery for obtaining pre and postoutcome data within a state wide infrastructure (Borkovec, Echemendia, Ragusea, & Ruiz, 2001). The results showed not only that significant improvement took place during treatment across a number of outcome dimensions (e.g., symptoms, role and interpersonal functioning), but also revealed interesting relationships between participant and treatment variables on the one hand, and outcome on the other. For example, female clients showed greater improvement than male clients in family relationships, and clients of male therapists showed more positive change in intimacy than clients of female therapists. Furthermore, while the therapist's caseload was negatively related to client outcome, greater improvement was associated with the number of sessions received and the client's initial expectancy for improvement (Borkovec et al., 2001).

The present paper reports the results of the second study (Phase II) conducted by the PPA-PRN. An adjacent paper (Castonguay, Nelson, et al., pp. 345–354, this issue) describes the experience of clinicians who participated in this study. Built on lessons learned from our first investigation, this study is the fruit of a collaboration between full-time clinicians and full-time researchers in central Pennsylvania, who were both fully involved not only in the implementation, but also in the design (which alone required regular meetings over several months) of this study.

The primary goal of this study was to assess what clients and therapists found helpful and/or unhelpful during each session of psychotherapy. This type of research is an example of what has been described as the "events paradigm" (Arnkoff, Victor, & Glass, 1993; Elliott, Slatick, & Urman, 2001), which is aimed at identifying especially meaningful aspects of psychotherapy, "with the notion that these are the events most likely to inform us about therapy process"

(Arnkoff et al., 1993, p. 37). The investigation of such significant events has been described elsewhere by Elliott (1983) as a research method that could "substantially close the gap between psychotherapy process research and the practice of psychotherapy" (p. 47). As described by Elliott (1983), the identification of critical processes of change is likely to be clinically relevant as it can help therapists recognize and foster unique opportunities for client change as they take place during psychotherapy. Using Greenberg and Pinsof's (1986) words, this type of process research "speaks to clinicians because it ultimately is dedicated to inform them how to become more effective agents of change" (p. 18). In addition, this type of research is also consistent with how clinicians think about and conduct therapy. Recognizing and understanding significant therapy events is indeed part of clinicians' cognitive activities (Elliott, 1983). On both conceptual and epistemological grounds, the investigation of clinically significant events was, thus, perfectly suited to the purposes guiding our psychotherapy PRN.

The present report focuses on five related questions:

1. What helpful and hindering events are reported by clients during psychotherapy sessions conducted within the context of outpatient independent practice?

2. What helpful and hindering events are reported by therapists during the same sessions?

3. What is the content or focus of therapeutic discussions between clients and therapist within these helpful and hindering events?

4. How helpful or hindering were these events and the contents of discussion, as reported by clients and therapists?

5. How similar are clients and therapists in their report of significant events (and the level of helpfulness or hindering of these events)?

In addressing these questions, this study was an attempt to replicate and extend a preliminary study conducted by Llewelyn (1988), who investigated the first two with sessions obtained from 40 therapist-client dyads in a naturalistic setting. Based on Llewelyn's (1988) findings, we predicted that the most frequently reported helpful events by clients would be (in descending order) the experience of both relief and a positive view of self, the resolution of specific problems, and the acquisition of insight. Also based on

Llewelyn's (1988) results, we predicted that the helpful events most frequently reported by therapists would be the facilitation of insight, the resolution of a problem, and the increase of client relief and a positive view of self. Primarily because of their low frequency in Llewelyn's (1988) study, however, we did not make specific predictions with regard to hindering events. Although we based our hypotheses on Llewelyn's (1988) findings, we were also interested in comparing our results with Llewelyn et al. (1988), who measured clients' perceptions of helpful and hindering events, not in a practice setting but in an RCT (Sheffield I Project; Shapiro & Firth, 1987) of cognitive–behavioral and psychodynamic-interpersonal therapies for depression.

Replication, as one of the goals of this study, provides a crucial safeguard for the validity of empirical knowledge. Although a number of constructs (e.g., alliance, experiencing) have been investigated by several researchers, the field is still replete with what Greenberg and Pinsof (1986) denounced 20 years ago, that is, ". . .one shot attempts to study process." As such, it is noteworthy that the present study, which was generated, designed, and conducted by full-time clinicians with the methodological assistance of clinical researchers, reflects one of the most important features of scientific research, namely, replication. Going beyond an exact replication, however, this study further attempted to extend Llewelyn's (1988) study by investigating the content of the helpful and hindering events, as well as their level of helpfulness or unhelpfulness.

Method

Participants

Thirteen therapists in independent practice participated in the design and implementation of this study. Seven therapists were female, and all were Caucasian. All participating therapists were doctoral level psychologists, with a mean of 17.5 years (range = 2–32 years) of posttraining experience. The majority of participating therapists (84.62%) reported two or more salient theoretical orientations, or approaches to psychotherapy. Six therapists (46.15%) identified cognitive–behavioral therapy (CBT) as their primary orientation, whereas four (approximately 30.77%) identified themselves primarily as psychodynamic. Of the remaining 23.08% of the sample,

one therapist identified primarily with humanistic therapy, one with constructivist therapy, and one with a family systems approach. For therapists who reported identifying with more than one orientation, three (27.27%) ranked CBT as their second most preferred theoretical orientation, and three (27.27%) ranked humanistic as their second most preferred. Of the remaining 45.46%, five therapists ranked cognitive, behavioral, interpersonal, dynamic, and family systems therapies, respectively, as their second most preferred orientation.

The current study involved 121 clients with a mean age of 32.63 years ($SD = 14.25$). Of these clients, 76.1% were female and the majority of the sample was Caucasian (88%), with less than 3% each of Hispanic, African American, Asian, American Indian/Alaska Native, East Indian, or other racial background. In terms of marital status, 40.8% were single, 43.9% were married, 10.2% were divorced, 3.1% separated, and 2% widowed. The mean number of years of education was 13.3 ($SD = 4.54$). Primary diagnoses, as provided by therapists, were as follows: adjustment disorders (40.5%), mood disorders (24.05%), anxiety disorders (18.99%), eating disorders (5.06%), disruptive/conduct problems (5.06%), sexual disorders (1.27%), sleep disorders (1.27%), psychotic spectrum disorders (1.27%), and deferred/unspecified (2.53%). Twenty-three adolescents or adults declined to participate in the study, and 11 others were not asked to participate because their clinical state at the first interview (e.g., risk for suicide) was judged to be contraindicated by the therapists. The mean number of clients in the study per therapist was 9.31 (range = 2 to 24), and the number of sessions per client ranged from 1 to 36 ($M = 8.04$; $SD = 8.30$).

Procedure

For a period of 18-months, all new clients in each therapist's independent practice were invited to participate in the study (unless therapists felt that such an invitation would be contraindicated by the client's clinical state). Clients were invited at the end of their first session, and those who agreed to participate in this study (and signed an informed consent approved by the Penn State University Office of Regulatory Compliance) were randomly assigned within each of three age groups (child, adolescent, adult) to either an experimental or a control condition. In order to compare our results to Llewelyn (1988; which included clients between the ages of 15 and 60 years old), the present study focused only on adolescent and adult clients (12 years-old and older).

In the experimental condition, both clients and therapists filled out (on an index card) parts of the Helpful Aspects of Therapy (HAT; Llewelyn, 1988), which asked them to report, describe, and indicate the level of helpfulness or unhelpfulness of particularly helpful and hindering events from the session they had just completed. Both clients and therapists did so at the end of every therapy session. Also, in the experimental condition, the therapist read the client's HAT card (after the client left the office and before the next session), but he or she was not instructed to follow any kind of protocol to address the helpful or hindering events mentioned by the client. Whether or not and how the therapist decided to respond to clients' HAT feedback was left to each therapist's judgment. Clients in the experimental condition were informed that their therapist would read the completed HAT before the next session. In the control condition only therapists were asked to fill out the HAT.[1]

---

[1] This experimental design was used because a second goal of our study was to investigate whether therapists' knowledge of clients' perceptions of both helpful and hindering events could have a causal impact on the effectiveness of therapy. In this context, the control condition was intended to control for the potential impact of therapists' increased attention to significant (positive or negative) events during therapy that could account for the anticipated effect of receiving feedback from clients about such events. However, our inability to secure sufficient posttreatment outcome data prevented us from addressing this question. Specifically, although participating clients filled out the Treatment Outcome Package battery (TOP, Kraus, Seligman, & Jordan, 2005), at pretreatment, only 31% (38 out of 121) of the participants in this study completed a TOP at the end of the treatment. This low response rate took place despite the fact that our study protocol included procedures to address anticipated difficulties with collecting postassessment data from clients in private practice. The study protocol required therapists to give clients a copy of the posttreatment TOP (along with a preaddressed and stamped envelope) at one of the first three sessions of therapy, along with instructions for the client to bring the TOP

Instruments

Helpful Aspects of Therapy (HAT). The HAT was developed by Llewelyn (1988) to provide descriptions of helpful and nonhelpful events in psychotherapy sessions. In the current study, we used two questions from the HAT, slightly modified by Elliott from its original version (see Elliott et al., 2001): "Did anything particularly helpful happen during this session?"; "Did anything happen during this session which might have been hindering?" When participants answered "Yes" to either of these questions, they were asked to briefly describe the event (s) and then to indicate on a scale from 1 to 4 the level of helpfulness (1 = *slightly helpful*; 4 = *extremely helpful*) or level of hindering (1 = *slightly hindering*; 4 = *extremely hindering*).[2]

Helpful Aspects of Experiential Therapy Content Analysis System (HAETCAS). The events described by participants were coded on two domains of Elliott's (1988) HAETCAS, which was designed to rate participant responses on the HAT. The first domain includes 13 helpful and 5 hindering "Impact" categories (henceforth referred to as "Event" categories). These categories were slightly modified by Elliott from his original set of trans-theoretical categories, called the Therapeutic Impacts Content Analysis System (TICAS; Elliott, James, Reimschuessel, Cislo, & Sack, 1985). In the revised version of the TICAS used for the present study, some items were divided to increase coding specificity (e.g., *insight* was divided into *self-insight* and *other-insight*; *reassurance/ relief* was divided into *relief* and *positive self*), others were aggregated into one global category (*understood*, *involvement*, and *personal contact* were combined into a new category called *alliance strengthening*), some were added to capture specific aspects of process-experiential therapy (*self-metaperception*, *other-metaperception*), and a number of hindering event categories were reformu-

lated, in part to increase reliability (see Table 1 for a complete list of the coding categories and descriptions).

Participant descriptions of helpful and hindering events were also coded on the "Content" domain of the HAETCAS. This domain includes seven categories aimed at capturing "what the event was about" (Elliott, 1988, pp. 1; see Table 1). We did not rate the HAT on the "Action/Technique" domain of the HAETCAS, as these categories measure behaviors or techniques that are specific to experiential therapy and, thus, did not permit the coding of events that can be expected to take place in a variety of theoretical orientations.

In previous studies, most of the event categories showed acceptable reliability. In Llewelyn et al. (1988), for example, only two helpful event (*problem clarification* and *involvement*) and two hindering event (*misperception* and *negative therapist reaction*) categories had a reliability (*alpha*) below .60 (however, all of these categories but *problem clarification* were reformulated in the revised version of the TICAS used in the present study). Reliability estimates for the content categories have not been published, but acceptable levels have been previously obtained (Elliott, personal communication, July 9, 2005).

All reported events were coded for each of the helpful and hindering event and content categories, which were developed as a set of overlapping scales (Elliott, 1988; Elliott et al., 1985). As in Llewelyn et al.'s (1988) study, each category was coded using a 4-point confidence rating scale that ranged from 0 (*clearly absent*) to 3 (*clearly or strongly present*)[3] (Elliott, 1988). It should be noted that in Llewelyn (1988), events were coded while considering the TICAS categories as mutually exclusive, and they were merely coded as present or absent. Thus, although we based our predictions on Llewelyn's (1988) study because

to the last session (if the termination session was known) and complete it immediately after the session, or complete it at home and mail it in. The research protocol also involved therapists calling clients who stopped coming to therapy to inquire whether or not they intended to terminate treatment and to asking them to fill out the TOP. The lack of success of these strategies has led us to build in financial incentives for returning the posttreatment assessment in our next study, thereby incorporating into naturalistic research a method of data collection frequently used in controlled studies.

---

[2] We modified the 5-point scale in Elliott's version of the HAT by removing one item (i.e., 0 = *neutral*) because it appeared to be inconsistent with the rating of an event described as particularly helpful or hindering.

[3] In Llewelyn et al. (1988), the coders' scores on the 0-3 rating scale for each event and event content category were transformed before analyses were conducted (.00, .33, .67, 1.0, respectively) "to yield scores comparable to proportions" (p. 108). The analyses (in terms of the reliability estimates and inferential statistics) for this paper were conducted with both the transformed and untransformed data, and led to identical findings. The results in the tables are reported with nontransformed data.

TABLE 1. Helpful Aspects of Experiential Therapy Content Analysis System (HAETCAS)

Category	Definition
	Helpful events
Self-insight	Client understands self (feelings, behaviors) better by seeing reasons, causes, connections, or parallels involving feelings or behavior.
Other-insight	Client understands another person better by seeing connections, causes or reasons for their behavior or experiences.
Self-awareness	Client more in touch with or clearer about presence or nature of personal feelings, own behaviors, physical states or perceptions of self.
Other awareness	Client becomes more aware of other's feelings or behaviors.
Positive self	Client comes to feel or think more positively and/or less negatively about self.
Positive other	Client comes to feel or think more positively and/or less negatively about specific or general other.
Self metaperception	Client comes to see self from another person's perspective.
Other metaperception	Client comes to see how a specific other views people or things other than client's self.
Problem clarification	Client identifies or becomes clearer about what his/her problems are or what s/he wants or wants to change, including tasks for therapy or in general.
Problem solution	Client figures out (realizes, comes closer to knowing) HOW to resolve a specific problem or achieve a specific goal or task.
Alliance strengthening	Client feels 1) understood, 2) supported, encouraged or reassured, 3) more involved or invested in therapy or its tasks (feels more able or freer to enter into therapeutic relationship), 4) closer to or better about therapist.
Relief	Client feels less negative: relieved, unburdened, relaxed, less depressed or hurt; or more positive: relaxed, safe, or confident or hopeful.
Other specific helpful	Impacts not described above.
	Hindering events
Unwanted thoughts	Client feels discomfort or pain as a result of being forced or stimulated to confront unpleasant experiences; client wants to withdraw from feelings or other experiences.
Therapist omission	Client describes therapist as failing to provide client with structure or with sufficient emotional support. Client experiences the absence of some desired therapist action.
Digression	Client describes straying or being deflected from important topics or tasks.
Poor fit	Client describes therapist trying something which doesn't fit the client's experience, which doesn't work, or which the client feels unprepared to deal with.
Other hindering	Hindering impacts which do not fit above categories.
	Content
Self only	Event exclusively involves feelings, wants, thoughts or actions of client, described as an object of experience.
Family of origin	Event involves client's parents or siblings, past or present; also childhood.
Marital family	Event involves client's spouse/ex-spouse/current lover or children/stepchildren.
Work	Event involves client's job or work situation; including career, school, and associated relationships.
Other relationships	Event involves other specific relationships (e.g., non-work friends, former lovers) or interpersonal issues in general.
Therapy	Therapist, therapy process or techniques or therapeutic relationship described as central to event.
Other specific content	Event involves types of content not included above.

Note. HAETCAS helpful and hindering event and event content coding categories and descriptions (Elliott, 1988).

it was conducted in a similar setting, it also appeared important to contrast our findings with those obtained by Llewelyn et al. (1988) who used a different setting (controlled trial of manualized therapies) but the same coding methods.

Each of three coders (two advanced graduate students in clinical psychology, and one individual with an undergraduate degree in psychology), who were unaware of the study predictions, coded all events. As in Llewelyn et al. (1988), we computed average ratings across the three coders. The coders were trained over a period of 7 months, meeting approximately 20 times for approximately one hour per meeting. The training involved reading and discussing Elliott's (1988) coding manual, coding events (provided by Elliott from previous studies or created for the sake of training) between meetings, calcu-

lating reliability estimates based on various waves of coding, and discussing coding discrepancies based on reliability estimates.

The length of training was longer than that in previous studies (e.g., six weeks in Llewelyn et al., 1988). The decision for a more extensive training period was based on the fact that less than optimal reliability was obtained for some categories (e.g., *problem clarification*) in previous investigations (Elliott et al., 1985; Llewelyn et al., 1988). The coding of HAT events collected for the study took place within a period of 18 months, with weekly meetings to avoid coder drift.

Sets of HAT responses (both therapist and client) were randomly assigned to all three coders. In other words, HAT data were assigned one client (or therapist) at a time, with all of the HAT responses given in the order in which they were collected during therapy (session by session). This was done to provide as much context as possible for the coding of each specific event, without contaminating the coding of the therapy events of one participant by the coding of the other. Also to provide appropriate context, coders were informed about whether therapy events were reported by therapists or clients and whether the events were reported as helpful or hindering. Before sets of HAT events were provided to coders, each description provided by participants was carefully reviewed by a research assistant to determine if it contained more than one event. When this was the case, the events were separated, allowing coders to rate each event reported for each session.

A total of 1,474 separate events (1,046 from therapists, 428 from clients) were coded (the larger number of events obtained from the therapists was due to the fact that they filled out HAT cards in both the experimental and control conditions included in the study design, whereas the clients filled them out only in the experimental condition, see *Procedure* section). Reliability was assessed with coefficient *alpha*. As shown in Table 2, acceptable to excellent reliability estimates were obtained for all event and event content categories, with the exception of *other metaperception*, which was almost never coded above a score of 0. Accordingly, this category was not included in our analyses.

Data Analytic Strategy

First, we conducted descriptive analyses to assess the mean ratings for each category of helpful and hindering events, as well as content, for both client-reported events and therapist reported events, respectively. Descriptive analyses were also conducted for participant-reported level of helpfulness (or hindering; i.e., How helpful or hindering was this event?). Second, we tested differences in HAETCAS category means using a one-way analysis of variance (ANOVA). Given that there were multiple data points per client and multiple clients within therapists, we needed to account for the multilevel structure of the data and possible dependency. To address this, we first examined intraclass correlations (ICCs) for three (multiple assessment points within clients within therapists) and two (multiple assessment points within clients) level models. Separate sets of ICCs were generated for client and therapist-reported events, as well as for all event categories (helpful and hindering) and content. Covariance parameter estimates and nonsignificant Wald statistics indicated relative independence of mean ratings at level three (i.e., there was a lack of similarity in mean ratings for clients seen by the same therapist). However, the assumption of nondependence was violated in the examination of level two variance estimates. Covariance parameter estimates and significant Wald statistics indicated that the variance at level two (similarity of assessment points from the same client) needed to be accounted for (e.g., 8% of the variance was attributable to the level two predictor in one of the derived ICCs, Wald = 3.44, $p < .01$). As such, we decided it would be most prudent to account for level two in subsequent analyses. Specifically, for the analysis determining whether event and event content category means could be differentiated, a series of two-level mixed-model ANOVAs were conducted. The category mean scores were treated as the dependent variable and the coding categories were treated as fixed effects, with random intercepts. A separate mixed-model analysis was conducted for therapist and client reported helpful and hindering events, as well as for the content of client and therapist reported helpful and hindering events.

Third, we conducted a series of logistic regression analyses to test for differences in level of helpfulness and hindering (how helpful or hindering the events were perceived to be by

TABLE 2. Mean Ratings and Coding Cronbach's Alpha Reliability

Category	Client ratings		Therapists ratings		Reliability
	M	*SD*	*M*	*SD*	*alpha*
Helpful events					
Self-insight	.275	.690	.191	.577	.886
Other-insight	.024	.194	.014	.128	.712
Self-awareness	.449	.719	.427	.755	.801
Other awareness	.117	.396	.065	.308	.798
Positive self	.156	.537	.249	.649	.914
Positive other	.023	.204	.020	.171	.819
Self-Metaperception	.018	.199	.004	.055	.873
Problem clarification	.312	.625	.288	.604	.770
Problem solution	.280	.624	.216	.577	.863
Alliance strengthening	.268	.640	.347	.774	.911
Relief	.134	.437	.084	.368	.851
Other specific helpful	.548	.945	.609	1.010	.809
Hindering events					
Unwanted thoughts	.011	.133	.024	.214	.889
Therapist omission	.015	.171	.082	.385	.873
Digression	.013	.149	.014	.133	.715
Poor fit	.044	.278	.044	.278	.838
Other hindering	.097	.476	.174	.631	.900
Content					
Self-only	.889	1.136	.675	1.035	.826
Family of origin	.284	.850	.308	.877	.978
Marital family	.196	.673	.273	.778	.924
Work	.029	.272	.045	.333	.929
Other relationships	.178	.569	.149	.517	.863
Therapy	1.172	1.150	1.460	1.190	.840
Other content	.021	.159	.004	.055	.347

Note. *M* = mean; *SD* = standard deviation. As in Llewelyn et al. (1988), average ratings were computed across the three coders. Each category was coded using a 4-point confidence rating scale that ranged from 0 (*clearly absent*) to 3 (*clearly or strongly present*).

the clients or the therapists) between different categories of events and their content. Since the dependent variable in this case was categorical (1 = *slightly helpful/hindering*, 2 = *moderately helpful/hindering*, 3 = *greatly helpful/hindering*, 4 = *extremely helpful/hindering*), we conducted the logistic regression analyses using SAS GENMOD, which fits generalized linear models. This is an extension of the general linear model that uses a link function to estimate probability (McCullagh & Nelder, 1989). This particular approach was chosen because it can account for correlated (clustered) data through the implementation of Generalized Estimating Equations (GEEs; Diggle, Liang, & Zeger, 1994) and the use of the REPEATED statement. This procedure utilizes Maximum Likelihood estimation (*mle*; McCulloch, 1997). We also specified a cumulative logit link function, which was chosen because the dependent variable is actually or-

dered (slightly helpful/hindering to extremely helpful/hindering) and we wanted to account for this ordering in the analysis.

Finally, in order to address the goal of examining the relationship between client and therapist reported events, we computed and tested rank-order coefficients using Kendall's tau-b statistic (τ^b), which does not penalize for ties. We conducted this analysis for both category ratings (of events and their content) and the level of reported helpfulness (or hindering) of events and their content.

Results

Helpful and Hindering Event Mean Ratings

Helpful events. As shown in Table 2, the three specific helpful event categories that had the highest mean ratings for the client reported events were *self-awareness*, *problem clarifica-*

tion, and *problem solution*.[4] Table 2 also indicates that the three helpful event categories with the highest mean ratings for the therapist reported events were *self-awareness*, *alliance strengthening*, and *problem clarification*. We conducted a series of mixed model ANOVAs to test if the coding category means could be statistically differentiated. Given the number of helpful event categories, we chose to focus on the top five categories for both client (*self-awareness*, *problem clarification*, *problem solution*, *self-insight*, and *alliance strengthening*) and therapist (*self-awareness*, *alliance strengthening*, *problem clarification*, *positive self*, and *problem solution*) reported events.[5]

The mixed model ANOVA for client reported events indicated a significant main effect for helpful event category, $F(4, 2102) = 5.681$, $p < .01$ (see Table 3 for estimated marginal means). All possible pairwise comparisons were examined, using the Bonferroni adjustment for multiple comparisons in order to help reduce the risk of Type I error. This was also the used in subsequent ANOVAs. Post hoc tests revealed that *self-awareness* ($M = .449$, $SE = .034$) occurred significantly more frequently than the other HAETCAS categories and was the only category that could be statistically differentiated from the other top categories (see Table 3 for comparisons).

The ANOVA for therapist reported events indicated a significant main effect for helpful event category, $F(4, 5154) = 16.159$, $p < .01$. Post hoc tests revealed that *self-awareness* ($M = .429$, $SE = .023$) was the most frequently coded category and it could be statistically differentiated from each of the remaining top categories, with the exception of *alliance strengthening* (see Table 3). *Alliance strengthening* was the category with the second highest mean ($M = .349$, $SE = .023$), and it could be statistically differentiated from *positive self* (fourth highest) and *problem solution* (fifth highest), but not *problem clarification* (third highest).

Hindering events. As predicted, the mean ratings for the hindering events coded from the client reported HATs were low. Of the four specific hindering categories, only *poor fit* had a mean rating above the three lowest helpful categories (*other insight*, *positive other*, and *self metaperception*). Similarly, the mean ratings for the therapist identified hindering events were very low, with the exception of *therapist omis-*

sion (which was still lower than most helpful categories).

The mixed model ANOVA conducted with client reported events revealed a significant main effect for hindering event category, $F(3, 1667) = 3.097$, $p < .05$. Although *poor fit* had the highest estimated mean ($M = .043$, $SE = .012$), post hoc tests revealed that it could not be statistically differentiated from the other hindering categories, although it approached significance when compared to *unwanted thoughts* ($M_{diff} = .033$, $SE = .012$, $p = .05$). For therapist reported events, there was a significant main effect for hindering event category, $F(3, 4103) = 13.045$, $p < .01$. Post hoc tests revealed that *therapist omission* ($M = .081$, $SE = .010$), the hindering category with the highest mean, was the only category that could be statistically differentiated from the others (see Table 3).

Content. As shown in Table 2, the highest mean ratings for the content of both client and therapist HAT cards were *therapy*, client *self*, and *family of origin* (closely followed by *marital family*). Whereas findings across events suggest that these types of content are significant foci of discussion in therapy, results should also be considered separately for helpful and hindering events in order to avoid lumping together themes that were discussed in good and bad moments of therapy (see Table 4 for separate means and standard deviations).

---

[4] As shown in Table 2, the most frequently coded category for both helpful and hindering events was *other*. One possible reason for these results is that we decided to code as *other* the events that only described the use of a specific technique (e.g., systematic desensitization), without a clear description or reference to the effect on the client. Because of their unclear conceptual and clinical relevance, the *other* categories (for helpful and hindering event categories, as well as content) were not included in inferential analyses and not addressed in the discussion of the results.

[5] We chose this strategy to increase our power to detect differences between the categories with the highest mean ratings. We considered reanalyzing the data with the inclusion of more categories, had differences been found between the last categories included in the analyses conducted. We judged that this reanalysis would be of little value if the third highest category was not statistically different from the fourth, or the fourth category was not statistically different from the fifth. We used the same strategy when conducting regression analyses on the participant-reported level of helpfulness or hindering of events (i.e., included only the top-five ranked categories in each regression).

TABLE 3. HAETCAS Event Category Mean Rating Comparisons

Coding category	Comparison category	M	M_{diff}	SE
	Client helpful events			
Self-awareness	Problem clarification	.312	.137*	.045
	Problem solution	.280	.169**	.045
	Self-insight	.276	.174**	.045
	Alliance strengthening	.269	.181**	.045
	Therapist helpful events			
Self-awareness	Alliance strengthening	.349	.080	.029
	Problem clarification	.290	.139**	.029
	Positive self	.251	.178**	.029
	Problem solution	.218	.211**	.029
Alliance strengthening	Problem solution		.131**	.029
	Positive self		.098**	.029
	Therapist hindering events			
Therapist omission	Poor Fit	.044	.037**	.012
	Unwanted thoughts	.024	.057**	.012
	Digression	.014	.067**	.012
	Client content helpful events			
Therapy	Self-only	.933	.198*	.059
	Family of origin	.261	.870**	.059
	Other relationships	.185	.945**	.059
	Marital family	.180	.950**	.059
	Work	.031	1.099**	.059
Self-only	Family of origin		.672**	.059
	Other relationships		.747**	.059
	Marital family		.753**	.059
	Work		.902**	.059
Family of origin	Work		.230**	.059
	Therapist content helpful events			
Therapy	Self-only	.744	.595**	.040
	Family of origin	.319	1.019**	.040
	Marital family	.271	1.067**	.040
	Other relationships	.168	1.171**	.040
	Work	.047	1.292**	.040
Self-only	Family of origin		.425**	.040
	Marital family		.473**	.040
	Other relationships		.576**	.040
	Work		.697**	.040
Family of origin	Other relationships		.151**	.040
	Work		.272**	.040
Marital family	Work		.224*	.040
	Client content hindering events			
Therapy	Family of origin	.573	1.115**	.212
	Marital family	.385	1.302**	.212
	Self-only	.344	1.344**	.212
	Other relationships	.094	1.594**	.212
	Work	.000	1.687**	.059
	Therapist content hindering events			
Therapy	Marital family	.282	1.978**	.081
	Family of origin	.235	2.025**	.081
	Self-only	.216	2.044**	.081
	Work	.029	2.230**	.081
	Other relationships	.022	2.238**	.081

TABLE 3 (*continued*)

Coding category	Comparison category	M	M$_{diff}$	SE
Marital family	Work		.252*	.081
	Other relationships		.260*	.081

Note. HAETCAS = Helpful Aspects of Experiential Therapy Content Analysis System (Elliott, 1988). *M* = estimated marginal mean; M$_{diff}$ = mean difference; *SE* = standard error. All post hoc tests conducted with Bonferroni adjustment. * $p < .05$. ** $p < .01$.

Our analysis of client reported helpful events revealed a significant main effect for content category, $F(5, 2370) = 121.044$, $p < .01$. Post hoc tests indicated that *therapy* ($M = 1.130$, $SE = .042$) could be statistically differentiated from each of the other content categories (see Table 3). *Self-only,* the category with the second highest mean, could be differentiated from the remaining content categories (*family of origin, marital family, work,* and *other relationships*). *Family of origin* ($M = .261$, $SE = .042$), could be statistically differentiated from *work*, the content category with the lowest mean. For therapist reported helpful events, a significant main effect for content category was found, $F(5, 5448) = 293.093$, $p < .01$. Post hoc tests revealed that *therapy* ($M = 1.339$, $SE = .028$) could be statistically differentiated from each of the other content categories. *Self-only,* the category with the second highest mean, could be differentiated from the remaining content categories. The category with the third highest mean, *family of origin,* could be statistically differentiated from *other relationships and work* (sixth highest). The category *marital family* could also be differentiated from *work* (see Table 3).

In the analysis of the content categories for client reported hindering events, a significant main effect for content category was found, $F(5, 186) = 16.557$, $p < .01$. Post hoc tests indicated that *therapy* ($M = 1.687$, $SE = .150$) could be statistically differentiated from each of the other content categories (see Table 3). For therapist reported hindering events, a significant main effect for content category was found, $F(5, 810) = 226.273$, $p < .01$. Post hoc tests revealed that *therapy* ($M = 2.260$, $SE = .058$) could be statistically differentiated from each of the other content categories. *Marital family,* the category with the second highest mean, could be differentiated from *other relationships* and *work* (the fifth and sixth highest, respectively; see Table 3).

Helpfulness and Hindering Levels

As described above, when clients and therapists reported events, they were also asked to rate these events on a 4-point scale (1 = *slightly*, 2 = *moderately*, 3 = *greatly*, and 4 = *extremely*), either in terms of level of helpfulness (for helpful events) or level of hindering (for hindering events). One goal of the study was to determine whether some observer-rated event categories were perceived as more helpful or hindering than others. To do so, we identified events for which one or

TABLE 4. Mean Content Ratings for Helpful and Hindering Events

	Mean ratings			
	Helpful events		Hindering events	
Category	Clients	Therapists	Clients	Therapists
Self-only	.933 (1.146)	.744 (1.067)	.344 (.827)	.216 (.628)
Family of origin	.261 (.821)	.319 (.890)	.573 (1.127)	.235 (.785)
Marital family	.180 (.647)	.271 (.780)	.385 (.932)	.282 (.772)
Work	.031 (.282)	.047 (.341)	.000	.029 (.271)
Other relationships	.185 (.576)	.168 (.549)	.094 (.473)	.022 (.163)
Therapy	1.131 (1.143)	1.339 (1.169)	1.688 (1.133)	2.260 (.998)
Other content	.016 (.146)	.005 (.059)	.083 (.268)	.000

Note. Standard deviations in parentheses. As in Llewelyn et al. (1988), average ratings were computed across the three coders. Each category was coded using a 4-point confidence rating scale that ranged from 0 (*clearly absent*) to 3 (*clearly or strongly present*).

more helpful, hindering, or content categories were coded as a 2 or a 3 on the HAETCAS's 4-point confidence rating scale (ranging from 0 *clearly absent* to 3 *clearly or strongly present*) by at least two of three coders. We then calculated the means of the helpfulness or hindering level (as rated by the clients or the therapists) for each category (see Tables 5 and 6).

We conducted a series of logistic regression analyses to test for differences in the level of helpfulness (or hindering) between different categories of events. Before proceeding with the analyses, however, some issues needed to be considered. For example, if too few cases exist relative to the number of variables (and levels within variables) or the expected frequencies are too small, the regression may produce abnormally large parameter estimates and standard errors and may compromise power (Tabachnick & Fidell, 2007). In order to test this, we used crosstabs to examine the expected frequencies before running each regression. Guidelines for interpretation include having at least 10 observations for each level of the categorical dependent variable. Additionally, all expected frequencies should be greater than 1 and no more than 20% should be less than 5 (Allison, 1999; Tabachnick & Fidell, 2007). If any of these conditions fail, the choices are to collapse categories for variables with more than two levels, or delete the offending category or discrete variable to reduce the number of cells. Based on these guidelines, we deleted *self-metaperception* from client reported helpful event category ratings, *other insight* from the therapist reported helpful event category ratings, and *digression* from therapist reported hindering category ratings. Additionally, we were forced to delete *work* from the analysis of the content of client reported category ratings for helpful events, as well as *work* and *other relationships* from the analysis of the content of therapist reported category ratings for hindering events.

For each of the regression analyses that we then conducted, the Type III test for the null hypothesis that all coefficients equate to zero was nonsignificant. These results indicated that the probability of a client (or therapist) rating an event (or content) as extremely helpful (or extremely hindering) was not statistically more or less likely given a particular category.[6]

Comparison of Client and Therapist Events

The comparison of the ranked means of client and therapist reported helpful event categories revealed a significant Kendall's tau-b correlation coefficient, $\tau^b = .84$, $p < .01$, indicating a significant relationship between client and therapist reported types of events. The comparison of the hindering event category rankings approached significance, $\tau^b = .74$, *ns*. For the content of helpful and hindering events, significant correlations were found, $\tau^b = .867$, $p < .05$ and $\tau^b = .733$, $p < .05$, respectively, indicating significant agreement between clients and therapist regarding the content of important events.

In the comparison of the perceived level of helpfulness of client and therapist reported helpful events, the correlation was nonsignificant ($\tau^b = -.12$). This was also the case for the level of hindering reported ($\tau^b = .40$). However, the relationship between the client and therapist reported level of helpfulness for the content of helpful events was significant $\tau^b = .60$, $p < .01$, as was the correlation for hindering events, $\tau^b = .60$, $p < .05$. This indicates significant agreement between clients and therapists regarding the level of helpfulness and hindering of different types of content in therapy.

Discussion

This paper presents the findings of what is, to our knowledge, the first process study in psychotherapy conducted within the context of a PRN. It is the product of a comprehensive and long-term collaborative effort, both in terms of the study design and implementation, between experienced clinicians of various theoretical orientations and full-time psychotherapy researchers. Based on a large sample of clients seen in outpatient independent practice settings, this study investigated close to 1,500 events described by clients and therapists as being particularly helpful or hinder-

---

[6] For the level of hindering in client reported hindering events, as well as the content of client reported hindering events, the cell frequencies were extremely low and a number of categories and levels of the dependent variable did not meet criteria for inclusion in the regression analyses. We judged that it would not be appropriate to conduct the regression analyses with this data, which totaled 31 observations for hindering items and 39 observations for the content of hindering items.

TABLE 5. Helpful and Hindering Level Means and Standard Deviations by Category

| | Mean helpful ratings | | | | Mean hindering ratings | | | |
| | Clients | | Therapists | | Clients | | Therapists | |
Category	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
	Helpful events							
Self-insight	3.11	.89	2.46	.96				
Other-insight	1.00	.00	2.57	.54				
Self-awareness	2.82	.82	2.26	.94				
Other awareness	2.94	.77	2.47	.97				
Positive self	2.79	.94	2.14	.93				
Positive other	2.50	.58	1.89	.78				
Self-metaperception	3.67	.58	1.50	.71				
Problem clarification	3.10	.83	2.32	.88				
Problem solution	3.23	.66	2.14	.85				
Alliance strengthening	2.98	.93	2.45	.90				
Relief	3.14	.79	2.27	.72				
Other specific helpful	2.77	.78	2.22	.84				
	Hindering events							
Unwanted thoughts					1.50	.71	1.38	.52
Therapist omission					2.00	1.73	1.43	.56
Digression					1.00	.00	1.00	.00
Poor Fit					2.00	1.00	1.42	.51
Other hindering					1.75	.68	1.60	.79

Note.   *M* = mean; *SD* = standard deviation. Event helpfulness/hindering level means are based on client and therapist ratings from 1 (*slightly helpful/hindering*) to 4 (*extremely helpful/hindering*).

ing in their psychotherapy. The examination of such events provides a direct window into what can (in the eyes of the therapy participants) facilitate or interfere with change, which in turn may lead to a better understanding and, ultimately, the improvement of psychotherapy. As such, this study addresses conceptual and clinical issues of high relevance. In addition, it is noteworthy that this investigation addressed an important issue of scientific knowledge that has received too little attention in psychotherapy process research (the crucial need for replication), while originating from a clinical interest. The idea for the study came during a PPA-PRN meeting designed to generate possible topics of investigation for the second phase of our collaborative effort when a clinician in our group voiced that she was most interested in learning from her clients, after each session, what they found helpful, as this might help her to be a better therapist. On this basis alone, this study is a testimony to how PRN initiatives can contribute to the growth

TABLE 6. Level of Helpfulness and Hindering for Content Categories

| | Mean helpfulness level | | | | Mean hindering level | | | |
| | Clients | | Therapists | | Clients | | Therapists | |
Content category	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Self-only	2.89	.87	2.33	.91	2.83	.75	1.56	.88
Family of origin	2.32	1.18	2.04	.95	2.38	1.06	1.36	.67
Marital family	3.06	.87	2.37	.86	2.00	.00	1.57	.76
Work	2.43	.79	1.85	.77	—		1.00	.00
Other relationships	3.06	.83	2.32	.98	1.00	.00	1.00	.00
Therapy	2.87	.79	2.28	.85	1.83	.94	1.42	.62
Other content	2.50	.71	—		—		—	

Note.   *M* = mean; *SD* = standard deviation. Level of helpfulness/hindering means are based on client and therapist ratings from 1 (*slightly helpful/hindering*) to 4 (*extremely helpful/hindering*).

(or at least the survival) of the scientific-practitioner model that is assumed by many to guide modern psychotherapy.

Our attempt at replication, for the most part, did not lead to predicted results. In Llewelyn's (1988) study, the three most frequent helpful event categories for client reported events were the experience of *relief* and *positive view of self* (aggregated into one category in Llewelyn's [1988] study but divided in two distinct categories in the revised TICAS used in the present study), *resolution of a specific problem*, and *insight*. In contrast, the three highest categories in the current study were *self-awareness*, *problem clarification*, and *problem solution*. As such, only one of the three predictions that we made based on Llewelyn's (1988) findings regarding client perceived helpful events was supported, that is, high rating for *problem solution*. Although, the *resolution of problems* also emerged as one of the three client reported helpful events with the highest ratings in the Sheffield controlled trial (Llewelyn et al., 1988), it failed to be significantly different from other client reported helpful events in the present study. In contrast with both Llewelyn (1988) and Llewelyn et al., (1988), however, the present study went beyond a listing and descriptive comparison of the frequency or mean ratings of helpful (and hindering) categories, and conducted inferential statistics to determine if these categories could be statistically differentiated. What our findings show is that only *self-awareness* had significantly higher mean ratings than the other client reported helpful events.

With regard to therapist perceived helpful events, none of our predictions were supported. Whereas the most frequently coded categories in Llewelyn (1988) were the facilitation of *insight*, *resolution of a specific problem*, and *reassurance/relief* (divided into *relief* and *positive self* in the present study), the categories with the highest mean ratings in the present study were *self-awareness*, *alliance strengthening*, and *problem clarification*. It was interesting that our statistical analyses indicated that *self-awareness* had significantly higher ratings than all other therapist reported events, with the exception of *alliance strengthening* (which in turn was significantly different from the next highest rated helpful events, with the exception of *problem clarification*).

Consistent with Llewelyn's (1988) study, hindering events were coded infrequently for both client and therapist reported events. However, the inferential analyses we conducted revealed that *therapist omission* on therapist reported cards had statistically higher mean ratings than the other therapist reported hindering events.

The general lack of consistency between Llewelyn (1988) and the present study could be due to a number of reasons. With regard to the therapist reported events, for example, the discrepancies observed may in part reflect different processes or interventions that are representing similar therapeutic intentions. In both studies, therapists frequently reported events that represent the fostering of exploration (self—awareness in this study and insight in Llewelyn's), provision of support (alliance strengthening in this study and reassurance in Llewelyn's) and a focus on specific problems (problem clarification in this study and problem resolution in Llewelyn's). This, as well as other interpretations, can in turn be based on or reflect several therapist differences, such as culture (United States vs. United Kingdom), professional background (only doctoral psychologists vs. a variety of mental health providers), and level of experience (mean of 17.5 years posttraining vs. mean of 7 years). While any attempt to explain the discrepancies between these studies should be highly tentative, it could be argued that our findings might be viewed with more confidence, not only because they are based on a larger sample, but because they have been derived from statistical, rather than purely descriptive analyses. However, to be consistent with what we have previously described as an imperative mission of science, replications of these statistical findings are required before further confidence in their reliability can be firmly established.

With this caveat in mind, what our results regarding self awareness indicate is that providing clients with opportunities to achieve a clearer sense of their experience (e.g., emotions, behaviors, and perceptions of self) is frequently reported as beneficial—by both clients and therapists. One client, for example, wrote that it "Helped to talk about what fears and thoughts are normal and how I am starting to feel better," while one of the therapists described the following event as helpful: "Exploring painful events and allowing her to experience and express feelings of sadness and shame." It may be that experiencing, symbolizing, naming or making more explicit some aspects of self may help clients to

obtain a greater sense of meaning (a new understanding of who they are, what they want, or what they need), purpose (fostering "a new way of perceiving and engaging the self, world, or other," Pascual-Leone & Greenberg, [2006]), p. 39), control and/or mastery (Frank, 1961).

Therapists also frequently identified as helpful those events that reflected and/or enhanced the development of the bond and collaboration between them and their clients. For example, one of the therapists noted that the client "opened up-expressed deep emotions. Seems to trust me a lot." High ratings for such events seem to indicate that in the eyes of therapists of different orientations, the relationship matters (as it may help clients engage in treatment and/or may provide, in and of itself, corrective experiences). Such a therapist perception of helpful events in psychotherapy is in line with the findings of process-outcome studies showing a robust correlation between alliance and improvement (Constantino, Castonguay, & Schut, 2002). Perhaps reflecting and/or contributing to alliance ruptures, the events that therapists most frequently reported as detrimental were those when they failed to be attuned to their clients needs (to adequately provide clients with needed structure, emotional support, or desired therapist action). It is interesting that it seems that when things went bad, therapists appeared not to blame the client, but instead to focus on what they had done (or, more accurately, failed to do). In one of these hindering events, for example, the therapist admitted that the client was initially upset because the session began 10 minutes late.

In addition to investigating types of helpful and hindering events, this study was also aimed at identifying the content or focus of these events. For both client and therapist helpful events, issues related to *therapy* had the highest mean ratings. For example, one client described the following as helpful "Sharing of responsibility of change. Honesty. Realness." Also emphasizing the positive impact of a therapeutic bond, one of the therapists reported this helpful event "We seemed to form a very close bond almost immediately." It is interesting that issues pertaining to *therapy* also showed the highest mean rating on both client and therapist reported hindering events. Among the hindering events reported, one client wrote sometimes feeling "as if I am under attack, at which point I tend to withdraw," and one therapist stated that her client's "Frustration

might be showing." Thus, as with many powerful tools, the relationship may well be a double edged sword: It might foster healing, yet it can also hurt, if not harm our clients (see Castonguay, Boswell, Constantino, Goldfried, & Hill, 2010).

Reflecting the importance of focusing on issues directly relevant to the client's needs, problems, resources, and/or progress, the content category of *self* had the second highest mean rating for both client (e.g., "Try to get an apartment on my own.") and therapist ("Identification of process of perfectionism.") reported helpful events. A focus on the client's *family of origin* also frequently appeared to be helpful for both client and therapists. As an example, one client described as helpful the fact that he "Talked about my dad and my needs to open up more and not keep my feelings hidden." Also described by a therapist as particularly helpful was a client revealing issues about her "mother's eating / body problems—something I never told anyone before." As noted by several scholars (e.g., Safran & Segal, 1990), one's core view of self is most likely to be tied to his or her relationship with early significant others. As such, exploring or examining thoughts and feeling about these important people in one's life and/or exploring and changing relationship patterns with them may be experienced as particularly significant and helpful.

It is interesting that *marital* and *family of origin* relationships evidenced particularly high mean ratings on therapist reported hindering events. In one of these events, for example, the therapist wrote "Patient became very frightful at the reality of possible divorce." This suggests that such meaningful issues should be addressed with optimal attunement to the client needs, as they can be perceived as painful or unhelpful topics (Grosse Holtforth & Castonguay, 2005). This is not to say that focusing on marital family and family of origin relationships is detrimental per se. As described elsewhere (Castonguay, Boswell, et al., 2010), unhelpful effects in therapy are not likely to be restricted to the use of what are now know as potentially harmful treatments (Lilienfield, 2007). Unhelpful effects may also be due to the less than optimal (in terms of timing and tact, for instance) use of interventions that may otherwise be helpful. The fact that the therapist detected, in the example above, the client's pain in relation to his potential divorce, suggests that she might be aware of the harm (and/or risk

of further toxic impact) of a premature, pertinacious or otherwise insensitive and inflexible way of addressing this important issue in client's life. In addition, what may be perceived as unhelpful at a specific time in therapy could ultimately end up being of therapeutic value later. To better understand what to do and not to do in therapy, however, qualitative (narrative) analyses may need to be conducted to explore what specific issues related to therapy and past and current family relationships were discussed (and both how and when they were discussed) during helpful and hindering events.

As recommended by Llewelyn et al. (1988), our study not only investigated the types of events clients and therapists find helpful and hindering, but also how helpful or unhelpful such events were perceived to be. However, the analyses conducted showed that not one single category had a significantly greater level of helpfulness or hindering than another. A possible explanation for this lack of finding is that what counts is whether or not a significant event, in the eyes of a client or therapist, has happened; the extent to which a given event is perceived as helpful or hindering (from *slightly* to *moderately*, e.g.) might be less meaningful than the fact that something important took place.

It is interesting that in assessing the agreement between client and therapists, the level of helpfulness helpful (or hindering) of events were the only comparisons that failed to approach or meet statistical significance. All other comparisons showed either a trend toward or indicated a general agreement between clients and therapists, and this included our examination of the content of significant events. Thus, it appears that although clients and therapists demonstrated general agreement regarding the relative occurrence of different types of significant events, there is less agreement on the perceived level of helpfulness or hindering of such events.

By identifying significant events experienced by clients and therapists during each therapy session, the use of process measures such as the HAT provides information that does not typically emerge from controlled trials. However, it is important to emphasize that outcome and process studies are both viable strategies for the advancement of knowledge. Whereas the former predominantly (though not always; see Borkovec & Castonguay, 1998) attempts to answer the question of what works, the latter focuses on how it works or

fails to work. In addition to not being incompatible (the fact that Llewelyn et al.'s, [1988] study was conducted within the Sheffield I controlled trial is a case in point), we would argue that the complementary answers they provide should lead to synergic advancements in the field. Specifically, as argued by Grawe (1997), we believe that outcome researchers should use process findings to improve the impact of effective interventions. Based on the findings of this study, for example, potential ways to improve many current forms of CBT could be to add interventions to foster clients' awareness of their emotional experience, enhance attention to and careful management of the therapeutic relationship, and increase therapists' focus on clients' family of origin. Examples of the development of new and integrative therapies based on process findings can be found elsewhere (Castonguay et al., 2004; Newman, Castonguay, Borkovec, & Molnar, 2004).

Perhaps one advantage that studies using such a tool as the HAT (and other methodologies under the umbrella of the "events paradigm") have over RCTs is that they may more readily provide ideographic as well as nomothetic information to the investigators, clinicians and researchers. Whereas the aggregate data provided by the current study inform us about potential processes of change across clients, the data collected and examined after each session provided therapists with information specific to the particular needs of each client. These reliable and repeated data can also be of great interest to researchers. For example, future research could investigate helpful and hindering events across the entire treatment of a small number of successful and nonsuccessful cases to determine if such differential outcomes are associated with specific developmental pathways or patterns of therapeutic events.

Perhaps even more importantly, this type of study is likely to foster the simultaneous integration of clinical and empirical tasks. It may well be that the full actualization of the scientific—practitioner model takes place when therapists conduct a task and do not know if they are doing research or clinical work when doing so. As described in the adjacent paper (Castonguay, Nelson, et al., this issue), one of the experiences shared by the participating therapists was that they felt that when they were filling out their HAT cards and reading their clients' responses, they were also preparing their session notes. This

raises an interesting question: In the precise moment when they were reading and writing the HAT cards, were the therapists collecting data or thinking about their evolving case formulation and treatment plan? The fact that the obvious answer to this question is "both" might explain why therapists eagerly implemented the study protocol (including the randomization and daily assessment of all their private clients) as part of their independent practice. The clinical relevance and feasibility, as well as the scientific rigor of such studies, provide hope for the viability of the Boulder model.

A number of this study's limitations need to be mentioned. First, the therapeutic events investigated were collected using a self-report measure. It is, therefore, possible that the observation of therapy sessions by nonparticipant observers would have led to different results in terms of the type, frequency, and/or rating of helpful and/or hindering events. It is also possible that our results were influenced by the fact that the therapists looked at the HAT responses of their clients after each session and that these clients were aware of this. As such, this idiosyncratic aspect of our design may place limits on the generalizability of this study's findings.

It is also important to note that the results no doubt reflect the conceptual lens (and, thus, biases) inherent to the instrument that we used to code the events reported by clients and therapists. Although the TICAS has been described as "an eclectic framework for analyzing the change process" (see Elliott et al., 1985, p. 628), it was created by a leading researcher identified with the humanistic/experiential orientation. Thus, it is likely that processes central to some forms of therapy (e.g., cognitive–behavioral, systemic) may not be captured as fully or adequately as other therapeutic factors by this instrument. One possible way of addressing this limitation would be to use different systems to code the events collected in this study. For example, an instrument assessing the use of techniques prescribed by different theoretical orientations might not only address this issue, but also reduce the number of events that were coded in this study as "other" helpful or hindering.

Finally, although this does not pertain to the goal of the present article, it should be reiterated that our difficulty in collecting posttreatment outcome data prevented us from testing the potential impact of receiving feedback from clients about helpful and hindering events in therapy (see Footnote 1). As argued elsewhere (Borkovec & Castonguay, 1998), experimental investigations conducted with large samples and within naturalistic settings are likely to provide optimal conditions for internally and externally valid studies. When designed by experienced clinicians and researchers, such investigations can simultaneously address the ultimate goal of empirical science (the discovery of cause and effect relationships) and answer clinically relevant questions. While the successful randomization of a large number of patients into two conditions demonstrated that such an experimental study can be conducted as part of day-to-day practice, our difficulty in collecting posttreatment outcome measures illustrates one of the major obstacles faced by this type of research. Fortunately, it also provides us with information that we can use to improve future PRN studies (such as providing financial incentives for filling out outcome measures after the end of treatment), or provide assistance to clinicians in a busy practice by monitoring and following up on client unilateral-termination.

Above and beyond its limitations, however, the present study appears to offer three primary contributions. First, it provides knowledge about the process of change, especially with regard to the importance, in the eyes of clients and therapists, of fostering awareness and of carefully paying attention to the therapeutic alliance and other significant interpersonal relationships in the client's life. It also illustrates how the process of data collection can be immediately relevant to clinical work (as a tool to guide and attune therapeutic interventions), thereby demonstrating the possibility of a simultaneous and seamless integration of science and practice. Third, its provides evidence of the feasibility of a long-term collaboration between clinicians and researchers in deciding what to investigate and how it should be studied, as well as in implementing the study and identifying problems to improve upon future efforts.

References

ALLISON, P. D. (1999). *Logistic regression using SAS: Theory and application.* Cary, NC: SAS Institute and Wiley.

ARNKOFF, D. B., VICTOR, B. J., & GLASS, C. R. (1993). Empirical research on factors in psychotherapeutic change. In G. Stricker & J. R. Gold (Eds.), *Compre-*

hensive handbook of psychotherapy integration (pp. 27–42). New York: Plenum Press.

BORKOVEC, T. D., & CASTONGUAY, L. G. (1998). What is the scientific meaning of "empirically supported therapy"? *Journal of Consulting and Clinical Psychology, 66,* 136–142.

BORKOVEC, T. D., ECHEMENDIA, R. J., RAGUSEA, S. A., & RUIZ, M. (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8,* 155–167.

CASTONGUAY, L. G., BOSWELL, J. F., CONSTANTINO, M. J., GOLDFRIED, M. R., & HILL, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist, 65,* 34–49.

CASTONGUAY, L. G., NELSON, D., BOUTSELIS, M., CHISWICK, N., DAMER, D., HEMMELSTEIN, N., . . . BORKOVEC, T. (2010). Psychotherapists, researchers, or both? A qualitative analysis of therapists' experiences in a practice research network. *Psychotherapy: Theory, Research, Practice, Training, 47,* 345–354.

CASTONGUAY, L. G., SCHUT, A. J., AIKINS, D., CONSTANTINO, M. J., LAURENCEAU, J. P., BOLOGH, L., & BURNS, D. D. (2004). Integrative cognitive therapy: A preliminary investigation. *Journal of Psychotherapy Integration, 14,* 4–20.

CONSTANTINO, M. J., CASTONGUAY, L. G., & SCHUT, A. J. (2002). The working alliance: A flagship for the scientist-practitioner model in psychotherapy. In G. Shick Tryon (Ed.), *Counseling based on process research* (pp. 81–131). New York: Allyn & Bacon.

DIGGLE, P. J., LIANG, K. Y., & ZEGER, S. L. (1994). *The analysis of longitudinal data.* New York: Oxford University Press.

ELLIOTT, R. (1983). Fitting process research to the practicing psychotherapist. *Psychotherapy: Theory, Research and Practice, 20,* 47–55.

ELLIOTT, R. (1988). *Helpful aspects of experiential therapy content analysis system (HAETCAS): Rating manual.* Unpublished manuscript. Department of Psychology, University of Toledo.

ELLIOTT, R., JAMES, E., REIMSCHUESSEL, C., CISLO, D., & SACK, N. (1985). Significant events and the analysis of immediate therapeutic impacts. *Psychotherapy: Theory, Research, Practice, and Training, 22,* 620–630.

ELLIOTT, R., SLATICK, E., & URMAN, M. (2001). Qualitative change process research on psychotherapy: Alternative strategies. In J. Frommer & D. L. Rennie (Eds.), *Qualitative psychotherapy research: Methods and methodology* (pp. 69–111). Lengerich, Germany: Pabst Science Publishers.

FRANK, J. D. (1961). *Persuasion and healing.* Baltimore: Johns Hopkins University Press.

GRAWE, K. (1997). Research-informed psychotherapy. *Psychotherapy Research, 7,* 1–19.

GREENBERG, L. S., & PINSOF, W. M. (1986). Process research: Current trends and future perspectives. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 3–20). New York: Guilford Press.

GROSSE HOLTFORTH, M., & CASTONGUAY, L. G. (2005). Relationship and techniques in CBT – A motivational approach. *Psychotherapy: Theory, Research, Practice, Training, 42,* 443–455.

KRAUS, D. R., SELIGMAN, D., & JORDAN, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61,* 285–314.

LAMPROPOULOS, G. K., GOLDFRIED, M. R., CASTONGUAY, L. G., LAMBERT, M. J., STILES, W. B., & NESTOROS, J. N. (2002). What kind of research can we realistically expect from the practitioner? *Journal of Clinical Psychology, 58,* 1241–1264.

LILIENFELD, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2,* 53–70.

LLEWELYN, S. P. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology, 27,* 223–237.

LLEWELYN, S. P., ELLIOTT, R., SHAPIRO, D. A., HARDY, G., & FIRTH-COZENS, J. (1988). Client perceptions of significant events in prescriptive and exploratory periods of individual therapy. *British Journal of Clinical Psychology, 27,* 105–114.

MCCULLAGH, P., & NELDER, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

MCCULLOCH, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association, 92,* 162–170.

NEWMAN, M. G., CASTONGUAY, L. G., BORKOVEC, T. D., & MOLNAR, C. (2004). Integrative therapy for generalized anxiety disorder. In R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), *Generalized anxiety disorder: Advances in research and practice* (pp. 320–350). New York: Guilford Press.

PASCUAL-LEONE, A., & GREENBERG, L. S. (2006). Insight and awareness in experiential therapy. In L. G. Castonguay & C. E. Hill (Eds.), *Insight in psychotherapy* (pp. 31–56). Washington, DC: American Psychological Press.

RAIMY, V. C. (Ed.). (1950). *Training in clinical psychology.* Englewood Cliffs, NJ: Prentice Hall.

SAFRAN, J. D., & SEGAL, Z. V. (1990). *Interpersonal process in cognitive therapy.* Northvale, NJ: Jason Aronson.

SHAPIRO, D. A., & FIRTH, J. (1987). Prescriptive v. exploratory psychotherapy: Outcome of the Sheffield Psychotherapy Project. *British Journal of Psychiatry, 151,* 790–799.

TABACHNICK, B. G., & FIDELL, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Allyn & Bacon.