8-2011

Enhancing therapists' clinical judgments of client progress subsequent to objective feedback

Michael M. Haderlie
University of Nevada Las Vegas

ENHANCING THERAPISTS' CLINICAL JUDGMENTS OF CLIENT PROGRESS SUBSEQUENT TO OBJECTIVE FEEDBACK

by

Michael M. Haderlie

Bachelor of Science
Brigham Young University
2005

Master of Science
Pacific University
2007

Master of Arts
University of Nevada, Las Vegas
2009

A dissertation submitted in partial fulfillment of
the requirements for the

Doctor of Philosophy in Psychology
Department of Psychology
College of Liberal Arts

Graduate College
University of Nevada, Las Vegas
August 2011

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

THE GRADUATE COLLEGE

We recommend the dissertation prepared under our supervision by

Michael M. Haderlie

entitled

Enhancing Therapists' Clinical Judgments of Client Progress Subsequent to Objective Feedback

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Psychology
Department Psychology

Christopher Heavey, Committee Chair

Jeffrey Kern, Committee Member

Russell Hurlburt, Committee Member

Stephen Fife, Graduate College Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies and Dean of the Graduate College

August 2011

ABSTRACT

Enhancing Therapists' Clinical Judgments of Client Progress Subsequent to Objective Feedback

By

Michael M. Haderlie

Dr. Chrisopher Heavey, Examination Committee Chair
Associate Professor of Psychology
University of Nevada, Las Vegas

Although it is intuitive that the judgments made by mental-health clinicians become increasingly accurate as they gain clinical experience, research has demonstrated only minimal effects of experience on clinical judgment. Feedback regarding the accuracy of judgments is widely considered to be an essential component in developing clinical judgment. However, very little research has systematically examined whether the provision of feedback following judgments leads to increased judgment accuracy. The current research explored the effects of providing feedback to therapists regarding client progress on the accuracy of therapists' judgments of change. The effect of feedback on therapists' confidence ratings regarding such judgments was also examined. Ten therapists at two on-campus outpatient clinics were randomly assigned to feedback (FB) or no-feedback (NFB) conditions. Immediately following each therapy session, therapists made judgments regarding the direction and magnitude of client progress. Therapists in the FB condition subsequently received feedback regarding clients' progress based upon a self-report measure of distress. The small size of the sample and correspondingly low statistical power made significance testing impractical. Thus the

results were examined in terms of effect sizes and should be considered exploratory. Results suggested that feedback did not improve judgment accuracy, as therapists in the NFB condition demonstrated greater improvement in accuracy over time. Therapists were found to be generally overconfident regarding the accuracy of their judgments. Feedback tended to reduce confidence ratings over time. Additionally, clients of therapists in the FB condition appeared to improve at a faster rate than clients of therapists in the NFB condition, consistent with previous research regarding the therapeutic effects of progress feedback. Finally, the number of judgments made by individual therapists was positively related to judgment accuracy, suggesting that repetition with the specific judgment task was beneficial. Results are discussed in terms of applications of feedback in training settings and directions for future research.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of many kind and dedicated individuals at the University of Nevada, Las Vegas. I am very grateful to Dr. Christopher Heavey and Dr. Jeffrey Kern for their time, support, and mentorship. I deeply appreciate the service of my committee members, Drs. Stephen Fife, Cortney Warren, and Russell Hurlburt. Their comments and suggestions were essential in shaping and strengthening the research. Dr. Hurlburt merits special thanks for serving as a last-minute committee member during two separate summers!

I am grateful to Drs. Colleen Peterson and Michelle Carro for allowing me to collect data at the Center for Individual, Couple, and Family Counseling. Dr. Peterson was especially instrumental in revising administrative procedures at the clinic in order to facilitate the research. I am also grateful to Dr. Phoebe Kuo-Jackson for her time and strong advocacy which allowed me to collect data at Counseling and Psychological Services. The selfless support and effort of these individuals was essential to the research and was very meaningful to me personally.

I am thankful to each of the therapists who volunteered their time, effort, and abilities as participants in the study. I am extremely grateful to have had the support of many talented undergraduate research assistants: Hannah Casares, Carla Farcello, Dmitriy Kazakov, Kim Lowe, Michaelangelo Miller, Sam Montano, Philipa Nothman, Laura Saunders, Laurie Smedley, and Chelsey Wilks.

Finally, I am forever thankful to my beautiful wife and children—Erica, Jordan, Evan, and Colette. Thank you for your unwavering support and the much-needed stress relief!

This research was supported by a grant from the Graduate & Professional Student Association of the University of Nevada, Las Vegas.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The clinical judgments made by mental-health clinicians, and the process by which such judgments are made, have long been a source of empirical scrutiny (e.g., Meehl, 1954). Clinical judgment entails a process of integrating data and observations that has been described as informal, subjective, and impressionistic (Bell & Mellor, 2009). A large body of research in the field of psychology is related to the relative merits of the clinical approach to decision making as compared to statistical techniques. In such research, the validity (accuracy) of decisions made on the basis of clinical judgment is compared to the validity of decisions reached through the application of statistical methods. Results have consistently favored formulaic, statistical techniques (e.g., Ægisdóttir et al., 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954), leading some to conclude that clinical judgment is seriously flawed and unreliable (e.g., Dawes, 1994; Faust & Ziskin, 1988).

A closely related research question has been the extent to which clinicians develop improved clinical judgment as they acquire clinical experience. Although it seems intuitive that clinicians become more accurate in their decisions with experience, most results have not generally supported such a relationship (Dawes, 1989, 1994; Faust, 1994; Faust & Ziskin, 1988; Garb, 1998; Garb & Boyle, 2003; Wedding & Faust, 1989). Didactic training specific to judgment tasks (e.g., training in the MMPI) does result in improved judgmental accuracy when clinicians are compared to lay judges with no such training; however, additional clinical experience beyond initial training has rarely resulted in subsequent increases in the validity of judgments (reviewed in Garb, 1998).

Recent meta-analytic results (Spengler, White, Ægisdóttir, Maugherman, et al., 2009) suggest that experience may in fact have a small positive effect overall on clinical judgment.

The relative inferiority of clinical judgment as compared to statistical techniques, coupled with the widely-accepted view that clinical judgment does not improve with experience (Lichtenberg, 1997), has led to skepticism toward clinical judgment among many researchers in the field of psychology. Westen and Weinberger (2004) coined the term *clinicism* (cynicism toward the clinician) to refer to a belief that little may be learned from clinical practice or experience. The degree to which such a belief is actually held has been debated (e.g., Garb & Grove, 2005). At any rate, researchers have increasingly recognized the need to move beyond descriptions of the limitations and flaws of clinical judgment in order to develop practices designed to improve clinical judgment. Such efforts are especially important given that clinical judgment is often considered the foundation of clinical practice (Ridley & Shaw-Ridley, 2009). Additionally, adequate statistical models have not been developed for a majority of the decisions made by clinicians. Therefore, the ability of clinicians to apply clinical judgment in making valid judgments for individual clients remains of utmost importance.

Relatively little research has evaluated systematic efforts to increase the accuracy of clinical judgment. Bell and Mellor (2009) noted that "although the debate over clinical versus statistical prediction may have achieved a significant degree of theoretical sophistication, it has largely failed to guide clinicians in improving their judgement [*sic*] accuracy" (p. 114). One potential method of improving clinical judgment is specific training regarding assessment procedures that incorporate research on judgment and

decision making. Although such methods have received only limited attention, some preliminary evidence suggests that such didactic training may lead to increased judgment accuracy (Meier, 1999; Spengler, White, Ægisdóttir, Maugherman, et al., 2009).

Perhaps the most frequently-suggested means of improving clinical judgment is the provision of corrective feedback during training on judgment tasks. Many authors (e.g., Faust, 1991; Garb, 1998; Lichtenberg, 1997) have argued that it is hard for clinicians to learn from their clinical experience because they do not receive feedback. Garb and Grove (2005) suggested that even when practitioners do receive feedback, it is often biased and subjective. Some decision-making models (e.g., the terminal insight model; Dawes, 1994) suggest that it is impossible for clinicians to improve the validity of their judgments without receiving feedback. Sapyta, Riemer, and Bickman (2005) compared clinical training that does not include feedback to attempting to learn archery while blindfolded. Spengler (1998) asserted, "To ensure judgment accuracy as a local clinical scientist, some form of feedback mechanism is needed" (p. 932).

Despite the widely-held belief that feedback is a necessary ingredient in the development of clinical judgment, research examining the effects of feedback on clinical judgment is surprisingly rare. Spengler, White, Ægisdóttir, Maugherman, and colleagues' (2009) meta-analysis of 75 clinical judgment studies included only two studies in which feedback about accuracy was included as a potential moderator of judgment accuracy. Additionally, judgment studies that systematically vary the availability of feedback have rarely been conducted (Spengler, White, Ægisdóttir, & Maugherman, 2009).

The purpose of the current research was to examine the effects of feedback on the accuracy of clinical judgments through an experimental design. The specific type of

clinical judgment examined was judgments about client progress over the course of psychotherapy. Although judgment research has largely focused on decisions made during the initial stage of therapy (e.g., diagnosis, treatment planning), the continuous and accurate evaluation of client response throughout therapy is essential to positive therapy outcomes (Hatfield & Ogles, 2006). Decisions regarding the course of therapy, alterations to the treatment plan, and the timing of termination are based on the clinician's judgment of client progress. Clinicians generally make such judgments on the basis of relatively subjective indicators including behavioral observations or clients' self-reports of change. However, the use of outcome measures as an aid in monitoring client progress has become increasingly common (Hatfield & Ogles, 2004). A significant body of research indicates that the use of outcome measures throughout therapy leads to improved clinical outcomes, especially for clients who initially do not respond well to treatment (Brodey et al., 2005; Lambert, 2007; Slade et al., 2006). Given such positive effects of feedback on client outcomes, as well as the widely-held belief that feedback is positively related to judgment, it was hypothesized that feedback regarding outcome questionnaires results in improvements in clinical judgment.

In the present study, the accuracy of clinicians' judgments regarding their clients' progress in therapy was evaluated. A standardized outcome questionnaire was utilized at each session as an objective indicator of the change in clients' overall distress and dysfunction. A randomized half of clinicians received feedback results from the outcome questionnaire, while the other half of clinicians did not receive feedback. It was anticipated that clinicians who received weekly feedback would make more accurate judgments regarding client progress than those who did not receive feedback. Additional

research questions included the effects of feedback on confidence, the effects of feedback on client progress, and the relationship between the accuracy of judgments made by a clinician and client outcomes.

CHAPTER 2

LITERATURE REVIEW

Clinical Judgment and Decision Making

From the moment of first contact with a new client to the conclusion of service provision, clinicians must weigh information in order to make decisions that will benefit the client. The ability of clinicians to make accurate and useful decisions is important to all clinical practice; indeed, it is the assumed ability of mental health professionals to make more valid decisions than untrained individuals that allows them to charge for their services. The validity of clinical decision making is dependent upon the clinical judgment of the clinician. Given its ubiquitous role in service provision, clinical judgment has been described as the foundation of psychological practice (Gambrill, 2005; Ridley & Shaw-Ridley, 2009). Clinical judgment is also among the most widely-researched topics in applied psychology. However, it is notable that judgment research in psychology is less advanced than similar research in other fields (e.g., medicine; Garb, 1998).

Defining Clinical Judgment

Clinical judgment and decision making are closely related, given that the applied product of clinical judgment is a decision. Research on both judgments and decisions is therefore generally referred to as *judgment research*. Clinical judgment is a process of critical thinking that is utilized to integrate all available data in order to make a valid decision. The term *clinical judgment* originated from early work comparing the judgments made by clinicians to decisions reached through the application of statistical, or actuarial, methods. Clinical judgment therefore refers to a process of data aggregation

that is informal and unstructured, as compared to more formal statistical methods (Meehl, 1954; Westen & Weinberger, 2004). Grove and Meehl (1996) described clinical judgment as relying "on human judgment that is based on informal contemplation and, sometimes, discussion with others" (p. 293). Clinical judgments therefore involve idiographic, multidimensional conceptualizations of an individual, rather than nomothetic or probabilistic generalizations (Bell & Mellor, 2009). Clinical judgment therefore incorporates sources of data such as relevant research evidence, clinicians' clinical experience, and various types of information about the client for whom a decision is to be made. A parallel term is *clinical prediction*, which refers to the application of clinical judgment to predict future or concurrent outcomes.

Clinical judgment also refers, more generally, to the judgments, inferences, observations, and practices of clinicians (Westen & Weinberger, 2004). Clinicians make decisions in varied contexts across all forms of clinical practice. The initial decision made in most therapeutic settings is the assignment of a clinical diagnosis. Decisions are also made early in therapy in order to determine whether treatment is needed, and if so, what kind of treatment may be beneficial. The treatment planning phase of therapy also requires several decisions, such as forming a case conceptualization, identifying appropriate goals for therapy, and selecting optimal interventions. Throughout the course of therapy, clinicians must make ongoing determinations regarding client progress and outcome. These evaluations are essential in order to decide whether to alter the treatment plan, or in the case of progress, whether to terminate therapy.

Judgment of Client Progress

Despite the importance of accurate clinical decisions throughout therapy, most clinical judgment research in psychology has focused on the accuracy of diagnostic assessments (Hatfield, 2007). The focus of clinical judgment in the current research is judgment regarding client change during the course of therapy. Evaluating client progress from week to week is essential to significant decisions. For example, the judgment that a client has not progressed or has deteriorated might lead a clinician to alter the treatment plan by shifting intervention strategies, recommending medication, or consulting. On the other hand, judgment that a client has improved is necessary in order to initiate the process of termination. In addition, judgment regarding client change from one week to another is important in evaluating the impact of specific interventions for a particular client. Monitoring of client progress throughout the course of therapy is especially important given that an estimated 5% to 10% of psychotherapy clients finish therapy in worse condition than when they began (Lambert & Ogles, 2004). Additionally, Kendall, Kipnis, and Otto-Salaj (1992) found that even when therapists considered clients to be deteriorated, a majority of therapists did not alter their treatment plans.

Several sources of information are available to clinicians in the context of evaluating client progress during therapy. The most ubiquitous data source is clients' own verbal reports regarding their relative change from one week to another. Additional sources of information may include behavioral observations during session, significant others' reports, process measures (e.g., standardized questionnaires assessing constructs like alliance or client satisfaction), and outcome measures.

It is apparent that various sources of data regarding client change yield different, and sometimes contradictory, perspectives on the progress made. Judgment of client progress is therefore highly influenced by the types of information that are available or sought by the clinician. For example, Hole (1972) compared clinician ratings of change with change in MMPI-2 results over the course of therapy for 50 depressed patients. Results indicated that there was no correlation between profile changes and clinical assessment. However, the study was constrained by methodological problems, including the use of one 3-point scale as the clinical rating of change. Weiss, Rabinowitz, and Spiro (1996) reviewed 23 studies which examined the relationship between client and therapist estimates of change. Results were variable, indicating that the relationship between client and therapist perspectives is not always strong.

Hatfield and Ogles (2006) mailed a clinical vignette to 810 practicing psychologists and asked them to rate the degree to which various information sources influenced their resulting treatment decisions. Clinicians self-reported that they weighed client verbal reports and their own observations of the client much more heavily than other data sources. However, quantitative analyses of responses to systematically manipulated variables suggested that client verbal reports and outcome measures had equal impacts on decisions. Negative information from any source was found to influence clinicians more than positive information. Consistently, therapists in other studies have rated negative feedback as more valuable than positive feedback (Brodey et al., 2005; Haderlie & Kern, 2009).

In addition to the source of information, judgment of client progress is dependent upon characteristics of individual clinicians. Therapists of different orientations

emphasize distinct criteria in evaluating progress. For example, behavioral therapists are likely to seek information regarding symptom frequency and severity as a means of evaluating change, whereas dynamic therapists may base judgments on observations of clients' interpersonal behaviors in session. Kendall and colleagues (1992) found that cognitive-behavioral therapists waited until 6 to 8 months of therapy to determine that clients were not making progress, but dynamic therapists did not make such a determination until 14 months of therapy, on average. In addition, cognitive-behavioral therapists were more likely to report an intention to try different treatment plans for cases in which clients were judged to have made no progress or deteriorated. This latter finding was replicated by Hatfield and Ogles (2006).

Evaluating Clinical Judgment

The primary criterion used to evaluate clinical judgment is validity (based on Chronbach's, 1971, discussion of the validation of decisions made on the basis of test results). In the context of clinical judgment, it is the validity of a set of judgments made by a clinician that is examined. Decisions are valid if they are reliable, unbiased, accurate, and ultimately, useful. Validity is often difficult to evaluate in judgment research because of the need to subjectively select a criterion of accuracy. For example, Garb (1998) noted that psychological diagnoses are difficult to evaluate because they are open and arbitrary concepts. The most common methods of studying the validity of diagnoses include the use of structured interviews, expert judgments based on all available data (the *LEAD* procedure; Spitzer, 1983), construct validity studies, latent class

analysis, and the Robins and Guze (1970) criteria. The latter criteria describe a research approach to establishing diagnoses as useful entities.

These methods have been applied extensively to examine the validity of clinical judgments in a variety of contexts. As noted previously, the value of clinical judgment is often evaluated by comparing decisions based on judgment to decisions based on actuarial methods. Results of such studies have consistently favored statistical methods. Meehl's (1954) seminal work summarized findings from 20 studies which compared the two methods. Statistical methods were more accurate in all but 1 study, leading Meehl to conclude that clinicians should leave prognosis and classification to statistical methods. Although highly controversial (e.g., Holt 1958), Meehl's conclusions have been generally accepted. Similar conclusions have been reported in subsequent narrative reviews (e.g., Dawes, Faust, & Meehl, 1989; Faust, 1989; Grove & Meehl, 1996) and meta-analyses (Grove et al., 2000; Ægisdóttir et al., 2006). Most recently, Ægisdóttir and colleagues examined 92 effect sizes in which clinical and statistical methods of prediction were compared. The mean effect size indicated a modest advantage for statistical methods in terms of judgment validity (Cohen's d = .12 to .16, depending upon the stringency of inclusion criteria). Such results underscore the fact that the clinical method of decision making is subject to biases which are not common to actuarial technique. Specific biases will be reviewed later. Attention is now turned to the effect of increased clinical experience on the validity of clinical judgments.

Experience and the Accuracy of Clinical Judgment

One of the most frequently examined variables in the clinical judgment literature is therapist experience. The relationship between experience and clinical judgment is also one of the most hotly-debated subfields of judgment research (Spengler, White, Ægisdóttir, Maugherman, et al., 2009). Although some studies have supported the hypothesis that increased experience is related to greater judgment accuracy, most reviewers of the literature have reached the same conclusion as Wiggins (1973): "There is little empirical evidence that justifies the granting of 'expert' status to the clinician on the basis of his [or her] training, experience, or information-processing ability" (p. 131). Howard Garb's (1998) book, *Studying the Clinician,* is generally considered the most comprehensive review of literature related to clinical judgment and experience (Hatfield & Ogles, 2006; Spengler, White, Ægisdóttir, Maugherman, et al., 2009). Garb stated, "among the most provocative results reported in the area of clinical judgment are those that indicate that presumed-expert clinicians are no more accurate than other clinicians" (p. 14), and furthermore that "clinical experience has generally not been related to validity, both when experienced clinicians have been compared to inexperienced clinicians and when clinicians have been compared to graduate students" (p. 110). Similar conclusions have been made so frequently (e.g., Dawes, 1989, 1994; Faust, 1994; Faust & Ziskin, 1988; Garb & Boyle, 2003; Wedding & Faust, 1989) that Lichtenberg (1997) surmised, "The fact that counselors' accuracy of clinical judgment does not increase with experience is now generally acknowledged" (p. 231).

Although the effects of clinical experience on judgment accuracy appear minimal, didactic training does appear to increase accuracy. Aronson and Akmatsu (1981)

evaluated the validity of 12 graduate students' judgments of MMPI profiles across several points in time. Following one year of graduate education in psychology, the mean validity coefficient for students' ratings was .20. However, the validity coefficient increased to .42 following the completion of a course on MMPI interpretation. The students were assessed again following the completion of a one-year assessment and therapy practicum; validity coefficients following the practicum were .44, reflecting no meaningful increase in validity following practicum experience. Garb and Boyle (2003) reviewed a significant body of literature indicating that both clinicians and graduate students outperformed lay judges (e.g., undergraduates, secretaries) on a variety of judgment tasks. However, comparisons between experienced and inexperienced clinicians generally reveal no differences in accuracy. Similarly, comparisons of clinicians with graduate students have not typically yielded significant differences (except in cases in which graduate students were just beginning their training). Garb and Boyle concluded that clinical judgment increases with didactic training relevant to judgment tasks, but that increased clinical experience following training is not consistently related to judgment accuracy.

Despite the common assertion that clinical experience does not correlate with increased validity of judgment, many authors (and clinicians) have been reluctant to accept such a conclusion. At an intuitive level, "It just seems right that clinical experience should beneficially affect clinical judgment and decision making" (Lichtenberg, 2009, p. 410). Citing research on implicit learning, Westen and Weinberger (2004) suggested that "clinicians would be a very peculiar species indeed if they showed no skill development over years of observing and treating psychopathology"

(p. 603). That such beliefs are held by many members of the profession of psychology may be inferred by examining the training model for new clinicians, who are required to complete a number of hours of practice under the supervision of a more experienced clinician before reaching eligibility for licensure. Similarly, the 2005 Presidential Task Force on Evidence-based Practice (Levant, 2005) identified "clinical expertise" as essential to identifying relevant research, integrating it with clinical data, considering patient characteristics, and providing services that have the highest probability of positive outcomes. The Task Force defined clinical expertise as "competence attained by psychologists through education, training, and experience resulting in effective practice" (p. 9). Beutler (1995) referred to the standard training model for psychotherapists as reflecting a "germ theory" of education, in that academic training programs "operate on the assumption that exposure to psychotherapy, through supervision and class instruction, over a finite period of time, will result in competence and expertise" (p. 490).

Some researchers have cited related research as potential evidence for positive effects of experience on clinical judgment. Spengler, White, Ægisdóttir, Maugherman, Anderson, Cook and colleagues (2009) noted that some indirect evidence for the benefit of experience has been reported in counseling psychology literature: as compared to novice counselors, experienced counselors have been found to differ in cognitive dimensions thought to be related to clinical decision making. For example, experienced counselors demonstrated broader knowledge structures, better short- and long-term memory for domain-specific information, greater efficiency of time spent on case conceptualizations, differing quality of schemata related to case material, and greater numbers of concepts generated. Advantages in such dimensions seem likely to result in

improved clinical judgment and decision making. However, as summarized by the various reviewers cited above, empirical studies have often failed to demonstrate such an effect. Garb and Grove (2005) noted that experienced clinicians appear to have some advantages in terms of generating hypotheses and structuring tasks.

Limitations in Experience-Judgment Research

Given the intuitive and theoretical reasons to believe that clinical judgment improves with experience, many authors have argued that the lack of consistent empirical support for such a relationship is a product of methodological deficiencies. Westen and Weinberger (2004) argued that experience in specific domains (e.g., patient populations, assessment measures, decision types) increases the validity of decisions in those domains, but that general clinical experience confers no advantages for domains in which a clinician has limited experience. However, studies of clinical judgment often require clinicians to make judgments regarding novel or analogue tasks with which they do not have previous experience. Holt (1970) noted that research comparing clinical and statistical prediction ought to utilize the best clinicians only, given that they would be competing with the best statistical formulae. Similarly, research examining the effects of experience on clinical judgment should use only clinicians with significant experience relevant to the task demands as part of the "experienced" sample.

Another methodological limitation in many studies of clinical experience and judgment is a restricted range of experience among raters. Skovholt, Rønnestad, and Jennings (1997) noted that the experience difference between "novice" and "experienced" clinicians is often too small, thereby diluting the effects of experience. This problem has similarly been noted in reviews of the effects of experience on outcome. Stein and

Lambert (1984) found that the average experience level in studies of practitioner expertise was 2.9 years. Skovholt and colleagues suggested that 10 to 15 years of experience may be necessary to develop expertise. It is therefore possible that studies utilizing samples that are more differentiated in terms of experience may produce greater discrepancies between groups in relation to judgment accuracy. A similar concern is inconsistency across studies in operationalizing "experience." Some researchers have utilized levels of training (e.g., master's versus doctoral students, graduate students versus professionals) while others have used continuous measures; even in the case of continuous measures, the units of measurement (e.g., years, number of clients seen) have varied. Experience is often measured by single-item measures, which tend to be low in reliability (Spengler, White, Ægisdóttir, & Maugherman, 2009). Several other potential methodological limitations were identified in the context of a recent meta-analysis, which is now described.

Meta-analysis of Experience and Clinical Judgment

Spengler, White, Ægisdóttir, Maugherman, and colleagues (2009) noted that reviews of the effect of clinician experience on clinical judgment generally take a narrative approach, which is subject to impressionistic biases. Spengler and colleagues therefore conducted a meta-analysis in order to examine whether judgments improve with experience. A thorough search strategy revealed 75 studies which provided sufficient information to calculate an effect size for clinical or educational experience related to clinical judgment accuracy by mental health professionals. Judgment types represented in the analysis included judgments of client problem type or severity, diagnosis, recommendations for treatment, and prognosis. After eliminating outliers, the mean

effect size for the relation of experience and the accuracy of judgments was $d = 0.12$. This effect was relatively homogenous in that potential moderator variables (e.g., experience type, experience breadth, and study design) were generally non-significant. However, the benefit of experience on judgment accuracy was more notable for decisions related to diagnosis and treatment recommendations. Additionally, experienced clinicians were better than inexperienced clinicians at decision tasks in which a highly valid criterion of accuracy was not available; this may suggest that experience improves accuracy when highly nuanced or uncertain decisions are required.

Spengler, White, Ægisdóttir, Maugherman, et al. (2009) concluded that their results indicated a small but reliable effect for the experience-accuracy relationship. They suggested that greater clinical or educational experience leads to a 13% increase in decision accuracy. The relatively small effect size found in the meta-analysis suggests that the effect of experience on training may be difficult to demonstrate in individual studies due to limited power. Spengler and colleagues calculated that even an atypically large study with 200 experienced clinicians and 200 novice clinicians would have a power of only .22, assuming an effect size of .12 and an alpha of .05, meaning that results would be statistically significant less than one in four times. Another possibility suggested by Spengler and colleagues is that research examining the relationship between experience and judgment accuracy may suffer due to contesting mediators. For example, experienced clinicians may be more likely to seek feedback, a potential mediator of a positive experience-accuracy relationship, but may also be more likely to engage in confirmatory hypothesis testing, which would be negatively related to judgment accuracy.

Such complex relationships may be examined by a shift toward path and mediation approaches to examining the relationship between experience and accuracy.

A significant limitation in the studies reviewed by Spengler, White, Ægisdóttir, Maugherman, et al. (2009) was that objective feedback was provided to raters in only two of the studies. Therefore, although feedback is frequently suggested to be an important means of improving judgment accuracy, it remains understudied. The influence of feedback on the accuracy of clinical judgment is a focus of this paper. Feedback is therefore discussed in more detail later. Attention is now turned to factors that influence the decision-making process in clinical judgment tasks.

Factors Affecting Clinical Judgment

Several features of human decision-making and cognitive processes have been identified as factors which make it difficult for clinicians to learn from experience. Dawes (1994) asserted, "there are good logical and empirical reasons why experience does not help in this context, even though we may all 'learn from experience' in other contexts" (p. 106). The "other contexts" referred to by Dawes include motor tasks such as riding a bicycle, driving a car, or even performing surgery. Such tasks are characterized by immediate feedback regarding success or failure. Additionally, they cannot be taught solely through verbal instruction, but rather require repetitive experience in order to achieve mastery. In contrast, clinical judgment tasks do not yield immediate physical feedback and involve more cognitive processes. They are therefore subject to cognitive biases and errors. Some of the most common "cognitive errors" related to clinical judgment are reviewed in this section.

Anchoring and Adjustment

Clinical judgment and subsequent decision-making is heavily influenced by the *anchoring-and-adjustment* heuristic, originally described by Tversky and Kahneman (1974): "people make estimates by starting from an initial value that is adjusted to yield the final answer" (p. 1128). Anchoring leads to errors in decision making because anchors are often based on insufficient or arbitrary information. Houts and Galante (1985) found that clinicians formed impressions of videotaped clients quickly, and that their final decisions regarding client status was influenced by those first impressions. Gambrill (2005) noted that clinicians "tend to believe in initial judgments, even when we are aware that the knowledge we have access to has been arbitrarily selected" (p. 232).

The anchoring heuristic influences decisions at each stage of therapy. For example, clinicians may overvalue information gained during the intake process and ignore additional, or even contrary, information revealed during the course of therapy (Turk, Salovey, & Prentice, 1988). It is also likely that anchoring occurs within individual sessions. For example, a client who states early in a therapy session that things have been "really good" may be rated by the clinician as improved from the previous session even if the client reports increased difficulties during the remainder of the hour.

Bias resulting from the anchoring heuristic is compounded by the fact that once new information is attained, corresponding adjustments are generally insufficient. Epley and Gilovich (2006) found that adjustments away from previous anchors tend to terminate once plausible values are reached. Various studies have found that the order in which information is presented influences clinicians' final ratings, even when all

clinicians received identical information by the time they made the ratings (Garb, 1998). For example, Pain and Sharpley (1989) presented "good," "bad," and "neutral" written information about hypothetical clients to clinicians in differing orders and subsequently asked them to rate the clients' global functioning. They found that when negative information was presented first it overshadowed good information presented later.

Confirmatory Bias

A related consideration is the *confirmatory bias*, the tendency to seek and overweigh evidence that supports our beliefs and to ignore and underweigh contradictory evidence (Gambrill, 2005). Confirmatory bias is an error of both memory processes as well as behavioral strategies. For example, after making a diagnosis, a clinician may selectively remember information that supports the judgment. Furthermore, the clinician may engage in assessment strategies that are more likely to confirm than to disconfirm the original judgment, a tendency known as *confirmatory hypothesis testing*. A dramatic example of confirmatory bias is Temerlin's (1968) study. An actor was commissioned to portray on videotape a happy and self-confident with no ostensible signs of maladjustment or distress. In the first condition, raters (psychologists and psychiatrists) had access to a senior clinician's suggestion that the man was a "healthy individual." Raters in this condition uniformly agreed that the man on the tape was, in fact, healthy. In a second condition, the senior clinician's suggestion was that the man appeared neurotic but was actually psychotic. Following viewing of the tape, only 6% of raters in the second condition considered that man to be healthy.

Strohmer, Shivy, and Chiodo (1990) gave three versions of a case history to master's degree counselors. One version of the history contained an equal number of

phrases describing good self-control and phrases describing a lack of self-control. The other two case histories included either more phrases denoting good self-control or more phrases indicating a lack of self-control. One week later, counselors working with the hypothesis that the client lacked self-control remembered more information that was consistent with this hypothesis, even when a greater number of contradictory statements had been made.

Pfeiffer, Whelan, and Martin (2000) examined the hypothesis-testing strategies of 72 psychology doctoral students. Participants were given a referral from a physician for a mock client who they would later view on videotape in an initial psychotherapy session. Referrals provided either a highly-plausible preliminary diagnosis (i.e., one that was consistent with content of the videotape), a low-plausibility diagnosis (i.e., largely inconsistent with the content of the videotape), or did not include a diagnosis. After reading the referral and viewing the videotape, students reported their preliminary hypotheses (diagnoses). They were then asked to recall nonverbal cues that they had attended to during the videotape and to provide 5 questions that they would ask the client if they were to continue with the client in psychotherapy. Participants who had received a highly-plausible initial hypothesis (diagnosis) and therapists who formulated their own hypotheses independently were more likely than those in the low-plausibility condition to utilize confirmatory strategies in their attending to cues as well as their follow-up questions.

Similarly, Owen (2008) asked 97 mental health counselors-in-training to review a hypothetical case study and to report their preliminary diagnostic impressions. Counselors were then asked to generate questions they would use in continued

assessment of the client. As hypothesized, counselors generated more confirmatory than disconfirmatory questions. Additionally, questions that were confirmatory tended to have more diagnostic clarity and more specificity than disconfirmatory questions, leading Owen to suggest that the counselors were better at developing questions that were likely to confirm their initial hypotheses. In the context of judging client progress during the course of therapy, it seems likely that clinicians who form the initial impression during a session that a client has improved will pursue topics related to client improvement; in contrast, the early impression that a client is worse-off may lead the clinician to ask about current difficulties in the client's life. These strategies would elicit information consistent with preliminary hypotheses, which may lead clinicians to gain an inaccurate view of client status and progress.

Hindsight Bias

Hindsight bias refers to "the tendency to believe, once the outcome is known, that the outcome could have been predicted more easily than is actually the case" (Wedding & Faust, 1989, p. 237). Knowledge of an outcome encourages the view that it was inevitable (Gambrill, 2005). Furthermore, when individuals engage in hindsight bias they tend to assume a direct relationship between the observed outcome and events or conditions prior to the outcome; explanations of an observed event are often therefore too deterministic. Hindsight bias is similar to confirmatory bias in that it leads to the overweighing of information that is consistent with the outcome.

Garb (1989) noted that a deterministic outlook can obstruct clinicians from learning by experience. For example, if clinicians construct deterministic explanations each time they receive feedback about a client, they will create incomplete explanations

given that they do not have all relevant information. Additionally, feedback may be influenced by measurement error, thereby causing clinicians to draw incorrect inferences. Dawes (1989, 1994) suggested that although biases in recall make it difficult for clinicians to learn from experience, they may also create an illusory correlation in which clinicians believe that the quality of their judgments increases with experience. For example, a clinician who notices that many of his depressed clients reported unhappy childhood memories may make deterministic assumptions about the influence of unhappy childhood events. He may therefore feel more confident in diagnosing future clients who report such events as depressed on the basis of his clinical experience. However, doing so would overlook the fact that many individuals with unpleasant childhood memories are not depressed.

Availability Heuristic

Confirmatory and hindsight biases are related to the *availability heuristic*, which occurs when clinicians are influenced by the ease with which instances or occurrences can be remembered (Garb, 1998). The ease of remembering instances varies depending upon factors such as retrievability, vividness of the instance, and the strength of association between two events. Tversky and Kahneman (1974) noted that the availability heuristic can account for illusory-correlation effects because when the cognitive association between two events is strong, one is likely to conclude that the events have been frequently paired. In the example in the previous paragraph, the availability of instances in which depression was paired with unhappy childhood memories may lead the clinician to perceive an illusory correlation between the two events. Additionally, the clinician is drawing only from an available sample of

individuals with depression, which is not representative of all individuals who had unhappy childhood experiences.

Biases Based on Client Characteristics

In addition to general cognitive biases, clinicians' decisions may be influenced by characteristics of clients themselves. Garb (1998) reviewed extant research regarding the extent to which clinical judgment is affected but various specific factors, only a few of which are summarized here. Garb concluded that client race was not generally associated with treatment decisions or with ratings of psychiatric symptoms and personality traits. However, some results suggested that White clinicians may be more prone than Asian American clinicians to describe Asian American clients as depressed or withdrawn (Li-Repac, 1980; Tseng, McDermot, Ogino, & Ebata, 1982). Garb reported some evidence for social class bias related to both personality assessment and treatment decisions. In cases in which bias was found, results most often indicated that lower class clients were rated more negatively than middle-class clients. Regarding treatment, lower class clients were less likely to receive psychotherapy than middle-class clients and were judged to be less likely to benefit from therapy. Lower class clients were also more likely to receive supportive, rather than insight-oriented, therapy.

Lichtenberg (1997) noted that clinicians share some of the same sex-role and sex-norm biases as society as a whole; such biases are likely to influence clinical decision making. Rosenfield (1982) examined archival data for admissions to a psychiatric hospital and found that for individuals diagnosed with personalities and substance-use disorders (considered to be associated with being male), 50% of females were hospitalized but only 18% of males. In contrast, among individuals diagnosed with

neurosis or depression (more often diagnosed in females), 66% of men and only 43% of women were hospitalized. Referrals for type of therapy may also be influenced by client gender. Two studies (Bowman, 1982; Fernbach, Winstead, & Derlega, 1989) indicated that men may be more likely to be recommended for couple therapy and group therapy, while women may be more likely to be referred for individual insight-oriented therapy.

Awareness of Cognitive Processes

The application of clinical judgment requires the clinician to carefully examine all available information and to assign relative weighting to the various sources in order to reach a decision. This type of complex integration of data, also referred to as *configural analysis*, is difficult to manage. Faust and Ziskin (1988) suggested that although clinicians believe that their decisions rest on a careful weighing of many variables, evidence suggests that only a few variables (two or three) make a significant impact in such decisions. Awareness of potential sources of bias and error is therefore an important factor in making these complex decisions. However, research indicates that clinicians frequently may be unaware of how they make judgments (Garb, 1998).

Rock (1994) surveyed 106 mental health professionals (mostly doctoral-level psychologists) in order to elicit their attitudes and knowledge regarding clinical judgment research. Although respondents largely agreed that judgment research was important and that it had meaningful applications for clinical practice, they reported a low level of personal familiarity with judgment literature. The average respondent had not read any of the (then) most significant books or articles on clinical judgment. It is therefore apparent that limited awareness of one's own internal processes, as well as limited

knowledge of cognitive biases in general, may further limit the ability to make complex judgments.

Confidence

An important factor related to clinical judgment is confidence. Decisions are based in part on the degree to which the decision-maker is confident that a chosen course of action will lead to a desired outcome (Stankov, Lee, & Paek, 2009). A large body of research suggests that individuals tend to be overconfident regarding the accuracy of their beliefs. This effect has been noted to have a wide effect in various financial, political, legal, and clinical contexts. Plous (1993) suggested that "no problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence" (p. 217). Awareness of the limitations of one's judgments is especially important for mental health practitioners, who are called upon to make important decisions in treatment, assessment, and forensic settings. In the context of assessing client progress from week to week, a clinician's overconfidence that a client has improved could lead to the premature termination of therapy that would still be beneficial. It could also deter the clinician from pursuing changes to the treatment plan in cases in which the client is experiencing little benefit. Additionally, clinicians who become overconfident fail to seek information that would allow them to revise their judgments (Einhorn, 1980).

Are Clinicians Overconfident?

Garb (1986, 1998) reviewed the accuracy ("appropriateness") of confidence ratings made by clinicians in clinical judgment studies. The majority of studies in this area have examined clinicians' confidence in their diagnoses. Garb concluded that

confidence often (but inconsistently) exhibits a weak positive relationship with the validity of judgments. Although validity and confidence are positively related in general, research has demonstrated that clinicians may frequently be overconfident. For example, Gaudette (1992) found neuropsychologists to be accurate on 62% of diagnoses related to potential cerebral impairment; neuropsychologists in the study had estimated their hit rate to be 77.5%. Arkes (1981) noted that the most confident clinicians tend to be the least accurate.

It should be noted that a positive overall correlation between confidence and validity does not ensure that raters are well *calibrated.* A judge is well calibrated if, over the long run, for all judgments assigned the same probability of being accurate, the proportion accurate is equal to the probability assigned (Lichtenstein & Fischoff, 1977). Therefore, if judgments are correct half of the time that a clinician rates his or her likelihood of being correct as .5, the clinician is well calibrated. Calibration, a common measure of the appropriateness of confidence ratings, is therefore not strictly a measure of the correlation between individual confidence estimates and the accuracy of related judgments.

Experience and Confidence

As opposed to the relationship between clinical experience and the validity of judgments, the relationship between clinician experience and the appropriateness of confidence ratings is generally established to be positive, albeit weakly. Garb (1986) summarized that experienced clinicians tend to make more appropriate confidence ratings than inexperienced clinicians as long as the information used to make decisions is valid. The relatively increased accuracy of confidence ratings by experienced clinicians has

been observed in judgments based on neuropsychological data, biographical data, objective personality measure data, and information observed in psychotherapy sessions. These findings are interesting given that research in other fields has indicated that experts, as compared to non-experts, are as likely or more likely to be overconfident (e.g., Lichtenstein & Fischoff, 1977; Mahajan, 1992; McKenzie, Liersch, & Yaniv, 2008). The reasons for the apparent benefit of experience are unclear. It is possible that clinicians gain appreciation for the complexity of decisions with experience. However, such a possibility is somewhat contrary to research regarding the effect of cognitive biases over time.

The Influence of Additional Information on Confidence

Oskamp (1965) examined overconfidence in judges consisting of graduate and undergraduate students as well as psychologists. Judges were given a case study and asked to answer multiple-choice questions about the case based on the information provided. They also indicated their level of confidence in each answer. After the first round of questions, the judges were provided with additional information regarding the case and were administered the same questions, with confidence ratings. This procedure was repeated across four stages. Oskamp found that clinicians' confidence increased at each stage as they received additional data. However, accuracy did not significantly increase from the first to the last stage. These results suggest that clinicians may become increasingly confident with more information. It is noteworthy that the additional information given at each round was consistent with previous information and added little new details.

Trueblood and Binder (1997) provided neuropsychological testing data to a sample of neuropsychologists and asked them to make diagnostic judgments. Some of the neuropsychologists were given results from forced-choice tests used to detect malingering. Judges who arrived at a diagnosis of malingering and who received the forced-choice test data were significantly more confident in their diagnoses than those who diagnosed malingering but without the use of such data.

In contrast to Oskamp's (1965) and Trueblood and Binder's (1997) results, Peterson and Pitz (1986) found that individuals demonstrated less overconfidence in decisions as the amount of information available increased. The decrease in overconfidence appeared to occur because as increased amounts of information were available, the accuracy of predictions increased while certainty about the precision of judgments decreased. Based on these results, Hatfield (2007) suggested that additional information may only increase confidence when it is consistent with previous information. Hatfield mailed brief paragraphs about the treatment progress of a hypothetical client to a large number of practicing psychologists. Among the paragraphs, the availability of information derived from client verbal reports and from outcome measures was systematically varied. Additionally, the content of verbal reports and outcome measure reports was varied to indicate progress or a lack of progress. After reviewing the information, clinicians indicated their treatment recommendations and rated their confidence about the accuracy of those decisions. Hatfield hypothesized that therapists who had access to outcome data and client verbal reports that were consistent would be more confident than those who had access to only one data source, but that therapists would be less confident in their estimations of client progress in cases of contradiction

between verbal reports and outcome data. However, no systematic differences in confidence were found among levels of the amount of information available to therapists. The effects of the amount of information available on confidence in clinical judgments remain unclear.

Improving Clinical Judgment

The judgment literature reviewed above indicates that clinical judgment is valid and reliable at times but is also vulnerable to a variety of errors due to its subjective and informal method of data integration. Decisions based on clinical judgment have generally been less valid than decisions made through the use of statistical algorithms. Additionally, clinical judgment appears to improve only marginally with increasing clinical experience. Given these conclusions, various researchers have suggested that it is essential to move beyond general examinations of the accuracy of judgment. Subsequent steps in judgment research include identifying conditions under which clinical judgment is optimized, developing a comprehensive theory of clinical judgment (e.g., Ridley & Shaw-Ridely, 2009), and developing methods of improving clinical judgment. Garb (1998) asserted that "improving the accuracy of judgments is the ultimate goal of judgment research" (p. 3). Although the general effect of training is small, individual clinicians appear to vary significantly in the accuracy of their judgments. For example, Haderlie and Kern (2009) found that the correlations between therapists' estimates of client change and an objective change measure varied from -.18 to .31. It is therefore important to consider methods of improving the judgment abilities of individual

therapists. An increasing amount of literature has therefore been devoted to suggesting means of increasing judgment accuracy. Some of these are now reviewed.

Judgment Task Considerations

The accuracy of clinical judgment is dependent upon the nature of the decision to be made. Some decisions, such as the prediction of violence or suicide, are extremely difficult (Garb, 1998). Such tasks may not be appropriate targets for clinical judgment, given that standardized decision algorithms tend to be much more accurate. Garb suggested that making ratings for difficult tasks may be not only dangerous, but unethical. Gambrill (2005) noted that judgment may at times be more difficult due to situational factors such as time pressures, limited resources, and conflicting goals. Such context factors should be attended to and addressed to the extent possible.

Another consideration related to judgment tasks is the data sources available. Several authors (e.g., Bell & Mellor, 2009; Garb, 1986) have noted that, like any assessment procedure, the validity of decisions based on clinical judgment is restricted by the reliability and validity of the information obtained. This observation is supported by the fact that results of judgment studies have been significantly more favorable when decisions were based on objective, as compared to projective, personality data. Furthermore, clinical judgment may vary based on the way in which clinical observations are aggregated. Westen and Weinberger (2004) suggested that clinical judgment may be most valid when it is used to rate and code (quantify) various sources of information, with these ratings then aggregated through more formal, statistical, procedures.

Suggestions for Clinicians

Clinicians may improve the validity of their judgments through careful attention to elements of a scientist-practitioner approach (Bell & Mellor, 2009; Spengler, White, Ægisdóttir, Maugherman, et al., 2009). For example, clinicians should attend to empirical research in order to select validated approaches to assessment and treatment. Garb (1998) suggested that when sound empirical research has been conducted, clinicians should weight it more heavily than their own clinical experience. Additionally, various authors have noted that clinicians should gain familiarity with literature regarding decision-making theories, social cognition, and cognitive biases (e.g., Nisbett & Ross, 1980; Tversky & Kahneman, 1974).

In addition to increasing their familiarity with relevant literature, clinicians may improve their clinical judgment by incorporating debiasing strategies into their decision making (Arkes, 1981). The most frequently-suggested strategy is to consider alternatives prior to making a decision. For example, a clinician who believes a client is depressed should consider alternative diagnoses and seek disconfirmatory evidence. Similarly, clinicians should consider client strengths in addition to deficits in order to form a more valid view of the client (Garb, 1998). Specific instruction to consider alternative explanations has been found to significantly reduce hindsight bias (Arkes, Faust, Gulmette, & Hart, 1988; Tutin, 1993). Hindsight bias may also be reduced by decreasing reliance on memory. Arkes noted that observations that are not recorded accurately and in detail at the time they are made can often be ignored due to hindsight bias. In other words, a clinician's beliefs influence the information that is remembered unless it is documented.

Training Considerations

The limited effects of experience on the validity of clinical judgments point to the essential role played by training in developing clinical judgment abilities. Didactic training may be especially useful in increasing the validity of clinical judgment in psychology, given that training effects have been significantly greater than experience effects on judgment accuracy. The suggestions noted above for clinicians may also be incorporated into training within graduate programs. In addition to skills necessary to make accurate judgments in certain tasks (e.g., interpretation of an MMPI profile), training programs should focus on the development of generalizable skills and attitude that will enhance clinical judgment among their graduates. Although introducing students to judgment research is an important component, Arkes (1981) noted that simply describing cognitive biases and encouraging individuals not to be influenced by them is not effective. Garb (1998) suggested that graduate programs might increase the accuracy of diagnostic decisions by training students to use instruments that are highly valid for that purpose, such as semistructured and structured interviews.

Spengler, Strohmer, Dixon, and Shivy (1995) developed a scientist-practitioner model of psychological assessment that was designed to reduce errors in judgment. Their model involves ongoing observation, inference, and hypothesis testing, which are utilized to develop a tentative and fluid conceptualization of the client. The model also focuses on developing scientific attitudes such as openness, self-awareness, and curiosity. Competing hypotheses are tested through multiple methods. Finally, Spengler and colleagues' model provides mnemonics designed to promote debiasing activites. These include the following: 1) consider probability and base rate data, 2) combine actuarial and

clinical prediction techniques, 3) delay judgments, 4) reduce overconfidence in one interpretation of the data, and 5) invoke a cognitively complex approach to data interpretation. Preliminary evidence suggests that such didactic techniques may increase students' confidence and abilities related to case conceptualizations and assessment (Meier, 1999).

In addition to increasing judgment accuracy, another potential goal for training programs regarding clinical judgment is to reduce overconfidence. Smith and Agate (2004) developed a 90-minute instructional module for trainees in a counseling program. Trainees were assigned to treatment and control groups. Those in the treatment group completed the instructional module in small group settings. The module included a hands-on judgment task, group discussion of the process by which decisions were made, and didactic instruction regarding cognitive heuristics. Although all trainees demonstrated overconfidence at pretest and posttest, confidence scores were found to decrease in the treatment group.

Decision-making Aids

Various aids are available to clinicians in making decisions. Specific aids depend upon the nature of the judgment task. An important consideration across judgment tasks is the consideration of base rates. Base rate information, such as that applied to statistical algorithms, can greatly enhance the validity of judgments. Arkes (1981) provided an instructive example: If the probability of a rare condition (e.g., "multiple personality") is 1 out of 100,000 in the general population, the base rate for the condition is .00001. Suppose a remarkably accurate test for the condition were available, such that a "positive" result for the test occurred 100 times more frequently in individuals with multiple

personality disorder than it did among individuals without the condition. A positive test therefore results in a *likelihood ratio* of 100/1, which would lead most clinicians to assume that a positive test almost proves the presence of multiple personality. However, in order to compute the true probability that a positive test is associated with multiple personality disorder for an individual in the general population, the likelihood ratio must be multiplied by the base rate (100 x .00001), resulting in a probability of only .1% even given a positive test result. The consideration of base rates therefore becomes increasingly important when base rates are extreme.

Similarly, clinicians should utilize statistical algorithms to the extent available. Local norms may be developed in order to establish base rates that are most valid in a given setting. Such activity is exemplary of the local clinical scientist model (Stricker, 2002). Clinicians should utilize decision aids such as diagnostic criteria, which improve interrater reliability and reduce the effects of bias (Garb, 1998). The application of many statistical decision aids has become more feasible across a variety of settings with advances in computers and statistical software.

Feedback

Reviewers of clinical judgment literature have almost universally emphasized the importance of feedback regarding decision accuracy. Meehl (1954) noted that if clinicians do not receive feedback about their decisions, the validity of judgment cannot be improved. This sentiment has been repeatedly echoed in more recent literature (e.g., Dawes, Faust, & Meehl, 1989; Garb & Boyle, 2003; Lichtenberg, 1997; Spengler, White, Ægisdóttir, Maugherman, et al., 2009). Garb (1998) suggested that "one reason why it can be difficult for mental health professionals to learn from their experiences is that they

do not receive feedback for some tasks" (p. 201). Dawes (1994) described a *terminal insight* model to suggest the central nature of feedback to true learning: individuals making categorical judgments will utilize a specific sorting rule until they are told that they are incorrect. They then abandon the original rule and try out another, until they identify the correct sorting rule and "stick with it." Dawes noted that "the terminal insight explanation implies that subjects will try new ways of sorting by distinguishing characteristics *only after they receive feedback that they have made an error*" (p. 114, italics in the original).

Despite the central importance afforded feedback in literature regarding the enhancement of clinical judgment, feedback in this context has been remarkably understudied. Spengler, White, Ægisdóttir, and Maugherman (2009) described the results of their literature review using the terms *clinical judgment* and *feedback*: "We found only a limited number of examples where feedback had been investigated in mental health decision making (we found several examples in medical decision making)" (p. 419). Spengler and colleagues concluded that "systematic research is sorely needed on methods and types of feedback" (p. 419). Given that it is a primary focus of the current research, the topic of feedback is now described in more detail.

Feedback on Clinical Judgment

The study of feedback originated in the fields of cybernetics and engineering and was adopted to the study of human behavior in the mid-20th century (Claiborn & Goodyear, 2005). In the context of psychology, feedback is defined as information that is provided to a person, from an external source, about that person's behavior or its effects

(Claiborn, Goodyear, & Horner, 2001). Although feedback is inherent in all human interaction, research on feedback has focused on that feedback which is deliberately provided with a certain objective, generally to improve performance in the behavior for which feedback is provided. Such deliberate feedback has been termed *feedback intervention* (Kluger & DeNisi, 1996). Deliberate feedback may be purely descriptive, involving a description of behavior with limited inference about implications of the behavior (Claiborn & Goodyear, 2005). However, in the context of enhancing clinical judgment, feedback is necessary evaluative. Evaluative feedback offers an assessment of behavior (in this case the decision made) in relation to some criterion (the validity of the decision). Feedback may be indirectly evaluative when it is provided regarding judgment tasks. For example, a clinician may estimate that a client has made progress during the course of therapy and then receive feedback indicating that the client has, in fact, deteriorated. In this case, feedback does not directly evaluate the clinician's judgment but nevertheless provides correction because it is compared by the clinician to his or her previous judgment.

Historically, it was widely assumed that the provision of feedback always results in improved performance. Latham and Locke (1991) noted that "few concepts in psychology have been written about more uncritically and incorrectly than that of feedback" (p. 224). However, researchers in the past thirty years have recognized that feedback interventions produce variable results. Kluger and DeNisi (1996) conducted a meta-analysis of data from 131 feedback studies, which included 607 effect sizes and reflected the data of more than 12,000 participants who received evaluative feedback for a variety of tasks. They found that over one third of the individual effect sizes were

negative (i.e., indicated that feedback had resulted in a deterioration of performance). Therefore, although the overall effects of feedback were positive and reflected a medium effect (d = .41) on performance, individual effects are not uniformly positive.

The effectiveness of feedback is strongly influenced by the context and manner in which feedback is delivered. Kluger and DeNisi (1996) found that the effectiveness of feedback interventions were augmented by increased frequency of feedback. Frequent, continuous provision of feedback allows the receiver to monitor the value of alterations and to continually refine the rules used to make decisions. Similarly, feedback is most effective when it is delivered as quickly as possible in relation to the judgment for which feedback is given. Feedback is also influenced by the validity of the criteria against which judgments are compared. Therefore, it is crucial that measures used to evaluate the validity of judgments be sensitive and valid themselves (Gambrill, 2005).

Effects of Feedback on the Validity of Judgments

As noted above, very little research has systematically examined the effects of feedback on the validity of clinical judgments. Goldberg and Rorer (1965, cited in Goldberg, 1968) studied the effects of training with feedback on judges' abilities to make valid differential diagnoses between neurosis and psychosis on the basis MMPI profiles. Their sample included 3 experienced psychologists, 10 graduate students in psychology, and 10 "naïve" participants (non-psychologists with no prior training in the MMPI). All participants engaged in extensive training in which they examined a profile, made a diagnosis, and turned the profile over to see the criterion diagnosis. Over 300 profiles were used, multiple times each. By the conclusion of the training period, judges had received feedback for the profiles more than 4,000 times each. All groups demonstrated

increased validity on the training profiles. However, on testing profiles (those not included in the training), Goldberg reported an increase in judgmental accuracy from 52% to approximately 58% for naïve judges; psychologists and graduate students demonstrated 65% accuracy at the beginning of training and did not experience gains following 17 weeks of training. Training therefore did not generalize beyond the sample of training profiles.

Graham (1971) conducted one of the only clinical judgment studies in which the availability of feedback was systematically varied. As in Goldberg (1968), the judgment task was to identify a Minnesota Multiphasic Personality Inventory (MMPI; Hathway & McKinley, 1943) profile as neurotic or psychotic. Participants included 21 clinical psychologists, 21 graduate students in clinical psychology who had completed coursework involving the MMPI, and 21 undergraduate students with no prior MMPI experience. Each participant made diagnoses for profiles in blocks of 10, with neurotic and psychotic profiles evenly divided in each block. Participants in one condition simply made diagnoses and received no feedback. Participants in the other two conditions turned profiles over following diagnosis in order to receive feedback; judges in one of the conditions received correct feedback, while those in the second feedback condition received random feedback (which was accurate only 67% of the time). Hit rates for PhD psychologists were 72% when they received correct feedback, 57% when they received random feedback, and 52% (just better than chance!) when receiving no feedback. Graduate student hit rates were 61% when receiving correct feedback and 58% in the other two conditions. Finally, undergraduates achieved a hit rate of 57% when receiving

correct feedback, 55% with random feedback, and 50% with no feedback. Receiving correct feedback therefore led to increases in judgment accuracy for each group.

The Nature of Feedback Regarding Client Progress

Therapists receive nearly constant feedback regarding their clients' progress in therapy. Client verbal reports (e.g., "I don't think what we're doing is helping me") and behaviors in session provide continuous data which may be integrated by the clinician to make judgments regarding client progress. Westen and Weinberger (2004) suggested that "psychotherapists tend to have much more direct and immediate feedback than most other medical practitioners, who may prescribe a medication or perform a procedure and not see the patient again for a year" (p. 603). However, feedback such as client verbal reports and behavioral observations are highly prone to a variety of biases on the part of both the clinician and the client. Garb and Grove (2005) argued that psychologists generally do not receive the same type of highly valid feedback that physicians, for example, receive through laboratory tests. Garb and Grove dismissed the type of verbal report feedback described by Westen and Weinberger by suggesting that "astrologists weigh the same types of feedback when they decide whether their interpretations are correct" (p. 658). Feedback such as client verbal reports is subject to socially desirable responding; many clients are likely hesitant to verbalize feelings that they are not progressing for fear of either hurting the therapist's feelings or personal discomfort discussing the topic. Furthermore, clinicians' interpretations of client reports may be biased due to processes such as confirmatory bias. Finally, feedback received regarding client progress may be misleading due to sampling bias, if clients who are not experiencing progress drop out of therapy (Garb, 1998).

Outcome Measures as a Source of Feedback

Due to the limited nature of standard sources of feedback regarding client progress in psychotherapy, researchers have increasingly sought to apply more standardized methods to the evaluation of client change throughout treatment. A particularly promising development is the use of outcome measures to monitor change at frequent intervals throughout therapy, rather than simply as pre- and post-treatment assessments. Although psychotherapists traditionally resisted the use of outcome measures in clinical practice (Gilbody, House, & Sheldon, 2002), they are quickly becoming more widely used. Hatfield and Ogles (2004) found that 37% of practitioners responding to their survey used outcome measures in some form or another. This number represented an increase as compared to similar surveys only a few years prior, in which the proportions of clinicians employing outcome measures were 29% (Phelps, Eisman, & Kohout, 1998) and 23% (Bickman et al., 2000). Outcome measure usage is even more common at psychology department training clinics (56%; Tyler, 2002) and internship training sites (47%; Mours, Campbell, Gathercoal, & Peterson, 2009), suggesting that the proportion of clinicians using such techniques will continue to grow. The increased use of outcome measures in therapy has developed from factors both within the field of psychology (e.g., advances in research data and methodologies) and outside of it (e.g., the requirements of many agencies and managed health care companies that clinicians provide evidence of treatment effectiveness).

The use of outcome measures as a means of monitoring client change over the course of therapy and providing feedback to therapists is one of the most significant recent trends in psychotherapy research and practice. Various systems of outcome

measurement have been developed within the last ten years (e.g., Kordy, Hannover, & Richard, 2001; Lueger et al., 2001; Miller, Duncan, Sorrell, & Brown, 2005; reviewed in Beutler, 2001). Much of the rapid increase in the usage of such techniques is due to evidence that providing clinicians with feedback from outcome measures may result in enhanced outcomes for their clients, especially in cases where clients are not progressing in treatment. For example, Brodey, Cuffel, McCulloch, Tani, Maruish, Brody, and Unutzer (2005) randomly assigned 1374 clients in a managed behavioral healthcare system to feedback or control conditions. All patients completed 11 items from the Symptom Checklist-90 (SCL-90; Derogatis, 1983) at intake and again 6 weeks later. Feedback regarding the results was only provided to therapists of patients in the feedback condition. The clients whose clinicians received feedback from their SCL-90 showed significantly greater improvement in total symptoms than clients whose clinicians did not receive feedback. Such a result is impressive given the minimal nature of the feedback data and the small number of administrations.

The use of outcome measures as a means of client monitoring and obtaining feedback also appears to have benefits in terms of the efficient use of community and professional resources. Slade, McCrone, Kuipers, Leese, Cahill, Parabiaghi, and colleagues (2006) administered monthly outcome and alliance measures to therapists and clients at a community outpatient clinic in London. Results of the measures were provided to a randomized half of therapists and clients. Slade and colleagues observed that clients in the feedback group averaged significantly fewer days as psychiatric inpatients over the course of the study (3.5 as compared to 16.4 in the control group). They concluded that the reduced inpatient care usage made the feedback a cost-effective

intervention. Similarly, Percevic (2002; cited in Percevic, Lambert, & Kordy, 2004) provided feedback to therapists of a randomly-selected group of clients, while providing no feedback on other clients. Clients in the feedback condition demonstrated reductions in their mean length of therapy before discharge at a clinically significant improved condition (46 days as compared to 57 days without feedback), suggesting that the provision of feedback may have fostered more rapid improvement during the course of treatment.

Although the effectiveness of outcome feedback as a means of enhancing outcomes has been well demonstrated, little is known regarding the mechanisms that lead to such improvements (Newnham & Page, 2007). One possibility is that receiving actuarial feedback regarding client outcomes may improve clinicians' judgments about treatment effectiveness when clients are not making gains (Spengler, White, Ægisdóttir, & Maugherman, 2009). This possibility will be examined in the current research utilizing the Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004). The OQ-45 and its associated feedback system, as well as relevant research results, are therefore reviewed in more detail.

The Outcome Questionnaire-45

The OQ-45 (Lambert et al., 2004) is a 45-item self-report questionnaire that measures general psychological distress and dysfunction. Originally developed in 1996 (Lambert et al., 1996), the OQ-45 has rapidly increased in popularity among clinicians and researchers. Recent surveys indicate that it is commonly administered at approximately 18% of psychology internship sites (Mours, Campbell, Gathercoal, &

Peterson, 2009) and 20% of psychology training clinics (Tyler, 2002). Administration of the OQ-45 requires approximately 5 minutes. Patients rate each of the items (e.g., "I feel hopeless about the future") on a 5-point Likert scale ranging from 0 (*never true*) to 4 (*almost always*), in regard to the prior week. The 45 items yield a Total Score ranging from 0 to 180, with higher scores indicating poorer functioning. The OQ-45 also includes three subscales, which are based on Lambert's (1983) conceptualization of psychotherapeutic progress as consisting of three domains of interest: 1) subjective discomfort, 2) the quality of interpersonal relationships, and 3) social role performance. However, examinations of the factor structure of the OQ-45 have provided limited support for a three-factor model (de Jong et al., 2007; Mueller, Lambert, & Burlingame, 1998). Umphress, Lambert, Smart, Barlow, and Clouse (1997) found high intercorrelations among the subscales, suggesting that they may represent variance of a single factor. A recent examination of the factor structure of an Italian translation of the OQ-45 (Lo Coco, Chiappelli, Bensi, Gullo, Prestan, & Lambert, 2008) found support for a bi-level solution with one general factor and three second-order factors (representing the three OQ-45 subscales).

Psychometric Properties

Test-retest reliability for the OQ-45 is adequate (.84 for the Total Score, .78 to .82 for the subscales) and decreases with increasing time intervals (Lambert et al., 1996). Internal consistency estimates are good, ranging from .90 to .93 for the Total Score (Lambert et al., 1996; Vermeersch et al., 2004). Umphress, Lambert, Smart, Barlow, and Clouse (1997) examined the construct and criterion validity of the OQ-45 by comparing OQ-45 scores among individuals with various diagnoses. They found that psychiatric

patients scored higher on the Total Score and Symptom Distress scale than did nonpatient samples. Among patient samples, those with Axis I diagnoses received higher scores than those with V-code diagnoses, suggesting that higher scores are indeed associated with greater distress and dysfunction. Umphress and colleagues also reported correlations between the OQ-45 and a variety of self-report scales (e.g., Beck Depression Inventory, State-Trait Anxiety Inventory), with coefficients ranging from 0.53 to 0.86.

Because the OQ-45 is often used to monitor client change during therapy, the sensitivity of the measure to change is important. Vermeersch and colleagues (Vermeersch, Lambert, & Burlingame, 2000; Vermeersch et al., 2004) have found the Total Score and all subscale to be sensitive in reflecting change experienced by clients following treatment. Additionally, 34 of the 45 individual items were found to be sensitive to treatment effects.

Clinical Significance and Reliable Change

The clinical cutoff score for the OQ-45 Total Score is 63/64 (Lambert et al., 2004). Scores of 63 or below are considered to be in the normative range (reflective of the level of distress typically experienced by individuals in the general population), whereas scores of 64 or above are classified as being in the clinical range (reflective of a level of distress experienced by individuals seeking treatment). This cutoff score was derived using Jacobson and Truax's (1991) methodology, which is the most commonly used method of calculating clinical cutoff scores (Ogles, Lunnen, & Bonesteel, 2001). Speer and Greenbaum (1995) performed a comparative analysis of several existing methods and endorsed the Jacobson and Truax approach due to its unambiguous calculation and supporting literature base. The cutoff value, described by Jacobson and Truax as *Cutoff*

C, represents a weighted midpoint between the means of a functional and dysfunctional sample.

Lambert and colleagues (2004) also applied Jacobson and Truax's (1991) formulae to calculate a reliable change index (RCI) for the OQ-45. The RCI represents the magnitude of change in any direction necessary to be considered reliable (i.e., not due to chance variation). The RCI is a function of the standard error of measurement, such that measures of increasing reliability require smaller magnitudes of change in order for the change to be classified as reliable. The RCI for the OQ-45 Total Score is 14. Therefore, clients whose Total Score decreases by 14 or more points are considered *reliably improved*, whereas clients demonstrating an increase of 14 points or more are classified as *reliably worsened* or *deteriorated*. Clients who begin therapy in the dysfunctional range (64 or above), demonstrate reliable improvement, and terminate therapy in the functional range (63 or below) are classified as *recovered*. Finally, clients whose Total Scores do not change in any direction by at least 14 points are classified as having made *no change*.

Jacobson and Truax's (1991) formulae, or similar methods, may be applied to determine clinical cutoffs and RCIs for any continuous-scale measure that has sufficient normative data. Distinct measures may therefore produce inconsistent classifications of clinical significance. Beckstead, Hatch, Lambert, Eggett, Goates, and Vermeersch (2003) examined the degree to which classifications for clinical significance based on the OQ-45 were consistent with those based on other common outcome measures (e.g., Symptom Checklist-90-R, Derogatis, 1983; Quality of Life Inventory, Frisch, Cornell, Villanueva, & Retzlaff, 1992; Client Satisfaction Questionnaire-8, Larson, Attkisson, Hargreaves, &

Nguyen, 1979). The average correspondence among measure classifications of patients in the functional or dysfunctional range was found to be 85%. Similarly, agreement for classification of patients as meeting criteria for clinically significant change among the measures averaged 65%. Classification into categories of clinical significance therefore appears to be similar across outcome measures, but to vary as a result of the instrument utilized. Beckstead and colleagues noted that the OQ-45 was most similar to the Symptom Checklist-90-Revised (Derogatis, 1983).

Feedback with the OQ-45

An innovative feature of the OQ-45 is a supporting software application that allows for the provision of rapid and standardized feedback to therapists regarding their clients' scores. The software, *OQ-Analyst* (www.oqmeasures.com), provides information regarding clients' current scores, distress level, and the presence of any critical item endorsement (e.g., endorsement of suicidal ideation). Scores are also tracked over time in order to monitor client change. The decision rules and format of the feedback is now described further.

Decision rules. Two separate sets of decision algorithms have been developed for the OQ-45. The first decision rules were developed based on theory and previous research and are therefore termed *rationally-derived.* The rationally-derived cutoffs (described in Lambert, Whipple, Bishop, et al., 2002) were based on client intake scores, the number of sessions completed, and change in the OQ-45 Total Score from intake. Later rules were developed statistically using a large normative database (described below); they are therefore termed *empirically-derived.* Three studies (Lambert, Whipple, Bishop, et al., 2002; Lutz et al., 2006; Spielmans, Masters, & Lambert, 2006) comparing

the two methods have found the empirical method to be superior in terms of overall predictive accuracy.

The empirical decision rules were developed by Finch, Lambert, and Schaalje (2001), utilizing OQ-45 data from over 11,000 clients seen at graduate training clinics, counseling centers, employee assistance programs, and managed behavioral health care settings. Each client in the sample had completed a course of therapy with at least two administrations of the OQ-45. The aggregate of OQ-45 scores at each session for the entire sample showed a pattern of decelerating recovery curves in which clients made rapid initial progress which then became more gradual. Finch and colleagues then divided the sample into 50 distinct groups based on intake score. Each group represented approximately 2% of the sample and was composed of at least 220 patients. Some groups represented one discrete score on the OQ-45, whereas groups at the extremes of the scoring range included patients with a range of intake scores. For each group, hierarchical linear modeling was used to generate expected recovery curves which indicate the mean OQ-45 Total Score at each session (1 through 20) for the corresponding group in the sample.

Following the development of the expected recovery curves, Finch and colleagues (2001) derived tolerance intervals around each curve based on the expected mean OQ-45 score at each session. For example, a two-tailed 80% confidence interval around the mean expected score allowed for identification of the 10% percent of clients that were making the least progress, and therefore might be expected to deteriorate or drop out of therapy (based on estimates that approximately 10% of patients become worse during treatment, Lambert & Ogles, 2004). Similarly, a two-tailed 68% tolerance interval was

calculated for each expected mean by session number. This provided a cutoff for patients whose rate of change was at least 1 standard deviation above or below the mean.

Alert status. The cutoffs described above form the basis for categorizations of *alert status* on the OQ-45 feedback. When a client's Total Score at a session falls above the 80% tolerance interval, the alert status of the resulting feedback is "Red," indicating that progress is significantly less positive than expected. Similarly, scores falling above the 68% tolerance interval (but below 80%) result in "Yellow" feedback. Scores in the middle 68% of scores result in "Green" feedback and scores below the 68% cutoff result in "White" feedback. The *OQ-Analyst* also generates written messages which accompany the alert status. Sample messages include the following:

> White feedback—'The client is functioning in the normal range. Consider termination.'
>
> Green feedback—'The rate of change the client is making is in the adequate range. No change in the treatment plan is recommended.'
>
> Yellow feedback—'The rate of change the client is making is less than adequate. Recommendations: consider altering the treatment plan by intensifying treatment, shifting intervention strategies and monitoring progress especially carefully. This client may end up with no significant benefit from therapy.'
>
> Red feedback—'The client is not making the expected level of progress. The chances are that he/she may drop out of treatment prematurely or have a negative treatment outcome. Steps should be taken to carefully review this case and decide upon a new course of action such as referral for medication or intensification of treatment. The treatment plan should be reconsidered. Consideration should also be

given to presenting this client at case conference. The client's readiness for change may need to be re-assessed.' (Lambert, Whipple, Bishop, et al., 2002, p. 153)

Accuracy of Prediction

Lambert and colleagues (e.g., Hannan et al., 2005; Lambert, 2007) have postulated that the accurate prediction of poor outcomes is essential to the effectiveness of feedback interventions with the OQ-45. Clients who receive yellow or red feedback at any time during treatment are considered signal-alarms and are predicted to deteriorate (operationalized by a demonstrated increase of 14 points or more on the OQ-45 from intake to termination). The accuracy of these predictions has been evaluated in several studies (Ellsworth, Lambert, & Johnson, 2006; Lambert, Whipple, Bishop, et al., 2002; Lutz et al., 2006; Spielmans, Masters, & Lambert, 2006). Overall hit rates (the percentage of all clients correctly predicted) have ranged from .68 to .83 for the empirical decision rules. The average sensitivity of the OQ-45 in correctly identifying patients who deteriorate is approximately .88 (Lambert, 2007). That is, if 100 patients deteriorate over the course of treatment, the OQ-45 will identify 88 of them before termination. The excellent sensitivity of the empirical method comes at the expense of specificity (approximately 0.82), as the OQ-45 generates a moderate proportion of "false alarms." Approximately 18% of clients who did not deteriorate were identified as signal-alarm cases. Although such patients did not deteriorate as predicted, they were found in two studies to be less likely than the other patients (who were not identified as alarm cases) to evidence reliable improvement (Hannan et al., 2005; Lambert, Whipple, Bishop, et al., 2002).

The validity of predictions made by the empirical decision rules may also be evaluated in terms of positive and negative predictive power. Positive predictive power refers to the proportion of patients who receive signal alarm feedback and actually deteriorate. Ellsworth, Lambert, and Johnson (2006) found the positive predictive power of the OQ-45 to be .27, a relatively unimpressive proportion due to the number of false positives generated by the empirical decision rules. The negative predictive power, referring to the proportion of those predicted not to deteriorate who in fact did not, is much higher. Ellsworth and colleagues found the negative predictive power of the OQ-45 to be .99. In other words, patients who do not receive Yellow or Red warnings at any point during therapy are very unlikely to have deteriorated at termination. However, it should be noted that these analyses are to be interpreted cautiously due to the use of OQ-45 data as both predictor and criterion variables. In other words, false negatives are precluded almost by definition; it is unlikely for clients who never produce Red or Yellow warnings based on OQ-45 data to be rated as deteriorated by the OQ-45. The degree to which OQ-45 predictions of change correspond with alternative criterion measures should be examined.

Lambert, Whipple, Bishop, Vermeersch, Gray, and Finch (2002) suggested that the large proportion of false alarms generated by the empirical decision rules has relatively little real-world cost. As compared to medical fields, where a false positive may result in intrusive procedures (such as surgery or medication), false positives in mental health practice are unlikely to create adverse consequences. A practical cost is that therapists may grow weary of frequent warning feedback (Ellsworth, Lambert, & Johnson, 2006); therapist frustration with warning feedback may lead them to discount it

or to undervalue it. Feedback is frequently Yellow or Red in alert status; the empirical method appears to label 22% to 24% of clients as signal alarms in university counseling centers. In other settings, this number is likely to be higher. For example, Hawkins and colleagues (2004) reported that approximately 50% of clients at a hospital-based outpatient clinic were identified as signal-alarm cases.

Hannan, Lambert, Harmon, Nielsen, Smart, Shimokawa, and Sutton (2005) compared the empirical prediction system to the clinical judgment of 48 therapists (graduate student trainees and professionals). Therapists were asked for three consecutive weeks to predict their patients' final status following treatment (recovered, improved but not recovered, no change, or deteriorated) and to rate patients' improvement up to that point in therapy. Of the 332 clients in the study, 26 were deteriorated at termination. Therapists predicted only 3 patients to deteriorate, 1 of whom did deteriorate (as measured by a reliable increase in OQ-45 score). In contrast, the empirical method based on OQ-45 results produced warnings for 55 patients, 20 of whom did in fact deteriorate. Based on the results, it appears that overall hit rates between therapists and statistical prediction may have been similar (due to the number of false alarms generated by the empirical method), but the statistical predictive method was much more likely to identify early in the course of treatment those patients at risk for no benefit from treatment. It bears noting that final outcome in the Hannan et al. study was measured by the OQ-45, increasing the probability that those patients with unusually high OQ-45 scores at some point in therapy would be rated by the same instrument as deteriorated at termination. The statistical prediction technique therefore had an inherent

advantage as compared to clinicians. The results would be more convincing had a concurrent outcome criterion been used.

Effects of OQ-45 Feedback

Lambert (2007) reviewed five studies (Harmon et al., 2007; Hawkins et al., 2004; Lambert et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002; Whipple et al., 2003) regarding the effects of OQ-45 feedback on client outcomes. The five studies shared many features: a) clients were seen in routine care and received a variety of clinical diagnoses; b) clients were randomly assigned to feedback or treatment-as-usual conditions (with the exception of Whipple et al., 2003); c) therapists provided treatment from a variety of theoretical orientations; d) postgraduate psychologists and graduate-students each represented about 50% of study therapists; e) therapists in each study saw patients in both the feedback and treatment-as-usual conditions; f) the OQ-45 was used as the outcome measure, and decision rules for identifying signal alarm cases remained constant; g) the length of therapy was determined by client and therapist without external constraints; and h) patient demographic characteristics were similar (with the exception of Hawkins et al, 2004, which was conducted in a hospital-based clinic).

Each of the five studies addressed the following main question: does feedback to therapists regarding client progress improve outcomes? The findings across studies are summarized below.

Effects of feedback on outcome. Each of the studies reviewed by Lambert (2007) found that, among "signal alarm" clients (i.e., clients receiving Yellow or Red feedback at some point during the course of therapy), those whose therapists received OQ-45 feedback achieved greater improvement than clients whose therapists did not.

This finding appears to be robust and has been well replicated, with effect sizes ranging from 0.34 to 0.92 (Lambert, 2007). Lambert noted that such effect sizes compare favorably to an average effect size of .20 in treatment outcome studies (Lambert & Ogles, 2004). Combined data across studies also demonstrated differences in regards to final treatment classification (based on Jacobson and Truax, 1991, criteria for clinically significant and reliable change). Lambert reported that 20% of alarm clients in no-feedback conditions were rated as deteriorated at termination, whereas the percentage of deterioration for clients whose therapists received feedback ranged from 8% to 15%. The percentage of clients classified as reliably improved was higher for alarm patients in the feedback conditions than for patients in treatment-as-usual conditions. These data provide evidence for the clinical utility of feedback, in addition to statistical significance.

Whereas all five studies demonstrated improved outcomes for signal-alarm clients, results were less conclusive across all clients. Indeed, only two studies (Harmon et al., 2007; Hawkins et al., 2004) found significant differences between the feedback and no-feedback groups when including on-track clients in the analyses. Lambert (2007) noted that "it appeared to make little difference in outcome for feedback (green or white messages) to have been given [to clients who progressed as expected in therapy]" (p. 10).

Effects of feedback on attendance. Examination of the effects of feedback on client attendance rates is important in order to evaluate whether feedback results in improvements in the cost-effective provision of services. Findings regarding attendance have been inconsistent. Three studies (Harmon et al., 2007; Lambert et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002) found an increased average number of

sessions for alarm clients in feedback conditions, whereas no such difference was found in the other studies. Observing this discrepancy, Harmon and colleagues stated, "it seems fair to conclude that the positive effects of feedback can be obtained with and without extending treatment length" (p. 390). Findings have also been mixed regarding attendance rates among on-track clients. Lambert (2007) reported that feedback decreased sessions by an average of 0.66 sessions in about half the studies. It is interesting to note that in the two studies (Harmon et al., 2007; Hawkins et al., 2004) that found improved outcomes for on-track groups, on-track clients who received feedback did not differ from their no-feedback counterparts in number of sessions attended. In conjunction with the finding (in Lambert et al., 2001, and Whipple et al., 2003) that on-track clients in the feedback conditions achieved equivalent results to their counterparts in fewer average sessions, the results suggest that clients deemed as on-track may improve at a slightly faster rate when therapists receive feedback.

Effects of feedback on clinical judgment. Haderlie (2009) examined the validity of 5 novice psychotherapists' judgments of client change over the course of 5 months of practicum experience during which therapists received weekly feedback from their clients' OQ-45 results. Following each session, but before receiving feedback, therapists estimated their clients' change from the previous session to the current session using a 7-point Likert-type scale. In order to evaluate change in judgment accuracy over time, the first and last 5 sessions of each therapist were examined as blocks. The mean correlation between individual therapists' estimates of client change from the previous session and change over the same period as indicated by the OQ-45 was .06 for the first block, reflecting virtually no validity of judgment. By the last block, the mean

correlation had increased to .56, a significant difference from the first block. The therapists therefore increased dramatically in their abilities to estimate client progress. However, the degree to which receiving feedback contributed to the observed gains cannot be known due to the uncontrolled nature of the design. However, given the consistent finding that clinical experience is at best weakly related to the validity of judgments, it seems likely that the feedback contributed significantly to the increased judgment accuracy.

Present Study

The primary purpose of the present research was to examine whether providing clinicians with feedback regarding client progress lead to increased accuracy of judgments regarding client progress. Specifically, therapists were asked to rate client progress immediately following each session. Client progress was rated in terms of change from the beginning of treatment as well as change from the previous session. Because the judgments were made on the basis of information and observations gained directly from sessions with the clients, therapists' judgments were a naturalistic reflection of the judgment of client change in psychotherapy. Clinical judgments of change were compared to changes on the OQ-45, a self-report measure of distress and dysfunction. The effect of feedback on clinicians' confidence in their clinical judgments regarding client progress was also examined. Therapists were randomly assigned to either receive feedback following each judgment or to receive no feedback. The current research therefore contributes to the clinical judgment literature by systematically examining the effects of feedback on the accuracy of clinical judgment. Although it has been widely

suggested that feedback is essential to improving the accuracy of judgment, very little research has examined this possibility.

The primary hypothesis examined in the present research concerned the effect of OQ-45 feedback on the validity of clinical judgment.

Hypothesis I: Clinicians who received OQ-45 feedback would demonstrate greater increases in judgment accuracy than clinicians who did not receive feedback.

Feedback is widely believed to be an important, and possibly necessary, means of enhancing clinical judgment. Additionally, some limited research has supported the positive effects of feedback on judgment accuracy (e.g., Goldberg, 1968; Graham, 1971). It was therefore expected that clinicians who received feedback would demonstrate more accurate judgments of client change at the conclusion of data collection than clinicians who did not receive feedback. In statistical terms, a group (feedback, no-feedback) by time (beginning, end of data collection) interaction was expected in which the rate of improvement in judgment accuracy would be more steeply positive for clinicians who received feedback. The effect was expected to be observed both for judgments of progress since the beginning of treatment and judgments of progress since the previous session.

Hypothesis II: Clinicians who received OQ-45 feedback would demonstrate greater increases in confidence regarding their judgments as compared to clinicians who did not receive feedback.

The second hypothesis concerned the effects of feedback on clinicians' confidence in their judgments. Although the provision of additional information does not

universally lead to increased confidence in judgments, it has been found to in some studies (e.g., Oskamp, 1965; Trueblood & Binder, 1997). It was expected that in the current study, receiving OQ-45 feedback for each session would result in increased confidence in the accuracy of judgments regarding client progress.

Hypothesis III: Clinicians who received OQ-45 feedback would demonstrate greater calibration of confidence ratings than clinicians who did not receive feedback.

In addition to examining the absolute value of clinician confidence ratings, the effects of feedback on the appropriateness of confidence ratings were of interest. The appropriateness of confidence ratings is most commonly evaluated by calculating each judge's calibration, which is equivalent to the absolute difference between the mean of confidence ratings and the overall proportion of correct decisions. Perfect calibration is reflected by a score of 0.0, whereas values moving away from 0.0 reflect poorer calibration. Although the effects of feedback on the appropriateness of clinicians' confidence ratings have received little examination, research on the confidence-accuracy relationship in eyewitness testimony has suggested that the accuracy of confidence may improve with feedback (Kassin, 1985; Perfect, Hollins, & Hunt, 2000). It was therefore hypothesized that clinicians who received feedback would demonstrate greater calibration than those who did not.

In addition to the hypotheses outlined above, several research questions received preliminary examination. One of these related to individual differences among therapists. Psychotherapy outcome research has indicated significant variability among individual clinicians in outcomes (e.g., Okiishi et al., 2006) and indicated that individual differences

among therapists account for 5% to 9% of the variance in outcomes (Crits-Cristoph et al., 1991). Spengler, White, Ægisdöttir, Maugherman, and colleagues (2009) noted that the degree of individual differences in judgment accuracy among therapists has not been examined. Additionally, it was unknown whether the accuracy of individual clinicians' judgment is related to the outcomes experienced by their clients. Given the primary importance of clinical judgment across the course of psychotherapy, it is possible that judgment accuracy will be related to client outcomes. "Outcome" data was unavailable for some clients in the present research given that the period of data collection did not always include their first or last sessions. This question was therefore examined in terms of the rate of change observed over the course of data collection.

CHAPTER 3

METHOD

Participants

Participants in the study included therapists at the Center for Individual, Couple, and Family Counseling (CICFC) and the student Counseling and Psychological Services (CAPS), both at the University of Nevada, Las Vegas. All therapists at CAPS and all psychology trainees at CICFC were invited to participate. Participating therapists were given the option of providing an email address in order to be entered in a drawing for one of two $25 gift cards to a bookstore. An original total of 15 therapists consented to participate in the study. In order to maximize confidentiality given the collegial relationship of the current researcher with many of the participants, demographic information was collected separately from all other data. Consenting therapists had a mean age of 32.6 years (SD = 11.2) and were predominantly female (66.7%). Regarding ethnicity, 60 percent of the sample was Caucasian. Other reported ethnic identities included Hispanic (n = 2), Asian-American, Pacific Islander, Indian, and mixed. Following recruitment, 4 therapists did not subsequently provide enough data for inclusion. Another therapist's data was not included because procedural errors had resulted in inconsistent provision of feedback. The resulting sample for data analyses included 10 therapists, with 5 in both conditions. Therapists in the final sample had been practicing psychotherapy for an average of 10.70 months (SD = 17.85). This figure was positively skewed, in that all therapists but one reported less than 12 months of clinical experience. Although the professional therapist in the sample had significantly more clinical experience than other participants, data from this therapist was consistent with

other data. Additionally, the exclusion of this therapist did not alter the observed trends in the data. The therapist was therefore included in analyses in order to maximize data. Two therapists reported having earned doctorate degrees, two held master's degrees, and six reported that their highest degree was a bachelor's degree. Participating therapists were predominantly doctoral students in a clinical psychology program; one therapist reported a counseling psychology background. Each therapist reported data for multiple sessions (M = 36.1, SD = 22.7, Range = 11 to 77) with multiple clients (M = 5.4, SD = 2.2, Range = 3 to 9). All therapists had some prior clinical experience with the OQ-45; only one reported having previously received formal feedback.

Measures

OQ-45 and Feedback

Client progress was measured with the Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004). The OQ-45 is a 45-item self-report measure of general distress and dysfunction. Total scores for the 45-item measure range from 0 to 180, with higher scores indicative of greater distress. As reviewed above, strong evidence has been compiled for both the reliability and validity of the OQ-45. Additionally, the OQ-45 total score as well as the majority of individual items are sensitive to client change over time. The measure generates three subscale scores (Symptom Distress, Interpersonal Relations, and Social Role). However, only the total score was used for data analyses in the current research due to limited support for the OQ-45 subscales. Feedback reports to therapists were generated based on clients' OQ-45 results utilizing the *OQ-Analyst* software. Feedback

was based on the empirical decision rules developed by Finch et al. (2001), as described above.

Therapist Registration Form

Information regarding therapist experience was collected at the beginning of data collection through a Therapist Registration Form (see Appendix). Therapists were asked to provide a 4-digit code number for use throughout data collection. Reviewers of clinical judgment literature (e.g., Spengler, White, Ægisdöttir, Maugherman, et al., 2009) have noted that studies are often limited by unreliable measurement of clinical experience. Spengler and colleagues suggested that "experience might be best measured using multiple items as well as with proximal and distal measures" (p. 422). The Therapist Registration Form therefore includes both categorical and continuous measures of experience, including the highest degree completed, field of practice, and the number of months for which a clinician has practiced psychotherapy. Additionally, whether clinicians have previously used the OQ-45, whether they received feedback, and for how long was assessed.

Therapist Estimate of Change

Therapists' judgments regarding client change were ascertained using a Therapist Estimate of Change form (see Appendix). Estimates of change from the previous session to the current session and from the beginning of treatment to the current session were made on separate one-item Likert-type rating scales. Each scale consists of 7 points (1 through 7). Therapists estimated the direction and the amount of change made by clients from the previous point of reference. The three points on the right side of both scales are labeled *much improved*, s*omewhat improved*, and *slightly improved*. The three points on

the left are equivalent except that *worse* is substituted for *improved*. The midpoints are labeled *no change*. Likert-type scales generally demonstrate adequate reliability; the validity of Likert scales has not been evaluated as extensively (Chang, 1994). The scales were designed to include a mid-point because it was necessary to include a point reflecting no change in either direction. There is no general consensus regarding the optimal number of scale points; some investigators (e.g., Brown, Widing, & Coulter, 1991; Matell & Jacoby, 1971) have found the reliability of items to be independent of the number of points. Others have noted that reliability increases with the variability of a scale; however, too many scale points may lead to method variance by invoking extreme response sets (Chang, 1994). Given that several authors have advocated 7-point scales as the most reliable (e.g., Cicchetti, Showalter, & Tyrer, 1985; Ramsay, 1973), 7 points were used for the Therapist Estimate of Change items. Although single-item measures have been criticized as low in reliability (Schmidt & Hunter, 1996), some researchers have found such measures to demonstrate adequate internal consistency (.66) and test-retest reliability (.82; Matell & Jacoby, 1971). Single-item measures were utilized in the present study due to the necessity of creating a brief measure of therapist change estimates that could be completed following each session. Additionally, it is likely that therapists make global judgments regarding client progress; a single-item estimate of change was therefore expected to adequately reflect clinical judgment of progress.

Therapists' level of confidence in both estimates was also ascertained through the Therapist Estimate of Change form. Therapists were instructed to base their confidence ratings solely on the *direction* of change without considering the magnitude of change. Focusing on the direction of change allowed estimates to be converted into a categorical

variable with three levels (improved, no change, worsened) that was subsequently utilized to examine the appropriateness of confidence ratings. Therapists indicate their level of confidence in each judgment from .30 (the rounded probability of being correct by chance) to 1.00, in increments of .10. The resulting data was used to calculate calibration for each judge.

Procedure

Therapists who opted to participate were oriented to the study and signed informed consent prior to beginning data collection. Clients at the CICFC are routinely administered the OQ-45 at each session for program evaluation and clinical training purposes and therefore experienced no variation from standard procedures. Thus, clients of participating therapists at CICFC who had consented to the use of archival research data were included without study-specific consent procedures. A majority of clients at CICFC were ongoing therapy clients who had met with therapists for multiple sessions prior to the beginning of data collection; baseline data regarding symptom severity was therefore unavailable. At CAPS, clients of participating therapists were invited by those therapists to participate at the time of their first sessions. Therapists provided an informed consent document describing the general research purposes and informing clients of study procedures, which included completing an OQ-45 upon arrival to the waiting room prior to each session. Participating clients self-identified as study participants upon checking in and OQ-45's were provided by CAPS front desk staff. Given that therapists were the primary subjects of interest, no client demographic information was collected at either clinic. Data were collected from a total of 47 clients

(38 at CICFC and 9 at CAPS). Individual clients provided data for multiple sessions, ranging from 1 to 22.

Participating therapists completed a registration form at the beginning of data collection in which each therapist provided a 4-digit code number that was used throughout the research. Therapist identities were therefore unknown to researchers at all stages of the research and study data is anonymous. The registration form also elicited basic information regarding therapists' training and clinical experience, as well as information regarding any previous experience with feedback from the OQ-45. Therapists completed two separate pages including a demographics form which was not associated with therapist code numbers, and an optional form in which they provided email addresses for consideration in a drawing for one of two gift cards. These forms were stored separately from therapist registration forms.

Following registration, each therapist was randomly assigned to either a Feedback (FB) or a No-Feedback (NFB) condition. Block random assignment was utilized to ensure equal sample sizes in the FB and NFB conditions and therefore preserve statistical power, which was important given the small number of therapists available for participation in the study.

Data collection occurred from March 2010 to January 2011. During the data collection phase, clinic staff administered the OQ-45 to clients as they arrived for therapy sessions. The measure takes approximately five minutes to complete on average, although administration may range from three minutes to 15 minutes in rare circumstances (Lambert et al., 2004). After clients completed the OQ-45, the questionnaires were set aside for researchers rather than delivered directly to clinicians.

Immediately following each session, therapists completed a brief Therapist Estimate of Change form in which they estimated client change from the previous to the current session and indicated their confidence in the accuracy of their estimation. The Therapist Estimate of Change and OQ-45 forms were collected by researchers twice weekly for data collection. The OQ-45 forms were scored and paper feedback reports were generated using *OQ-Analyst* software (www.OQmeasures.com). OQ-45 forms and associated feedback reports were delivered to therapists in the FB condition within two days of each session in order to allow them to review the reports prior to their next session with the client. Thus, feedback was essentially delayed one session: feedback reports received by therapists reflected client change over the time period from two sessions previous to the most recent session. Additionally, therapists' estimates of client progress were made following sessions, whereas client OQ-45 data was obtained prior to the session. This schedule was necessary in order to elicit therapist estimates of change before providing feedback. The schedule was also consistent with those of previous research studies that demonstrated beneficial effects of OQ-45 feedback (e.g., Harmon et al., 2007). OQ-45 forms for therapists in the NFB condition were stored securely by researchers until the conclusion of data collection and then returned to client files.

Data Analyses

In order to examine changes in criterion variables (e.g. accuracy of judgments, confidence ratings, calibration) from the beginning of data collection to the conclusion of data collection, it was necessary to identify a baseline and final group of judgments for each therapist. This was complicated by the variability in the number of judgments made

by each therapist during data collection. The first 5 judgments made overall by each therapist were therefore considered baseline performance (i.e., criterion variables were calculated using only those sessions). Similarly, the last 5 judgments made by each therapist were utilized to calculate outcome performance. A 5-session cutoff was utilized in the current study because it was considered to be a minimum number of sessions in order to produce reliable estimates, and allowed for the inclusion of the maximum number of therapists possible (e.g., a cutoff of 10 sessions would have necessitated the exclusion of 4 additional therapists). The baseline and outcome blocks therefore represented the first and last 5 judgments made by the therapist during the course of data collection; these 5 sessions included multiple sessions with the same client in some cases.

Given the differences in the number of judgments and subsequent feedback administrations per therapist, the correlation between total number of sessions completed and each criterion variable was examined; where this correlation was significant, the number of sessions completed was utilized as a covariate for analyses.

The accuracy of therapists' judgments of client progress was estimated in two manners. The first was to examine the "hit rate" of therapists' judgments of the direction of change. A judgment was considered a "hit" if the therapist estimated that client progress had been in the same direction indicated by the client's OQ-45 self-report (i.e., comparing the OQ-45 Total score at current session to the Total score at the previous session). Reducing estimates to the direction of change allowed for a standard criterion of accuracy.

The accuracy of judgments was also estimated by calculating the correlation between therapists' estimates of client progress (on a Likert scale from 1 to 7) and client

progress as measured by the OQ-45. This correlation was calculated for each therapist individually based on all ratings made by the therapist. Although this method of operationalizing judgment accuracy does not provide an absolute criterion of accuracy, it examines the degree to which therapists had a "feel" for client progress and altered their judgments in conjunction with client change.

All estimates of judgment accuracy were calculated on the basis of judgments of progress since the previous session. This judgment was considered a more difficult and more important judgment, given that it typically requires finer-grade distinctions in client functioning and status. Additionally, data from each therapist's first session with each client was unavailable, making impossible the examination of accuracy of judgments regarding change from the first session.

The appropriateness of confidence ratings was computed using Lichtenstein and Fischoff's (1977) formula for calibration, which is equal to the absolute difference between the mean of the probability (confidence) responses and the overall proportion correct (hit rate). In order to increase the validity of the outcome criterion (i.e., the direction of change on the OQ-45), the Standard Error of Measurement (SEM) of the OQ-45 was placed around the value 0 on the OQ-45 total score. Change values falling within this range were considered to reflect no change. An SEM of 6.57 was estimated for the OQ-45 Total Score based on a reported internal consistency coefficient of .93 (Lambert et al., 1996) and a standard deviation of 24.84 at pre-test in outpatient settings (Lambert et al., 2004). Therefore, change scores falling within 3 points of 0 were coded as *no change*.

Because "outcome data" (i.e., estimates of client status at the beginning and end of treatment) were not available for each client, the effectiveness of therapy was considered by examining the rate of change for each client during the course of data collection. A rate of change was calculated for each client by dividing observed change on the OQ-45 Total Score from the beginning to the end of data collection by the number of sessions attended by the client. Experience variables were not utilized in the analyses, given the limited variability of experience among therapists in the current sample.

The limited number of therapists available for the current study led to small groups (n = 5 for both conditions) and therefore low statistical power. Given the limited power, statistical significance was unlikely for most analyses. Results are therefore considered exploratory and are discussed in terms of observed trends and effect sizes.

CHAPTER 4

RESULTS

Preliminary Analyses

In order to evaluate potential pre-existing differences between therapists in the FB and NFB conditions, group means were compared utilizing independent-samples *t* tests. For all criterion variables (e.g., judgment accuracy, confidence ratings, calibration), preliminary differences were examined by utilizing the first 5 data points for each therapist. No significant differences were found for therapist experience, judgment accuracy (as measured by hit rate or by correlation), confidence ratings, calibration, or the number of sessions completed during data collection. See Table 1 for a summary of these preliminary analyses. There was some notable difference between conditions in the initial judgment hit rate, as therapists in the feedback condition demonstrated greater accuracy than those in the no-feedback condition. Additionally, initial calibration was better (lower) among therapists in the feedback condition. The mean number of months of clinical experience also differed somewhat between groups, primarily due to the inclusion of one professional clinician in the feedback group, whereas all other study therapists were practicum students (as reflected in the discrepancy between standard deviations for experience means in the FB and NFB conditions).

Descriptive Statistics

Mean ratings for the Therapist Estimate of Change (TEC) form are reported in Table 2. Means were calculated by therapist first and then combined as grand means in

Table 1

Initial Differences between Feedback (FB) and No-Feedback (NFB) Groups

	FB		NFB		
Variable	*M*	*SD*	*M*	*SD*	Cohen's *d*
Experience (mos.)	17.20	24.23	4.20	4.87	0.74
Initial Judgment Accuracy					
Hit rate[a]	0.50	0.20	0 .25	0.19	1.28
Correlation[a]	0.08	0.59	0.14	0.25	0.13
Initial Confidence Ratings					
From first[a]	0.83	0.08	0.81	0.09	0.23
From prev.[a]	0.81	0.04	0.83	0.09	0.29
Initial Calibration[a]	0.30	0.23	0.56	0.26	1.06
Number of sessions	37.40	23.08	34.80	23.08	0.11

[a]Value calculated using the first 5 data points for each therapist

order to account for variability in the number of ratings made by each therapist. On average, therapists were slightly more positive than "neutral" in estimating change from the previous session. When considering progress since their first session with a client, therapists tended to rate change as "slightly improved" to "somewhat improved" on average. Therapists reported high confidence ratings for their judgments of client progress.

Table 2

Mean Scores for Therapist Estimate of Change Across all Therapists

Item	Mean	SD	Range
In your judgment, how has your client's status changed since the previous session?	4.65[a]	0.59	1 to 7
How confident are you in the accuracy of your judgment regarding the *direction* of change (worse, no change, or improved) since the previous session?	80.34	6.70	50 to 100
In your judgment, how has your client's status changed since your first session with the client?	5.46[a]	0.60	1 to 7
How confident are you in the accuracy of your judgment regarding the *direction* of change (worse, no change, or improved) since your first session?	83.58	8.52	50 to 100

Note. Means (and accompanying standard deviations) reflect the grand mean of individual therapist means.

[a]Responses based upon a 7-point scale with 1 = *much worse*, 2 = *somewhat worse*, 3 = *slightly worse*, 4 = *no change*, 5 = *slightly improved*, 6 = *somewhat improved*, and 7 = *much improved*.

Examination of raw data (without collapsing by therapist) indicated that when judging progress since the previous session, therapists rated clients as improved at 44.5 percent of sessions; clients were rated as having made no change at 39.2 percent of sessions and as having deteriorated at 16.2 percent of sessions. Corresponding OQ-45 data reflected improvement at 37.2 percent of sessions, no change at 28.5 percent of sessions, and deterioration at 34.3 percent of sessions. For judgments of change since the first session, therapists' estimates were as follows: 79.4 percent improved, 16.8 percent no change, and 3.8 percent deteriorated. Corresponding OQ-45 data were not available given that many clients in the study had already been attending sessions before their

inclusion in the study and therefore data regarding their initial severity had not been obtained.

Judgment Accuracy

Hit Rate

The accuracy of therapist estimates regarding the direction of client change from one session to another was estimated by calculating hit rates for each therapist. The mean hit rate across all therapists was 0.36 (*SD* = 0.14). On average, therefore, therapists performed at approximately chance level when making estimates regarding the direction of change (as there were three available response categories: positive change, negative change, or no change). Hit rates for individual therapists ranged from 0.20 to 0.64. Judgment accuracy as measured by hit rates was unrelated to the number of sessions completed by individual therapists, $r(8) = .19$.

The effect of feedback on judgment accuracy was evaluated through examination of accuracy as a function of group assignment and time. The time factor was operationalized by comparing hit rates for each therapist's first and last 5 sessions. Although judgment accuracy increased slightly across the course of the study, from a total mean hit rate of .38 (*SD* = .23) to .42 (*SD* = .13), this did not appear to be a meaningful increase (*partial eta squared* = .05) and it was not significant. On average, judgment accuracy increased among therapists in the no-feedback condition from the first 5 to the last 5 sessions. In contrast, judgment accuracy decreased slightly among therapists in the feedback condition from the first 5 to the last 5 sessions. This interaction effect was relatively large (*partial eta squared* = .33) and was contrary to the

hypothesized results, in that feedback was found to be detrimental to judgment accuracy. See Table 3 for a summary of all between-group comparisons.

Correlations

Judgment accuracy was also evaluated by examining the correlation between therapists' judgments of client change and observed change on the OQ-45 at each specific session. The grand mean of individual therapist correlations was -.01 (*SD* = .41), indicating that therapists' estimates of client change did not tend to covary with change as measured by the OQ-45. Correlations for individual therapists varied widely, from -.51 to .61. There was a moderate positive relationship between these correlations and the number of sessions completed by the therapist, $r(8) = .57$, indicating that correlations tended to improve as therapists completed more sessions. Estimates of judgment accuracy based on correlation were not related to estimates of judgment accuracy using the hit rate criterion, $r(8) = -.03$.

The correlation between therapists' estimates of client change from the previous session and change as measured by the OQ-45 was utilized as a dependent measure to examine the effect of feedback on clinical judgments. A repeated-measures ANCOVA was conducted, with feedback condition as a between-groups factor. The number of sessions completed by the therapist was utilized as a covariate for the analysis, given the relationship between sessions and judgment accuracy as measured by correlation. Baseline and outcome periods were operationalized as above. Mean correlations increased somewhat from the first 5 (*M* = .05) to the last 5 sessions (*M* = .19); this effect was very small (*partial eta squared* = .01). Inspection of means revealed that average correlations increased in the NFB condition, whereas correlations for therapists who

Table 3

Condition by Time Interactions for Outcome Variables

	FB		NFB		
Variable	*M*	*SD*	*M*	*SD*	*partial eta squared*
Judgment Accuracy					
Hit rate					
First 5	0.50	0.20	0.25	0.19	
Last 5	0.40	0.00	0.45	0.19	0.33
Correlation					
First 5	0.08	0.59	0.03	0.10	
Last 5	0.03	0.60	0.40	0.38	0.19[a]
Confidence Ratings					
From first					
First 5	0.83	0.08	0.81	0.09	
Last 5	0.81	0.12	0.89	0.06	0.60[a]
From previous					
First 5	0.81	0.04	0.83	0.09	
Last 5	0.74	0.08	0.83	0.07	0.47[a]
Calibration					
First 5	0.30	0.23	0.56	0.26	
Last 5	0.33	0.09	0.36	0.20	0.21

Note. partial eta squared effect size statistic is for interaction of variables *condition* (Feedback, FB, or No-Feedback, NFB) and *time* (first 5 and last 5 sessions completed by each therapist).

[a]The number of sessions completed by the therapist was utilized as a covariate for calculation of effect size.

received feedback slightly decreased on average. This effect was moderate in size (*partial eta squared* = .19). ANCOVA results were not statistically significant.

Rate of Therapeutic Change

Across all clients, the mean rate of change in OQ-45 scores was -0.53 (*SD* = 2.86). In other words, the average client's OQ-45 Total Score decreased by approximately one half of a point per session during the period of data collection.

In order to evaluate the possibility that therapists who made more accurate judgments would also demonstrate better therapeutic outcomes, the correlation between judgment accuracy and client rate of change was examined. The rate of change experienced by clients of individual therapists was unrelated to therapists' judgment accuracy as measured by either hit rates, $r(8) = .11$, or the correlation of therapist estimates and OQ-45 change, $r(8) = -.19$.

The relationship between judgment accuracy and rate of change was also examined at the individual client level (without collapsing means by therapist). The hit rate of therapist judgments for individual clients was not related to the rate of change experienced by the clients, $r(37) = -.04$. Therefore, clients whose therapists made more accurate judgments about their progress were no more likely to experience positive change.

Clients of therapists in the feedback condition demonstrated more rapid symptom reduction ($M = -0.77$, $SD = 1.48$) as compared to clients of therapists who did not receive feedback ($M = -0.04$, $SD = 0.90$). The observed effect of feedback condition was moderate, Cohen's $d = 0.61$.

Confidence

Therapists tended to report high levels of confidence in their estimates of client progress since the previous session, and since their first session with the client. The mean confidence rating for judgments regarding change from the first session was 0.84 (*SD* = .09, Range = .68 to .95). Similarly, confidence ratings for judgments regarding change from the previous session averaged .80 (*SD* = .07, Range = .66 to .88). The correlation between confidence and judgment accuracy was investigated through examination of the correlation between therapists' confidence ratings and hit rates for judgments regarding change from the previous session. A negative correlation was observed, $r(8) = -.42$. Overall, therapists who reported greater confidence in their judgments tended to be less accurate. Confidence ratings for judgments of change from the previous session were negatively correlated with the number of sessions completed by the therapist, $r(8) = -.60$. A negative relationship was also observed between the number of sessions completed and confidence ratings for judgments of change from the first session, $r(8) = -.44$. Results indicated that as the number of sessions completed by a therapist increased, confidence scores tended to decrease.

The effect of feedback on therapists' confidence ratings was explored by examining mean confidence ratings by group and time. Each therapist's first and last 5 confidence ratings were utilized as endpoints for the time factor. The number of sessions completed by each therapist was used as a covariate for the analysis. Confidence ratings regarding estimates of change from the first session with the client to the time of judgment were examined first. Confidence ratings increased slightly from the first 5 (*M* = .82, *SD* = .08) to the last 5 sessions (*M* = .85, *SD* = .10). This effect was relatively

large, *partial eta squared* = .25. Results indicated that therapists in the no-feedback condition made higher confidence ratings at the last 5 sessions than at the first 5 sessions. In contrast, confidence ratings for therapists who received feedback slightly decreased from the first 5 to the last 5 sessions. The effect of this interaction was large, *partial eta squared* = .60.

The effect of feedback on confidence ratings was also evaluated utilizing confidence ratings regarding client change from the previous session. Confidence ratings tended to decrease from the first 5 sessions (M = .82, SD = .07) to the last 5 sessions (M = .78, SD = .09), representing a large effect of time, *partial eta squared* = .54. Examination of means indicated that confidence ratings tended to decrease for therapists who received feedback, whereas ratings remained similar for therapists who did not receive feedback. This effect was relatively large, *partial eta squared* = .47. Results were consistent for both confidence ratings, in that final ratings were lower for therapists who received feedback than for those in the no-feedback condition. In both cases, the time by condition interaction effect was large.

Calibration

In addition to the raw value of confidence ratings, the appropriateness of confidence ratings was examined by calculating calibration scores for each therapist. Calibration scores reflect the absolute difference between a therapist's overall judgment accuracy (hit rate) and average confidence rating related to the specific judgment (estimating the direction of client progress from the previous session). Lower scores therefore indicate better calibration, with a score of 0.0 denoting perfect calibration. The

mean calibration score across therapists was .45 (*SD* = .18), which reflected a general pattern of confidence ratings (*M* = .80) being greater than hit rates (*M* = .36). Individual therapists' calibration scores varied considerably, from near-perfect calibration in one therapist (.02) to very poor calibration (.65). The relationship between calibration scores and the number of sessions completed by the therapist was negative, $r(8) = -.38$, indicating that calibration scores tended to improve (decrease) slightly as the number of sessions increased.

Calibration scores were examined as a function of group and time. On average, calibration improved from the first 5 (*M* = .43, *SD* = .26) to the last 5 sessions (*M* = .35, *SD* = .15; this effect was moderate, *partial eta squared* = .13. Calibration improved among therapists in the no-feedback condition from baseline to the conclusion of data collection; for therapists in the feedback condition, calibration remained similar from baseline to the conclusion of data collection. Although this effect was notable, *partial eta squared* = .21, it was primarily due to pre-intervention differences between groups; calibration was similar between groups at outcome.

Stability of Variables

In order to examine the degree to which individual differences in study variables were stable over time, the correlation of individual therapists' scores at the first and last 5 sessions was examined. Results are summarized in Table 4. Hit rates at baseline were not predictive of hit rates at outcome; in contrast, there was a moderate relationship between accuracy as measured by correlation from baseline to outcome, indicating that therapists who were more responsive to fluctuations in client progress tended to remain

so over time. Confidence was highly stable for judgments regarding client progress from both the most recent and from the first session. Calibration (which is calculated on the basis of hit rate and confidence and therefore not independent) was moderately stable over time.

Table 4

Stability of Variables within Therapists from First 5 Sessions to Last 5 Sessions

Variable	*r*	*df*
Judgment Accuracy		
Hit Rate	-.17	(6)
Correlation	.39	(5)
Confidence		
From First	.60	(8)
From Previous	.68	(8)
Calibration	.35	(6)

CHAPTER 5

DISCUSSION

The primary purpose of the present study was to examine the effect of standardized feedback on the accuracy of therapists' judgments regarding client progress. Judgment accuracy was estimated in two distinct methods: 1) calculation of a "hit rate" for each therapist (i.e., the proportion of total estimates in which the direction of change estimated by the therapist was congruent with the direction of change indicated by the OQ-45); and 2) examination of the correlation between therapist estimates of change and OQ-45 change scores. Both methods therefore relied upon OQ-45 data as an indicator of client status; the limitations of this approach are discussed below. Somewhat surprisingly, these two methods of operationalizing judgment accuracy were not related; improvements in the overall correlation between therapist and OQ-45 estimates did not predict improvement in hit rates. In some cases, therapists' estimates paralleled changes in the OQ-45 but remained inaccurate in the absolute direction of change specified. In other cases, therapists' estimates were more accurate in terms of hit rate but did not tend to covary with OQ-45 changes. Generally, judgment researchers have used the "hit rate" method as a primary measure of judgment validity (accuracy), in part because many studies examine categorical judgments (e.g., diagnosis); however, research regarding more dimensional or quantitative judgments, such as client progress, may also utilize the correlation method. Given that both methods appear to measure different outcomes, it is unclear whether one method is to be favored; future research may benefit from incorporating both methods. It is notable that the correlation method was more stable in the current study and may therefore be preferred, particularly for research involving small

sample sizes. An alternate approach would be to have therapists estimate the client's current status at the time of session using a standardized measure (e.g., Global Assessment of Functioning), without directly comparing the status to a previous point. This would allow for comparison of parallel judgments made by clients and therapists; however, this approach would be a less direct means of examining the extent to which therapists estimate accurately client progress, which requires not only a judgment of current status but also judgment and recall of client status at a previous point in time.

Overall judgment accuracy was surprisingly low in the current study as measured by either method of assessing accuracy. The overall hit rate for estimates of the direction of client change was 0.36, which represents approximately chance performance (as the three possibilities were improvement, deterioration, or no change). Similarly, the overall correlation between therapists' estimates and progress as measured by the OQ-45 was -.01. Accuracy increased only slightly from baseline to the conclusion of data collection; however, accuracy did increase as the number of judgments made by the individual therapists increased. This relationship was particularly notable for accuracy as measured by correlation, which was strongly related to the number of sessions completed by the therapist ($r = .57$). It is possible that accuracy as measured by hit rates is limited by therapists' tendency to be overly optimistic regarding the direction of change, and particularly by therapists' hesitance to rate clients as deteriorated. Therapists may have become more attuned to fluctuations in client progress over time while continuing to overestimate the absolute amount of change being made. (And given the slow progress that is typical for many psychotherapy clients, such "naïve optimism" on a therapist's part may not be entirely undesirable as prevention against burnout.)

In a previous study (Haderlie, 2009), judgment accuracy as measured by correlation increased dramatically over time among novice therapists at the beginning of their first practicum experience. Therapists in the current study were primarily first-year practicum students as well, although they had been practicing for approximately 6 months prior to the beginning of data collection. It is possible that therapists orient relatively quickly to client status as they begin to accrue clinical experience, and therefore improvements were less notable in the current study. Another possibility is that repetition with the judgment task itself (i.e., repeatedly engaging in the process of considering available data and making a judgment regarding client change) leads to improved accuracy; this hypothesis is supported by the relationship between the number of judgments made and accuracy in the current study, although the relationship was relatively weak when looking at hit rates. The overall effect of time may have been restricted due to the fact that therapists completed varied numbers of sessions, and in some cases therapists completed relatively few sessions (minimum was 11).

Feedback did not have a positive effect on judgment accuracy in the current study as measured by either method of assessing accuracy, and was in fact detrimental to accuracy in both cases. As estimated by either method, judgment accuracy decreased slightly across time for therapists who received feedback. Conversely, therapists in the no-feedback condition demonstrated relatively large improvements in judgment accuracy over time, as measured by either method. It is clear that feedback did not improve judgment accuracy in the current sample, contrary to the original hypothesis. Given the extended theoretical literature suggesting that feedback is a crucial element of improving clinical judgment, the observed pattern was unexpected. It should be noted that the

trends in the current study are based on small sample sizes and would need to be replicated. It is possible that the relatively delayed nature of feedback in the current methodology decreased its effectiveness. Additionally, timing differences in the administration of the OQ-45 (before session) and the TEC (after session) may have limited the validity of OQ-45 results as a means of providing feedback regarding the accuracy of therapists' judgments, given that therapists may have been estimating client status at a different point in time than clients were reporting it. Current results were also influenced by initial differences between groups at baseline. Judgment accuracy as measured by hit rates was higher at the outset for therapists in the feedback condition; outcome accuracy was similar between groups. The observed effect was therefore shaped largely by chance variability at baseline.

Individual therapists varied widely in judgment accuracy, by either measure of accuracy. Individual differences in judgment accuracy were moderately stable when accuracy was assessed through the correlation method (but not when assessed by hit rates). It remains unclear, on the basis of current data, whether the ability to make accurate judgments of client progress has applied benefit in psychotherapy. Presumably, therapists who are more attuned to client progress would have advantages in recognizing poor responses early and adjusting treatment plans. Additionally, judgment regarding progress is important in deciding when to discontinue treatment; accurate judgment would therefore be expected to contribute to the efficient use of therapy resources. However, these possibilities remain theoretical and are in need of further investigation.

The overall rate of change observed in the current sample was relatively flat; clients demonstrated a mean reduction in OQ-45 Total Scores of 0.53 points per session,

after controlling for unequal observations across therapists. It is worth noting that a majority of clients in the current study were seen at the CICFC, a low-cost community mental health clinic in which clients are often seen over the course of one or more years. Therefore, a high proportion of clients in the sample were not in the initial phase of treatment and treatment response might therefore have reached a plateau, consistent with research on the dose-effect relationship in psychotherapy (e.g., Howard et al., 1986; 1993). Despite the relatively limited rate of change, a moderate effect of feedback was observed; clients whose therapists received feedback demonstrated a mean reduction of 0.77 points per session, in comparison to a mean reduction of 0.04 points for clients in the no-feedback condition. This trend was consistent with a burgeoning research literature supporting the treatment utility of progress feedback to therapists (e.g., Brodey et al., 2005; Lambert, 2007; Slade et al., 2006).

The current study examined the relationship between clinical judgment and client outcomes. It is notable that although there was a trend for feedback to lead to improved rates of therapeutic change, feedback did not lead to improved clinical judgment accuracy. Furthermore, the accuracy of therapists' judgments regarding the progress of individual clients was statistically unrelated to the rate of change experienced by the clients. Clinical judgment, therefore, was not shown in the current study to contribute meaningfully to client outcomes. It seems likely that the therapeutic effects of feedback owe to separate mechanisms, which have not yet been demonstrated empirically. One possibility is that therapists spend more time thinking about clients for whom they receive feedback; Percevic, Lambert, and Kordy (2004) described this possibility as the "attention effect." Increased attention may also lead to increased empathy or alliance.

This possibility may be examined by administering process measures in addition to outcome measures and examining the effect of feedback on process variables.

The most notable effect of feedback in the current study was on therapists' ratings of confidence in the accuracy of their judgments. When making estimates regarding change from both the first session and the most recent session, therapists who received feedback reported lower confidence over the course of the study, as compared to therapists who did not receive feedback. The statistical magnitude of this effect was large in both cases, although the absolute magnitude of change in confidence ratings was modest. Regarding overall confidence levels, therapists tended to be highly overconfident regarding their judgments, with mean confidence ratings above 80 percent. The decrease in confidence ratings therefore appears to have been an appropriate response to feedback, given that confidence ratings were consistently too high, even after they decreased somewhat due to feedback.

Interestingly, confidence ratings were negatively related to judgment accuracy; therapists who reported higher confidence ratings tended to demonstrate lower hit rates. Given that confidence was negatively related to judgment accuracy (consistent with some previous research; Arkes, 1981), and that therapists tended to be overconfident, this appears to represent a desirable outcome of feedback administration. Particularly among therapists in training, the reduction of overconfidence may provide clinicians with a degree of humility regarding the limitations of clinical judgment and inference. Somewhat inconsistently, therapist calibration did not improve as a function of receiving feedback. However, the reliability of this latter finding may be limited, given that both components of calibration scores (hit rates and confidence ratings) were calculated on the

basis of small sets of 5 judgments. It seems likely that a general trend of decreased confidence ratings would lead to better calibration over a longer period of observation.

Limitations

Several limitations should be noted in interpreting current results. A primary limitation was the small sample size of therapists who provided sufficient data for analyses (N = 10, 5 therapists in both conditions). A repeated-measures design was utilized in order to maximize the number of observations and increase the reliability of individual variables for each therapist; however, the small sample size limited statistical power. Thus the results were presented primarily based on effect sizes and other descriptive statistics. All conclusions are therefore tentative, awaiting a study of sufficient size to employ inferential statistics.

Another consideration germane to the generalizability of results is the nature of the current sample of therapists. With one exception, therapists were first-year practicum students in a doctoral clinical psychology program. The sample was reasonably diverse in terms of gender and ethnicity; however, the extent to which current results would generalize to clinicians at varied levels of training and experience, and in various professional disciplines, remains unknown.

The accuracy of therapists' judgments was measured in two methods in the current study. However, both methods rely upon clients' self-reported OQ-45 data as the criterion measure of progress. Although self-report measures of distress have the advantages of objectivity and statistical reliability, no single measure of outcome or progress is completely sufficient. Measurement of client progress would ideally be based

upon a multi-method assessment incorporating various approaches (e.g., self-report, observation, other-report, external behavioral data, or other-clinician ratings). Reliance upon the OQ-45 was necessary in the current study due to practical limitations.

In order to evaluate criterion variables such as judgment accuracy and calibration, it was necessary to select baseline and outcome groups of observations for each individual therapist; for example, the correlation between therapists' and OQ-45 estimates of change could not be meaningfully examined by simply grouping one observation from each therapist within conditions. The first and last 5 observations for each therapist were therefore utilized as baseline and outcome periods. The selection of 5 sessions was somewhat arbitrary and was largely based upon the limited number of sessions available for some therapists. Although more reliable than a single observation, correlations utilizing only 5 pairs of data are highly influenced by single values; current results are therefore considered preliminary. Future studies should incorporate larger sample sizes with more sessions completed.

Conclusions and Future Directions

The current results did not provide support for the hypothesis that feedback would improve clinical judgments regarding client progress. However, there was a trend for clients whose therapists received feedback to improve at a faster rate than clients whose therapists did not receive feedback. It appears unlikely that improvement in clinical judgment was a significant mechanism in the therapeutic effects of feedback. Future studies should continue to explore potential mechanisms of change.

Although feedback did not improve the accuracy of clinical judgments, it did impact therapists' confidence ratings regarding their judgments. Although study therapists remained overconfident across the course of data collection, those who received feedback demonstrated reduced confidence ratings, in comparison to therapists in the no-feedback condition. As noted above, the reduction of confidence ratings appears to have been a desirable outcome. Programs designed to improve clinical judgment, and more specifically to reduce overconfidence in judgments, would benefit from the incorporation of feedback regarding the accuracy of judgments made. Additionally, the clinical impact of overconfidence among clinicians is relatively unexamined. Future studies may examine potential clinical correlates of therapists' confidence regarding clinical judgments. This might be addressed through a process-oriented approach examining a variety of variables over time (e.g., outcomes, number of sessions attended, number of sessions completed before making significant treatment decisions, proportion of clients referred to outside providers, frequency of consultation, etc.).

The number of sessions completed (and subsequent judgments made) by each therapist was related to several other variables. As the number of judgments increased, judgment accuracy increased (more notably for accuracy as defined by correlation), confidence ratings decreased, and calibration improved. It appears therefore that repetition with the judgment task was beneficial. Repetition (i.e., making repeated judgments regarding client progress, across several clients) appeared to be more influential than feedback in increasing judgment accuracy. On the basis of such results, it appears that providing clinicians with experience related to the specific judgment to be

made is important in developing clinical judgment. Although experience has been inconsistently related to the accuracy of clinical judgment across judgment literature, the current results support Westen and Weinberger's (2004) position that experience with the specific judgment task, rather than general clinical experience, is essential in making valid clinical judgments. Experience effects may therefore be more notable when experience is considered in relation to the specific judgment task, rather than as a raw amount of time spent in clinical training and service. The effect of repetition with a judgment task on the accuracy of judgments may be examined through experimental methods in which the number of judgments made is systematically varied.

In summary, although the current study did not find support for feedback as a means of improving judgment accuracy, feedback did affect confidence ratings in an apparently desirable manner. Feedback may therefore be an important aspect of training programs related to clinical judgment and reducing overconfidence. Current results were also consistent with a growing body of research supporting feedback to therapists as a positive intervention for clients. Additionally, the present data suggest promising lines of further research on clinical judgment. In particular, the effect of repeated experience with a specific judgment task on judgment accuracy should be examined.

APPENDIX

MEASURES

Therapist Registration Form.

Please think of a 4-digit code number (e.g. 2345) that you will remember and use throughout the study. Creating a code number will allow us to keep your responses completely anonymous.

1. Code number: _____

2. What is your highest level of education obtained in a mental health field?

□ Doctorate (Ph.D., Psy.D., Ed.D.)
□ Masters (M.A., M.S., M.F.T., Ed.M.)
□ Bachelors (B.A./B.S.)

3. What is your mental health field? (check one)

□ Clinical Psychology
□ Counseling Psychology
□ Educational Psychology
□ Marriage and Family Therapy
□ Other

4. For how long (in months or years) have you conducted psychotherapy (including training experience)? _____

5. Have you previously utilized the OQ-45 in clinical practice? □ Yes □ No

IF YES: For approximately how long? _____
Did you receive computer-generated feedback? □ Yes □ No

Please provide your demographic information on the following page but DO NOT write your code number on that page.

Participant Demographics

Age: _____

Gender: ☐ Female ☐ Male

Ethnicity: _____

Please see the next page regarding a drawing for a gift card. After completing all pages, separate the three pages and place them in the research drop box in random order.

Entry for Gift Card Drawing

Thank you for your participation in this research. If you would like to be eligible to win one of two $25 gift cards to Barnes and Noble, please provide your email address below. The drawing will be held following data collection and winners will be notified via email.

Email address: _____

Therapist Estimate of Change

Client #: _____ Therapist ID (4-digit code): _____
Date of Session: _____ Number of sessions you have had with this client: _____
(include current session)

If this was your first session, check here _____ and do not complete the remainder of form.

1. In your judgment, how has your client's status changed since the previous session?

1	2	3	4	5	6	7
Much Worse	Somewhat Worse	Slightly Worse	No Change	Slightly Improved	Somewhat Improved	Much Improved

How confident are you in the accuracy of your judgment regarding the *direction* of change (worse, no change, or improved) since the previous session?

30% 40% 50% 60% 70% 80% 90% 100%

2. In your judgment, how has your client's status changed since your first session with the client?

1	2	3	4	5	6	7
Much Worse	Somewhat Worse	Slightly Worse	No Change	Slightly Improved	Somewhat Improved	Much Improved

How confident are you in the accuracy of your judgment regarding the *direction* of change (worse, no change, or improved) since your first session?

30% 40% 50% 60% 70% 80% 90% 100%

REFERENCES

Ægisdottir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34,* 341-382.

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49,* 323-330.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73,* 305-307.

Aronson, D. E., & Akamatsu, T. J. (1981). Validation of a Q-sort task to assess MMPI skills. *Journal of Clinical Psychology, 37,* 831-836.

Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goates, M. K., & Vermeersch, D. A. (2003). Clinical significance of the Outcome Questionnaire (OQ-45.2). *The Behavior Analyst Today, 4,* 86-97.

Bell, I., & Mellor, D. (2009). Clinical judgements: Research and practice. *Australian Psychologist, 44,* 112-121.

Beutler, L. E. (1995). The germ theory myth and the myth of outcome homogeneity. *Psychotherapy, 32,* 489-494.

Beutler, L. E. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. Journal of Consulting and Clinical Psychology, 69, 197-204.

Bickman, L., Rosof-Williams, J., Salzer, M. S., Summerfelt, W. T., Noser, K., Wilson, S. J., & Karver, M. S. (2000). What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice, 31,* 70–74.

Bowman, P. R. (1982). An analog study with beginning therapists suggesting bias against "activity" in women. *Psychotherapy: Theory, Research, and Practice, 19,* 318-324.

Brodey, B. B., Cuffel, B., McCulloch, J., Tani, S., Maruish, M., Brodey, I., et al. (2005). The acceptability and effectiveness of patient-reported assessments and feedback in a managed behavioral healthcare setting. *The American Journal of Managed Care, 11,* 774-780.

Brown, G., Widing, R. E., II, & Coulter, R. L. (1991). Customer evaluation of retail salespeople utilizing the SOCO scale: A replication, extension, and application. Journal of the Academy of Marketing Science, 9, 347-351.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18,* 205 215.

Chronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-509). Washington, DC: American Council on Education.

Cicchetti, D. V., Showalter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A monte carlo investigation. *Applied Psychological Measurement, 9,* 31-36.

Claiborn, C. D., & Goodyear, R. K. (2005). Feedback in psychotherapy. *Journal of Clinical Psychology, 61,* 209-217.

Claiborn, C. D., Goodyear, R. K., & Horner, P. A. (2001). Feedback. *Psychotherapy, 38,* 401-405.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Pery, K., et al. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 2*, 81-91.

Dawes, R. M. (1989). Experience and validity of clinical judgment: The illusory correlation. *Behavioral Sciences & the Law, 7,* 457-467.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: The Free Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1674.

de Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology and Psychotherapy, 14,* 288-301.

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring and procedures manual-II.* Baltimore, Md: Clinical Psychometric Research.

Einhorn, H. J. (1980). Overconfidence in judgment. In R. A. Shweder & D. W. Fiske (Eds.), *New directions for methodology of behavioral research: Fallible judgment in behavioral research.* San Francisco: Jossey-Bass.

Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A comparison of the Outcome Questionnaire-45 and Outcome Questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology and Psychotherapy, 13,* 380-391.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17,* 311-318.

Faust, D. (1989). Data integration in legal evaluations: Can clinicians deliver on their premises? *Behavioral Sciences & the Law, 7,* 469-483.

Faust, D. (1991). What if we had really listened? Present reflections on altered pasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology. Volume I: Matters of public interest* (pp. 185-217). Minneapolis: University of Minnesota Press.

Faust, D. (1994). Are there sufficient foundations for mental health experts to testify in court? No. In S. A. Kirk & S. D. Einbinder (Eds.), *Controversial issues in mental health* (pp. 196-201). Boston: Allyn & Bacon.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241,* 31-35.

Fernbach, B. E., Winstead, B. A., & Derlega, V. J. (1989). Sex differences in diagnosis and treatment recommendations for antisocial personality and somatization disorders. *Journal of Social and Clinical Psychology, 8,* 238-255.

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy, 8,* 231-242.

Frisch, M. B., Cornell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the quality of life inventory: A measure of life satisfaction for use in treatment planning an outcome assessment. *Psychological Assessment, 4,* 92-101.

Gambrill, E. (2005). *Critical thinking in clinical practice: Improving the quality of judgments and decisions.* Hoboken, NJ: John Wiley & Sons, Inc.

Garb, H. N. (1986). The appropriateness of confidence ratings in clinical judgment. *Journal of Clinical Psychology, 42,* 190-197.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association.

Garb, H. N., & Boyle, P. A. (2003). Understanding why some clinicians use pseudoscientific methods: Findings from research on clinical judgment. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 17-38). New York: The Guilford Press.

Garb, H. N., & Grove, W. M. (2005). On the merits of clinical judgment. *American Psychologist, 60,* 658-659.

Gaudette, M. D. (1992). Clinical decision making in neuropsychology: Bootstrapping the neuropsychologist utilizing Brunswik's lens model (Doctoral dissertation, Indiana University of Pennsylvania, 1992). *Dissertation Abstracts International, 53,* 2059B.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Psychiatrists in the UK do not use outcome measures. *British Journal of Psychiatry, 180,* 101-103.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23,* 483-496.

Graham, J. R. (1971). Feedback and accuracy of clinical judgments from the MMPI. *Journal of Consulting and Clinical Psychology, 36,* 286-291.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, & Law, 2,* 293-323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19-30.

Haderlie, M.. (2009). *Examining the psychotherapist as a feedback receiver.* (Master's Thesis). Retrieved from ProQuest Dissertations and Theses. (Accesion Order No. AAT 1474381).

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61,* 155-163.

Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., et al. (2007). Enhancing outcome for potential treatment failures: Therapist-client feedback and clinical support tools. *Psychotherapy Research, 17,* 379-392.

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L., & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research, 14,* 308-327.

Hatfield, D. R. (2007). The influence of outcome measures in assessing client change and treatment decisions. *Dissertation Abstract International, 67, 8-B.* (UMI No. 3230534).

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35*, 485– 491.

Hatfield, D. R., & Ogles, B. M. (2006). The influence of outcome measures in assessing client change and treatment decisions. *Journal of Clinical Psychology, 62,* 325-337.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory* (Rev. ed.). Minneapolis: University of Minnesota Press.

Hole, G. U. (1972). A comparison between clinical judgment and change in MMPI profiles in observing the progress of depressed patients. *Psychologie, 31,* 341-350.

Holt, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology, 56,* 1-12.

Holt, R. R. (1970). Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? *American Psychologist, 25,* 337-349.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The does-effect relationship in psychotherapy. *American Psychologist, 41,* 159-164.

Howard, K. I., Leuger, R. J., Maling, M. S. & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology, 61,* 678-685.

Houts, A. C., & Galante, M. (1985). The impact of evaluative disposition and subsequent information on clinical impressions. *Journal of Social and Clinical Psychology, 3,* 201-212.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19.

Kassin, S. M. (1985). Eyewitness identification: Retrospective self-awareness and the accuracy–confidence correlation. *Journal of Personality and Social Psychology*, *49*, 878–893.

Kendall, P. C., Kipnis, D., & Otto-Salaj, L. (1992). When clients don't progress: Influences on and explanations for lack of therapeutic progress. *Cognitive Therapy and Research, 16,* 269-281.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284.

Kordy, H., Hannover, W., & Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg model. *Journal of Consulting and Clinical Psychology, 69,* 173-183.

Lambert, M. J. (1983). Introduction to the assessment of therapy outcome: Historical perspectives and current issues. In M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 3-32). New York: Wiley.

Lambert, M. J. (2007). Presidential address: What have we learned from a decade of research aimed at improving psychotherapy outcome in routine care? *Psychotherapy Research, 17,* 1-14.

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., et al. (1996). The reliability and validity of the outcome questionnaire. *Clinical Psychology and Psychotherapy, 3,* 249-258.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., et al. (2004). *Administration and scoring manual for the Outcome Questionnaire-45.* Salt Lake City, UT: OQ Measures.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change, 5th ed.* (pp. 139-193). New York: John Wiley & Sons, Inc.

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology and Psychotherapy, 9,* 149-164.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11,* 49-68.

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., et al. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy, 9,* 91-103.

Larson, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction in human service programs: Development of a general scale. *Evaluation and Program Planning, 2,* 197-207.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50,* 212-247.

Levant, R. F. (2005, July 1). *Report of the 2005 Presidential Task Force on Evidence-Based Practice*. Retrieved August 18, 2009, from http://www.apa.org/practice/ebpreport.pdf

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

Lichtenberg, J. W. (2009). Comment: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37,* 410-415.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20,* 159-183.

Li-Repac, D. (1980). Cultural influences on clinical perception: A comparison between Caucasian and Chinese-American therapists. *Journal of Cross-Cultural Psychology, 11,* 327-342.

Lo Coco, G., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., & Lambert, M. J. (2008). The factorial structure of the Outcome-Questionnaire-45: A study with an Italian sample. *Slinical Psychology and Psychotherapy, 15,* 418-423.

Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology, 69,* 150-158.

Lutz, W., Lambert, M. J., Harmon, C., Tschitsaz, A., Schurch, E., & Stulz, N. (2006). The probability of treatment success, failure, and duration—what can be learned from empirical data to support decision making in clinical practice? *Clinical Psychology and Psychotherapy, 13,* 223-232.

Mahajan, J. (1992). The overconfidence effect in marketing management predictions. *Journal of Marketing Research, 29,* 329-342.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31,* 657-674.

McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes, 107,* 179-191.

Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meier, S. T. (1999). Training the practitioner-scientist: Bridging case conceptualization, assessment, and intervention. *The Counseling Psychologist, 27*(6), 846-869.

Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61,* 199-208.

Mours, J. M., Campbell, C. D., Gathercoal, K. A., & Peterson, M. (2009). Training in the use of psychotherapy outcome assessment measures at psychology internship sites. *Training and Education in Professional Psychology, 3,* 169-176.

Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70,* 248-262.

Newnham, E. A., & Page, A. C. (2007). Client-focused research: New directions in outcome assessment. *Behaviour Change, 24(1),* 1-6.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* New York: Prentice Hall.

Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, definitions and applications. *Clinical Psychology Review, 21,* 421-446.

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, S. L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62,* 1157-1172.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261-265.

Owen, J. (2008). The nature of confirmatory strategies in the initial assessment process. *Journal of Mental Health Counseling, 30,* 362-374.

Pain, M. D., & Sharpley, C. F. (1989). Varying the order in which positive and negative information is presented: Effects on counselors' judgments of clients' mental health. *Journal of Counseling Psychology, 36,* 3-7.

Percevic, R., Lambert, M. J., & Kordy, H. (2004). Computer-supported monitoring of patient treatment response. *Journal of Clinical Psychology, 60(3),* 285-299.

Perfect, T. J., Hollins, T. S., & Hunt, A. L. R. (2000). Practice and feedback effects on the confidence-accuracy relation in eyewitness memory. *Memory, 8,* 235-244.

Peterson, D. K., & Pitz, G. F. (1986). Effects of amount of information on predictions of uncertain quantities. *Acta Psychologica, 61,* 229-241.

Pfeiffer, A. M., Whelan, J. P., & Martin, J. M. (2000). Decision-making bias in psychotherapy: Effects of hypothesis source and accountability. *Journal of Counseling Psychology, 47,* 429-436.

Phelps, R., Eisman, E. J., & Kohout, J. (1998). Psychological practice and managed care: Results of the CAPP practitioner survey. *Professional Psychology: Research and Practice, 29,* 31–36.

Plous, S. (1993). *The psychology of judgment and decision making.* New York: McGraw-Hill.

Ramsay, J. O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika, 38,* 513-533.

Ridley, C. R., & Shaw-Ridley, M. (2009). Clinical judgment accuracy: From meta-analysis to metatheory. *The Counseling Psychologist, 37,* 400-409.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry, 126,* 107-111.

Rock, D. L. (1994). Clinical judgment survey of mental health professionals: I. An assessment of opinions, ratings, and knowledge. *Journal of Clinical Psychology, 50,* 941-950.

Rosenfield, S. (1982). Sex roles and societal reactions to mental illness: The labeling of "deviant" deviance. *Journal of Health and Social Behavior, 23,* 18-24.

Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology, 61,* 145-153.

Skovholt, T. M., Rønnestad, M. H., & Jennings, L. (1997). Searching for expertise in counseling, psychotherapy, and professional psychology. *Educational Psychology Review, 9*, 361-369.

Slade, M., McCrone, P., Kuipers, E., Leese, M., Cahill, S., Parabiaghi, A., et al. (2006). Use of standardized outcome measures in adult mental health services. *British Journal of Psychiatry, 189,* 330-336.

Smith, J. D., & Agate, J. (2004). Solutions for overconfidence: Evaluation of an instructional module for counselor trainees. *Counselor Education & Supervision, 44,* 31-43.

Speer, D. C., & Greenbaum, P. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63,* 1044-1048.

Spengler, P. M. (1998). Multicultural assessment and a scientist-practitioner model of psychological assessment. *The Counseling Psychologist, 26*(6), 930-938.

Spengler, P. M., Strohmer, D. C., Dixon, D. N., & Shivy, V. A. (1995). A scientist-practitioner model of psychological assessment: Implications for training, practice, and research. *The Counseling Psychologist, 23*, 506-534.

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2009). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37,* 350-399.

Spengler, P. M., White, M. J., Ægisdóttir, S., & Maugherman, A. S. (2009). Time keeps on ticking: The experience of clinical judgment. *The Counseling Psychologist, 37,* 416-423.

Spielmans, G. I., Masters, K. S., & Lambert, M. J. (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology and Psychotherapy, 13,* 202-214.

Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry, 24,* 399-411.

Stankov, L., Lee, J., & Paek, I. (2009). Realism of confidence judgments. *European Journal of Psychological Assessment, 25,* 123-130.

Stein, D. M., & Lambert, M. J. (1984). On the relationship between therapist experience and psychotherapy outcome. *Clinical Psychology Review, 4*, 1-16.

Stricker, G. (2002). What is a scientist-practitioner anyway? *Journal of Clinical Psychology, 58,* 1277-1283.

Strohmer, D. C., Shivy, V. A., & Chiodo, A. L. (1990). Information processing strategies in counselor hypothesis testing: The role of selective memory and expectancy. *Journal of Counseling Psychology, 37,* 465-472.

Temerlin, M. K. (1968). Suggestion effects in psychiatric diagnosis. *Journal of Nervous and Mental Disease, 147,* 349-353.

Trueblood, W., & Binder, L. M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology, 12,* 13-27.

Tseng, W., McDermott, J. F., Jr., Ogino, K., & Ebata, K. (1982). Cross-cultural differences in parent-child assessment: U.S.A. and Japan. *International Journal of Social Psychiatry, 28,* 305-317.

Turk, D. C., Salovey, P., & Prentice, D. A. (1988). Psychotherapy: An information-processing perspective. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference, and judgment in clinical psychology* (pp. 1-14). New York: The Free Press.

Tutin, J. (1993). The persistence of initial beliefs in clinical judgment. *Journal of Social and Clinical Psychology, 12,* 319-335.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tyler, J. D. (2002). Treatment outcome assessment practices of psychology training clinics. *The Behavior Therapist, 25,* 144-147.

Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the Outcome Questionnaire. *Journal of Psychoeducational Assessment, 15,* 40-55.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74,* 242-261.

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome Questionnaire: Is it sensitive to changes in counseling center clients? *Journal of Counseling Psychology, 51,* 38-49.

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4*, 233-265.

Weiss, I., Robinowitz, J., & Spiro, S. (1996). Agreement between therapists and clients in evaluating therapy and its outcomes: Literature review. *Administration and Policy in Mental Health, 23,* 493-511.

Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59,* 595-613.

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50,* 59-68.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

VITA

Graduate College
University of Nevada, Las Vegas

Michael M. Haderlie

Degrees:

Bachelor of Science, Psychology, 2005
Brigham Young University

Master of Science, Clinical Psychology, 2007
Pacific University

Master of Arts, Psychology, 2009
University of Nevada, Las Vegas

Publications:

Haderlie, M., & Kern, J. (submitted). Beck Anxiety Inventory. In C. Reynolds, R. Kamphaus, & C. DeStefano (Eds.). *Encyclopedia of Psychological Testing*. Oxford University Press.

Allen, D. N., Donohue, B. C., Sutton, G., & Haderlie, M. (2010). Neuropsychology of substance use disorders in forensic settings. In A. M. Horton, Jr., & L. C. Hartlage (Eds.), *Handbook of Forensic Neuropsychology, 2nd Edition*, pp. 507 – 539. New York: Springer.

Allen, D. N., & Haderlie, M. (2010). Trail Making Test. In I. Weiner & E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology, 4th Edition*. John Wiley and Sons: New York.

Allen, D. N., Donohue, B. C., Sutton, G., Haderlie, M., & LaPota, H. (2009). Application of a standardized assessment methodology within the context of an evidence-based treatment for substance abuse and its associated problems. *Behavior Modification, 33*, 618-654.

Allen, D. N., Haderlie, M., Kazakov, D., & Mayfield, J. (2009). Construct Validity of the Comprehensive Trail Making Test in Children and Adolescents with Traumatic Brain Injury. Child Neuropsychology, 15, 543-553.

Haderlie, M.. (2009). *Examining the psychotherapist as a feedback receiver.* (Master's Thesis). ProQuest Dissertations and Theses. (Accesion Order No. AAT 1474381).

Dissertation Title: Enhancing Therapists' Clinical Judgments of Client Progress Subsequent to Objective Feedback

Thesis Examination Committee:

Chairperson, Christopher Heavey, Ph.D.
Committee Member, Jeffrey Kern , Ph.D.
Committee Member, Russell Hurlburt, Ph.D.
Graduate Faculty Representative, Stephen Fife, Ph.D.