Examining the Relationship Between Therapist Perception of Client Progress and Treatment Outcomes

Scott Waltman
Pacific University

Recommended Citation

This Thesis is brought to you for free and open access by the College of Health Professions at CommonKnowledge. It has been accepted for inclusion in School of Graduate Psychology by an authorized administrator of CommonKnowledge. For more information, please contact CommonKnowledge@pacificu.edu.
Examining the Relationship Between Therapist Perception of Client Progress and Treatment Outcomes

Abstract

The goal of therapy is to have a positive outcome, but how should practitioners determine when a positive outcome occurs? Diverse methods of studying and defining treatment outcomes have been developed; however, research suggests that the different methods do not measure the same construct. Numerous psychometrically sound tools for assessing treatment outcomes have been developed, but many clinicians do not administer such tools, and research suggests that many clinicians who do administer outcome tracking measures do not utilize the results. This study compared therapist judgment regarding outcome to client self-reported scores on an outcome measure. Clients of a psychological training clinic in the Pacific Northwest were administered the standardized outcome measure at the beginning and end of treatment; changes in scores were evaluated with the criteria of reliable change and clinically significant change. These scores were then compared to therapist judgment regarding the success of clients. Reliable change was found to be significantly related therapist description of treatment outcome, with a low level of agreement (Φ = .321). This finding has specific implications for the training of future psychologists as defining treatment outcomes is an important part of assessing treatment effectiveness and a critical part of improving treatment. Suggestions are made for maximizing the accuracy and practicality of client outcome assessment.

Degree Type
Thesis

Degree Name
Master of Science in Clinical Psychology (MSCP)

Committee Chair
Alyson M. Williams, Ph.D,

Subject Categories
Psychiatry and Psychology

This thesis is available at CommonKnowledge: https://commons.pacificu.edu/spp/120
Copyright and terms of use

If you have downloaded this document directly from the web or from CommonKnowledge, see the “Rights” section on the previous page for the terms of use.

If you have received this document through an interlibrary loan/document delivery service, the following terms of use apply:

Copyright in this work is held by the author(s). You may download or print any portion of this document for personal use only, or for any use that is allowed by fair use (Title 17, §107 U.S.C.). Except for personal or fair use, you or your borrowing library may not reproduce, remix, republish, post, transmit, or distribute this document, or any portion thereof, without the permission of the copyright owner. [Note: If this document is licensed under a Creative Commons license (see “Rights” on the previous page) which allows broader usage rights, your use is governed by the terms of that license.]

Inquiries regarding further use of these materials should be addressed to: CommonKnowledge Rights, Pacific University Library, 2043 College Way, Forest Grove, OR 97116, (503) 352-7209. Email inquiries may be directed to: copyright@pacificu.edu
EXAMINING THE RELATIONSHIP BETWEEN THERAPIST PERCEPTION OF CLIENT PROGRESS AND TREATMENT OUTCOMES

A THESIS
SUBMITTED TO THE FACULTY OF
SCHOOL OF PROFESSIONAL PSYCHOLOGY
PACIFIC UNIVERSITY
HILLSBORO, OREGON

BY
SCOTT WALTMAN

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE IN CLINICAL PSYCHOLOGY

MAY 27, 2010

APPROVED:

ALYSON M.WILLIAMS, PH.D, ASSISTANT PROFESSOR
Abstract

The goal of therapy is to have a positive outcome, but how should practitioners determine when a positive outcome occurs? Diverse methods of studying and defining treatment outcomes have been developed; however, research suggests the different methods do not measure the same construct. Numerous psychometrically sound tools for assessing treatment outcomes have been developed, but many clinicians do not administer such tools, and research suggests that many clinicians who do administer outcome tracking measures do not utilize the results. This study compared therapist judgment regarding outcome to client self-reported scores on an outcome measure. Clients of a psychological training clinic in the Pacific Northwest were administered the standardized outcome measure at the beginning and end of treatment; changes in scores were evaluated with the criteria of reliable change and clinically significant change. These scores were then compared to therapist judgment regarding the success of clients. Reliable change was found to be significantly related therapist description of treatment outcome, with a low level of agreement ($\Phi = .321$). This finding has specific implications for the training of future psychologists as defining treatment outcomes is an important part of assessing treatment effectiveness and a critical part of improving treatment. Suggestions are made for maximizing the accuracy and practicality of client outcome assessment.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>LITERATURE REVIEW</td>
<td>3</td>
</tr>
<tr>
<td>Methods for Investigating Treatment Outcomes</td>
<td>3</td>
</tr>
<tr>
<td>Treatment outcomes and premature termination</td>
<td>4</td>
</tr>
<tr>
<td>Premature termination in training clinics</td>
<td>5</td>
</tr>
<tr>
<td>Problems with Investigating Treatment Outcomes</td>
<td>7</td>
</tr>
<tr>
<td>Clinicians’ use of outcome measures</td>
<td>8</td>
</tr>
<tr>
<td>The need for more reliable methods</td>
<td>9</td>
</tr>
<tr>
<td>A Statistical Method of Defining Treatment Outcome</td>
<td>11</td>
</tr>
<tr>
<td>Reliable change and clinically significant change</td>
<td>11</td>
</tr>
<tr>
<td>Using reliable change and clinically significance to determine outcome</td>
<td>12</td>
</tr>
<tr>
<td>Clinically significant change may be too stringent</td>
<td>14</td>
</tr>
<tr>
<td>METHOD</td>
<td>15</td>
</tr>
<tr>
<td>Participants</td>
<td>15</td>
</tr>
<tr>
<td>Measures</td>
<td>16</td>
</tr>
<tr>
<td>Procedure</td>
<td>17</td>
</tr>
<tr>
<td>RESULTS</td>
<td>18</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>21</td>
</tr>
<tr>
<td>Examining the Reliable Change and Clinically Significant Change Rate</td>
<td>22</td>
</tr>
<tr>
<td>Limitations</td>
<td>24</td>
</tr>
<tr>
<td>Implications for the Disparity in Success Rates</td>
<td>24</td>
</tr>
<tr>
<td>Implications for Reliable Change Rate</td>
<td>25</td>
</tr>
<tr>
<td>Implications for Going Forward</td>
<td>27</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>29</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>35</td>
</tr>
</tbody>
</table>
Examining the Relationship between Therapist Perception of Client Progress and Treatment Outcomes

How do practitioners determine if psychotherapy has been successful? Do they rely on their intuition or more objective methods? Researchers have used divergent criteria for determining when treatment is successful (Hatchett & Park, 2003). Numerous psychometrically sound tools for assessing treatment outcomes have been developed, but many clinicians do not administer such tools (Lambert & Hawkins, 2004), and many clinicians who do administer outcome tracking measures do not utilize the results (Garland, Kruse, & Aarons, 2003). Researchers have found that, therapists’ perceptions of treatment outcomes may differ from clients’ experiences (Swift, Callahan, & Levine, 2009). This is a problem because accurately assessing client outcomes allows for determining whether treatment was effective. It is additionally a critical component of improving treatment (American Psychological Association, 2006). In psychological training clinics, it is especially important to study the methods of defining treatment outcomes because the training and competence of future psychologists is affected by their ability to assess client outcomes (Aubuchon-Endsley & Callahan, 2009; Hatcher & Lassiter, 2009; Kaslow et al., 2009; Rodolfa et al., 2005).

In a recent study involving treatment outcomes in a psychological training clinic (Swift et al., 2009), researchers compared the traditional methods of determining treatment outcomes (i.e., using arbitrary cutoff points and therapist perception) with a different method, using the criteria of reliable change and clinically significant change to determine treatment outcome. Jacobson, Follette, and Revenstorf (1984) introduced the
concept of reliable change and clinically significant change. Reliable change indicates that changes on a standardized outcome measure are large enough to be considered not due to chance (Jacobson et al., 1984; Wise, 2004). Clinically significant change indicates that reliable change has occurred and the change reflects recovery, meaning an individual’s score on a standardized outcome measure has moved from the clinical range of functioning to the nonclinical range of functioning (Jacobson et al., 1984; Wise, 2004). Comparison between the newer method of using the criteria of clinically significant change and the older methods of defining treatment outcome based on a predetermined minimum number of therapy sessions and therapist judgment showed a low level of agreement (Swift et al., 2009). This suggests that therapists who rely on their judgment alone are likely to describe treatment as having been successful when reliable change or clinically significant change has not occurred. Further investigation is needed to determine if this low level of agreement between therapist perception of treatment outcomes and client self-reported changes in functioning is found in other psychological training clinics.

Therapists may have accurate perceptions of treatment outcomes when the outcome is positive, but their perception of therapy may be discrepant from the clients’ experience when the treatment outcome is negative (Hunsley, Aubry, & Verstervelt, 1999). Although various methods exist for determining treatment outcome, these methods appear to measure different phenomena (Hatchett & Park, 2003; Swift et al., 2009). This study will examine client treatment outcomes in a training clinic setting. A traditional method of determining treatment outcomes will be compared to an alternative method suggested by Jacobson et al. (1984).
Literature Review

It is important for practitioners to be able to determine if psychotherapy has been successful. Diverse objective measures exist as means for evaluating treatment outcomes (Lambert & Hawkins, 2004), but many practitioners rely on their judgment alone to decide when treatment has been successful (Garland et al., 2003). There are mixed findings as to whether there is a relationship between treatment outcomes as defined by changes in scores on standardized outcome measures and treatment outcomes as defined by therapist judgment. In examining the relationship between the different methods of determining treatment outcomes it is important to review a brief history of the study of treatment outcomes, as this history continues to influence current practices.

Methods for Investigating Treatment Outcomes

As early as 1909, psychologists began studying treatment outcomes (Munsterberg, 1909). In 1952, Eysenck published a qualitative review of the existing treatment outcome research and literature (Eysenck, 1952). He concluded that the studies he reviewed failed to demonstrate that psychotherapeutic treatment facilitated recovery, and that any treatment gains could be explained by the simple passage of time (Eysenck, 1952). It has been suggested that Eysenck’s findings were influenced by poor research designs and low internal validity (Truax & Thomas, 2003). However, Eysenck’s findings influenced researchers and led to the development and utilization of more refined research methods (Truax & Thomas, 2003). Following the development of more refined methods, Smith and Glass (1977) conducted a meta-analysis of 375 studies and found that the average client receiving psychotherapy was better off than 75% of the untreated control group (Smith & Glass, 1977). This demonstrated that treatment gains could be
explained by more than just the simple passage of time. Over the years, researchers have
developed more refined methods for examining treatment outcomes. Although multiple
methods now exist, the degree of agreement between these diverse methods is variable
and understudied. A few common measures of treatment outcome that have traditionally
and more recently been used are explained below.

**Treatment outcomes and premature termination.** One historical method for
determining treatment outcome has been to evaluate whether a client ended therapy early
or received a “full dose” of therapeutic treatment (Howard, Kopta, Krause, & Orlinsky,
1986). Premature termination has been described as occurring when a client ceases
therapy prior to achieving predetermined treatment goals or attaining a requisite level of
improvement (Hatchett & Park, 2003). Researchers have used different terms to describe
premature termination, including: attrition, dropout, early termination, early withdrawal,
and unilateral termination (Swift et al., 2009). These terms are used synonymously
throughout this paper. Much of the research on therapy outcomes has been examining
client attrition.

The current methods commonly used to designate if a client has prematurely
terminated are: therapist judgment, termination by failure to attend the last scheduled
appointment, median split procedure, and failure to return after intake (for review, see
Hatchett & Park, 2003). A recent study concluded that these four methods of determining
dropout rates do not measure the same construct (Hatchett & Park, 2003). Wierzbicki and
Pekarik (1993) conducted a meta-analysis of 125 studies on psychotherapy dropout.
Wierzbicki and Pekarik (1993) found different rates of attrition depending on the method
used to measure premature termination. They concluded that therapist judgment may be
the best method for determining dropout, because therapist judgments about attrition have face validity. However, a problem with using therapist judgment to determine dropout is the issue of reliability (Wierzbicki & Pekarik, 1993). Not all therapists may decide a client has prematurely terminated from therapy based on the same criteria (Rodolfa, Rapaport, & Lee, 1983; Swift et al., 2009). Therapists’ descriptions of client termination may differ from a client’s description of the same termination (Hunsley et al., 1999).

**Premature termination in training clinics.** This project will examine the relationship between different methods of treatment outcome assessment in a training clinic setting. The extant literature reveals elevated rates of attrition in training clinics (Aubuchon-Endsley & Callahan, 2009; Callahan, Aubuchon-Endsley, Borja, & Swift, 2009; Richmond, 1992). A meta-analysis of the existing literature on premature termination found a mean dropout rate of 46.8% in general (Wierzbicki & Pekarik, 1993). In psychological training clinics, researchers have examined the rate of premature termination. Richmond (1992) examined the clinical and demographic variables of clients who prematurely terminated from a psychological training clinic; the researcher determined 59% of the participants prematurely terminated by the 10th session of treatment. In a training clinic, Callahan et al. (2009) conducted a research project comparing clients’ pretreatment expectations with premature termination rates. Of the 42 participants, 77.5% were determined to have prematurely terminated (Callahan et al., 2009). Aubuchon-Endsley and Callahan (2009) examined the pretreatment expectations and dropout rates in a training clinic. They determined that 67.92% of 53 participants prematurely terminated from treatment (Aubuchon-Endsley & Callahan, 2009). The reported dropout rates in training clinics, 59% (Richmond, 1992), 77.5% (Callahan et al.,
2009), and 67.92% (Aubuchon-Endsley & Callahan, 2009), are higher than the general mean dropout rate of 46.8% found by the meta-analysis of Wierzbicki and Pekarik (1993), which encompassed research conducted in a multitude of settings. Due to the negative implications of elevated attrition rates in training clinics, it is important to examine treatment outcomes in training clinic settings for psychologists in training and their clients.

Various researchers have suggested the elevated dropout rates of psychological training clinics have different implications. First, the elevated level of attrition in psychological training clinics may in part be accounted for by training clinic clients not seeing student therapists as experts (Aubuchon-Endsley & Callahan, 2009). Also, clients of therapists in training have been found to have slower rates of improvement when compared to clients of professional therapists (Reese, Norsworthy, & Rowlands, 2009). It is possible that clients become frustrated with their rate of improvement and dropout from treatment. Researchers have stated that the high levels of attrition in training clinics may limit the training opportunities of psychologists. Due to the high rates of early termination in a training clinic, psychologists in training are likely to gain more experience and competencies associated with starting psychotherapy than the competencies associated with the middle and end stages of therapy (Callahan et al., 2009). This may affect the development of the necessary skills competencies to become a competent psychologist (Hatcher & Lassiter, 2007; Kaslow et al., 2009; Peterson, Peterson, Abrams, & Stricker, 2006; Rodolfa et al., 2005). The impact of the training of future psychologists is an important reason to examine treatment outcomes and dropout rates in training clinics (Aubuchon-Endsley & Callahan, 2009).
Another reason it is especially important to study treatment outcomes in psychological training clinics is, many psychological training clinics operate on a sliding fee scale and provide services for people who otherwise may not be able to afford psychological services (Aubuchon-Endsley & Callahan, 2009; Callahan et al., 2009). Inspecting treatment outcomes of psychological training clinic clients is important, because these individuals may be underserved and at risk for poor outcomes. Clients of psychological training clinics who have a lower socioeconomic status are more likely to prematurely terminate from therapy (Richmond, 1992). In addition, as sites where future psychologists are gaining valuable experience, premature client termination must be understood in order to be minimized.

In order to understand premature termination and other treatment outcomes it is essential that reliable methods exist for determining the outcome of therapy. Of the different methods for determining treatment outcomes, researchers have suggested that therapist judgment may be the best (Wierzbicki & Pekarik, 1993), but some researchers have suggested that using standardized measures may be a more reliable than therapist judgment (Lambert & Hawkin, 2004; Swift et al., 2009; Jacobson & Truax, 1991). This project will examine the relationship between these two methods of determining outcomes.

**Problems with Investigating Treatment Outcomes**

As mentioned above, diverse methods have been developed to describe whether treatment has been successful. Each method has strengths and drawbacks. Therapist judgment and the use of standardized outcome measures have corresponding advantages and disadvantages when defining therapy outcomes. The usefulness of standardized
outcomes measures is dependent upon the clinician to interpret and utilize the results of those measures, and the reliability of therapist judgment has been questioned (Wierzbicki & Pekarik, 1993).

**Clinicians’ use of outcome measures.** In examining the relationship between therapist judgment of treatment outcomes and standardized outcome measures, it is necessary to examine therapist attitude toward and use of outcome measures. Numerous outcome measures have been developed to help researchers and clinicians identify when treatment has been successful (Hatfield & Ogles, 2004; Garland et al., 2003; Mours, Campbell, Gathercoal, Peterson, 2009; Slade et al., 2006); however, it has been suggested that less than one third of all clinicians use outcome tracking measures in their practice (Lambert & Hawkin, 2004). Garland et al. (2003) examined the attitudes clinicians have about the use of outcome measures. They found many clinicians reported viewing outcome measures as cumbersome or intrusive. Many clinicians reported the use of standardized outcome measures did not produce data they would not have already known (Garland et al., 2003). The use of outcome tracking measures can be expensive to practitioners and burdensome to clients (Garland et al., 2003), and there are conflicting findings as to whether the use of outcome measures lead to improvements in treatment and treatment outcomes (Reese et al., 2009; Slade et al., 2006). Many clinicians reported using “real world functional indicators” (e.g., school grades, disciplinary actions), clinician intuition, and client reported satisfaction more frequently than standardized measures in evaluating the effectiveness of treatment (Garland et al., 2003). Paradoxically, many of the responding clinicians, who reported that outcome measures
had little clinical utility, went on to give examples of how the use of such measures had been helpful to treatment in some way.

In her study of 50 mental health providers in San Diego County, as many as 92% of the clinicians surveyed by Garland et al. (2003) reported they had never used scores from standardized measures in their practice; it is unclear if this percentage is representative of the entire population, as they were mandated by the state to use the outcome measures. In a different survey 37% of respondents reported using some sort of standardized assessment in their practice, in general the respondents who used assessment tools were younger, provided more hours of therapy per week, worked in an institutional setting, and practiced therapy from a cognitive-behavioral perspective (Hatfield & Ogles, 2004). It is clear that some therapists may view their clinical judgment as being more accurate or useful than the information that is obtained from the use of outcome measures. It is unclear whether this trend also exists in a training clinic setting.

Understanding the relationship between therapist judgment about treatment outcomes and client changes in scores on outcome measures, will be informative as to whether some student therapists view their clinical judgment as being more accurate and useful than the information generated from standardized outcome measures.

The need for more reliable methods. It has been suggested that therapist judgment may be the best method of determining when a client has prematurely terminated (Wierzbicki & Pekarik, 1993); this statement does not necessarily support the conclusion that therapist judgment is the best method of determining whether treatment was successful. In fact, it has been suggested that in describing the effectiveness of treatment, therapist perception of progress and client self-report of satisfaction have a low
level of agreement (Mours et al., 2009). Yet multiple surveys of practicing clinicians have found that many clinicians are likely to rely on their own intuition and the client’s self-report to determine if treatment has been successful (Garland et al., 2003; Mours et al., 2009).

There is a danger to therapists relying solely on their judgment. Research has demonstrated that therapists are more likely to be accurate in judgments about treatment outcomes when there is a positive outcome, but are likely to be less accurate when a negative outcome occurs (Hunsley et al., 1999). This may be due to biases in the judgment of the clinicians. It has been argued that therapists, just like other humans, have tendencies to make errors in judgment (Ruscio, 2007). Specifically therapists may be likely to attend to positive information about treatment outcomes more than negative information about treatment outcomes (Ruscio, 2007). This is consistent with the finding that therapists are more accurate in their judgments about treatment outcomes when the outcome is positive. Biases in therapist judgment are problematic and decrease the accuracy and reliability of therapist judgment.

Hatchett and Park (2003) stated that researchers should develop more reliable methods of determining treatment outcomes. They suggested that clients should be administered a standardized measure at the beginning and end of treatment, and clients whose scores drop into the nonclinical range would be considered as treatment successes. The use of psychometrically sound assessment measures to define treatment outcomes is consistent with the suggestions of Meehl (1957). In a seminal article, Meehl (1957) rhetorically asked when psychologists should use their heads instead of an objective
measure. He answered his own question, by stating that except in rare and obvious exceptions, psychologists should trust the measures over their clinical judgment. To apply the writings of Meehl to the assessment of treatment outcomes, unless a treatment outcome is as obvious as a “broken leg” (Meehl, 1957), clinicians and researchers should use standardized outcome measures in determining treatment outcomes, at least as a supplement to relying on their intuition. Hatchett and Park (2003) suggested determining treatment outcome based on a standardized measure may be the most reliable and valid method.

**A Statistical Method of Defining Treatment Outcome**

This study will examine client treatment outcomes in a training clinic setting. Therapist judgment, a traditional method of determining treatment outcomes, will be compared to a more objective method introduced by Jacobson et al. (1984).

**Reliable change and clinically significant change.** The concept of reliable change and clinically significant change was introduced by Jacobson et al. (1984) as a statistical approach of determining treatment outcomes. Reliable change indicates that an individual’s change in scores on a measure is sufficiently larger than the standard error of the difference between the two scores, and thus is not due to chance (Jacobson & Truax, 1991). As noted previously, clinically significant change is more stringent than reliable change as a criterion for determining treatment outcome. When an individual’s score on an outcome measure moves from the dysfunctional to functional range and that change is large enough to be designated reliable, it is determined that the change is clinically significant (Jacobson & Truax, 1991). Reliable change indicates a real and substantial
change in functioning. Clinically significant change indicates a return to a normative level of functioning.

The use of the criteria of reliable change and clinically significant change has advantages over other statistical methods. A traditional statistical method of comparing different groups is comparing group means (e.g., ANOVA) which has two major limitations (Jacobson, Roberts, Berns, & McGlinchey, 1999). First, comparing group means provides little information about the variability within those groups, or the proportion of group members who improved or recovered. Second, a large effect size does not mean the effect was clinically meaningful (Jacobson et al., 1999). Clinically significant change indicates a return to a normal level of functioning; in some cases returning to a normal level of functioning may not be an appropriate goal of treatment (Jacobson et al., 1999). However, providers and consumers of mental health services are still interested in knowing how often a return to normal functioning occurs. Reliable change indicates change is large enough to be considered not due to chance, and clinically significant change indicates the change is meaningful (Jacobson & Truax, 1991; Wise, 2004). The use of reliable change and clinically significant change to define treatment outcomes can be useful to determine if improvement is large enough to be considered reliable and meaningful (Jacobson et al., 1999).

**Using reliable change and clinical significance to determine outcomes.** In a recent study on premature termination, researchers determined clients had prematurely terminated when they discontinued treatment prior to obtaining reliable change and prior to obtaining clinically significant change (Swift et al., 2009). They compared the
traditional methods of determining treatment outcomes with the statistical methods of 
reliable change and clinically significant change suggested by Jacobson and Truax (1991). That study examined the data from 135 clients who received therapy in a 
university training clinic (Swift et al., 2009). Each client’s outcome was evaluated using 
traditional and newer statistical methods; different rates of dropout were found for each 
method used. Using the criteria of clinically significant change the investigators labeled 
77% of the clients as dropping out, and using the reliable change criteria they classified 
63% of the clients as dropping out. Using traditional methods different dropout rates were 
found: median split criteria 50% of clients, intake-only criteria 8.1% of clients, missed-
appointment criteria 48.1% of clients, and therapist judgment classified 74.1% of clients 
as prematurely terminating (Swift et al., 2009). The researchers concluded that traditional 
methods often labeled treatment as successful when the client had not recovered, and 
sometimes traditional methods labeled treatment as unsuccessful when the client had 
recovered (Swift et al., 2009).

Other studies have used the criteria of reliable change and clinically significant 
change to determine when treatment has been successful (Callahan & Hynan, 2005;
Callahan, Swift & Hynan, 2006). Callahan and Hynan (2005) examined the treatment 
outcomes of 61 clients who were treated in a psychological training clinic, comparing the 
clients’ changes in scores on the Outcome Questionnaire 45.2 (OQ; Lambert et al., 2004) 
from the beginning to end of treatment. The researchers used the recommended minimum 
change of 14 or more to designate reliable change and a cutoff score of 63 to designate 
clinically significant change. Of the 61 clients, 18% achieved clinically significant
change, 33% demonstrated reliable change and 67% showed no positive reliable change (Callahan & Hynan, 2005).

Clinically significant change may be too stringent. In order for an individual to obtain clinically significant change, they must begin treatment in the clinical range of functioning and move into the nonclinical range of functioning (Wise, 2004). Due to the fact that some clients enter treatment with scores in the nonclinical range, this may be too stringent of a criterion for determining treatment outcomes. One study found that 28.4% of their sample began treatment with scores in the nonclinical range of functioning (Reese et al., 2009). In instances where a sizeable portion of the population being served begins treatment in the nonclinical range, the criterion of reliable change may be a more appropriate method of determining when treatment has been successful. Also, in some instances, such as the treatment of chronic conditions, reaching a non-clinical level of functioning may not be an appropriate goal of treatment (Jacobson et al., 1999). In these instances reliable change may be more appropriate than clinically significant change as a goal of treatment.

This investigation will examine clients’ changes in OQ scores to determine whether reliable change and/or clinically significant change has occurred. This study will investigate if there is a relationship between therapists’ description of the outcome of therapy in a psychological training clinic and the clients’ changes in OQ scores, and if a relationship is found, the magnitude of the relationship will be examined.
Method

The following research questions guided this examination: (a) What is the nature of the sample? (b) How often is reliable change and clinically significant change occurring? (c) What is the relationship between therapist judgment of treatment outcome and client changes in OQ scores? In order to answer these research questions, the following methods were employed.

Participants

Clients. Archival data from 144 clients seen for therapy during the 2007-2008 academic school year at a pacific northwestern university-based psychological training clinic were used in this project. Of the 144 clients, the data from 45 of the clients were judged to be inappropriate for this study due to missing information or the client having only attended an intake session. Clients who were identified as only attending an intake session were not used for this examination, as it has been suggested that clients who did not return for therapy after the intake never started treatment (Hatchett & Park, 2003). Of the remaining 99 clients, demographic data were available for 66 clients (i.e., 67%) and missing for the rest. Demographic information about the clients was recorded on the Closing File Outcome Tracking Form (see Appendix for Closing File Outcome Tracking Form).

Therapists. Clients were seen by doctoral students in clinical psychology who were supervised by licensed psychologists. This training program follows a practitioner-scholar model, and student-therapists were either in their first, second, or third year of training. In addition, three of the approximate forty-five therapists who recorded data for this study were advanced level pre-doctoral interns who were either in their fifth or sixth
year of training. The therapists were not identifiable individually from the recorded data, so the number of therapists was approximated due to an effort to protect their privacy.

**Measures**

**Outcome Questionnaire.** The outcome measure used in this clinic and one commonly used in psychological training clinics is the Outcome Questionnaire 45.2 (OQ) (Anderson & Lambert, 2001; Callahan et al., 2006; Callahan & Hynan, 2005, Lambert et al., 2004, Lambert & Hawkins, 2004; Swift et al., 2009). The OQ is a 45 item self-report measure that is easy to administer and score. The OQ is administered to every client at every session at this psychology training clinic. Other studies have used the OQ to measure reliable change and clinically significant change and to identify when individual treatment has been successful (Callahan & Hynan, 2005; Callahan et al., 2006). On the OQ an overall score above 63 is indicative of symptoms in the clinical range (Anderson & Lambert, 2001). The OQ has been found to have adequate reliability and validity (Lambert et al., 2004). The OQ manual states that no gender differences exist between male and female scores, also no significant differences have been found between the total scores of Caucasians, Hispanics, and African-Americans.

**Closing File Outcome Tracking Form.** The Closing File Outcome Tracking Form (see Appendix) was created for use in this particular training clinic. It is a brief form that is filled out by the client’s assigned therapist as a part of the file closing procedure. This form contains information about diagnosis, client scores on the OQ, therapist description of therapy outcome, therapy duration, and demographic information. For this study, only a subset of information from the form was obtained: data related to
OQ scores from the first and last administration of the measure, therapist description of the outcome of therapy, and demographic information.

**Procedure**

The current project utilized archival data that were collected as part of a previous project (Brown, Williams, Waltman, & Sutton, 2010). Clients were informed that their information may be used for research purposes during the informed consent process at the beginning of therapy. The data were taken from the Closing File Outcome Tracking Form (see Appendix); this information included length of treatment, diagnosis, client scores on the OQ, therapist description of therapy outcome, and demographic information. Client data were de-identified and coded into a database by the previous researchers. Approval for the study was obtained from the Institutional Review Board (IRB) and the database containing the de-identified client data was utilized for statistical examination. Client data include therapist description of treatment outcome. Therapists described the outcome of therapy as: successful completion of treatment, substantial progress without successful completion of treatment, incomplete or moderate progress, or no progress.

Pearson’s Chi-Square ($\chi^2$) was utilized to determine if there is a relationship between therapist description of treatment outcome and client changes in OQ scores. A phi correlation ($\Phi$) was conducted to measure the relationship between changes in OQ scores and therapist rating of treatment successes. Data were coded into three variables: therapist judgment based on therapist perception, positive reliable change based on whether the client’s score on the OQ dropped by 14 or more points, and clinically significant change based on whether positive reliable change occurred and whether the
client’s OQ score dropped from above 63 to 63 or below (i.e., from a clinical to a nonclinical range).

**Therapist Judgment.** The information about therapist judgment of the treatment outcome was recorded on the Closing File Outcome Tracking Form by each therapist practicing in the training clinic (see Appendix for Closing File Outcome Tracking Form). Therapist judgment was coded as it was recorded on the form: “successful completion of treatment,” “substantial progress without successful completion of treatment,” “incomplete or moderate progress,” or “no progress.”

**Positive Reliable Change and Clinically Significant Change.** A decrease in total OQ scores of 14 or more was coded as “positive reliable change.” A change in total OQ scores of 13 or less will be coded as “no positive reliable change.” Clinically significant change requires that reliable change has occurred and that a predetermined cutoff point has been crossed. An decrease in total OQ scores of 14 or more and a drop from a total score of 64 and above to 63 or below was coded as “clinically significant change” a failure to attain a drop in OQ scores of at least 14 points or a failure to cross the cutoff point of 63 was coded as “no clinically significant change.”

**Results**

In order to determine the nature of the sample, descriptive statistics were utilized; it was revealed that the average client age was 32 years ($\bar{X} = 32.24; M_p = 29$; range: 17 to 57); 56.1% of the clients were female, 43.9% were male, and 0% were coded as transgendered or other. Of clients, 87.9% were described as being Caucasian while 4.5% were from an Asian background, 1.5% were African-American, 4.5% were Multi-Ethnic, and 1.5% were of an unknown ethnicity.
Descriptive statistics were used to compute the frequency of reliable change, clinically significant change, and therapist description of treatment outcomes. Overall, from the beginning to the end of treatment, the mean change in total OQ score for the entire sample was a decrease of 9.67 points ($\bar{X} = 9.67$). It was found that only 35.4% of treatment outcomes had achieved positive reliable change and 64.6% had not achieved positive reliable change. In determining clinically significant change, it was found that only 18.2% of the clients had achieved clinically significant change; it is worth noting that 47.5% of the clients began treatment with an OQ score in the nonclinical range, so they could not achieve clinically significant change. It was found that therapists described 10.1% of treatment outcomes as no progress, 33.3% of treatment outcomes as incomplete or moderate progress, 21.2% of treatment outcomes as substantial progress without successful completion of treatment, and 35.4% of treatment outcomes as successful completion of treatment. Table 1 contains the percentages of treatment outcomes that achieved positive reliable change or clinically significant change and the percentage of treatment outcomes according to therapist description.

In order to evaluate whether therapist description of treatment outcome was related to whether client changes in OQ scores reflected positive reliable change, a two-way contingency table analysis was created. The two variables were therapist judgment with four levels (successful completion of treatment, substantial progress without successful completion of treatment, incomplete or moderate progress, or no progress.) and positive reliable change (positive reliable change and no positive reliable change). The two variables were found to be significantly related, Pearson $\chi^2 (3, N = 99) = 10.18,$
A Phi correlation ($\Phi$) of .32 is consistent with a medium effect size. These results support the conclusion that therapist judgment of treatment outcome are related to whether reliable change occurs.

Table 1

*Client Treatment Outcomes by Percentages: Comparing Therapist Judgment and Reliable Change*

<table>
<thead>
<tr>
<th>Therapist Judgment of Treatment Outcome</th>
<th>Total&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Positive Reliable Change</th>
<th>Clinically Significant Change&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Success Completion of Treatment</td>
<td>35.4%</td>
<td>18.2%</td>
<td>17.2%</td>
</tr>
<tr>
<td>Substantial Progress Without Successful Completion of Treatment</td>
<td>21.2%</td>
<td>9.1%</td>
<td>12.1%</td>
</tr>
<tr>
<td>Incomplete or Moderate Progress</td>
<td>33.3%</td>
<td>7.1%</td>
<td>26.3%</td>
</tr>
<tr>
<td>No Progress</td>
<td>10.1%</td>
<td>1.0%</td>
<td>9.1%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>35.4%</strong></td>
<td><strong>64.6%</strong></td>
<td><strong>18.2%</strong></td>
</tr>
</tbody>
</table>

<sup>a</sup> n = 99. <sup>b</sup> 47.5% of clients began treatment with OQ scores in the nonclinical range, and subsequently could not achieve clinically significant change.

In order to determine if therapist judgment was related to whether client changes in total OQ scores meet criteria for clinically significant change, a second two-way contingency table analysis was done. The two variables were therapist judgment and clinically significant change (clinically significant change and no clinically significant change). The two variables were not found to be significantly related, Pearson $\chi^2 (3, N = 99) = 5.075$, $p = .05$. These results indicate that therapist judgment of treatment outcome and whether clinically significant change occurs is not related. In general these results
support the conclusion that therapist judgment and the criterion of reliable change are related; also, the criterion of reliable change may have more utility than the criterion of clinically significant change.

**Discussion**

These results suggest that therapist judgment of treatment outcomes are related to positive reliable change on an outcome measure, but not related to clinically significant change on the outcome measure used in this study with this sample. Although these results suggest that therapist judgment and reliable change are related, a Phi correlation of .32 suggests that there is a low level of agreement between these two methods of determining treatment outcomes. It is unclear from these analyses why a therapist would label treatment as being successful when reliable change had not occurred, or label treatment as unsuccessful when reliable change had occurred. Further research is needed to explore why this low level of agreement exists.

The rates of reliable change and clinically significant change found in this project were consistent with the rates reported in other psychological training clinics. In this study it was found that 35.4% of clients achieved positive reliable change. Swift et al. (2009) found that 37% of the clients of a training clinic in their study experienced positive reliable change. Callahan and Hynan (2005) found that 33% of the clients in the training clinic they studied reached positive reliable change. In this study it was found that 18.2% of clients achieved clinically significant change; it is worth noting that 47.5% of the clients in this project could not achieve clinically significant change because they began treatment with a score in the nonclinical range. Swift et al. (2009) found that 23% of the clients from the training clinic they examined reached clinically significant change.
Callahan and Hynan (2005) found that 18% of the clients of the training clinic they examined achieved clinically significant change. As such, the rates of reliable change and clinically significant change found in this study are consistent with the rates reported by other training clinics.

These results support the conclusion that reliable change may have more utility than clinically significant change as criteria for describing the outcome of therapy in this type of setting. Jacobsen et al. (1999) acknowledged that clinically significant change may not always be an appropriate goal of treatment. With 47.5% of the clients beginning treatment with an OQ score of 63 or below and thus being unable to achieve clinically significant change, it would seem that the criteria of clinically significant change is not always an appropriate method to determine the outcome of treatment. The less stringent criteria of reliable change, which can apply to all clients, whether they start in “clinical” or “nonclinical” ranges of functioning, may be a more appropriate method of determining when treatment has been successful.

**Examining the Reliable Change and Clinically Significant Change Rate**

It has been suggested that therapist perception of progress and client self-report of satisfaction have low levels of accuracy in describing the effectiveness of treatment (Mours et al., 2009). However, in this project therapist perception of treatment outcome was found to be significantly related to the criteria of reliable change on a standardized measure. In this study, therapists judged that 56.6% of clients achieved successful completion of treatment or made substantial progress; whereas, 35.4% of clients were found to have met the criteria for reliable change. When comparing treatment success rates, there is a disparity of over 20%. It is possible that this disparity is due to errors in
therapist perception (Ruscio, 2007); it is also possible that in some cases treatment was successful despite there being no reliable change on the total OQ score.

The OQ is a brief outcome measure that captures many aspects of a person’s life and generates an overall score (Lambert et al., 2004). For clients assessed in this study, the mean change in total OQ scores was a decrease by 9.67 points; this change is larger than the standard error of the difference between pre and post test scores, but not sufficiently larger than that error to be classified as reliable change. A decrease in overall score by 9.67 could indicate reliable change on one of the subscales of the OQ, but subscale data were not collected as part of this study. It is possible that if subscale data of the OQ were collected that individuals who did not demonstrate reliable change on the OQ may demonstrate reliable change on one of the subscales of the OQ. The therapists who rated the successfulness of treatment had this subscale data, and that may in part account for the disparity between therapist judgment and reliable change rates.

Another possibility is that if these clients were administered a standardized measure specific to their presentation that perhaps reliable change may have been found. The OQ, which is a very useful tool, is a measure of general distress. For this study, client OQ scores were the only outcome measure data collected. In many instances, as in this clinic, additional standardized measures are utilized. For example clients being treated for an Anxiety Disorder are often administered an inventory specific to anxiety, such as the Beck Anxiety Inventory (BAI; Beck & Steer, 1990). If a client were repeatedly administered an additional measure, the scores from that measure may have influenced the therapists’ description for the outcome of therapy; however, data from any
additional measures were not collected as a part of this investigation. This may account for some of the disparity in treatment success rates.

**Limitations**

There are limitations to the findings of this project. This study utilized archival data, and there are several limitations which are associated with the nature of archival data. The type of data that was available was limited to what was previously collected; the information about therapist judgment of outcome was categorical which limited the types of analyses which could be conducted. This study utilized nonparametric statistics. Nonparametric statistics have no parameters to describe and so it can be difficult to estimate how the findings of this study represent the general populations. Because nonparametric statistics do not have to meet any assumptions such as the assumption of normality, it will be difficult to generalize these finding to the general population. Subscale scores for the OQ were not collected as part of the previous study; had this information been collected further analysis could have been conducted. This study was conducted in the Pacific Northwest with a sample of clients that may not represent the cultural or ethnic diversity of other regions. Of the 99 clients whose information was utilized for this study, demographic data were only available for 66 of those clients; this limits generalizability of the findings. Also no demographic data about the therapists were collected; it is possible that some therapist characteristic (e.g., years of training) may reveal interesting information.

**Implications for the Disparity in Success Rates**

There are implications for the disparity between treatment success rates found in this study. Student therapists should be trained to recognize that their perception of the
progress the client is making may not accurately reflect the client’s self report.

Psychologists in training may be, in some cases, selectively attending to information which confirms their beliefs that the client is making progress, and ignoring any conflicting information (Ruscio, 2007). A good practice for any practitioner would be for them to purposefully look for information which is not consistent with their perception of treatment. The use of standardized outcome measures can and should be a preferred method to evaluate the accuracy of a clinician’s intuition. Psychologists in training are forming the habits which will guide their future practice, and it is imperative that they form habits which will consistent with best and ethical practice guidelines (APA, 2006).

**Implications for Reliable Change Rate**

It has been suggested that when a client discontinued treatment prior to obtaining a requisite level of improvement it should be designated that client prematurely terminated from treatment (Hatchett & Park, 2003). In this study 64.6% of clients discontinued treatment prior to obtaining reliable change, and subsequently may by some definitions be labeled as prematurely terminating from therapy (Swift et al., 2009). This rate is dramatically higher than the mean dropout rate of 46.8% found by Wierzbicki & Pekarik (1993). There are many implications to consider regarding the elevated dropout rates in training clinics. As discussed earlier, psychologists in training who experience a high level of attrition are likely to gain more experience learning how to initiate treatment than how carry out the middle and end stages of treatment. Training clinic directors and clinical supervisors are then forced to maximize and enhance training about the middle and end stages of treatment. Many common practices such as group supervision, grand rounds, case presentations, role-plays, and co-therapy may help to enhance the training of
future psychologists and subsequently aid in the development of the necessary skills and competencies (Hatcher & Lassiter, 2007; Kaslow et al., 2009; Peterson et al., 2006; Rodolfa et al., 2005). It is important to examine how the elevated drop-out rates found in training clinics affect psychologists in training, and how these effects can be compensated for.

The elevated drop-out rate found in training clinics affects psychologists in training, and it may also impact clients who receive services from psychological training clinics. In discussing the elevated dropout rates of training clinics, Callahan et al. (2009) suggested it should be considered whether clients seen in a training clinic should be informed at intake that progress may be slower in a training clinic, than in another setting. Another point to consider is whether clients seen in a psychological training clinic should be informed about the rate of reliable change and clinically significant change of such clinics. The directors of psychological training clinics may want to consider adding this to the informed consent process; though additional research is needed to establish that the rate of progress is slower in training clinics when compared to other therapy settings. The amount of research necessary to support adding this information to the informed consent process is unclear, but administrators should consider how accurately their informed consent reflects the therapy experience. It is possible that if clients knew that progress may be slower in a training clinic, they may become less inpatient at the rate of improvement, and subsequently be less likely to drop out of treatment.
Implications for Going Forward

This study found both situations where reliable change occurred and the therapist did not describe treatment as being successful and situations where reliable change did not occur and the therapist did describe treatment as being successful. In comparing the different methods for determining treatment outcomes, without knowing the specific details of each case, it is unclear if one method is more accurate than another. It is also unclear what a clinician should do when these two methods produce different descriptions for the same outcome. Should clinicians follow the counsel of Meehl (1957) and disregard their own judgment? Or should the clinicians trust their judgment and intuition? The answer to these questions is beyond the scope of this paper, but it would seem imperative to recognize the discrepancies in the descriptions of outcomes. To account for the low level of agreement between the two methods, therapists should draw information about the disposition of treatment from a number of sources; a Multitrait Multimethod type approach (MTMM; Campbell & Fiske, 1959) may be appropriate in determining whether treatment was successful. Gathering information about the outcome of treatment from multiple sources may provide a clearer and more comprehensive description of the outcome of therapy.

One method of determining treatment outcomes which could incorporate both therapist perception and standardized measures is the use of the client’s treatment plan as the standard for evaluating the successfulness of treatment. A good treatment plan should include realistic goals which are measurable. If clients meet their treatment goals then treatment has been successful. If clients do not meet their treatment goals then treatment has not yet been successful. A treatment plan should include what the client expects to
get out of treatment. Research has found that many clinicians prefer using “real world functional indicators” instead of standardized measures in the evaluation of the effectiveness of treatment (Garland et al., 2003). Using the client’s treatment plan to evaluate the successfulness of treatment may be a valuable addition to the use of therapist judgment and standardized outcome measures. It is possible to construct treatment plans that measure progress based on “real world functional indicators” (e.g., improvements in school or work), scores on outcome measures, and client reported improvement (e.g., subjective units of distress scales). Research on the use of treatment plans to evaluate when treatment has been successful will demonstrate whether a client’s treatment plan is an appropriate method for determining the outcome of treatment. It is likely that research may find the essential components needed for a treatment plan to be an appropriate means of determining when treatment is successful.
References


and scoring manual for the Outcome Questionnaire (OQ-45.2). Orem, UT:
American Professional Credentialing Service.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? Journal of
Counseling Psychology, 4, 268-273.

use of psychotherapy outcome assessment measures at psychology internship
sites. Training and Education in Professional Psychology, 3(3), 169-176.


Council of Schools and Programs of Professional Psychology educational model.
Training and Education in Professional Psychology, S(1), 17-36.

system improve psychotherapy outcome? Psychotherapy Theory, Research,
Practice, Training, 46(4), 418-431.

Richmond, R. (1992). Discriminating variables among psychotherapy dropouts from a
psychological training clinic. Professional Psychology: Research and Practice,
23(2), 123-130.

model for competency development: Implications for psychology educators and

Ruscio, J. (2007). The clinician as subject: Practitioners are prone to the same judgment errors as everyone else. In S. O. Lilienfeld & W. T. O’Donohue (Eds.), *Great ideas of clinical science: 17 principles that every mental health professional should understand* (pp. 29-47). New York: Routledge.


## Pacific Psychological Services: Closing File Outcome Tracking

<table>
<thead>
<tr>
<th>OQ DATA</th>
<th>Date</th>
<th>OQ Overall Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earliest Administration w/current therapist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Administration w/current therapist</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnostic DATA</th>
<th>Date</th>
<th>Diagnosis/Diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intake or Earliest Diagnosis Axis I and II this treatment episode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Axis I and II Diagnosis this treatment episode</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Treatment Duration w/current therapist in weeks: __________

Did this client have treatment with a different therapist previously (circle one)?  YES  NO

- If “yes,” how long was most recent prior treatment episode (in weeks)? __________
- If “yes,” how many total previous therapists did the client have at this clinic? __________

**Reason for Closing (circle one):**
- a) No attempts to engage in therapy
- b) Never showed after intake
- c) Dropped out after session number:
- d) Completed treatment

**Outcome of Therapy with this Therapist (circle one):**
- a) No progress
- b) Incomplete or moderate progress
- c) Substantial progress without successful completion of treatment
- d) Successful completion of treatment
### e) Demographic Information:

Please describe to the best of your ability your client’s stance on these variables. They are not things to be asked of your client, but simply demographic information to capture from existing records regarding clients’ background. Fill in as much as you can using existing client records and your knowledge of the client. Also note where you got the information recorded on this form (i.e. intake report, progress notes, admin electronic records, etc.)

<table>
<thead>
<tr>
<th>DEMOGRAPHIC VARIABLES</th>
<th>CLIENT INFORMATION (describe)</th>
<th>Where Information Came From (if “other,” please describe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at Termination</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Disability status</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Sexual orientation</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Indigenous heritage</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Therapy Language</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Nationality</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>SES: Fee Rate at Termination</td>
<td></td>
<td>Intake Report Other:</td>
</tr>
<tr>
<td>SES: Reported Monthly Income (if reported in records, if not, put “NA”)</td>
<td>Administration Electronic Records Other:</td>
<td></td>
</tr>
<tr>
<td>Religious/Spiritual Background</td>
<td>Intake Report Other:</td>
<td></td>
</tr>
<tr>
<td>Other diversity considerations</td>
<td>Intake Report Other:</td>
<td></td>
</tr>
</tbody>
</table>