Psychological Assessment
2009, Vol. 21, No. 1, 79–88

1040-3590/09/$12.00 DOI: 10.1037/a0014643

Evaluating the Validity of Computerized Content Analysis Programs for Identification of Emotional Expression in Cancer Narratives

Erin O'Carroll Bantum
University of Hawaii at Manoa

Jason E. Owen
Loma Linda University

Psychological interventions provide linguistic data that are particularly useful for testing mechanisms of action and improving intervention methodologies. For this study, emotional expression in an Internet-based intervention for women with breast cancer ($n = 63$) was analyzed via rater coding and 2 computerized coding methods (Linguistic Inquiry and Word Count [LIWC] and Psychiatric Content Analysis and Diagnosis [PCAD]). Although the computerized coding methods captured most of the emotion identified by raters (LIWC sensitivity = .88; PCAD sensitivity = .83), both over-identified emotional expression (LIWC positive predictive value = .31; PCAD positive predictive value = .19). Correlational analyses suggested better convergent and discriminant validity for LIWC. The results highlight previously unrecognized deficiencies in commonly used computerized content-analysis programs and suggest potential modifications to both programs that could improve overall accuracy of automated identification of emotional expression. Although the authors recognize these limitations, they conclude that LIWC is superior to PCAD for rapid identification of emotional expression in text.

Keywords: linguistic analysis, emotion, cancer

The proliferation of expressive writing and web-based interventions, coupled with increasingly sophisticated speech-recognition technology, has added to the need for valid methods of rapidly analyzing text-based data for content relevant to therapeutic processes. These interventions provide researchers and clinicians with textual data that can be used objectively to measure emotional expression and supplement self-report measures. Extensive qualitative analysis has the potential to provide behavioral data relevant to understanding physical and psychological adjustment to a diagnosis and treatment for cancer. However, the time required to conduct a thorough and reliable qualitative analysis and to validate the results makes such analysis impractical for many potential applications. Computerized text-analysis programs exist, although they have not been well validated for the purpose of evaluating emotional expression in therapeutic discourse. The goal of the current study was to create a manual coding system that could be used to evaluate the validity of two widely used text-analysis programs for identification of emotional expression.

Emotional expression has been suggested as a target for psychological treatments, but the putative mechanism of action is poorly understood (Greenberg & Safran, 1989). Emotional expression may serve to help individuals acknowledge and synthesize emotions that were previously unavailable to conscious awareness, lead to habituation to intense emotions, access state-dependent core beliefs or modify maladaptive emotional responses (Greenberg & Safran, 1989). The idea of a cathartic release once emotion is expressed has long been theorized as a crucial aspect of beneficial therapy (Breuer & Freud, 1895/1966). Stanton and colleagues (e.g., Austenfeld & Stanton, 2004) suggested that coping through emotional expression is most useful when it is done in a social context that is receptive, when it helps frame goals that can then lead to action, and when it facilitates habituation to a stressor. Along similar lines, Kennedy-Moore and Watson (2001) proposed that emotional expression might produce benefits by alleviating anguish about distress and facilitating insight, which in turn leads to opportunities to respond to the environment.

Focused expressive writing was one of the first systematically evaluated types of emotionally expressive interventions (Pennebaker, 1997; Pennebaker & Beall, 1986). Meta-analyses have generally suggested beneficial outcomes associated with focused expressive writing intervention (Frisina, Borod, & Lepore, 2004; Sloan & Marx, 2004; Smyth, 1998). Emotional expression has also played a key role in intervention studies for cancer (Graves, Carter, Anderson, & Winett, 2003; Lieberman & Goldstein, 2006; Smith, Anderson-Hanley, Langrock, & Compas, 2005). Examples include unstructured journaling for women with newly diagnosed breast cancer (e.g., Smith et al., 2005), online support groups (e.g., Lieberman & Goldstein, 2006), face-to-face support groups (e.g., Graves et al., 2003), and supportive-expressive group therapy (Spiegel, Bloom, Kraemer, & Gottheil, 1989). Expressive writing and group-based intervention studies provide researchers with a wealth of textual data from both written essays and transcribed individual or group sessions. The resulting textual (or video) data often provide detailed accounts of the ways in which individuals cope with diagnosis and treatment of cancer. If valid methods for quantifying psychological processes in text were available, re-

Erin O'Carroll Bantum, Cancer Research Center of Hawaii, University of Hawaii at Manoa; Jason E. Owen, Department of Psychology, Loma Linda University.

Erin O'Carroll Bantum acknowledges her dissertation committee members for their help on this project: Jason E. Owen (chair), Michael Galbraith, Mark Haviland, Patricia Pothier, and Kendal Boyd.

Correspondence concerning this article should be addressed to Erin O'Carroll Bantum, Cancer Research Center of Hawaii, Prevention and Control Program, 1960 East-West Road, Biomed C-105, Honolulu, HI 96822. E-mail: ebantum@crch.hawaii.edu

searchers and clinicians would be able to explore relationships between specific emotional factors (e.g., description of anger in text) and outcomes. Such methods would enhance our understanding of what types of emotional expression might be related to improved psychological adjustment.

There have been a few studies designed to evaluate the effect of emotional expression on both physical and psychological symptoms in breast cancer (Stanton, Danoff-Burg, Cameron, Snider, & Kirk, 1999, Stanton, Danoff-Burg, & Huggins, 2002; Walker, Nail, & Croyle, 1999; Zakowski, Ramati, Morton, Johnson, & Flanigan, 2004). Although in one study expressive writing was not associated with differences in positive affect, negative affect, intrusive thoughts, or avoidance (Walker et al., 1999), in the other three studies, fewer physical symptoms (Stanton et al., 1999; Stanton, Danoff-Burg, Sworowski, et al., 2002) and fewer social constraints after writing (Zakowski et al., 2004) were found in participants who were asked to write either about their feelings regarding breast cancer or benefits resulting from their experience with breast cancer. Supportive-expressive group psychotherapy for metastatic breast cancer patients has been related to a significant reduction in suppression of primary negative affect; improvement in restraint of inconsiderate, irresponsible, impulsive, aggressive behavior (Giese-Davis et al., 2002); and a decline in traumatic stress symptoms (Classen et al., 2001). Lieberman and Goldstein (2006) found that increased expression of anger was associated with improved quality of life (as identified by the Functional Assessment of Cancer Therapy—Breast Cancer Form [FACT–B]; Cella, 1994) outcomes and lower levels of depression (as identified by the Center for Epidemiological Studies—Depression scale; Radloff, 1977), although expression of fear and anxiety was associated with decreased quality of life and higher levels of depression. This complicates the picture and further brings into question the notion that emotional expression is a mechanism leading to psychological and physical betterments for cancer patients. Similar findings by Smith et al. (2005) revealed that the prevalence of negative emotion in journal entries of breast cancer patients was related to increased anxiety and depression.

Computational text analysis provides a set of tools for analyzing the large volume of text that is produced as a result of expressive writing studies or other types of interventions. In the psychological literature, only a few computational text-analysis programs have been described (see Pennebaker, Mehl, & Niederoffer, 2003, for review). Although the General Inquirer (Stone, 1966) is considered by some as the first computational text-analysis program, more recent programs, such as Mergenthaler's (1993) TAS/C and DICTION (North, Langerstrom, & Mitchell, 1984), also have applicability for coding emotional expression. Others have attempted to capture the multidimensional nature of emotionally relevant words. Lang, Bradley, and Cuthbert's (1999) Affective Norms for English Words provides an assessment of the degree of pleasure, arousal, and dominance evoked by each of over 3,000 words.

One of the more recently developed and used computational text-analysis programs that has been used to provide information regarding linguistic correlates of psychological and health outcomes after writing is Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001). Pennebaker and Francis devised this tool to analyze text on a word-by-word basis. LIWC calculates a percentage of words falling into 74 different categories, ranging from emotion words to words about social context and religion. Pennebaker (1997) found four factors to correlate with the greatest health benefits. These include more positive emotion words used, a moderate number of negative emotion words used, and an increasing number of causal and insight words used over the course of writing (Pennebaker, 1997; Pennebaker & Chung, 2007).

LIWC was initially validated for content and construct validity by the creators of the program (Pennebaker & Francis, 1992; Pennebaker et al., 2001). Interrater reliability discrimination of category word elements has been found to range from 86% to 100%, depending on the dimension being assessed (Pennebaker et al., 2001), suggesting content validity. To assess construct validity, four judges rated 210 essays on several LIWC dimensions (Pennebaker, Mayne, & Francis, 1997). Moderate to strong correlations between LIWC and judges' global ratings of written essays were found for most emotion categories (0.22–0.75; Pennebaker et al., 1997). LIWC does not take context into consideration when identifying emotion words. It simply identifies words that are deemed by the developers of the program to appropriately fit into a number of different categories. It has been recognized that computer programs such as LIWC are inadequate for distinguishing between different meanings of the same word (Chung & Pennebaker, 2007). Very little work regarding psychometric properties has been done outside of the research conducted by those who developed LIWC. In one known study assessing psychometric properties of LIWC with breast cancer patients, researchers analyzed text for overall valence in a number of different categories (Alpers et al., 2005). Low to moderate correlations were found between rater codes and LIWC codes (Alpers et al., 2005), although the sample size was small ($n = 9$) and rater codes were assigned to each text file, rather than to individual words.

Another computer program that has been used to measure emotional expression in text and dialogue is Psychiatric Content Analysis and Diagnosis (PCAD). PCAD arose from the Gottschalk-Gleser scales (Gottschalk, Winget, & Gleser, 1969), in which verbal output is transferred into text and analyzed by extensively trained coders on a number of different facets, most of which are geared toward psychiatric diagnoses. Adequate construct validity and reliability have been established for all of the Gottschalk-Gleser scales (Gottschalk, 1995), but little is known about the psychometric properties of PCAD. One major difference between this program and LIWC is that PCAD takes context into consideration. Although this is an attractive aspect of PCAD, the types of scoring rules used by PCAD for coding emotion are unclear. This scoring ambiguity makes it difficult to critically evaluate the decisions and resulting scores PCAD provides for a number of different emotion domains.

In this study, we analyzed data collected by Owen et al. (2005) with the goal of comparing rater-coded emotional expression with emotion coded via LIWC and PCAD computerized text-analysis programs. We used signal-detection theory to guide the inclusion of signal-detection indices to help distinguish between signal and noise (Green & Swets, 1966). For our purposes here, a *signal* is emotional expression, and *noise* is the lack of emotional expression. The signal-detection indices used here were sensitivity, specificity, positive predictive value, and negative predictive value. In this study, sensitivity was the probability that a word that is actually representative of emotional expression would be characterized by LIWC as an emotion word (Portney & Watkins, 2000).

Specificity was the probability that a word that is not indicative of emotional expression would be characterized by LIWC as a non-emotion word (Portney & Watkins, 2000). Positive and negative predictive values were used to assess the probability that a word characterized by LIWC as an emotion word is truly representative of an emotion word and the probability that a word characterized as not being indicative of emotional expression by LIWC is, in fact, absent of emotional expression, respectively (Portney & Watkins, 2000).

There were two primary aims of the study. The first aim was to assess the accuracy of LIWC and PCAD for the detection of emotional expression. To accomplish this aim, we created a reliable coding system to identify emotional expression. It was hypothesized that, relative to PCAD, which uses context to disambiguate meanings of words, LIWC would exhibit low sensitivity to nuanced emotional expression. Because of the text-independent coding rules employed by LIWC, it was also hypothesized that LIWC would systematically over-identify emotional expression (i.e., exhibit low positive predictive values). Thus, the goals of the study were to evaluate sensitivity and positive predictive power of LIWC and PCAD relative to human raters. Although we had no specific hypotheses about specificity and negative predictive value, these signal-detection indices are mathematically related to sensitivity and positive predictive power and provide useful information about the types of overlap between computerized text-analysis methods and human raters. For sake of completeness in reporting, we elected to present the data from specificity and negative predictive values alongside sensitivity and positive predictive values. The second aim was to explore the relationship between coding methods and self-report measures of emotional well-being. It was hypothesized that rater-coded emotional expression would be more closely associated with self-report measures of emotional functioning than would LIWC- or PCAD-identified emotional expression.

Method

Participants

Participants in the initial study included 49 women with Stage 1 or 2 breast cancer, recruited from a hematology/oncology outpatient clinic at a large academic medical center in the southeastern United States. Participants were not excluded on the basis of time since diagnosis or medical treatment. They were recruited to participate in a randomized 12-week clinical trial of an Internet-based support group. Those who consented to participate completed a baseline assessment and were then randomized into a wait-list control group ($n = 19$) or an Internet-based support group ($n = 30$). Wait-listed participants were able to join a support group after completing a baseline questionnaire and waiting approximately 12 weeks for their group to begin. Additional details about the sample have been previously reported (Owen et al., 2005). For the current study, we included data from 14 additional participants. Thirteen women with Stage 3 or 4 breast cancer who were enrolled in a nonrandomized pilot of the online intervention were included in these analyses. Of those participants, 6 of 13 provided baseline data. In addition, text from 1 participant with Stage 2 breast cancer did not include any LIWC emotion words and was therefore not included in the findings displayed in the initial paper. For the analyses that included solely textual coding, 63 participants were included. For the analyses that included baseline assessment data, 55 participants were included. Participants had a mean age of 49.8 years ($SD = 11.0$), were largely college educated ($M = 15.4$ years; $SD = 2.4$), and were primarily Caucasian (93%).

Procedures

Those who were assigned to the online support group and those who later crossed over from the wait list into an online support group ($n = 19$) were encouraged to interact with one another remotely using a discussion board for general topics of conversation and a series of interactive coping-skills training exercises. All messages and responses to coping exercises were digitally stored in separate text files for each participant. Combined with the self-report survey data obtained at baseline and 12 weeks after beginning the support group ($n = 55$), the text files ($n = 63$) served as the primary source of data for the present study.

Rater coding of emotional expression. Developing a set of manual coding rules designed to identify emotional expression in text was a primary goal of the study. Definitions and coding rules for identification of emotional expression were developed iteratively. The initial coding rules were derived from a review of the literature on verbal and nonverbal behavioral indicators of emotional expression (Clore, Ortony, & Foss, 1987; Ekman, 1992; Mergenthaler, 1996; O'Rorke & Ortony, 1994; Ortony, Clore, & Foss, 1987; Ortony & Turner, 1996). Particularly helpful to this process were reports by Ortony and colleagues describing dictionaries of emotion words and decision rules for classifying words as indicators of emotional expression (Clore et al., 1987; O'Rorke & Ortony, 1994; Ortony et al., 1987). Their work on the affective lexicon provided a framework for considering different properties of emotion words and whether they should be included. An example of this was making the decision to code as *emotion words* words that referenced either internal feeling states or external feeling states (i.e., referencing emotion experienced by someone else; Clore et al., 1987).

After carefully reviewing the literature, we codified simple rules for identifying emotional expression and classifying various dimensions of positive and negative emotional expression in text. We then attempted to apply these rules to a subsample of approximately 33% of the available textual data, not to apply final coding decisions but, rather, to identify instances where the coding rules were unable to account for clear instances of emotional expression in the text. When such instances were identified, the coding rules were modified. After reviewing 33% of the text, we finalized coding rules and developed a brief training program used to teach blinded raters to apply the coding rules. Then, with the help of the blinded raters, we reevaluated all text data.

Brief description of coding rules. For raters, the initial decision to be made was regarding the presence of emotional expression. If it was determined that emotional expression was present, the decision was then made regarding the best fitting category of emotional expression: positive feelings, optimism, anxiety, anger, sadness, other positive emotion, or other negative emotion. Positive feeling was coded when it was deemed that the person was expressing a state of happiness, peacefulness, or gratitude. Optimism was coded when there was a demonstration in the text of the tendency to express the best possible outcome or the feeling that something would turn out well. The code of *other positive emotion* was given when a word and the adjoining phrase was representative of positively valenced emotion that was not better captured in the *positive feelings* or *optimism* categories (e.g., feelings of excitement). If a word or phrase represented uneasiness and apprehension, stress or tension, a feeling of being out of control, rumination, or a sense of hyperarousal, it would be coded in the *anxiety* category. *Anger* was coded when there was a demonstration of displeasure, hostility, frustration, or a dissatisfaction/discrepancy between an ideal and actual outcome. *Sadness* was coded when the word and/or phrase was characterized by sorrow or unhappiness. *Sadness* was also coded when the writer expressed a pessimistic sense of inadequacy, despondent lack of activity, or described behavior clearly associated with feelings of sadness (e.g., tears). The final emotion category coded was *other negative emotion*. If the word and/or phrase represented a negative emotion other than anxiety, anger, or sadness, it was placed in this category. An example of this instance is when there was a representation that was negative but ambiguous (e.g., "I felt really bad" when context made clear that the reference was not to a physical state). Categories mirrored dimensions of emotion identified in the literature (e.g., Clore et al., 1987) and emotion categories identified by LIWC, facilitating direct comparisons with LIWC emotional coding.

Rater-coding procedure. We replicated LIWC scoring rules using the Practical Extraction and Report Language (PERL) programming language and the default LIWC 2001 dictionaries. All text files were processed by our PERL code: Each word of each file was compared with the LIWC scoring dictionary, and words identified by the LIWC dictionary as emotion words were tagged (i.e., the word "sad" became "***sad***"). Tagged text files were then saved separately and made available for review by trained raters. Raters then reviewed each tagged text file and scored all tagged words using the coding rules described above. Raters were also instructed to score instances of emotional expression that had not been tagged. Words were coded into a specific emotion category (or coded as being absent of emotion). As with the trained raters, blinded raters were asked to review the text again for any missed emotion (words that were not tagged). Raters were blind to scoring decisions made by LIWC. The trained coders each independently coded all text files and then reconvened to assess congruency in codes. Discrepancies in coding were handled by describing the reasons for original coding decisions, and final codes were assigned by consensus. Interrater reliability between the two trained coders (EB, JO) was very good (κ = .80). Two additional raters who were blinded to the aims of the study underwent three training sessions, each lasting 2 hours, to learn the coding rules. Each of these blinded raters then independently reviewed 33% of the text, and interrater reliability was also established between two coders (κ = .69). The kappa statistic is very sensitive, and it is suggested that kappa scores of .61–.80 demonstrate "substantial" reliability, and scores of .81–1.0 demonstrate "almost perfect" reliability (Landis & Koch, 1977).

PERL text manipulation. PERL, which is an open-source programming language useful for evaluating patterns (i.e., pattern matching) with text-based data, was used to perform several key procedures in the present study. First, using the LIWC word library, we developed PERL code to reproduce LIWC scoring of emotion words, and this process allowed us to generate transcript files for manual coding in which all words coded by LIWC as emotion words were tagged by a common symbol. These tagged words were used to prompt coders to categorize the presence, absence, or type of emotional expression without indicating the specific code that LIWC would have assigned. Second, once rater coding was complete, we developed PERL code that matched each instance of emotional expression identified by rater coding with both LIWC-assigned codes and PCAD-assigned codes. Because PCAD assigns codes on the basis of clauses, rather than individual words, rater-assigned codes were matched with the clauses that contained the word(s) associated with the code. Our PERL code read each line of PCAD summary data for each text file, identified the clauses scored by PCAD, extracted the scoring decisions made by PCAD, and matched each clause with a Microsoft Excel spreadsheet containing the words and scoring decisions made by the human raters. We checked output from the PERL code for each subject to verify matching accuracy, and no errors were identified.

Materials

Self-report measures. Self-report measures for quality of life, cancer-related trauma, anxiety, and depression were administered to the treatment group (n = 30) at the beginning of the online support group and to the crossover control group (n = 32) 12 weeks after they entered the study (this is equivalent to the start date of their group).

Health-related quality of life. The Functional Assessment of Cancer Therapy—Breast Cancer Form (FACT–B) is a 27-item questionnaire that was used as a measure of perceived quality of life, with items of quality of life tapping into physical, social, emotional, and functional domains (Cella, 1994). This instrument has adequate internal consistency and good concurrent validity with the Eastern Cooperative Oncology Group performance status (Brady et al., 1997). The FACT–B has also been sensitive to change over time in persons with cancer. Participants were also asked to rate their overall health status on a 0–100 scale (using the EuroQuol 5–D; Brooks, 1996). On this scale, 0 = *the least desirable state of health you can imagine* and 100 = *perfect health* (Llach, Herdman, Schiaffino, & Dipstat, 1999).

Cancer-related trauma. We used the Impact of Events Scale (IES; Horowitz, Wilner, & Alvarez, 1979), which uses 22 items with a 5-point Likert-type scale to assess intrusiveness and avoidance of cancer-related thoughts and stimuli, to assess cancer-related anxiety. This instrument has been demonstrated to have good internal consistency and has been shown to be sensitive to effects of psychological intervention (Edgar, Rosberger, & Nowlis, 1992).

Anxiety and depression. We used the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), a 14-item measure that provides summary scores for depression and anxiety, to measure anxiety and depression. This is a self-report measure that was created to be used in medical populations, in that the scale does not overestimate for a mood disorder on the basis of somatic symptoms. It was found to have good interrater reliability and good construct validity for distinguishing between patients with and without mood disorder (Zigmond & Snaith, 1983). This measure is also sensitive to the effects of adjuvant psychotherapy treatment (Greer et al., 1992).

Self-reported positive and negative emotion. To compare coding methods with self-reported measures of both positive and negative emotion, we created two scales. We took individual items from the self-report measures previously described and used them in the initial study (Owen et al., 2005). The items were standardized, and Cronbach's alphas were assessed and found to be good for both self-reported positive emotion (Cronbach's α = .93) and negative emotion (Cronbach's α = .81). For the self-reported positive emotion scale, the following items were included from the FACT–B (Cella, 1994): "I feel close to my friends and family," and "I feel close to my partner (or the person who is my main support)." The HADS items that were used in the scale were "I still enjoy the things I used to enjoy, I can laugh and see the funny side of things, I feel cheerful, I can sit at ease and feel relaxed, I look forward with enjoyment to things," and "I can enjoy a good book or radio or TV program." For the self-reported negative emotion scale, the following items were included from the FACT–B (Cella, 1994): "I feel sad, I am losing hope in the fight against my illness, I feel nervous, I worry about dying," and "I worry that my condition will get worse." The HADS items that were included in this scale were "I feel tense or 'wound up,' I get a sort of awful feeling as if something is about to happen, worrying thoughts go through my mind, I feel as if I am slowed down, I get a sort of frightened feeling like 'butterflies' in the stomach, I feel restless as if I have to be on the move," and "I get sudden feelings of panic." Four items from the IES (Horowitz, Wilner, & Alvarez, 1979) were included in the scale: "I had waves of strong feelings about cancer, I felt watchful and on guard, I felt jumpy and easily startled," and "I felt irritable and angry."

LIWC analysis. Text was analyzed across all participants in the website (n = 63). This included all of the text submitted by each participant. The LIWC computer program analyzes written text on a word-by-word basis and is available in Spanish, German, Dutch, Norwegian, Italian, and Portuguese. This tool was developed by a process in which groups of judges reviewed 2,000 words or word stems and decided how the reviewed words related to dozens of categories (e.g., word count, total first-person usage, negative emotion; Pennebaker & Francis, 1992). Every word of a text file is compared with "dictionaries" of 74 dimensions. The dimensions include (a) standard linguistic dimensions, such as words per sentence; (b) psychological constructs, such as positive emotions; (c) dimensions related to relativity, such as past-tense verbs; and (d) personal concern categories, such as the use of job- or work-related words. A word might fit and be placed into more than one category. The categories that were used in this study included all of the emotion categories. LIWC separates emotion into positive emotion and negative emotion. It then more specifically identifies the word as being either indicative of positive feelings or optimism in the positive emotion category and into anxiety, anger, or sadness in the negative emotion category. There are cases in which LIWC identifies a word as being positive or negative emotion but does not further identify the word as fitting into one of the more specific categories.

Psychiatric Content Analysis and Diagnosis. We also compared PCAD with the rater-coding system. PCAD was developed from the Gottschalk-Gleser scales (Gottschalk et al., 1969). The program was designed to transfer verbal output into text and analyze it on a number of different facets, most of which are geared toward psychiatric diagnoses (e.g., anxiety, dependency, and social alienation). In large part, the program was designed to assess transcripts from psychotherapy. Adequate construct validity and reliability have been reported for all of the scales (Gottschalk, 1995). PCAD is different from LIWC in that it evaluates the word in the context of the surrounding clause and scores it on a number of different dimensions. A clause that contained one of the rater-coded emotion words was identified, and scores given by both PCAD and rater codes were compared for agreement. To link PCAD scores with rater-defined emotion categories, we matched PCAD scores with the most appropriate rating category as follows: positive feelings (human relations codes A2 and D1), optimism (hope H1, H2, H3, and H4), anger (hostility outward, human relations D2, and frustrated dependency codes A and D), anxiety (death, mutilation, separation, guilt, shame, and diffuse anxiety), and sadness (hostility inward, any depression score, or any ambivalent hostility score).

Results

There were two primary aims of the study. The first aim was to assess the accuracy of LIWC and PCAD for the detection of emotional expression. The second aim was to explore the relationship between coding methods and self-report measures of emotional well-being. To accomplish these aims, we recruited 63 people to participate in the research website and engage in text-based interaction with other participants. The entire transcript available for analysis consisted of 165,754 words (278 pages of single-spaced text, 12-point font). The transcripts were spell checked, and the text that was used in the analyses included all of the available text for each participant. Writing samples averaged 2,631 words per subject, and there was considerable variation across subjects (SD = 2,868). On average, LIWC identified 1.43% (SD = 0.6) of total words as negative emotion and 3.2% (SD = 1.1) of total words as positive emotion. Raters identified less negative emotion (0.8%; SD = 0.5) and less positive emotion (0.8%; SD = 0.4) than did LIWC. PCAD identified an average of 2.5% (SD = 0.7) of total words as negative emotion and 2.9% (SD = 0.8) as positive emotion. For the more specific categories of emotion, the average LIWC codes were as follows: positive feeling (posfeel) = 1.0%, optimism = 0.8%, anxiety = 0.5%, anger = 0.2%, and sadness = 0.3%. For specific emotion categories, manually coded averages were as follows: posfeel = 0.8%, optimism = 0.03%, anxiety = 0.2%, anger = 0.1%, and sadness = 0.3%. Specific categories of emotion, as coded by PCAD, were as follows: posfeel = 0.2%, optimism = 0.8%, anxiety = 1.5%, anger = 0.6%, and sadness = 2.3%. Each instance of emotion was counted as one point, and frequency of a given emotion was divided by total words for that participant, resulting in a percentage of a given emotion for each participant. This was true for LIWC, manual codes, and PCAD.

Sensitivity

Sensitivity captures the proportion of total emotion words identified by raters as being indicative of emotional expression that were captured by either LIWC or PCAD. Sensitivity for overall emotional expression was good for both LIWC (0.88) and PCAD (0.83). LIWC sensitivity for both positive and negative emotion was also good, although LIWC did not perform as well in the subcategories of positive feelings, anger, and sadness (see Table 1). PCAD sensitivity for both total positive (0.77) and negative (0.78) emotion was good; however, there was considerable variability between specific types of emotion. PCAD sensitivity was poor for positive feelings (0.15) and

Table 1
LIWC and PCAD Sensitivity and Specificity With 95% Confidence Intervals (CI; N = 63)

Type of emotion	LIWC sensitivity (95% CI)	PCAD sensitivity (95% CI)	LIWC specificity (95% CI)	PCAD specificity (95% CI)
Emotional expression	.88[a] (0.864–0.888)	.83[b] (0.82–0.85)	.97[c] (0.9624–0.966)	.58[d] (0.57–0.59)
Total positive emotion	.89[a] (0.870–0.902)	.77[b] (0.74–0.79)	.97[c] (0.972–0.974)	.74[d] (0.74–0.75)
Positive feelings	.47[a] (0.441–0.493)	.15[b] (0.13–0.17)	.993[c] (0.993–0.994)	.98[d] (0.976–0.980)
Optimism	.82 (0.724–0.916)	.65 (0.52–0.77)	.993[c] (0.993–0.993)	.72[d] (0.71–0.72)
Total negative emotion	.78 (0.763–0.805)	.78 (0.75–0.80)	.991[c] (0.991–0.991)	.78[d] (0.77–0.78)
Anxiety	.83[a] (0.798–0.860)	.59[b] (0.55–0.63)	.998[c] (0.998–0.998)	.86[d] (0.85–0.86)
Anger	.66[a] (0.596–0.730)	.40[b] (0.33–0.48)	.998[c] (0.998–0.998)	.95[d] (0.945–0.950)
Sadness	.70[b] (0.649–0.749)	.84[a] (0.80–0.88)	.997[c] (0.997–0.997)	.78[d] (0.77–0.78)

Note. LIWC = Linguistic Inquiry and Word Count (Pennebaker, Francis, & Booth, 2001); PCAD = Psychiatric Content Analysis and Diagnosis (Gottschalk, Winget, & Gleser, 1969). Superscript a denotes significantly greater sensitivity than does superscript b; superscript c denotes significantly greater specificity than does superscript d.

anger (0.40) but was substantially higher for sadness (0.84). Sadness was the only category for which PCAD was significantly more sensitive than LIWC. LIWC performed significantly better in categories of overall emotional expression, positive emotion, positive feelings, and anger (see Table 1).

Specificity

Specificity measured the proportion of nonemotional words that were accurately coded by LIWC or PCAD (see Table 2). Specificity for LIWC was exceptional in all emotion categories (0.97–0.999). PCAD specificity was poor for overall emotional expression (0.58) but was slightly better for overall positive (0.74) and negative emotion (0.78). In all eight of the identified emotion categories, LIWC was significantly more specific than PCAD (see Table 1).

Positive Predictive Value

Positive predictive value measured the probability that a word identified by LIWC and PCAD as being indicative of emotional expression was in agreement with rater codings of emotional expression. For LIWC, only 32% of words identified as any type of emotion, 24% of words identified as positive emotional expression, and 43% of words identified as negative emotion were in agreement with rater codes (i.e., 69% of words identified by LIWC as indicators of emotional expression were not thought by raters to be instances of emotional expression). PCAD performed poorly in all emotion categories (0.005–0.26). LIWC's positive predictive value was significantly better than PCAD in all emotion categories (see Table 2).

Negative Predictive Value

Negative predictive value measured the probability that a word not identified as emotion by LIWC or PCAD agreed with raters' judgment that the word was not associated with emotional expression. Both LIWC and PCAD had excellent negative predictive value across all emotion categories. Negative predictive values for LIWC ranged from 0.995 for positive feelings to 0.999 for total positive emotion, optimism, anxiety, anger, and sadness. For PCAD, negative predictive values ranged from 0.96 for positive feelings to 0.998 for optimism and sadness. As with specificity and positive predictive value, LIWC performed significantly better than PCAD on each emotion category (see Table 2).

Relationship Between Coding Methods and Self-Report Measures

To assess the relationship between coding methods, we calculated Pearson product-moment correlations to compare rater

Table 2
LIWC and PCAD Positive and Negative Predictive Power With 95% Confidence Intervals (CI; N = 63)

Type of emotion	LIWC positive predictive value (95% CI)	PCAD positive predictive value (95% CI)	LIWC negative predictive value (95% CI)	PCAD negative predictive value (95% CI)
Emotional expression	.32[a] (0.308–0.328)	.19[b] (0.18–0.19)	.998[c] (0.998–0.998)	.97[d] (0.96–0.97)
Total positive emotion	.24[a] (0.224–0.246)	.15[b] (0.14–0.15)	.999[c] (0.9990–0.999)	.98[d] (0.980–0.984)
Positive feelings	.38[a] (0.353–0.399)	.26[b] (0.23–0.29)	.995[c] (0.995–0.995)	.96[d] (0.95–0.96)
Optimism	.04[a] (0.030–0.052)	.005[b] (0.004–0.007)	.999[c] (0.999–0.999)	.998[d] (0.998–0.9993)
Total negative emotion	.43[a] (0.415–0.453)	.16[b] (0.15–0.17)	.998[c] (0.998–0.998)	.98[d] (0.982–0.986)
Anxiety	.60[a] (0.565–0.635)	.08[b] (0.006–0.103)	.999[c] (0.999–0.999)	.990[d] (0.989–0.991)
Anger	.31[a] (0.269–0.359)	.05[b] (0.04–0.07)	.999[c] (0.999–0.999)	.995[d] (0.995–0.996)
Sadness	.31[a] (0.272–0.338)	.04[b] (0.04–0.05)	.999[c] (0.999–0.999)	.998[d] (0.997–0.998)

Note. LIWC = Linguistic Inquiry and Word Count (Pennebaker, Francis, & Booth, 2001); PCAD = Psychiatric Content Analysis and Diagnosis (Gottschalk, Winget, & Gleser, 1969). Superscript a denotes significantly greater sensitivity than does superscript b; superscript c denotes significantly greater specificity than does superscript d.

Table 3
Correlations Between Rater-Coded Emotional Expression and Emotional Expression as Coded by LIWC (N = 63)

Variable	MANUAL POSEMO	MANUAL POSFEEL	MANUAL OPTIM	MANUAL NEGEMO	MANUAL ANX	MANUAL ANG	MANUAL SAD
LIWC POSEMO	.75**	.76**	.25	.02	−.03	.15	−.06
LIWC POSFEEL	.64**	.66**	.15	−.04	−.11	.22	−.04
LIWC OPTIM	.54**	.53**	.07	−.33*	−.26*	−.08	−.23
LIWC NEGEMO	.05	.01	.07	.54**	.41**	.40**	.37**
LIWC ANX	−.004	.002	−.08	.45**	.69**	.07	.09
LIWC ANG	−.24	−.25	.11	.07	.01	.25*	.03
LIWC SAD	.12	.09	.06	.50**	.37**	.55**	.33**

Note. LIWC = Linguistic Inquiry and Word Count (Pennebaker, Francis, & Booth, 2001); MANUAL POSEMO = manually coded positive emotion; MANUAL POSFEEL = manually coded positive feelings; MANUAL OPTIM = manually coded optimism; MANUAL NEGEMO = manually coded negative emotion; MANUAL ANX = manually coded anxiety; MANUAL ANG = manually coded anger; MANUAL SAD = manually coded sadness; LIWC POSEMO = LIWC-coded positive emotion; LIWC POSFEEL = LIWC-coded positive feelings; LIWC OPTIM = LIWC-coded optimism; LIWC NEGEMO = LIWC-coded negative emotion; LIWC ANX = LIWC-coded anxiety; LIWC ANG = LIWC-coded anger; LIWC SAD = LIWC-coded sadness (n = 55).
* $p < .05$. ** $p < .01$.

codes with both LIWC and PCAD codes. Tables 3 and 4 display the results. All but one of the LIWC codes were significantly positively correlated with corresponding rater codes. Optimism was the only emotion category that was not correlated with rater codes of the construct ($r = .07$; $p > .05$). The strongest correlations were found for positive emotion ($r = .75$; $p < .01$) and anxiety ($r = .69$; $p < .01$). Among PCAD emotion codes, only anxiety was significantly correlated with its corresponding rater codes ($r = .27$; $p < .05$).

It was anticipated that self-reported emotional well-being and psychological disturbance in women with breast cancer would be associated with some level of expression of those emotional experiences, particularly in the context of a structured coping-skills training program that encouraged active emotional coping efforts. To facilitate the examination of intercorrelations between emotional expression (positive and negative), as identified by different coding systems and self-report outcome measures, we used a multitrait, multimethod matrix to display these correlations. As previously stated, rater-coded positive and negative emotion and LIWC positive and negative emotion were highly correlated (for positive emotion, $r = .77$, $p < .01$; and for negative emotion, $r = .78$, $p < .01$). PCAD negative and positive emotion were not significantly correlated with corresponding categories in manual codes or LIWC. None of the coding systems was correlated with the created scales of self-reported positive and negative emotion, although the self-reported positive and negative scales were highly correlated ($r = -.76$; $p < .01$). Details of this analysis can be found in Table 5. In addition, we examined correlations between coding methods and the original scales. The correlations examined were as follows: Manual positive emotion and scores on FACT ($r = -.04$), manual positive emotion and scores on health status ($r = .04$), and manual positive emotion and scores on IES ($r = .11$) were not significant. Manual negative emotion and scores on FACT ($r = .06$), manual negative emotion and health status ($r = -.01$), and manual negative emotion and IES ($r = -.07$) were also not significant. It should be noted that the sample size is smaller in this table ($n = 55$) because not all participants returned the questionnaires that were included in this analysis. The text of the 55 participants who returned self-report data was included in these analyses.

Table 4
Correlations Between Manually Coded Emotional Expression and Emotional Expression as Coded by PCAD (N = 63)

Variable	MANUAL POSEMO	MANUAL POSFEEL	MANUAL OPTIM	MANUAL NEGEMO	MANUAL ANXIETY	MANUAL ANGER	MANUAL SAD
PCAD POSEMO	−.18	−.14	−.21	−.14	−.07	−.27*	−.07
PCAD POSFEEL	−.13	−.08	−.10	−.22	−.09	−.18	−.18
PCAD OPTIM	−.19	−.15	−.21	−.14	−.07	−.29*	−.07
PCAD NEGEMO	.25*	.29*	−.02	.10	.06	.23	.02
PCAD ANXIETY	−.03	.03	−.16	.37**	.27*	.21	.14
PCAD ANGER	.29*	.33**	−.03	.01	−.10	.16	−.001
PCAD SAD	−.23	−.19	−.04	.19	.14	.18	.12

Note. PCAD = Psychiatric Content Analysis and Diagnosis (Gottschalk, Winget, & Gleser, 1969); MANUAL POSEMO = manually coded positive emotion; MANUAL POSFEEL = manually coded positive feelings; MANUAL OPTIM = manually coded optimism; MANUAL NEGEMO = manually coded negative emotion; MANUAL ANXIETY = manually coded anxiety; MANUAL ANGER = manually coded anger; MANUAL SAD = manually coded sadness; PCAD POSEMO = PCAD-coded positive emotion; PCAD POSFEEL = PCAD-coded positive feelings; PCAD OPTIM = PCAD-coded optimism; PCAD NEGEMO = PCAD-coded negative emotion; PCAD ANXIETY = PCAD-coded anxiety; PCAD ANGER = PCAD-coded anger; PCAD SAD = PCAD-coded sadness (n = 55).
* $p < .05$. ** $p < .01$.

Table 5
Multitrait, Multimethod Matrix Correlating Coding Systems, Types of Emotion, and Outcome Measures (N = 55)

Variable	1	2	3	4	5	6	7	8
1. Manual positive emotion	—							
2. Manual negative emotion	.06	—						
3. LIWC positive emotion	.77**	.03	—					
4. LIWC negative emotion	−.10	.78**	−.03	—				
5. PCAD positive emotion	−.09	.03	.13	.14	—			
6. PCAD negative emotion	.16	.04	.23	−.08	.41**	—		
7. Self-report positive emotion	−.01	−.05	.17	−.18	−.20	−.05	—	
8. Self-report negative emotion	.11	.11	−.09	.25	.02	.12	−.76**	—

Note. LIWC = Linguistic Inquiry and Word Count (Pennebaker, Francis, & Booth, 2001); PCAD = Psychiatric Content Analysis and Diagnosis (Gottschalk, Winget, & Gleser, 1969).
** $p < .01$.

Discussion

Our hypothesis that high interrater reliability would be established between both trained and blinded coders was supported. Our second hypothesis, that PCAD would be more sensitive to emotional expression than LIWC, was not supported. LIWC's test characteristics were comparable to PCAD or significantly better than PCAD for all emotion categories. Both LIWC and PCAD exhibited poor positive predictive value, suggesting that they substantially over-identified emotional expression. However, PCAD's attempts to use context to disambiguate word meanings did not appear to be particularly effective. Although both LIWC and PCAD measure a number of domains other than emotional expression, our findings suggest that LIWC is superior to PCAD for identification of emotion in text.

Test characteristics for LIWC were more desirable for general emotion categories (i.e., overall emotion, negative emotion, positive emotion) than for specific types of emotion (i.e., positive feelings, optimism, anxiety, anger, and sadness), although this was not the case for PCAD. LIWC's stronger performance in general emotion categories indicated that, although it accurately captured that a given word was indicative of some kind of positive or negative emotion, it often did not adequately identify the specific types of emotional expression. LIWC uses a very simple strategy to place words into a given category on the basis of lists ("dictionaries") of related words and does not take context into account. The simplicity of the strategy is appealing and works well in general, but accuracy of LIWC for identifying emotional expression could be improved using more sophisticated computational linguistic strategies, such as word disambiguation (Agirre & Edmonds, 2006) or key word in context (Weik, 1996). Revising simple programs, such as LIWC, to include other, more sophisticated computational linguistic strategies could be fruitful.

With the exception of optimism, there was good convergent and discriminant validity between LIWC codes and rater codes. Agreement between LIWC and trained coders was quite high, with large effects for the general emotion categories of positive and negative emotion and the subcategories of positive feelings and anxiety and small effects in the subcategories of anger and sadness. There was no correspondence between LIWC and rater codes for optimism, and raters noted that optimism was particularly difficult to examine. Raters did not code a given word as being indicative of optimism if it was thought to express *desire* (e.g., "I hope I get better" or "Hopefully, the weather will clear up"), rather than a *feeling* of desire (e.g., "His response was one of hope and caring"). *Hope* is an example of a word that LIWC automatically codes as optimism, but human coders identified a great deal of context-dependent variability for this word in particular.

A very different pattern of results was observed for the correlations between PCAD and rater-coded emotions. Across nearly all categories of emotion, there was little agreement between PCAD and trained coders on corresponding types of emotion. Agreement between PCAD and raters existed only for anxiety. Expected convergent and discriminant validities were not obtained. For example, PCAD codes of negative emotional expression and anger were significantly and unexpectedly positively correlated with rater codes of overall positive emotion and positive feelings. The original Gottschalk-Gleser content-analysis scales (Gottschalk et al., 1969) were developed with the goal of having a clinician use transcripts of therapy sessions to better understand psychological functioning, as well as to assist the clinician with making a diagnosis. Although the therapist was instructed to use clinical judgment to score specific aspects of psychological disturbance related to an emotion category for anxiety (e.g., death anxiety, mutilation anxiety, guilt anxiety, separation anxiety, shame anxiety, and diffuse anxiety), it is unclear how the PCAD computer program makes these decisions. This makes it difficult to identify potential improvements in the way PCAD assigns emotion-related scores.

Our hypothesis that rater-coded emotional expression would be more closely associated with self-report measures of positive and negative emotion than would LIWC was not supported. Self-reported positive and negative emotion were not highly correlated with rater, LIWC, or PCAD codes of positive or negative emotional expression. Other literature supports this finding (e.g., Owen et al., 2006). This finding suggests that behavioral linguistics and subjective self-report measures of emotional well-being provide different perspectives on the individual experience of emotion. Because so much of our understanding of how emotional experiences influence and are influenced by behavior rests on studies involving self-report measures of the emotional experience, it is imperative that we think more carefully about how objective measures of emotional expression could contribute to the measurement and conceptualization of emotion. We suggest that behavioral linguistics should be considered as a supplement to self-report data in subsequent studies.

LIWC appears to be a relatively useful instrument for the identification of emotional expression in text. However, LIWC over-identifies emotional terms, suggesting that LIWC's ability to disambiguate words that are often used to convey quite different meanings could be substantially improved. Specifically, there were approximately 6,000 unique instances where LIWC identified emotional expression that was at odds with human ratings of the word. LIWC failed to identify emotional expression in only approximately 500 instances in which raters had identified the presence of emotion. Thus, LIWC is 12 times more likely to make errors of over-identification than it is to make errors of under-identification. There were also a number of words that were frequently coded by LIWC as emotion that could be removed from the emotion dictionaries. For instance, the word "good" is coded as positive emotion by LIWC, yet in almost every instance (94% of cases), it was not deemed by raters to be representative of emotion. Other words that were frequently coded as emotion by LIWC but not coded as emotion by manual codes were "hope" (97% of cases found to demonstrate the absence of emotion), "like" (97% of cases found to demonstrate the absence of emotion), "beautiful" (100% of cases found to demonstrate the absence of emotion), and "best" (100% of cases found to demonstrate the absence of emotion). In general, there were a few consistent errors that, if addressed, could improve the ability of LIWC to accurately code emotional expression. Some of these errors can be corrected by making simple changes to the way LIWC codes words. To our knowledge, the ability to add additional guidelines or rules to LIWC, a feature built into the software, has not been tested. If the addition of simple qualifiers or guidelines to LIWC could improve the accuracy with which emotional expression is identified, the ability of researchers and clinicians to characterize associations between emotional expression and of physical and mental health benefits could be enhanced. Recently, an updated version of LIWC has become available (Pennebaker, Booth, & Francis, 2007). In this version, the emotion categories of positive feelings and optimism have been removed from the program. Additional studies are needed to independently evaluate the test characteristics of this new version of LIWC and perhaps to evaluate the contribution of other LIWC categories to the identification of emotional expression.

There were a few noteworthy limitations of the study. Because this study evaluated emotional expression in a population of women with breast cancer using an Internet-based group, it is unclear how these specific findings will generalize to other populations. Because women with breast cancer engage in more frequent emotional expression than do people with other cancer types (e.g., men with prostate cancer; Owen, Klapow, Roth, & Tucker, 2004), we believe breast cancer patients to be a particularly useful population for the study of emotional expression. In addition, findings might differ for face-to-face groups or groups that are instructed to engage in the prototypical expressive writing paradigm. It should also be noted that PCAD and LIWC emotion categories are not identical. Because the default PCAD scoring was used and scores were summed to create composite scores consistent with our taxonomy of emotions, our findings may somewhat underestimate PCAD's validity. Scoring rules used by PCAD were not readily available and so could not be used to tag our text files prior to qualitative analysis. Although the fairly small sample size can be noted as a limitation, the volume of text analyzed in the study provided a more than adequate sampling frame for characterizing the accuracy of LIWC and PCAD.

A final limitation concerns base rates. Base rates are known to affect signal-detection indices. It is worth noting that the base rate of emotional expression observed in this study (1.8% of words identified by raters) may not generalize to other types of writing or verbal expression. The base rate of emotional expression identified by LIWC was comparable in the present study (5.2% of overall affect) to that found across expressive writing samples (5.3%; Pennebaker et al., 2001). However, our results should be further evaluated in samples with different base rates of emotional expression.

Extensive qualitative analysis can be considered the gold standard for characterizing the emotional content of text-based data. In this study, we have successfully and reliably coded the emotional content of a large volume of text from an Internet-based psychosocial intervention for women with breast cancer. Although LIWC has notable limitations, it appears to have acceptable convergent and divergent validity for the identification of emotional expression in text. Given the importance of being able to analyze available behavioral data and the sheer volume of text-based communication in an increasingly digital world, additional efforts to improve the accuracy of automated identification of emotional expression and other behavioral markers are much needed.

References

Agirre, E., & Edmonds, P. (2006). *Word sense disambiguation.* Dordrecht, The Netherlands: Springer.

Alpers, G. W., Winzelberg, A. J., Classen, C., Dev, P., Koopman, C., Roberts, H., et al. (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior, 21,* 343–358.

Austenfeld, J., & Stanton, A. L. (2004). Coping through emotional approach: A new look at emotion, coping, and health-related outcomes. *Journal of Personality, 72*(6), 1335–1363.

Brady, M. J., Cella D. F., Mo, F., Bonomi, A. E., Tulsky, D. S., Lloyd, S. R., et al. (1997). Reliability and validity of the Functional Assessment of Cancer Therapy—Breast quality-of-life instrument. *Journal of Clinical Oncology, 15,* 974–986.

Breuer, J., & Freud, S. (1966). *Studies on hysteria.* New York: Avon. (Original work published 1895)

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy, 37*(1), 53–72.

Cella, D. F. (1994). *F.A.C.T. Manual: Functional Assessment of Cancer Therapy (FACT) Scales.* Chicago: Rush-Presbyterian-St. Luke's Medical Center.

Chung, C. K., & Pennebaker, J. W. (2007). The psychological function of function words. In K. Fiedler (Ed.), *Social communication: Frontiers of social psychology* (pp 343–359). New York: Psychology Press.

Classen, C., Butler, L., Koopman, C., Miller, E. DiMiceli, S., Giese-Davis, J., et al. (2001). Supportive-expressive group therapy and distress in patients with metastatic breast cancer. *Archives of General Psychiatry, 58,* 494–501.

Clore, G., Ortony, A., & Foss, M. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology, 53,* 751–766.

Edgar, L., Rosberger, D., & Nowlis, D. (1992). Coping with cancer during the first year after diagnosis: Assessment and intervention. *Cancer, 69*(3), 817–828.

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99,* 550–553.

Frisina, P. G., Borod, J. C., & Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *Journal of Nervous and Mental Disease, 192*(9), 629–634.

Giese-Davis, J., Koopman, C., Butler, L. D., Classen, C., Cordova, M., Fobair, P., et al. (2002). Change in emotion-regulation strategy for women with metastatic breast cancer following supportive-expressive group therapy. *Journal of Consulting and Clinical Psychology, 70,* 916–925.

Gottschalk, L. (1995). *Content analysis of verbal behavior: New findings and clinical applications.* Hillsdale, NJ: Erlbaum.

Gottschalk, L., Winget, C., & Gleser, G. (1969). *Manual of instructions for using the Gottschalk-Gleser content analysis scales: Anxiety, hostility, and social alienation-personal disorganization.* Berkeley: University of California Press.

Graves, K. D., Carter, C. L., Anderson, E. S., & Winett, R. A. (2003). Quality of life pilot intervention for breast cancer patients: Use of social cognitive theory. *Palliative Supportive Care, 1*(2), 121–134.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greenberg, M. A., & Safran, J. D. (1989). Emotion in psychotherapy. *American Psychologist, 44,* 19–29.

Greer, S., Moorey, S., Baruch, J. D., Watson, M., Robertson, B. M., Mason, A., et al. (1992). Adjuvant psychological therapy for those with cancer: A prospective randomized trial. *British Medical Journal, 304*(6828), 675–680.

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of events scale: A measure of subjective stress. *Psychosomatic Medicine, 41,* 209–218.

Kennedy-Moore, E., & Watson, J. C. (2001). How and when does emotional expression help? *Review of General Psychology, 5*(3), 187–212.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International Affective Picture System (IAPS): Technical manual and affective ratings.* Gainesville: Center for Research in Psychophysiology, University of Florida.

Lieberman, M., & Goldstein, B. (2006). Not all negative emotions are equal: The role of emotional expression in online support groups for women with breast cancer. *Psycho-Oncology, 15*(2), 160–168.

Llach, X. B., Herdman, M., Schiaffino, A., & Dipstat, A. (1999). Determining correspondence between scores on the EQ-5D "thermometer" and a 5-point categorical rating scale. *Medical Care, 37,* 671–677.

Mergenthaler, E. (1993). *TAS/C user manual.* Ulm, Germany: Ulmer Textbank.

Mergenthaler, E. (1996). Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology, 6,* 1306–1315.

North, R., Langerstrom, R., & Mitchell, W. (1984). *DICTION computer program.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research, University of Michigan.

O'Rorke, P., & Ortony, A. (1994). Explaining emotions. *Cognitive Science, 18,* 283–323.

Ortony, A., Clore, G., & Foss, M. (1987). The referential structure of the affective lexicon. *Cognitive Science, 11,* 341–364.

Ortony, A., & Turner, T. (1996). What's basic about basic emotions? *Psychological Review, 97,* 315–331.

Owen, J. E., Giese-Davis, J., Cordova, M., Kronenwetter, C., Golant, M., & Spiegel, D. (2006). Self-report and linguistic indicators of emotional expression in narratives as predictors of adjustment to cancer. *Journal of Behavioral Medicine, 29*(4), 335–345.

Owen, J. E., Klapow, J. C., Roth, D. L., Shuster, J. L., Bellis, J., Meredith, R., & Tucker, D. C. (2005). Randomized pilot of a self-guided internet coping group for women with early stage breast cancer. *Annals of Behavioral Medicine, 30*(1), 54–64.

Owen, J. E., Klapow, J. C., Roth, D. L., & Tucker, D. C. (2004). Use of the Internet for information and support: Disclosure among persons with breast and prostate cancer. *Journal of Behavioral Medicine, 27*(5), 491–505.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8,* 162–166.

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95,* 274–281.

Pennebaker, J. W., Booth, R. E., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC-2007 operations manual.* Austin, TX: LIWC.net.

Pennebaker, J. W., & Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. In H. Friedman & R. Silver (Eds.), *Handbook of health psychology* (pp. 263–284). New York: Oxford University Press.

Pennebaker, J. W., & Francis, M. F. (1992). Putting stress into words: The impact of writing on physiological, absentee, and self-reported emotional well-being measures. *American Journal of Health Promotion, 6*(4), 280–287.

Pennebaker, J. W., Francis, M. F., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC-2001 operations manual.* Austin, TX: LIWC.net.

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology, 72*(4), 863–871.

Pennebaker, J. W., Mehl, M. R., & Niederoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547–577.

Portney, L. G., & Watkins, M. P. (2000). Foundations of clinical research: Applications to practice. Upper Saddle River, NJ: Prentice Hall Health.

Radloff, L. (1977). The CES–D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 3,* 385–401.

Sloan, D. M., & Marx, B. P. (2004). A closer examination of the structured written disclosure procedure. *Journal of Consulting and Clinical Psychology, 72*(2), 165–175.

Smith, S., Anderson-Hanley, C., Langrock, A., & Compas, B. (2005). The effects of journaling for women with newly diagnosed breast cancer. *Psycho-Oncology, 14,* 1075–1082.

Smyth, J. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology, 66,* 174–184.

Spiegel, D., Bloom, J. R., Kraemer, H. C., Gottheil, E. (1989). Effect of psychosocial treatment on survival of patients with metastatic breast cancer. *Lancet, 2,* 889–891.

Stanton, A., Danoff-Burg, S., Cameron, C. L., Snider, P. R., & Kirk, S. B. (1999). Social comparison and adjustment to breast cancer: An experimental examination of upward affiliation and downward evaluation. *Health Psychology, 18*(2), 151–158.

Stanton, A., Danoff-Burg, S., & Huggins, M. (2002). The first year after breast cancer diagnosis: Hope and coping strategies as predictors of adjustment. *Psycho-Oncology, 11,* 93–102.

Stanton, A., Danoff-Burg, S., Sworowski, L., Collins, C., Branstetter, A., Rodrigues-Hanley, A., et al. (2002). Randomized, controlled trial of written emotional expression and benefit finding in breast cancer patients. *Journal of Clinical Oncology, 20,* 4160–4168.

Stone, P. J. (1966). *The General Inquirer: A computer approach to content analysis.* Cambridge, MA: MIT Press.

Walker, B. L., Nail, L. M., & Croyle, R. T. (1999). Does emotional expression make a difference in reactions to breast cancer? *Oncology Nursing Forum, 26*(6), 1025–1032.

Weik, M. (1996). *Communications standard dictionary* (3rd ed.). New York: Chapman & Hall.

Zakowski, S. G., Ramati, A., Morton, C., Johnson, P., & Flanigan, R. (2004). Written emotional disclosure buffers the effects of social constraints on distress among cancer patients. *Health Psychology, 23*(6), 555–563.

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*(6), 361–370.

Received February 21, 2008
Revision received September 30, 2008
Accepted October 22, 2008 ■