



2014-06-01

Predicting Youth Treatment Failure: An Investigation of Clinical Versus Actuarial Judgment

Tessa Nicole Salisbury
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Psychology Commons](#)

BYU ScholarsArchive Citation

Salisbury, Tessa Nicole, "Predicting Youth Treatment Failure: An Investigation of Clinical Versus Actuarial Judgment" (2014). *All Theses and Dissertations*. 5266.
<https://scholarsarchive.byu.edu/etd/5266>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Predicting Youth Treatment Failure:
An Investigation of Clinical Versus
Actuarial Judgment

Tessa Nicole Salisbury

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Jared S. Warren, Chair
Michael J. Lambert
Mikle D. South
Dawson W. Hedges
Chad D. Jensen

Department of Psychology
Brigham Young University
June 2014

Copyright © 2014 Tessa Nicole Salisbury
All Rights Reserved

ABSTRACT

Predicting Youth Treatment Failure: An Investigation of Clinical Versus Actuarial Judgment

Tessa Nicole Salisbury
Department of Psychology, BYU
Doctor of Philosophy

Research investigating clinical versus actuarial prediction and judgment has consistently demonstrated the superiority of actuarial (statistical) methods. Little research to date has directly compared clinical and actuarial predictions in the context of patient-focused psychotherapy outcomes. The most relevant study on this issue was completed with an adult population and results indicated that the actuarial method was significantly more accurate at predicting client treatment failure compared to clinician's predictions. This study examined clinical versus actuarial prediction of client deterioration in a sample of children and adolescents receiving treatment in a managed care and community mental health setting. Predictions of treatment failure made by the actuarial method were found to be significantly more accurate than predictions of treatment failure made by clinicians. More specifically, participating clinicians did not make a single prediction of treatment failure. These findings add further evidence to support the use of actuarial methods in enhancing clinical decision-making in community-based mental health services for children and adolescents.

Keywords: deterioration, clinical judgment, actuarial judgment, and Youth Outcome

Questionnaire

ACKNOWLEDGEMENTS

My experience as a graduate student at Brigham Young University was great largely because of the faculty with whom I had the opportunity to interact. They truly care for each student and have a strong desire for each student to succeed. I can recall several instances when I personally felt them rallying around me, providing care and support.

I would like to acknowledge and thank my chair and mentor, Dr. Jared S. Warren, for having confidence in me. There were definitely occasions when I relied on his confidence to buoy me up. Additionally, I am extremely grateful for the countless examples he set for me. I respect and admire him as a professor, researcher, clinician, and individual.

Whether it was my young ambition to be a ‘tornado chaser’ when I grew up, or my post-adolescent desire to journey across the globe to India just so I could ride an elephant, my parents never wavered in their support of me, even if it caused a handful of gray hairs and a few wrinkles. For this I am eternally grateful.

Although he would deny it, this accomplishment belongs equally to my husband and best friend, Chetan Prasad. It was with his steady confidence, encouragement, and patience, that I was able to successfully scale the small (and sometimes large) mountains needed to complete this project. I hope he can one day understand all that he means to me.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	vi
List of Figures	vii
Introduction	1
Actuarial Versus Clinical Judgment	2
Actuarial Methods for Identifying Treatment Failure	5
OQ-45 Feedback Studies	6
Clinical Versus Actuarial Methods of Predicting Treatment Failure	7
Current Study	8
Method	9
Participants	9
Measures	11
Youth Outcome Measure	11
Therapist Prediction Questionnaire	12
Clinical Significance and Reliable Change on the Y-OQ	13
Procedure	14
Results.....	16
Clinical Versus Actuarial Methods for Predicting Deterioration	19
Agreement between Y-OQ Progress Ratings and Therapist Progress Ratings	23
Can Therapist Progress Ratings Be Used to Predict Treatment Outcome?	25

Can Therapists Estimate Y-OQ Results? Do Therapist's Estimations of Y-OQ Results Differ from Their Own Predictions?	27
Discussion	31
Implications for Clinical Practice	35
Limitations	36
Conclusion	39
References	40
Appendix A: Therapist Prediction Questionnaire	45
Appendix B: Y-OQ Feedback Messages	46

List of Tables

1. Table Depicting Therapist Predictions of Treatment Failure and Y-OQ Final Treatment Status	17
2. Associations Between Therapist Predictions of Final Treatment Status and Y-OQ Final Treatment Status	18
3. Comparison of Therapist Predictions and Actuarial Predictions of Youth Client Deterioration	20
4a. Relationship Between Therapist Predictions of Deterioration and Actual Outcomes ...	20
4b. Relationship Between Y-OQ Actuarial Predictions of Deterioration and Actual Outcomes	21
5. Relationship Between Therapist Predictions and Actuarial Predictions of Client Deterioration	22
6. Therapist Progress Estimations Compared with Y-OQ Session Signal Ratings	24
7. Therapist Ratings of Client Worsening and Actual Outcomes	26
8. Therapist Predictions of Final Outcome Compared to Therapist Predictions of Y-OQ Final Outcome Status	28
9. Therapist Progress Estimations Compared to Therapist Estimations of Y-OQ Progress Ratings	28
10. Therapist Predictions of Y-OQ Final Treatment Status Compared to Actual Y-OQ Final Treatment Status	30
11. Therapist Estimate of Y-OQ Progress Ratings Compared to Actual Y-OQ Progress Ratings	30

List of Figures

1. Therapist Predictions of Final Treatment Status Versus Y-OQ Final Treatment
Status 19
2. Comparison of Therapist-Predicted, Actuarially-Predicted and Actual Client Outcome
Status 23

Predicting Youth Treatment Failure:
An Investigation of Clinical versus
Actuarial Judgment

Although psychotherapy has been shown to be effective for individuals presenting with a broad range of mental illness, ample evidence indicates that a significant number of clients experience *deterioration* or *treatment failure*—ending therapy in a worse state than when they started therapy (Lambert & Bergin, 1994; Smith, Glass, & Miller, 1980; Warren, Nelson, & Burlingame, 2009). Most studies indicate that the rate of treatment failure among adult clients receiving psychotherapy is between 5-10% (Hansen, Lambert, & Forman, 2002; Hatfield, McCullough, Frantz, & Krieger, 2010; Lambert & Ogles, 2004; Ogles, Lambert, & Sawyer, 1995).

Deterioration rates among children and adolescents have typically been found to be significantly higher than deterioration rates for adults. For example, Warren, Nelson and Burlingame (2009) conducted a study with data taken from youth receiving treatment in a community mental health setting. This study found youth deterioration rates to be around 21%. A subsequent study (Warren, Nelson, Mondragon, Baldwin & Burlingame, 2010) found a significant difference between the deterioration rates of youth receiving treatment in a community mental health clinic compared to those receiving treatment in a managed care setting. The deterioration rate for youth receiving treatment in the community mental health setting was found to be 24% and the deterioration rate was 14% for youth receiving treatment in a managed care setting. These findings highlight the need for increased attention to deterioration and treatment failure, particularly in youth treatment settings.

Past research examining the reasons for deterioration in treatment suggests that there are many potential explanations for treatment failure. According to Lilienfeld (2007), clients who experience deterioration or treatment failure may experience “heightened concern regarding extant symptoms” (p. 56). Additional research indicates that motivation for treatment is another factor that plays a role in treatment failure (Nelson, Warren, Gleave, & Burlingame, 2013). Furthermore, factors specific to youth deterioration may also include variables associated with parents such as parent self-efficacy, parent stress level, and parent psychopathology (Warren, Nelson, Burlingame, & Mondragon, 2012). While the reasons for youth deterioration in treatment seem to be many, the need remains for accurate identification of these cases. Such identification would provide clinicians with an opportunity to make necessary treatment adjustments that could potentially lead to improved outcomes and avoidance of treatment failure.

Actuarial Versus Clinical Judgment

The debate in the mental health field regarding clinical judgment versus actuarial judgment has been ongoing for several decades and was first introduced by Paul Meehl in 1954. In this context, actuarial judgment methods involve the use of formal, reliable, statistical methods employed to make decisions or predictions (Meehl, 1954). Clinical judgment methods, on the other hand, rely on human judgment that tends to be based on informal contemplation and consultation with others (Grove & Meehl, 1996). Meehl (1954) argued that a therapist’s ability to use their clinical judgment was inferior to statistical methods and therefore, psychologists and therapists should be using statistical methods to guide them in their clinical-decision making. More specifically, Meehl believed that therapist judgments about their clients were often inaccurate and more likely to be affected by subjective views. Following Meehl’s assertions, results of dozens of studies comparing these two methods confirmed the superiority of statistical

methods over clinical judgment for many tasks (Dawes, Faust, & Meehl, 2002; Garb, 1998; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Wiggins, 1973).

For example, Garb's research (1998, 2005) combined results from more than 1,000 controlled studies examining clinical judgment and decision making regarding personality assessment, psychopathology, case formulation, treatment planning, psychodiagnosis, and neuropsychological assessments, lending further support to the statistical (empiricist) tradition.

Garb (2005) observed:

A large body of literature or research indicates that it can be surprisingly difficult to learn from informal observations, both because clinician's cognitive processes are fallible and because accurate feedback on the validity of judgments is frequently not available in clinical practice. (p. 17)

Garb (2005) further noted that in his review of the studies assessing the ability of psychologists to make accurate and consistent diagnoses it was found that when psychologists used and adhered to diagnostic criteria or utilized semi-structured interviews they were more likely to have acceptable to good levels of interrater reliability. This appears to add further support to the evidence that when clinicians are guided in a more systematic way, their clinical judgments are more likely to be valid and accurate.

Similarly, in a landmark study, Grove and Meehl (1996) surveyed 136 studies that examined the comparison between actuarial and clinical methods of prediction. They found that 64 of the studies yielded results favoring the actuarial method, 64 indicated the two methods were equally accurate and only eight concluded that clinical judgment was superior to actuarial methods. Grove and Meehl (1996) attributed the results of the eight studies favoring clinician judgment to a combination of random sampling errors and to clinicians having the advantage of

being provided with more information than the actuarial method. Consequently, the bulk of the literature supports the use of actuarial methods by clinicians and other mental health professionals.

Although there is ample evidence to suggest that actuarial methods are superior to clinical judgments for many tasks, this evidence has done little to change routine clinical practice, as many therapists continue to ignore this large body of research (Dawes et al., 2002; Grove & Meehl, 1996; Marchese, 1992). Dawes et al. (2002) concluded that mental health professionals continue to rely more on their own clinical judgment than actuarial evidence. Furthermore, Grove and Meehl (1996) indicated that the years of or amount of experience a clinician has is not related to predictive accuracy. Therefore, when clinicians indicate that they are using their “years of clinical experience” to make a clinical judgment or prediction (a common explanation offered by mental health professionals to justify their clinical decision-making; Dawes, 1989) they are operating under a false sense of security.

Finally, mental health professionals tend to be overly confident in the clinical decisions they make even though evidence indicates that their decisions are often not reliable or accurate (Garb, 1998). A study was conducted in 2003 that surveyed how mental health professionals felt about the outcomes of their clients (Dew & Riemer, 2003). Results indicated that most mental health professionals believe they are well above average when it comes to producing clients with improved outcomes. For their study 143 therapists were asked to grade their performance as a clinician on a scale from A+ to F. It was found that 65% of the therapists gave themselves a grade of A or better and no therapist rated themselves below average. More recent research conducted by Walfish, McAlister, O'Donnell, & Lambert (2012) indicated that 129 mental health professionals rated their clinical skills to be above average compared to other clinicians

with similar credentials. Furthermore, the clinicians in this study, on average, viewed their clinical skills to be at the 80th percentile and none of the participating clinicians rated their skills as being below the 50th percentile.

Actuarial Methods for Identifying Treatment Failure

One of the most frequently researched actuarial methods used to keep track of client progress and change in psychotherapy is the Outcome Questionnaire 45 (OQ-45; Lambert et al., 1996). The OQ-45 is a 45-item measure that assesses three domains in adults receiving mental health treatment: symptom distress, interpersonal relations, and social role. It was specifically developed for tracking psychotherapy clients on a weekly basis and uses reliable statistical recovery curves to do so. Research using the OQ-45 (and accompanying OQ Analyst system) as a method of identifying those clients who are at-risk for treatment failure demonstrates that this method is accurate at identifying such clients (Lambert, et al., 2002; Spielmans, Masters, & Lambert, 2006). The results of Lambert et al. (2002) indicated that the OQ Analyst system for the OQ-45 had a 100% accuracy rate for identifying treatment failure and was able to identify 85% of patients on track for treatment deterioration by the end of the third treatment session.

In the youth treatment outcome literature, research examining the accuracy of actuarial methods at identifying cases of treatment failure has also been conducted using the Youth Outcome Questionnaire (Y-OQ; Bishop et al., 2005; Bybee, Lambert, & Eggett, 2007; Cannon, Warren, et al., 2009; Warren, et al., 2012). The results of this research indicate that such methods are accurate at identifying between 63% and 77% (sensitivity rates) of the youth therapy cases that eventually go on to deteriorate. This appears to suggest that the statistical formulas developed in these studies can be applied in clinical practice with youth and are comparable to those used with adult clients. Even though the accuracy rates of the warning system are lower in

youth, this finding is consistent with the observation that treatment failure rates differ among these populations and tend to be significantly higher for children and adolescents. Lilienfeld (2007) has emphasized the importance of monitoring client outcomes as a potential antidote to harmful treatments. These results, and the results found in the adult psychotherapy outcome literature, emphasize the likely value of using this type of feedback system in order to improve clinical decision-making and therapeutic outcomes for children and adolescents.

OQ-45 Feedback Studies

A substantial amount of research indicates that employing actuarial methods for identifying therapy cases that are at-risk for treatment failure can lead to improved psychotherapy outcomes by providing clinicians with timely feedback regarding their clients (Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Lambert et al., 2001; Lambert, et al., 2002; Shimokawa, Lambert, & Smart, 2010). The majority of this research has been done with adult psychotherapy clients. In 2001, Lambert et al. carried out the first feedback study using the OQ-45. Clients from a university counseling center were randomly assigned to either a control or experimental condition. As indicated by the OQ-45, both groups had equivalent intake scores. Of the clients in the sample, 12% were identified by the OQ-45 to be “signal alarm” cases, or likely to experience treatment failure. Therapists in the experimental group were given weekly feedback from the OQ-45, which informed them of those clients who were “signal alarm” cases. As such, the outcomes of the experimental group were significantly improved compared to the outcomes of those in the control group. It should also be noted that the “signal alarm” cases from the experimental group ended up remaining in therapy significantly longer than those participants in the control group. In 2002, this study was replicated adding further support to the use of actuarial feedback by clinicians in therapy (Lambert, et al., 2002).

Hawkins et al. (2004) also conducted a feedback study where in one group feedback from the OQ-45 was given to both therapists and their clients. The researchers were interested in examining the effect this might have on client outcomes. The two remaining groups consisted of a group where only the therapist was given feedback and a group where neither the therapist nor client was given feedback. The results of the study indicated that participants in either feedback group experienced significantly greater improvement at termination than those in the no feedback group. Thus, it appears that informing patients of their progress in therapy likely has a positive effect on outcome.

A recent meta-analysis conducted by Shimokawa et al. (2010) that analyzed the effects of progress feedback on patient outcome found more evidence supporting the use of feedback systems with clinicians. In this meta-analysis the data for over 6,000 patients was reanalyzed. The results indicated that the feedback received by the clinicians from the OQ-45 was effective in improving patient psychotherapy outcomes. These results lend further evidence to support the importance of being able to monitor progress and predict deterioration.

Clinical Versus Actuarial Methods for Predicting Treatment Failure

Studies comparing clinical versus actuarial judgment in the context of psychotherapy outcomes have been very few. In fact, only one such known study has been completed with adult therapy clients and it is the study after which the present study is modeled. In 2005, Hannan et al. investigated clinical versus actuarial prediction of patient deterioration by asking therapists to use their best clinical judgment and predict which of their adult patients would have a negative treatment outcome (deteriorate). Therapists at a large western university counseling center were recruited to participate in the study. The participating clinicians were asked to focus specifically on identifying those clients they believed would be treatment failures. Clients took

the OQ-45 before each therapy session. Their OQ-45 data were stored electronically and not shared with the clinician or client. At the end of each session therapists completed a prediction questionnaire that was used to assess their ability to estimate their client's progress as compared to the OQ-45 as well as predict their client's outcome at termination. Hannan et al. (2005) then compared the predictions made by the therapists with an actuarial method (i.e. the OQ-45 and OQ Analyst system).

The findings of the study yielded interesting and important results. Not only were the therapists unable to predict which of their clients would deteriorate or experience treatment failure, they very rarely predicted that their clients would experience a negative outcome (3 out of 550) even though they had been informed of what the typical treatment failure rate was in that specific counseling center (8%). In contrast, the actuarial method was found to over-predict deterioration, but accurately identified 20 of 26 (77%) clients who had deteriorated by treatment termination. The therapists, on the other hand, accurately identified only one of the clients who eventually deteriorated. Hannan et al. (2005) conclude that therapists need to use actuarial outcome methods if they want to become aware of clients at risk of treatment failure.

Current Study

Due to the higher frequency of treatment failure among youth receiving mental health services, the possible inability of therapists to recognize such cases during the treatment process, the tendency of therapists to be overly confident in their clinical judgments, and the lack of studies examining a clinician's ability to accurately predict treatment failure among youth clients, a strong need remains to examine the accuracy of clinical versus actuarial prediction of treatment failure in child and adolescent mental health treatment. Therefore, this study investigated how

well clinicians working with children and adolescents can accurately identify which of their clients will experience treatment failure.

The study hypotheses were as follows: 1) Actuarial predictions will be more accurate than clinician predictions of youth deterioration. It should be noted that this is the primary hypothesis of interest as the study aims to identify whether the clinicians or actuarial method will be more successful at identifying those youth clients who eventually end up as treatment failures. 2) Therapists will be inaccurate at identifying a youth patient that is worsening in treatment. In other words, they will not be able to identify when a patient is worse off than when they started treatment. 3) There will be significant disagreement between clinician prediction of client progress and actuarial prediction of client progress. 4) Clinician treatment progress estimates will not correspond to actuarial final treatment outcome status. 5) Clinician prediction of treatment progress and final outcome will disagree with clinician prediction of actuarial treatment progress and final outcome status. This hypothesis is expected to show a lack of agreement between the therapist's own clinical judgment and his/her assessment of the actuarial method. 6) Years of experience and theoretical orientation on the part of the clinicians in the study will not be related to prediction accuracy.

Method

Participants

Participants for this study were therapists of youth clients (ages 4-17) receiving mental health services in a community managed care setting and a community mental health clinic, and their parents/caregivers. Both the community managed care and community mental health clinic were located in the western United States, on the outskirts of a major western city. Furthermore, both clinics offered a combination of psychiatric medication management and various therapy services including individual, group, and family therapy for children and adolescents of all ages.

Additionally, the community mental health clinic provided case management services to children whose families were likely to benefit from such services. The community managed care clinic accepted most private insurance plans while the community mental health clinic provided services to individuals with Medicaid. A total of 221 youth clients and their parents were invited to participate; 196 consented yielding a response rate of 88%. Fifty-eight clients were not included in the final sample because they did not return for a third session ($n = 38$) or did not have Y-OQ session data allowing treatment progress and outcome to be estimated ($n = 1$). A total of 12 clients were not included due to being transferred to a non-participating clinician after intake and a total of seven clients were not included due to their Y-OQs being scored before the cessation of their third appointment. This allowed clinicians access to the Y-OQ feedback data and may have given clinicians an unfair advantage in making treatment predictions.

Youth clients who had a previous course of treatment before the dates of the study were considered to have a new course of treatment if it had been at least six months since their previous course of treatment. Child and adolescent clients included in the study presented for individual therapy only. No clients participating in family therapy or group therapy were included in the current sample, although it should be mentioned that individual therapy for youth clients often involves a therapist meeting with parents/guardians in order to work on behavior management skills in addition to working one-on-one with the youth.

The final youth client sample ranged in age from 4-17 ($M = 10.9$, $SD = 3.6$) and was 59% male, 41% female. The ethnicity for the youth client sample was as follows: 76% Caucasian, 15% Hispanic, 5% other, 2% African American, and 2% mixed race. Referents for treatment were typically the parents/guardians of the youth, but in some cases teachers, primary care physicians, and the Division of Child and Family Services were the main referents. Youth

clients and their parents/guardians gave informed consent for their participation in the present study and agreed to complete the Y-OQ prior to each session during their course of treatment.

A total of 29 clinicians were invited to participate; 21 consented yielding a response rate of 72%. Two of the participating clinicians were doctoral level psychologists, 10 were master's level clinicians (i.e. LCSWs and CMHCs), seven were doctoral level interns, and two were master's level therapist interns in a social work program. Twelve therapists were recruited from the community managed care setting and nine therapists were recruited from the community mental health clinic described above. The group of participating clinicians included 12 females (57%) and 9 males (43%), ranging in age from 26 to 58 ($M = 36.9$, $SD = 10.2$). Years of experience varied from 1 to 30 ($M = 10.6$, $SD = 9.3$). Therapists adhered to a variety of theoretical orientations with most subscribing to Cognitive Behavioral Therapy (CBT). The most common orientations were CBT (52.4%), eclectic/integrative (23.8%), systems (9.5%), other (9.5%), and interpersonal (4.8%). Participating clinicians gave their informed consent and agreed to predict, following each of the first three sessions of treatment, which of their youth clients they believed were likely to end treatment worse off than when they began (i.e., in the "deteriorated" category).

Measures

Youth Outcome Questionnaire: Following routine practice in this treatment setting, parents/caregivers of the youth participating in the study completed the Youth Outcome Questionnaire (Y-OQ; Burlingame et al., 1996; Burlingame, et al. 2001) before each therapy session. Completion of the Y-OQ takes between 5-10 minutes. The Y-OQ is a 64-item questionnaire completed by the parent/caregiver of youth between the ages of 4-17. It is used as a measure of treatment progress in mental health settings to track change in client functioning

over the course of treatment. The 64 items are measured on a 5 point Likert scale: 0 = never, 1 = rarely, 2 = sometimes, 3 = frequently, 4 = almost always; with the total score yielding a range of possible scores from -16 to 240 (eight of the items allow for a negative score). Higher scores indicate the endorsement of more severe distress and pathology. The Y-OQ assesses five domains in youth receiving mental health treatment: (1) interpersonal distress, (2) somatic distress, (3) interpersonal relationships, (4) social problems, and (5) behavioral dysfunction. A final subscale suggests critical items that aid in determining further client stabilization (e.g. paranoia, obsessive-compulsive behaviors, hallucinations, delusions, suicide, mania, and disordered eating).

The Y-OQ has adequate internal consistency ($r = .97$), criterion-related validity ($r = .84$) and 4-week test-retest reliability ($r = .81$) (Burlingame et al., 2001). It has also been demonstrated by previous research that algorithms built into the OQ Analyst software for the Y-OQ are successful at identifying and predicting youth clients who end therapy as treatment failures (Bishop et al., 2005; Warren et al. 2009; Cannon, et al., 2010; Warren, et al., 2012). More specifically, these algorithms make predictions based on the deviations of a client's Y-OQ score that is relative to the progress of other clients that began treatment at similar distress levels. In addition, session number (i.e. session 3, session 7, session 9) is another factor influencing the predictions created by the algorithms. For more information, interested readers are encouraged to refer to the article that discusses the creation of these algorithms in more detail (Bishop, et al., 2005).

Therapist Prediction Questionnaire: Clinicians made outcome predictions for their youth clients and attempted to identify their at-risk cases using a brief questionnaire that was developed and used in a previous study for the same purpose (Hannan et al., 2005). The questionnaire was

completed by clinicians following the first three sessions of treatment completed during the dates of the study. The length of the questionnaire was shortened when the therapist was seeing a youth client for the first time. A copy of the Therapist Prediction Questionnaire is provided in the appendix (see Appendix A).

Clinical Significance and Reliable Change on the Y-OQ

The youth participant's final treatment status as well as their progress in therapy was determined by the Y-OQ completed by their parent/caregiver. Based on procedures developed by Jacobson and Truax (1991), the Y-OQ provides cutoff scores that allow the clinician to assess patient change. These two cutoff scores are the Reliable Change Index (RCI) and clinically significant change. Reliable change on the Y-OQ is recognized as youth patients "who change in a positive or negative direction by at least 13 points" (Bishop et al., 2005). This amount is considered large enough to be confident that it exceeds measurement error. The cutoff score for indicating the point when a youth patient's score is more similar to individuals from the dysfunctional or clinical population than a functional or normal population is 46/45. Thus, when a youth patient's score falls below 46 their functioning is considered to more accurately resemble non-patients (Burlingame et al., 1996). When a youth client improves by 13 or more points on the Y-OQ and their Y-OQ score falls below 46, the client has met criteria for both reliable and clinically significant change (i.e. recovery; Burlingame et al., 1996).

For the present study, youth clients were given progress ratings and final treatment classifications based on algorithms created for the OQ Analyst system for the Y-OQ that have been used by Bishop et al. (2005) in previous research. In the study conducted by Bishop et al. (2005) where the base rate of deterioration was 7.3%, the algorithms were able to correctly identify approximately 77% of youth clients who ended treatment deteriorated, which is

comparable to the results reported by Lambert et al. (2002), who used a similar approach with adult clients. After a youth's parent/caregiver completes the Y-OQ, the algorithms produce a message that informs the clinician of the youth client's current status and progress in therapy. See Appendix B for a depiction of the four color-coded feedback messages given to clinicians during the course of treatment with a youth client. It should be noted that these feedback messages were withheld from the clinicians during the first three weeks the youth client received mental health services in order to put participating clinicians and the OQ Analyst system on similar ground.

Procedure

Youth participants and their parents were recruited to participate in the study at intake. The parents/caregivers of the youth participants completed the Y-OQ at intake and before each therapy session their youth attended during the course of treatment. The algorithms developed by the OQ Analyst system were applied following the second administration of the measure. The clinicians participating in the study completed the Therapist Prediction Questionnaire following the first three therapy sessions. The feedback the Y-OQ provides to clinicians was withheld for the first three weeks of treatment and Y-OQ questionnaires were stored and not revealed to the therapist, youth client, or their parent/caregiver until the three-week mark was met.

Clinicians participating in the study were informed that the purpose of the study was to examine their ability to evaluate their youth client's treatment progress and their prediction of their client's final outcome status at termination. They were also informed that another purpose of the study was to evaluate whether their projection of how their client would fare at the end of treatment is similar to or different from statistical methods of assessing therapy progress and predicting outcome cessation. Clinicians participating in the study were informed about typical

youth deterioration rates which were found by Warren et al. (2010) to be around 14% in a community managed care setting and as high as 24% in a community mental health setting. They were informed that the main interest in having them complete the Therapist Prediction Questionnaire was to examine whether clinicians can predict the small, but significant percentage of youth clients who worsen during treatment and therefore end treatment worse off than where they started, as indicated by the Y-OQ scores. Clinicians participating in the study were informed of all research procedures verbally and in writing so they could refer to them as needed.

Y-OQ scores for participating youth were tracked throughout the course of therapy until termination or six months, whichever came first. A common course of treatment in a managed care setting is typically between eight to 12 weeks, or about two to three months, due to insurance restrictions. A similar course of treatment is found in community mental health settings. Clinicians typically end therapy as a joint decision with their client, at the request of a client, or because the number of sessions they are allowed in the managed care setting has been expended. After the first three weeks of treatment, the therapists were no longer asked to complete the Therapist Prediction Questionnaires and access to the Y-OQ data was provided. The youth client's first and last Y-OQ scores were used to determine final treatment status and compared with therapist predictions to obtain the results.

Participating clinicians made between one and three predictions for the participating youth clients, resulting in a total of 406 predictions for the 196 clients. Of the 138 clients who had more than two predictions made, if a client was predicted to be a treatment failure at any prediction point, it was counted as a negative outcome prediction (i.e. deterioration). More specifically, if a therapist predicted at any point during the first three weeks of treatment that

their client would make ‘no change’ by the end of treatment, but later predicted that their client would ‘improve’ by the end of treatment, the most negative prediction was used in the reporting of predictive results (i.e. ‘no change’). This guideline was followed because the algorithms created for the OQ Analyst system apply the same standards for the Y-OQ. This was also done in order to put participating clinicians and the OQ Analyst system on equal ground. It should be noted further that this was the same method used by Hannan et al. (2005) for the same reasons described above.

The participating youth clients finished treatment with a wide range of exposure to treatment, ranging from 1 to 20 sessions ($M = 6.1$, $SD = 3.4$). In total, data were gathered for 138 youth clients who attended at least three sessions of treatment. For the results of the study deterioration or treatment failure was operationally defined as: reliable negative change in the clinical range (i.e. a 13-point or more increase on the Y-OQ and remaining status in the dysfunctional range). Recovered/improved was operationally defined as: clinically significant or reliable improvement (i.e. a 13-point or more decrease on the Y-OQ or both reliable improvement and moving from the dysfunctional range into the functional range). No change was operationally defined as any Y-OQ total score that did not demonstrate reliable improvement or reliable deterioration (i.e. -12 to +12 points). Hannan et al. employed the same operational definitions in their 2005 study.

Results

None of the 21 participating clinicians ever predicted that any of their youth clients would deteriorate in treatment. More specifically, zero treatment failure predictions were made by any of the participating clinicians at any point during the first three weeks of a client’s treatment (i.e. the time they were asked to predict treatment failure). Because no therapist

predictions of treatment failure were made, the planned McNemar chi square equivalent test (used by Hannan et al., 2005) to examine the relationship between therapist predicted deterioration and actual treatment outcome could not be conducted. Instead, a one-sample chi square test was conducted to examine the question of therapist prediction accuracy of treatment failure. Results for this analysis are provided in Table 1, using the sample of 138 clients.

Table 1.

Therapist Predictions of Treatment Failure and Y-OQ Final Treatment Status.

Therapist Predicted Final Treatment Status	Y-OQ Final Treatment Status		Total
	Not Deteriorated	Deterioration or Reliable Worsening	
Not Deteriorated	113	25	138
Deterioration or Reliable Worsening	0	0	0
Total	113	25	138

$$\chi^2 = 56.116, p < .001$$

Table 1 and the significant one-sample chi square test reflect that therapists never made a prediction that their youth clients would deteriorate in treatment. More specifically, 0 (0%) of the 138 clients (based on 406 total predictions) were predicted to worsen at the end of therapy by participating clinicians. The one-sample χ^2 test found a significant difference in prediction ($\chi^2 = 56.116, p < .001$). These results indicate that clinicians do not predict deterioration in agreement with its occurrences based on a parent-report measure (Y-OQ). Actuarial final outcome data indicated that 25 clients (18%) deteriorated by the end of therapy, a percentage consistent with estimates of treatment failure found by previous research (Warren, Nelson, & Burlingame, 2009 and Warren et al. 2010) indicating that 14% of youth clients in a managed care setting and up to 24% of youth clients in a community mental health setting experienced treatment failure.

Table 2 presents the results from Table 1 with the predictive task of final outcome status arranged into more refined categories (i.e. organizing youth clients who did not deteriorate into those who did not reliably change and those who showed reliable change). Due to the lack of treatment failure predictions, this table was included in order to shed more light on the predictive patterns of the participating clinicians. Similar to Hannan et al. (2005), these data were not used for hypothesis testing.

Table 2.

Association Between Therapist Predictions of Final Treatment Status and Y-OQ Final Treatment Status.

Therapist Predicted Final Treatment Status	Y-OQ Final Treatment Status			Total
	Deterioration or Reliable Worsening	No Change	Improve or Recovered	
Deterioration or Reliable Worsening	0	0	0	0
No Change	0	4	4	8
Improve or Recovered	25	40	65	130
Total	25	44	69	138

Due to zero treatment failure predictions the researcher was not able to use the matched pairs sign test that had been utilized by Hannan et al. (2005) to examine the relationship between outcome status predicted by the therapists and final treatment status as indicated by the Y-OQ. An examination of Table 2 indicates that the participating clinicians predicted that 130 youth clients (94.2%) would end treatment as reliably improved or recovered whereas 69 youth clients (50%) ended therapy improved or recovered according to the Y-OQ data. Therapists also predicted that 8 youth clients (6%) would experience no progress or change at the conclusion of

therapy while Y-OQ data scores showed 44 clients (32%) experienced no change at the end of treatment. Figure 1 displays a bar chart depicting therapist predicted end of treatment status versus actual end of treatment status as determined by the Y-OQ.

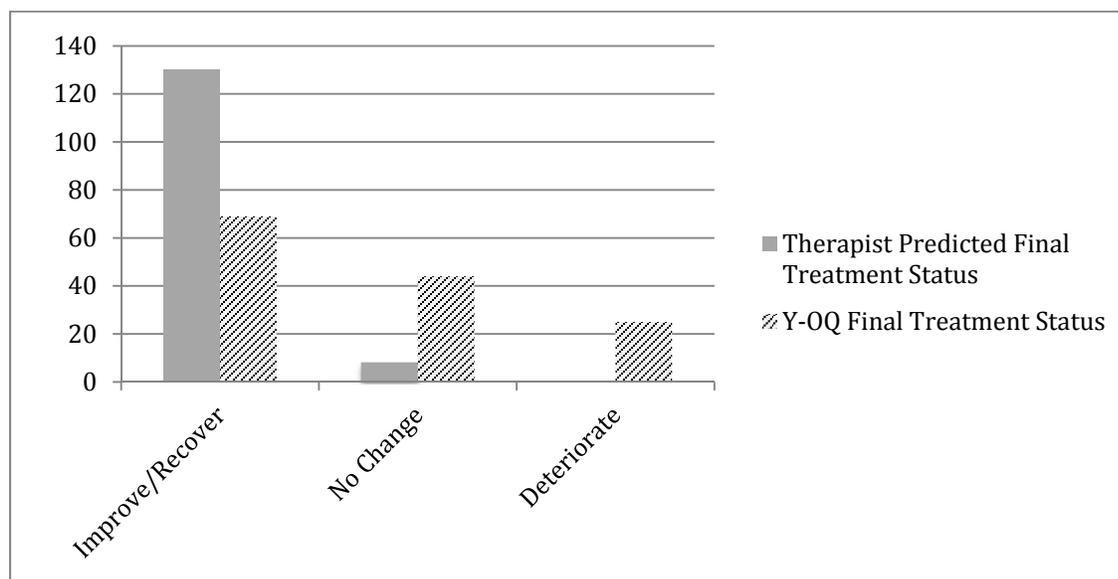


Figure 1. Therapist Predictions of Final Treatment Status Versus Y-OQ Final Treatment Status.

Clinical Versus Actuarial Methods for Predicting Deterioration

Table 3 presents the data for therapist-predicted deterioration versus Y-OQ-predicted deterioration for the 138 clients who had end-of-treatment predictions made. These data were used to examine the primary hypothesis of the study, which was that clinical judgment (i.e. therapist predictions) would be less accurate than an actuarial method (i.e. OQ Analyst system) in predicting youth deterioration or treatment failure. For the sample, 25 clients (18%) ended treatment as reliably and significantly worse than when they started treatment. As is evident in Table 3, the two methods yielded very different predictions.

Table 3.

Comparison of Therapist Predictions and Actuarial Predictions of Youth Client Deterioration

Therapist and Y-OQ Predicted Final Treatment Status		Y-OQ Final Treatment Status		
		Deterioration	No Deterioration	Total
Deterioration or Reliable Worsening	Therapist	0	0	0
	Y-OQ	21	21	42
No Deterioration	Therapist	25	105	130
	Y-OQ	4	92	96
Total	Therapist/Y-OQ	25	113	138

Tables 4a and 4b provide the same information from Table 3, this time in separate tables for each prediction method showing the hit rates of each method.

Table 4a.

Relationship Between Therapist Predictions of Deterioration and Actual Outcomes

Therapist Outcome Prediction	Y-OQ Final Treatment Status		
	Deterioration <i>n</i> (%)	No Deterioration <i>n</i> (%)	Total <i>n</i> (%)
Deterioration	0 (0%) hit	0 (0%) false positive	0 (0%)
No Deterioration	25 (100%) false negative	105 (93%) correct rejection	130 (94%)
Total	25 (18%)	113 (82%)	138 (100%)

Table 4b.

Relationship Between Y-OQ Actuarial Predictions of Deterioration and Actual Outcomes

Y-OQ Outcome Prediction	Y-OQ Final Treatment Status		
	Deterioration <i>n</i> (%)	No Deterioration <i>n</i> (%)	Total <i>n</i> (%)
Deterioration	21 (84%) hit	21 (19%) false positive	42 (30%)
No Deterioration	4 (16%) false negative	92 (81%) correct rejection	96 (70%)
Total	25 (18%)	113 (82%)	138 (100%)

Table 4a indicates that therapists predicted 0% of their youth clients would deteriorate and had a hit rate of 0% (0/25) for their predictions of deterioration, while correctly estimating no deterioration for 105 of the 138 clients. Furthermore, clinicians missed detection of 25/25 of the deteriorated cases—the main occurrence they were asked to identify. Clinicians failed to identify 100% of the client sample that deteriorated. More specifically, 0/25 (0%) of the youth patients that failed treatment were identified by their therapist as likely to deteriorate during the first three weeks of treatment. Table 4b displays the hit rate of the Y-OQ in predicting which clients would end treatment deteriorated. Results show that the OQ Analyst system identified 30% of clients as potential treatment failures (42/138), and attained a hit rate of 84% for correctly identifying end-of-treatment deterioration. However, the OQ Analyst system classified 21 clients as likely to deteriorate who did not, a false positive rate of 19% (21/113). It should be

noted that the OQ Analyst system missed identifying four clients who went on to deteriorate, a false negative rate of 16%.

Table 5 illustrates the results from Table 3 broken into the following outcome categories: deterioration, no change, and improve/recover. Similar to Table 2, Table 5 provides more insight into the predictive patterns of the participating clinicians given that zero predictions of treatment failure were made. Additionally, Table 5 sheds more light on the final outcomes for youth clients who were identified as potential treatment failures by the OQ Analyst algorithms. For the 42 youth cases classified as signal-alarms, 21 (50%) actually did get worse, 10 (24%) made no change in treatment, and 11 (26%) ended treatment reliably improved. Figure 2 presents the information in Table 5 in a bar chart to provide another reading of the data depicting therapist predicted final treatment status, Y-OQ predicted final treatment status, and actual Y-OQ final treatment status.

Table 5.

Relationship Between Therapist Predictions and Actuarial Predictions of Client Deterioration

Therapist and Y-OQ Predicted Final Treatment Status		Y-OQ Final Treatment Status			
		Deterioration	No Change	Improve/Recover	Total
Deterioration or Reliable Worsening	Therapist	0	0	0	0
	Y-OQ	21	10	11	42
No Change	Therapist	0	4	4	8
	Y-OQ	4	5	5	14
Improve or Recover	Therapist	25	40	65	130
	Y-OQ	0	28	54	82
Total		25	44	69	138

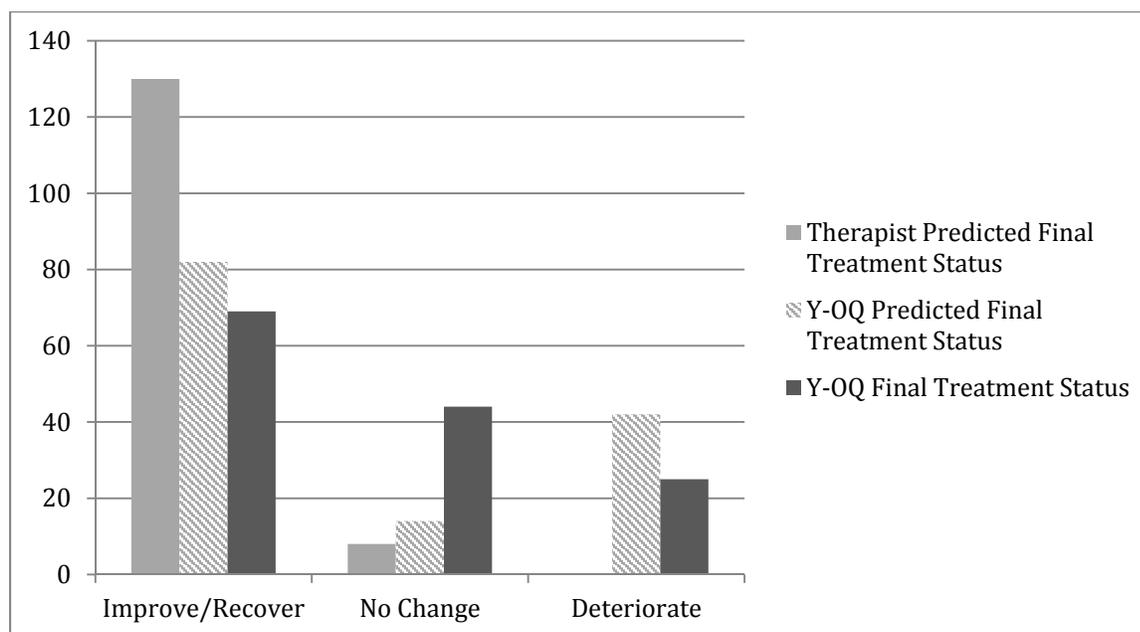


Figure 2. Comparison of Therapist-Predicted, Actuarially-Predicted, and Actual Client Outcome Status.

Agreement between Y-OQ Progress Ratings and Therapist Progress Ratings

Results examining agreement between therapist judgment of progress and Y-OQ progress classification can be viewed in Table 6. These data were based on question three of the Therapist Prediction questionnaire that participating clinicians completed at the end of the first three sessions of treatment with their youth clients. The question reads: “Considering this client’s initial session with you, rate this client’s progress as of today’s session. Base your rating on your clinical judgment and clinical experience alone, disregarding what Y-OQ scores might indicate. Today this client is (choose one rating)...recovered and ready for termination; improving as expected, but in need of continued treatment; making no progress or poor progress; getting worse.” The reader may remember that these judgments correspond with Y-OQ classifications of “White, Green, Yellow, or Red.” In order to put the Y-OQ and participating

clinicians on equal footing when it came to making predictions, only red signals were used in relationship with therapist endorsements of “getting worse,” although it should be noted that yellow signals are also considered to be potential treatment failures by the Y-OQ’s statistical algorithms. The results in Table 6 are presented in binomial form where Green, White, and Yellow signals are classified as “Not Worsening” and therapist ratings of “recovered,” “improving,” and “no/poor progress” are also classified as “Not Worsening”. This table presents data related to the hypothesis that actuarially based treatment progress scores would not be congruent with therapists’ progress estimates of treatment failure.

Table 6.

Therapist Progress Estimations Compared with Y-OQ Session Signal Ratings

Therapist Estimation of Client Progress	Y-OQ Signal Rating at Measured Session of Interest		
	Worsening	Not Worsening	Total
Worsening	1	3	4
Not Worsening	39	94	134
Total	40	97	138

$\kappa = .02$, not significant.

When compared against each other, treatment progress estimated by clinicians and actual Y-OQ progress signals show little agreement. More specifically, the kappa value showed extremely low agreement ($\kappa = .02$, not significant) between therapist progress estimations and actual Y-OQ progress ratings. An examination of Table 6 and the kappa value indicate that therapists rarely judge a youth client to be failing in therapy. Four youth clients (3%) were predicted to have worsened in treatment according to participating therapists at the measured sessions of interest, while the OQ Analyst system generated 40 (30%) red signal-alerts. This means that 39 clients (98%) were not identified by therapists as failing treatment, or being on

track to worsen in therapy, although they had received red signals at this point in treatment. Red Y-OQ signals are the strongest indicator of potential treatment failure a clinician can receive.

Can Therapist Progress Ratings Be Used to Predict Treatment Outcome?

The next question examined therapist progress ratings in order to determine if any youth clients identified as worsening in treatment according to clinicians did, in actuality, end up failing treatment. As was previously mentioned, therapists could indicate their belief that a youth client was worsening in treatment, but could still predict a positive or reliably improved outcome. Similar to Hannan et al. (2005), it was believed that clinicians might have been aware that their youth clients were worsening, but were hesitant to predict deterioration for whatever reason. The purpose of this analysis was to see exactly how aware of treatment worsening clinicians were for the 25 youth clients that did end up experiencing treatment failure. Table 7 summarizes data displaying the correctness of therapist-estimated progress and actual treatment outcome. These data correspond best to the hypothesis that therapist progress estimates would disagree with actuarial final outcome status. For a review and ease of comparison, the reader is encouraged to refer to Table 4b, considered earlier.

As was previously stated, this study found a deterioration rate of 18%. More specifically, out of 138 clients whose progress could be rated, 25 ended treatment worse than they began. The results in Table 7 demonstrate only 4 clients (3%) were predicted by therapists to be worsening or failing treatment. None of the four clients predicted to be worsening in therapy at that time were included in the 25 that did eventually deteriorate. This gives therapists a hypothetical hit rate of 0% (0/25), which is identical to their actual hit rate. Had therapists' judgments of treatment progress been used to predict a negative outcome they would have

Table 7.

Therapist Ratings of Client Worsening and Actual Outcomes

Therapist Progress Rating	Y-OQ Final Treatment Status		
	Deterioration <i>n</i> (%)	No Deterioration <i>n</i> (%)	Total <i>n</i> (%)
Deterioration	0 (0%) hit	4 (4%) false positive	4 (3%)
No Deterioration	25 (100%) false negative	109 (96%) correct rejection	134 (97%)
Total	25 (18%)	113 (82%)	138 (100%)

resulted in 4 (4%) false positive ratings of negative treatment outcome. These same judgments would have resulted in therapists identifying 100% of those youth clients who did deteriorate as not deteriorating. Once again, these data suggest that clinicians aren't likely to judge youth clients as worsening during the course of mental health treatment and as a result they are likely to miss those who are in fact worsening in treatment.

In contrast, Table 4b displays the Y-OQ's rating of 42 clients (30%) as potential treatment failures, and the hit rate of 84% (21/25) in correctly identifying those youth clients who eventually experienced treatment failure. The actuarial method identified 21 (19% false positive rate) clients as treatment failures who did not in actuality end treatment deteriorated. Furthermore, they missed identifying 4 (16% false negative rate) clients who did experience treatment failure.

Can Therapists Estimate Y-OQ results? Do Therapist's Estimations of Y-OQ Results Differ from Their Own Predictions?

Concerning the hypothesis that clinician prediction of treatment progress and final outcome will disagree with clinician prediction of Y-OQ treatment progress and final outcome category, clinicians were asked to provide predictions of the Y-OQ that were separate from their clinical predictions. This included estimates of progress (Y-OQ signal) and outcome based on the following questions from the Therapist Prediction Questionnaire (Appendix A): “Where do you guess today’s Y-OQ scores fell compared to this client’s initial session with you? White, today’s Y-OQ score will show improvement and be in the non-clinical range; Green, today’s Y-OQ score will show adequate change for this point in treatment; “Yellow, today’s Y-OQ score will show less than adequate change for this point; Red, today’s Y-OQ score will show worse than expected change for this point.” And, “Predict what this client’s Y-OQ scores will show at the end of treatment. This client’s Y-OQ scores will show (choose one prediction)...recovery, scores will drop reliably, move into, and stay in the non-clinical range; improvement, scores will drop reliably, but not move into the non-clinical range; no benefit from treatment, scores will not drop to a reliable degree; worsening, scores will increase to a reliable degree.”

The purpose for these questions was to investigate whether clinician’s judgments were significantly different from what they believed a statistical method would detect. In other words, it was of interest to examine whether clinicians ‘trust’ the method and if they believed their judgments would correspond to the actuarial judgments. Table 8 displays results that appear to indicate that clinician prediction of final outcome status and clinician judgment of what the Y-OQ’s final outcome status are in very strong agreement. More specifically, the result of the Spearman Rank Correlation Coefficient ($r_s = .904, p < .001$) indicated better than chance

agreement and was found to be statistically significant. Table 9 presents results that also seem to indicate that clinician judgment of progress and their judgment of the Y-OQ progress estimates also agreed at better than chance levels ($r_s = .822, p < .001$). Tables 8 and 9 display relationships that are not congruent to those hypothesized for this data. It was hypothesized that clinicians would be skeptical of the Y-OQ and not believe their judgments to be congruent with Y-OQ data.

Table 8.

Therapist Predictions of Final Outcome Compared to Therapist Predictions of Y-OQ Final

Outcome Status

Therapist Predicted Final Treatment Status	Therapist Prediction of Y-OQ Final Outcome Status			Total
	Deterioration or Reliable Worsening	No Change	Improve or Recover	
Deterioration or Reliable Worsening	0	0	0	0
No Change	0	8	0	8
Improve or Recover	0	0	130	130
Total	0	8	130	138

$r_s = .904, p < .001$

Table 9.

Therapist Progress Estimations Compared to Therapist Estimations of Y-OQ Progress Rating

Therapist Progress Estimation	Therapist Prediction of Y-OQ Progress Estimation			Total
	Deterioration or Reliable Worsening	No Change	Improve or Recover	
Deterioration or Reliable Worsening	3	1	0	4
No Change	5	35	4	44
Improve or Recover	1	9	80	90
Total	9	45	84	138

$r_s = .822, p < .001$

Table 8 showcases the interesting finding that clinicians predicted that the actuarial method would also not make any predictions of treatment failure. Of further interest is the fact that clinicians thought their predictions of no change and improve or recover would correspond perfectly with OQ Analyst system predictions for the Y-OQ. In other words, there was no disagreement between therapist predictions of final treatment status and therapist prediction of Y-OQ final treatment status. Table 9 appears to demonstrate that therapists are confident that their judgments regarding treatment progress will not be different from the OQ Analyst system's prediction of treatment progress. In other words, their agreement of treatment progress was found to be significantly better than chance. Overall, therapists estimated that the Y-OQ would show nine clients worsening while they only estimated four clients to actually be worsening. Their estimates of 'no change' were nearly identical and therapists estimated the Y-OQ would show 84 youth clients as improving/recovering while they estimated that 90 youth clients would fall in this category.

Finally, this study examined the level of congruence between therapist prediction of Y-OQ final treatment status and actual Y-OQ final treatment status. This analysis examined the hypothesis that clinician prediction of treatment progress and final outcome would disagree with actuarial treatment progress and final outcome status. Table 10 demonstrates that even though clinicians were confident that their predictions of the Y-OQ's final treatment status would be similar to their predictions of final treatment status, agreement between their predictions of the Y-OQ and actual Y-OQ classifications demonstrated little to no agreement ($r_s = -.081$, not significant).

Table 10.

*Therapist Predictions of Y-OQ Final Treatment Status Compared to Actual Y-OQ Final**Treatment Status*

Therapist Prediction of Y-OQ Final Treatment Status	Y-OQ Final Treatment Status			Total
	Deterioration or Reliable Worsening	No Change	Improve or Recover	
Deterioration or Reliable Worsening	0	0	0	0
No Change	0	0	8	8
Improve or Recover	25	44	61	130
Total	25	44	69	138

$r_s = -.081$, not significant.

Table 11.

Therapist Estimate of Y-OQ Progress Ratings Compared to Actual Y-OQ Progress Ratings

Therapist Prediction of Y-OQ Progress Rating	Y-OQ Progress Rating			Total
	Deterioration or Reliable Worsening	No Change	Improve or Recover	
Deterioration or Reliable Worsening	4	1	4	9
No Change	4	9	32	45
Improve or Recover	24	8	52	84
Total	32	18	88	138

$r_s = .019$, not significant.

The comparisons of therapist prediction of Y-OQ progress ratings and actual Y-OQ progress ratings are shown in Table 11. Overall, little to no or low agreement was observed for therapists' estimates of Y-OQ progress ratings and actual Y-OQ progress ratings ($r_s = .019$, not significant). An examination of the results of Tables 10 and 11 seem to indicate that

participating clinicians were slightly more accurate when it came to estimating Y-OQ progress ratings as opposed to Y-OQ final outcome status. However, their prediction accuracy overall is significantly limited due to a lack of any treatment failure predictions and very few predictions of 'no change'. Results of these analyses appear to suggest that clinicians strongly believe that their judgments and predictions will be similar to Y-OQ classifications. In actuality, therapist judgments and predictions of Y-OQ classifications were incongruent with actual Y-OQ classifications.

The final study question was to examine the difference in prediction accuracy between experienced and inexperienced clinicians as well as the difference in prediction accuracy based on theoretical orientation when it came to successfully predicting treatment failure. Due to the nonexistence of negative treatment predictions, such analyses could not be conducted.

Discussion

The purpose of this study was to examine clinicians' ability to predict treatment failure (negative change) among their youth clients based on their clinical judgment, and compare that to the accuracy of an actuarial method's (i.e. OQ Analyst system and Y-OQ) ability to do the same. As has been previously mentioned, this study was a replication of a previous study completed in a university counseling center with adult patients (Hannan et al., 2005). Overall, the results of the current study and the results of the prior study completed by Hannan et al. (2005) appear to be very similar. The results of the current study will now be discussed further.

A surprising finding of this study was the fact that none of the participating clinicians made any predictions of treatment failure for their participating youth clients despite 18% of the youth clients eventually experiencing treatment failure. Therapists therefore failed to identify clients who did deteriorate in treatment and over-predicted recovery/improvement. This

occurred despite clinicians knowing the rate of deterioration for the setting in which they worked (i.e. managed care setting or community mental health setting). Prior to the beginning of the study, clinicians were provided with this information both verbally and in writing by the primary researcher. In addition, participating clinicians were also familiar with the actuarial measure that was used. Both the community managed care setting and community mental health setting where data were collected had been using this actuarial method for the past several years. This result suggests that clinicians are unable to accurately predict youth client deterioration, a finding that is troubling considering youth treatment failure rates have been found to be as high as 24% in some settings (Warren et al., 2010).

Furthermore, this main finding also seems to indicate that clinicians are hesitant to make treatment failure predictions and have the tendency to believe almost all of their clients will reliably improve. This seems to be a finding consistent with Walfish et al. (2012) whose results indicated that clinicians tend to overestimate client improvement. While it is likely helpful for clinicians working with mentally ill children to be optimistic about their treatment outcomes, it appears that clinicians would benefit from having slightly more realistic psychotherapy outcome expectations.

Due to there being zero predictions of treatment failure, some of the planned analyses could not be conducted, including the questions of whether years of clinical experience and theoretical orientation would be related to prediction accuracy. An important finding of the study is that participating clinicians did not expect their youth clients to experience negative treatment outcomes. Even though there were only three negative treatment predictions made in the Hannan et al. (2005) study, it was expected that there would be more treatment failure predictions in the current study due primarily to the fact that youth clients tend to experience

significantly higher rates of deterioration than adult clients participating in psychotherapy (Warren, et al., 2009; Warren et al., 2010).

Examining the main hypothesis of the current study further, it is apparent that the actuarial method was significantly more accurate at predicting treatment failure. In contrast to therapists' predictions, the actuarial method obtained a hit rate of 84% for correctly identifying 21/25 youth clients who experienced treatment deterioration.

This finding clearly seems to support the main hypothesis of the study, that actuarial prediction of treatment failure would be more accurate than clinician prediction of treatment failure. Furthermore, this finding provides more evidence for and highlights the need for clinicians to use and rely on actuarial outcome methods when conducting psychotherapy with youth clients. As previous research with adults has shown, and future research with youth is hoping to show, use of actuarial outcome measures to monitor progress in psychotherapy tends to lead to improved outcomes (Hawkins et al., 2004; Lambert et al., 2001; Lambert et al., 2002; Lambert et al., 2003; & Shimokawa et al., 2010).

Overall, the OQ Analyst system over-predicted deterioration and had a false positive rate of 19% (21/138). However, 10 of the 21 youth clients the OQ Analyst system misidentified ended up making no significant change in treatment suggesting increased attention to these cases was warranted and would have likely been helpful at improving outcome. As indicated by past research, the OQ Analyst system of the Y-OQ only alerts therapists to the need for reconsidering the course of treatment and does not appear to be related to any overly damaging treatment consequences (Hannan et al., 2005). As a result it is believed that the signal-alarm produced by the OQ Analyst can be viewed as a way to enhance a clinician's ability to make treatment decisions, not hinder it. According to results of patient-focused research, a higher false positive

rate is preferred over a high false negative rate. Consequences associated with a higher false positive rate are believed to be less serious than the consequences associated with missing or not identifying a client that is experiencing treatment failure (Hannan et al., 2005; Lambert, et al., 2003; Nelson et al., 2013). Therefore, the fact that the OQ Analyst system produced a false-positive rate of 19% in this study should not be viewed as too troubling.

The current examination found that a small number of youth clients are singled out by clinicians as getting worse during the course of treatment; however, results indicate that these clients do not correspond very well with Y-OQ progress ratings. In other words, clients identified as worsening by clinicians were not actually worsening according to the Y-OQ algorithms. Clinicians in this study identified 3% (4/138) of the youth clients as worsening in treatment, however, only one out of four (25%) of those predictions was congruent with Y-OQ progress ratings. Clinicians missed identifying 98% of the youth clients who had in fact received a signal-alarm code at that point in treatment. Even though clinicians only identified 4 clients as worsening at the measured session of interest, it appears that they would have improved their prediction accuracy slightly by interpreting their judgment of worsening as a sign of treatment failure. However, even if clinicians had interpreted these worsening judgments as a sign of potential treatment failure, it would not have improved their prediction accuracy to a significant degree. Once again, these findings highlight the tendency clinicians in this study had to not predict client worsening in therapy as well as their inaccuracy when worsening predictions were made and compared to the Y-OQ algorithms.

This study also examined the hypothesis that clinician prediction of treatment progress and outcome would not be congruent with clinician prediction of Y-OQ treatment progress and final outcome status. In other words, clinicians were also asked to provide predictions of the Y-

OQ that were separate from their clinical predictions and it was believed that these predictions would not be similar. This hypothesis was included as a way to examine or measure the level of ‘trust’ clinicians had when it came to the actuarial method as previous research has indicated that clinicians tend to distrust such measures (Hannan et al., 2005). Overall, the results of this data do not seem to support the hypothesis described above. More specifically, clinicians had the tendency to believe that their predictions would correspond very well with the Y-OQ’s ratings of progress and final treatment status. Additional analyses were conducted to examine the level of agreement between clinicians’ judgments of Y-OQ progress and final outcome compared to actual progress ratings and outcome. The results of these analyses indicated low levels of agreement and were not statistically significant. These results suggest that even though therapists believe their judgments will be similar to the progress ratings and final treatment status of the Y-OQ they tend to be more dissimilar than similar. Once again, these findings highlight the inaccuracies of therapists’ predictions of treatment progress and outcome.

Implications for Clinical Practice

The results of this study add further evidence to support the use of actuarial methods for identifying cases at risk for deterioration in psychotherapy. More specifically, the results of this study indicate that use of actuarial outcome methods in youth psychotherapy is needed to assist clinicians when it comes to identifying youth clients who are on track to experience treatment failure. This is primarily due to the findings of this study indicating that clinicians are not able to successfully identify such clients using their clinical judgment alone. As the OQ Analyst system of the Y-OQ was found to be significantly more accurate at identifying treatment failures, use of such measures by clinicians will likely improve their clinical decision making and give them an opportunity to make treatment adjustments for those youth clients identified as potential

treatment failures. In other words, use of actuarial methods will likely enhance a therapist's clinical abilities to a reliable degree. As previous research has found with adults, this is likely to help keep such youth clients in treatment longer and improve their chances of experiencing an improved outcome (Hawkins et al., 2004; Lambert et al., 2001; Lambert, et al., 2002; Shimokawa et al., 2010). This would also logically seem to lead to lowered rates of treatment failure in youth psychotherapy, where rates already tend to be more than double that of adult rates of treatment deterioration (Warren et al., 2009; Warren et al., 2010).

Limitations

Similar to other research studies carried out in 'real-world' settings, this study had a number of limitations. One limitation of the current study was the use of only one parent-report measure of treatment progress and final treatment outcome. This provided only one reflection of the youth client's experience in therapy. While the Y-OQ is considered to be a good measure of treatment progress and outcome, it is not infallible. It would have been more ideal to use multiple measures, as this is likely to provide more comprehensive indications of treatment progress and outcome. It is possible that different results would have been achieved had other measures of treatment progress and outcome been used in the current study. Therefore, future research using multiple measures is recommended, although it should be noted that using multiple outcome measures in 'real-world' settings is not likely to be feasible due to the demands on time it would require. In addition, this study compared clinicians' judgments with an actuarial method that used only one measure to alert to possible treatment failure.

A second limitation involves the use of the 'homemade' Therapist Prediction Questionnaire. Due to it being specifically created for one study prior to the current study, no validity or reliability estimates exist (Hannan et al., 2005). This confines any conclusions that

can be drawn from the use of this measure. It would have been more statistically sound to establish the psychometric properties of this questionnaire prior to its use in the current study.

Another limitation of the current study is the fact that some youth clients were still receiving treatment when the primary researcher stopped collecting their Y-OQ data. This is because they had been in treatment for six months, the previously identified discontinue criterion. It is possible that different results would have been achieved had such clients been followed until treatment termination. However, this limitation can be described as a weak limitation due primarily to previous research indicating that the most change in psychotherapy occurs during the first few sessions (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). In other words, it is likely that this study captured most of the change experienced by the youth participants.

This study did not attempt to collect any follow-up measures from the youth clients as a way to examine if the final outcome statuses achieved by the Y-OQ remained unchanged after treatment had ended. The results of the current study may have changed had the primary researcher used follow-up measures. More specifically, the use of such follow-up measures may have found more evidence to support the accuracy of clinicians' predictions of treatment failure. As a result, future research could attempt to gather follow-up data as a way to further examine long-term treatment outcomes.

There were a few participating clinicians who expressed concern over the Y-OQ feedback data being withheld from them during the course of data collection. They indicated their belief that the cost of withholding the Y-OQ might be greater than any significant research findings and explained that they had come to rely on the Y-OQ as a way to gain critical information regarding suicidal ideation. As a result of this concern, therapists were encouraged

to ask their clients about any critical items directly and the length of time the feedback data were withheld was kept as brief as possible.

Another limitation of the current study is the exclusive use of red alarm signals in comparison to therapist estimates of client worsening in treatment. Future studies could also use yellow alarm signals as they are considered to be softer indications of risk for treatment failure. Utilizing yellow alarm signals in the future would likely require small modifications to be made to the Therapist Prediction questionnaire. Due to therapists only being told to focus on the client worsening in therapy, including yellow alarm signals would have been unfair to the clinicians and only red signals were included as a result. This helped create more equal footing between the therapists and actuarial method.

The participating clinicians made only very early treatment predictions. More specifically, predictions were only made at the end of the first three sessions of treatment. It is possible that the clinicians' accuracy of treatment failure would have improved had they been given a chance to make such predictions at later treatment sessions. This also needs to be considered a limitation of the current study, however, this appears to also be a weak limitation. Previous research indicates that the most change in psychotherapy occurs during the first few sessions (Baldwin et al., 2009). Therefore, the argument could be made that having the participating clinicians make early treatment predictions was justified.

In youth psychotherapy it is common for the Y-OQ respondent to be inconsistent from week to week. More specifically, a youth client's mother might bring him/her to treatment one week and complete the Y-OQ while the youth client's father, or stepmother, or stepfather, etc., might complete the Y-OQ the following week. It is likely that this trend has an effect on the Y-OQ total scores from week to week as well as the signals created by the algorithms. This was a

common occurrence in the current study and it is likely that this had an effect on both the Y-OQ progress ratings and Y-OQ final outcome status. However, the argument can be made that this is normal in ‘real-world’ settings and strengthens the overall external validity of this study. It is therefore considered to be a relatively ‘weak’ limitation.

Finally, evaluation apprehension can also be considered another possible, although unlikely limitation of the current study. Participating clinicians may have experienced anxiety at the thought of being compared to a statistical treatment tracking measure with which they were already familiar. It seems unlikely, but this may have had an impact on how the therapists answered the questions on the Therapist Prediction questionnaire. Future research could be done in a setting where there is little familiarity with the Y-OQ in order to avoid this limitation, although this also seems like a ‘weak’ limitation.

Conclusion

This was the first study to examine clinical versus actuarial judgment in the context of youth psychotherapy in routine clinical practice and is a replication of a previous study completed with adult psychotherapy clients (Hannan et al., 2005). The results of this study are very consistent with past research on clinical versus actuarial judgment in the context of adult psychotherapy. Similar to the results found by Hannan et al. (2005), it appears that clinicians remain reluctant or unable to identify youth clients worsening in treatment, despite knowing the rates of treatment failure in the settings in which they work. Clinicians should be encouraged to utilize any signal-alarms they receive as a sign of a potential negative outcome and take steps to adjust treatment so deterioration can be avoided. With the results of the current study, it is hoped that more clinicians will be persuaded to use reliable outcome measures as a way to improve the quality of care they provide to their patients.

References

- Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J., & Nielsen, S. (2009). Rates of change in naturalistic psychotherapy: Contracting dose effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77*, 203-211. doi: 10.1037/a0015235
- Bishop, M. J., Bybee, T. S., Lambert, M. J., Burlingame, G. M., Wells, M. G., & Poppleton, L. E. (2005). Accuracy of a rationally derived method for identifying treatment failure in children and adolescents. *Journal of Child and Family Studies, 14*(2), 207-222. doi: 10.1007/s10826-005-5049-1
- Burlingame, G. M., Wells, M. G., Hoag, M., Hope, C., Nebeker, S., Konkel, K., Reisinger, C. (1996). *Administration and scoring manual for the Y-OQ.2*. Setauket, NJ: American Professional Credentialing Services.
- Burlingame, G. M., Mosier, J. I., Wells, M. G., Atkin, Q. G., Lambert, M. J., Whoolery, M., ... Latkowski, M. (2001). Tracking the influence of mental health treatment: The development of the Youth Outcome Questionnaire. *Clinical Psychology and Psychotherapy, 8*, 361-379. doi: 10.1002/cpp.315
- Bybee, T. S., Lambert, M. J., & Eggett, D. (2007). Curves of expected recovery and their predictive validity for identifying treatment failure. *Tijdschrift voor Psychotherapie (The Dutch Journal of Psychotherapy), 33*(6), 419-434.
- Cannon, J. A. N., Warren, J. S., Nelson, P. L., & Burlingame, G. M. (2010). Change trajectories for the Youth Outcome Questionnaire self-report: Identifying youth at risk for treatment failure. *Journal of Clinical Child and Adolescent Psychology, 39*(3), 289-301. doi: 10.1080/15374411003691727

- Dawes, R. M., Faust, D., & Meehl, P. E. (2002). Clinical versus actuarial judgment. In Gilovich, T. Editor, & Griffin, D. Editor (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. (pp. 857). New York, NY, US: Cambridge University Press.
- Dew, S., & Riemer, M. (2003). Why inaccurate self-evaluation of performance justifies feedback interventions. In L. B. Chairperson (Chair), *Improving outcomes through feedback intervention*. Symposium conducted at the 16th Annual Research Conference, A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, University of South Florida, The Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health.
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington DC: American Psychological Association.
- Garb, H. N., (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67-89. doi: 10.1146/annurev.clinpsy.1.102803.143810
- 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 and 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.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W. & Shimokawa, K. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*, 155-163. doi: 10.1002/jclp.20108

- Hansen, N. B., Lambert, M. J., & Forman, E. W. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329-343.
- Hatfield, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy, 17*, 25-32. doi: 10.1002/cpp.656
- Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K., & Tuttle, K. (2004). The therapeutic effects of providing client progress information to patients and therapists. *Psychotherapy Research, 10*, 308-327. doi: 10.1093/ptr/kph027
- 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*(1), 12-19.
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53-70. doi: 10.1111/j.1745-6916.2007.00029.x
- Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A.E. Bergin and S.L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143-189) New York: John Wiley and Sons.
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G., ... Reisinger, C. W. (1996). *Administration and scoring manual for the outcome questionnaire (OQ 45.2)*, Wilmington, DE: American Professional Credentialing Services.

- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeesch, 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*(1), 49-68.
- 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. doi: 10.1002/cpp.333
- Lambert, M. J., Whipple, J. J., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., and Smart, D. W. (2003) Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288-301. doi: 10.1093/clipsy/bpg025
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press. Reprinted with new Preface, 1996.
- Nelson, P. L, Warren, J. S., Gleave, R. L, and Burlingame, G. M. (2013). Youth psychotherapy change trajectories and early warning system accuracy in a managed care setting. *Journal of Clinical Psychology, 69*(9), 880-895. doi: 10.1002/jclp.21963
- Shimokawa, K., Lambert, M. J., and Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298-311. doi: 10.1037/a0019247
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: John Hopkins University Press.

- Spielmanns, 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(3), 202-214.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2010). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, 110(2), 1-6. doi: 10.2466/02.07.17.PRO.110.2
- Warren, J. S., Nelson, P. L., & Burlingame, G. M. (2009). Identifying youth at risk for treatment failure in outpatient community mental health services. *Journal of Child and Family Studies*, 18, 690-701. doi: 10.1007/s10826-009-9275-9
- Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*, 78(2), 144-155. doi: 10.1037/a0018544
- Warren, J. S., Nelson, P. L., Burlingame, G. M., & Mondragon, S. A. (2012). Predicting patient deterioration in youth mental health services: Community mental health vs. managed care settings. *Journal of Clinical Psychology*, 68(1), 24-40. doi: 10.1002/jclp.20831
- 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, 1, 59-68. doi: 10.1037/0022-0167.50.1.59
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Appendix A

Therapist Prediction Questionnaire

1. In your clinical judgment alone, ignoring what Y-OQ scores might indicate, predict this client's end-of-treatment outcome. This client will (choose one prediction)...
 - a. Recover
 - b. Improve, but not recover
 - c. Make no progress in treatment
 - d. Get (significantly) worse.

2. Predict what this client's Y-OQ scores will show at the end of treatment. This client's Y-OQ scores will show (choose one prediction)...
 - a. Recovery-scores will drop reliably, move into, and stay in the non-clinical range
 - b. Improvement-scores will drop reliably, but not move into the non-clinical range
 - c. No benefit from treatment-scores will not change to a reliable degree
 - d. Worsening-scores will increase to a reliable degree.

3. Considering this client's initial session with you, rate this client's progress as of today's session. Base your rating on your clinical judgment and clinical experience alone, *disregarding what Y-OQ scores might indicate*. Today this client is (choose one rating)...
 - a. Recovered and ready for termination
 - b. Improving as expected, but in need of continued treatment
 - c. Making no progress or poor progress
 - d. Getting worse

4. Where do you guess today's Y-OQ scores fell compared to this client's initial session with you?
 - a. White, today's Y-OQ score will show improvement and be in the non-clinical range
 - b. Green, today's Y-OQ score will show adequate change for this point in treatment
 - c. Yellow, today's Y-OQ score will show less than adequate change for this point
 - d. Red, today's Y-OQ score will show worse than expected change for this point

Appendix B

Y-OQ Feedback Messages

White Feedback- “The youth client is functioning in the normal range. Consider termination.”

Green Feedback- “The rate of change the youth client is making is in the adequate range. No change in the treatment plan is recommended.”

Yellow Feedback- “The rate of change the youth 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 youth client may end up with no significant benefit from therapy.”

Red Feedback- “The youth client is not making the expected level of progress. Chances are 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 on a new course of action.”