

Administration and Scoring Manual for the

Y-OQ™ 2.01 (Youth Outcome Measures)

**Gary M. Burlingame, M. Gawain Wells, Jonathan C. Cox, Michael J. Lambert,
Brigham Young University**

**Mark Latkowski
Primary Children's Medical Center**

**David Justice
Valley Mental Health**

OQ Measures L.L.C.

(Toll Free 1.888.647.2673)

Revised July 2005

ADMINISTRATION AND SCORING MANUAL

FOR THE

Y-OQ™ 2.01
(YOUTH OUTCOME QUESTIONNAIRE)

JULY 2005

OQ MEASURES L.L.C.
(Toll Free) 1-888-MH SCORE (1-888-647-2673)
E-MAIL: info@oqfamily.com
WEB: www.oqmeasures.com

ACKNOWLEDGEMENTS

We wish to recognize individuals and organizations that have contributed and acted as partners in the development of the Y-OQ™. Funding for this project came from the College of Family, Home, and Social Sciences, Brigham Young University. Without the kind support of the University, a project of this size could not have been undertaken. Thanks also to the Ben B. and Iris M. Margilis Foundation for their support of studies testing the Y-OQ™. The members of Primary Children's Medical Center Outcome Research team have demonstrated unabated commitment to a variety of projects related to the Y-OQ™.

We would also like to thank the many students and colleagues who helped with data collection and analysis; without their painstaking efforts the Y-OQ™ would only be an idea. Thanks also to the many people who have taken the Y-OQ™, for their time and effort, especially those patients and non-patients who filled it out on a weekly basis.

Largely because of the unselfish support of these organizations and people, the Y-OQ™ is now ready for use with the public. It is a pleasure to offer it at low cost to the professional community for unlimited use. We ask that Y-OQ™ users carefully follow our licensing requirements. We would appreciate your support in encouraging your colleagues to properly license and use this tool. With this type of support we will be able to continue to offer the Y-OQ™ as one of the most competitively priced mental health outcome tools anywhere.

Gary M. Burlingame, Ph.D.

For the authors

TABLE OF CONTENTS

	PAGE
Introduction	1
Intersection of Research, Clinical Work, and Managed Health Care	1
Research	1
Clinicians	2
Health Care Corporations	3
Test Development	3
Initial Literature Review	4
Focus Groups	4
Hospital Records	4
Description of Subscales	5
Intrapersonal Distress (ID)	5
Somatic (S)	5
Interpersonal Relations (IR)	5
Critical Items (CI)	5
Social Problems (SP)	5
Behavioral Dysfunction (BD)	6
Y-OQ™ Total	6
Administration	6
Time	6
Scoring	6
Computer Scoring	7
Missing Data	7
Normative Data	7
Cutoff Score	10
Reliable Change Index	12
Gender Differences	13
Age Differences	17
Other Normative Groups	21
Psychometric Properties	22
Reliability	22
Internal Consistency	22
Test-retest Reliability	22
Frequency of Administration	24
Validity	24
Criterion-Related Validity	24
Construct Validity	27

	PAGE
Sensitivity to Change	28
Sensitivity to Psychopathology	32
Sensitivity and Specificity	32
Basic Interpretive Strategy	34
Use of the Instrument for Treatment Planning	35
General Treatment Planning Issues	35
Applications to Treatment Planning Issues	36
Identification of Primary and Secondary Problems	36
Appropriate Level of Care	36
Potential Use or Limits for Treatment Planning in a Managed Care Setting	36
Use of the Instrument for Treatment Monitoring	36
Purpose of Treatment Monitoring	36
How to Use the Instrument for Treatment Monitoring	38
Use of the Y-OQ™ for Treatment Outcomes Assessment	38
Use of Findings from the Y-OQ™ with Other Evaluation Data	39
Provision of Feedback Regarding Outcomes Assessment Findings	39
Limitations in the Use of the Y-OQ™ for Outcomes Assessment	40
Additional Versions and Related Products of the Y-OQ™	40
Self-Report Version	40
Shortened Version	40
Y-OQ™ 12	40
Software	41
Conclusion	41
References	43
Appendix A: Y-OQ™ Tracking Sheet	49
Appendix B: Outpatient and Inpatient Age Differences	51
Appendix C: Application for a Y-OQ™ License	54

LIST OF TABLES

	PAGE
Table 1A: Normative Group Data for the Y-OQ™ Total Score	7
Table 1B: Normative Group Raw Score Data for the Y-OQ™ Domains	8
Table 2A: Validation Group Data for the Y-OQ™ Total Score.....	8
Table 2B: Validation Group Raw Score Data for the Y-OQ™ Domains.....	9
Table 3A: Total and Subscale Y-OQ™ Raw Scores for Outpatients from Three Health Maintenance Organizations	10
Table 3B: Total and Subscale Y-OQ™ Raw Scores for Outpatients from Two Health Maintenance Organizations	13
Table 4A: Comparison of Gender Scores on the Y-OQ™ - Total Score	13
Table 4B: Comparison of Validation Gender Scores on the Y-OQ™ - Total Score	14
Table 5A: Comparison of Gender Scores on the Y-OQ™ - Domain Scores	14
Table 5B: Comparison of Validation Gender Scores on the Y-OQ™ - Domain Scores.....	15
Table 5C: Comparison of Respondent Scores on the Y-OQ™ - Domain Scores	16
Table 6A: Y-OQ™ Total Score by Age in Community Normal Sample.....	17
Table 6B: Y-OQ™ Subscale Scores by Age in a Sample of Community Normals	17
Table 6C: Y-OQ™ Subscale Scores by Age and Clinical Samples.....	19
Table 6D: Means According to Race.....	20
Table 7: Internal Consistency Values for the Y-OQ™ Total and Domain Scores.....	21
Table 8A: Internal Consistency Values for the Y-OQ™ Total and Domain Scores	22
Table 8B: Internal Consistency Values for the Y-OQ™ Total and Domain Scores Across Treatment Settings	22
Table 9: Reliability Estimates for the Y-OQ™ Using Normal Sample.....	25
Table 10: Criterion-Related Validity Estimates for the Y-OQ™ Using A Normal Elementary School Sample (N = 41).....	25
Table 11: Criterion-Related Validity Data from Patient Populations (N = 80)	25
Table 12: Criterion-Related Validity Estimates for the Y-OQ™ Using A Normal Elementary School Sample (N = 423).....	26
Table 13: Criterion-Related Validity Data from Patient Populations (N = 61)	26
Table 14: Comparison of Level of Psychopathology as Measured by the Total Y-OQ™ Score Across Patient and Non-patient Samples.....	27
Table 15: Formulas Used in Measuring Sensitivity to Change	28
Table 16: Cross-Tabulation of EN and RCI Classification Methods	29
Table 17: Comparison of Level of Psychopathology as Measured by the Total Y-OQ™ Score Across Patient and Non-patient Sample	31
Table 18: AUC Values for Y-OQ™ Total and Subscale Scores	32

INTRODUCTION

The question of how to assess psychotherapy outcome presents unique difficulties due to the apparent conflict of interests of three competing influences: the demands of rigorous research, the concerns of clinicians in practice, and the call for accountability from customers and third-party payers (Burlingame, Wells, Lambert, & Cox, 2004; Burlingame, Mosier et al., 2001). This manual describes the Youth Outcome Questionnaire™ (Y-OQ™; Wells, Burlingame, & Lambert, 1999), which was created by mutual effort from individuals representing each of these competing sectors--a university-based research team, administrators from a large managed health organization, parents and child patients, and clinicians in practice. A continuation of a popular adult outcome instrument, the Outcome Questionnaire® (OQ®-45; Lambert, Gregerson & Burlingame, 2004), the Youth Outcome Questionnaire™ (Y-OQ™) is a parent-report measure of treatment progress for children and adolescents (ages 4 - 17) receiving psychological or psychiatric treatment. In contrast to traditional diagnostic measures oriented to the measurement of psychopathology, it is specifically constructed to assess the occurrence of observed behavior change. Because of this unique purpose, the "intersect" of influences on its development will be briefly reviewed below.

INTERSECTION OF RESEARCH, CLINICAL WORK, AND MANAGED HEALTH CARE

Research

Although research on the effectiveness of psychotherapy has long demonstrated significant gains (Lambert & Ogles, 2003), the field of child and adolescent behavioral treatment research has traditionally lagged far behind (Kazdin, 1988, 1991, 1993; Kazdin, Bass, Ayers, & Rodgers, 1990; Shirk & Russell, 1992; Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, & Donenberg, 1992). However, recently meta-analytic reviews (Barnett, Docherty, & Frommelt, 1991; Casey & Berman, 1985; Hoag & Burlingame, 1997; Kazdin, 1990; Kovacs & Paulaskas, 1986; Weisz et al., 1987) have concluded that the average child treated in psychotherapy shows improvement (effect size = .71) over 76% of children in comparable untreated samples (Hoag & Burlingame, 1997), child treatment appears to be as effective as adult psychotherapy (Brown, 1987), and behavioral treatments appear to be the most effective in this population. Although methodological flaws may explain the last stated conclusion, aggregate findings indicate that psychological treatment for children and adolescents is effective (Kazdin, 1993).

Despite the increasing focus on child psychotherapy effectiveness research, demonstrated by more than 1500 studies of treatment outcome for children and adolescents, the available options and methods for treating children is multiplying rapidly, and most of these have never been subjected to empirical investigation (Kazdin, 2003). Kazdin (1995) reports that current therapy research in general is restricted and narrow. He called for more careful and diverse methodological research to explore and address a number of problems. He noted that youth recruited for research tend to demonstrate lower severity of symptomatology compared to children and adolescents in clinical practice. This methodological problem limits the generalizability of current research (Kazdin, 2003). Interestingly, it may actually underestimate the level of improvement achieved with clients in clinical practice since greater gains have been shown from more severely distressed youth on the Y-

OQ (Brown, Burlingame, Lambert, Jones & Vaccaro, 2001). Another methodological problem noted by Kazdin (1995) is the lack of standard outcome assessment protocols in the current research literature, preventing the aggregation of findings needed to build a common knowledge base. He suggests that standardized measures would allow researchers to compare and integrate findings on specific types of problems by profiling youth in a consistent way.

Despite Kazdin's cautions, Durlak, Wells, Cotten, and Johnson (1995) found that, in a study of 516 clinical interventions with children under the age of 13, their data suggested that a majority of studies employed adequate controls, random assignment to treatment groups, multiple outcome measurements, and assessment of generalization of treatment effects. They asserted that the methodologies employed in youth outcome research have improved significantly. However, a number of important neglected variables, including the need for normed outcome measures, allowing assessment of clinical significance (as opposed to statistical significance) of treatment effects, were noted. They assert, "There is, however, certainly room for improvement. In general, child psychotherapy research would be strengthened by greater use of normed outcome measures, assessing the general as well as the specific impact of treatment, using attention-placebo controls, and collecting follow-up data" (p. 143).

Clinicians

Ideally, research on psychotherapy informs the practicing clinicians, who subscribe to the current literature in the field to remain current regarding accepted practices. Unfortunately, clinicians are often hesitant to participate in research for two reasons. First, many practitioners feel that the practices described in the current literature are dissimilar to their practice. Kazdin (1991) asserts, "Research and clinical practice differ in who is treated, how they are treated, with what treatment approaches, how long they are treated, and range of persons involved in the child's treatment. There is genuine risk that the generalizability of research findings to clinical practice is quite strained. A critical research issue for the field is the deep chasm between research and practice on child treatments" (p. 795). Therefore, it is understandable for clinicians to consider research as irrelevant to their practical needs.

The second reason for the discrepancy between current research and practice is that the results of research are not for third-party payers. Therapists often feel pressure from third-party payers to produce substantial improvement in more severe clients and chaotic families, following shorter treatment periods than they see in the literature (Brown, et al., 2001).

Durlak et al. (1995) investigated this issue by reviewing practitioner concerns, such as: (a) that there have been few studies of nonbehavioral treatments; (b) that child research has seldom been conducted on subjects from minority populations; (c) that long-term treatments have not been studied; and (d) that few treatments have been investigated. They found that approximately 40% of the recent studies evaluate nonbehavioral interventions, 18% of the studies treated minority children, 20% of the reports evaluated treatments of 20 sessions or longer, and that 19 categories were required to describe the types of treatment entailed. Durlak and associates assert that the quality of child psychotherapy research is important to the practicing clinician.

Health Care Corporations

Although it is difficult to estimate the cost of mental health treatment with children and adolescents, several sources provide evidence that justifies the current move in the healthcare world towards cost containment and accountability in quality care (e.g., Burlingame, Lambert, Reisinger, Neff, & Mosier, 1995). The total cost of health care in the U.S. in 1996 was estimated to be \$943 billion (Department of Health and Human Services, 1999). In the same year healthcare costs for the treatment of psychiatric disorders in the U.S. accounted for 10.5% of total health costs (Department of Health and Human Services, 1999). While, mental health costs grew at an alarming rate of 30% to 40% during the early 1980s (Cummings, 1987) this rate appears to have decreased considerably. Specifically, the U.S. Department of Health and Human Services (1999) reported that "between 1986 and 1996, mental health expenditures grew at an average annual growth rate of more than 7 percent" which was equivalent to the growth seen in total health care costs during the 1990s.

In response to spiraling costs, the health care industry has instituted managed care and the so-called era of accountability. Thus, third-party payers require health-care providers to document therapeutic progress despite suspicions of practitioners. This has created a niche for psychotherapy outcome research. Additionally, in 1992, the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) adopted the CQI (continuous quality improvement) model, making continual tracking of patient care a further requirement in traditional outcomes assessment. In this setting, standardized outcome assessment becomes vital, as described by Burlingame et al. (1995): "Continuous monitoring of outcome ideally requires standardized data to profile reliable and valid patterns of improvement across time, providers, programs, and patient groups rather than data generated solely from professional judgment that tends to be more variable and unstable" (p. 227).

This manual describes the development and uses of the Youth Outcome Questionnaire™ (Burlingame, Wells, & Lambert, 1996), a 64-item parent-report measure constructed specifically to track treatment progress. Because the Y-OQ™ was created to track behavioral change in children in treatment, parents or others with reasonably extensive interaction with the client complete the questionnaire at intake to establish a severity baseline and then complete it repeatedly at regular intervals to track the child's progress. As will be described below, psychometric calculations from the normative database permit determination of the client's behavioral similarity at each measurement interval to in-patient, residential, in-home, and out-patient populations, and a large untreated community sample (Gillman, 1998; Burlingame, Mosier et al. 2001; Burlingame, Wells et al., 2001). Utilizing cut-off scores and a reliable change index (RCI), clinicians and/or administrators can determine if and when the client's behavior has entered the "normal" range of behavior.

The Y-OQ™ is composed of 64 items that comprise six separate subscales designed to tap several behavioral domains of children and adolescents experiencing behavioral difficulties. In addition, reverse-scoring items are also included to describe elements of healthy behavior increasing the range of the total Y-OQ™ score. Most parents require from 5 - 7 minutes to complete the measure, although particularly careful parents may take longer. Each item is to be rated on a five point Likert scale (0-4).

TEST DEVELOPMENT

Originally conceived as the child and adolescent equivalent of the OQ@45.2 (Burlingame et al., 1995; Lambert & Burlingame, 1996; Lambert et al., 2000), the Y-OQ™, like its predecessor, was

constructed to be brief, sensitive to change over short periods of time, and available at a nominal cost, while maintaining high psychometric standards of reliability and validity. Inasmuch as psychotherapy outcome research for children and adolescents has been considered as less “mature” than the research in adult outcome (Kazdin, 1995), a comprehensive instrument construction was undertaken to identify the most salient content domains to include in the instrument.

Initial Literature Review

Initially, a comprehensive review of the literature was performed of both narrative and meta-analytic reviews of the general psychotherapy treatment literature for children and adolescents having heterogeneous diagnoses. This review’s aim was to search for content domains that had empirical support as being sensitive to change in clinical work with children and adolescents. From recent meta-analytic reviews content domains were considered in which a .5 effect size was obtained (e.g., Baer & Nietzel, 1991; Casey & Berman, 1985; Grossman & Hughes, 1992; Prout & DeMartino, 1986; Roberts & Camasso, 1991; Russel, Greenwald, & Shirk, 1991; Shirk & Russell, 1992; Weisz, Weiss, Alicke, & Klotz, 1987). In other words, content domains in which the average treated child or adolescent was one-half a standard deviation better than untreated children or adolescents were initially considered. An effect size of .5 has been classified by Cohen (1988) as being in the medium range.

Focus Groups

The second element of the Y-OQ™ development protocol entailed gathering information from focus groups made up of consumers (former clients and their parents), as well as inpatient and outpatient provider focus groups (psychologists, psychiatrists, and other support staff). Both consumers and providers were drawn from a large western healthcare corporation. Professional focus group leaders led ten separate focus groups to identify characteristics of change thought to be the direct results of treatment. Audio tapes of the focus groups were transcribed and became part of the material reviewed. Focus groups from providers of inpatient care were particularly valuable in the construction of the Critical Items subscale which assesses characteristics necessitating inpatient treatment and which must be ameliorated or stabilized prior to referral to outpatient services.

Hospital Records

Hospital records were examined to assess the characteristic behavior change goals that were being addressed in treatment planning for both inpatient and outpatient clients. A manifest content analytic process was used to delineate the most frequently occurring change themes noted by providers in these two settings. A list of these themes was then compared to domains generated by the literature reviewed and focus groups. In some cases, item content for the Y-OQ™ directly reflects change described in hospital records.

The intersection of content from the literature reviews, focus groups, and hospital charts constituted the sources from which the final content domains for each of the six subscales were developed. Frequently, final item wording directly reflected change terminology used by parents in their description of a child or adolescent. The description of each subscale follows. Note that, while most of the items are formatted to describe various levels of behavior difficulty, 7 of the 64 items are written positively to describe and track healthy child/adolescent behaviors.

DESCRIPTION OF SUBSCALES

Intrapersonal Distress (ID)

The purpose of this scale is to assess the amount of emotional distress in the child/adolescent. Anxiety, depression, fearfulness, hopelessness, and self-harm are aspects measured by the ID subscale. Since depression and anxiety are frequently correlated in assessment instruments (Burlingame et al., 1995) no attempt was made at differentiating these symptoms. High scores indicate a considerable degree of emotional distress in the patient.

Somatic (S)

This scale assesses change in somatic distress that the child/adolescent may be experiencing. Items address symptoms that are typical presentations, including headaches, dizziness, stomachaches, nausea, bowel difficulties, and pain or weakness in joints. High scores indicate that the patient's caregiver is aware of a large number of somatic symptoms while low scores indicate either absence or unawareness of such symptoms.

Interpersonal Relations (IR)

The purpose of this scale is to assess issues relevant to the child's/adolescent's relationship with parents, other adults, and peers. Assessment is made regarding their attitude towards others, communication and interaction with friends, cooperativeness, aggressiveness, arguing, and defiance. High scores indicate that the care giver is reporting significant interpersonal difficulty while low scores reflect a cooperative, pleasant interpersonal demeanor.

Critical Items (CI)

This scale describes features of children and adolescents often found in inpatient services where short-term stabilization is the primary change sought. It assesses change in paranoia, obsessive-compulsive behaviors, hallucination, delusions, suicide, mania, and eating disorder issues. High scores are indicative of those who may need immediate intervention beyond standard outpatient treatment (inpatient, day treatment or residential care). A high score on any single item should receive serious attention by the provider.

Social Problems (SP)

This scale assesses problematic behaviors that are socially related. Many of the items describe delinquent or aggressive behaviors that are frequently the cause for bringing a child or adolescent into treatment. Although aggressiveness is also assessed in IR scale, aggressive content found in this scale is of a more severe nature, typically involving the breaking of social mores. Items include truancy, sexual problems, running away from home, destruction of property, and substance abuse. Another feature of items in this scale is that they are slow to change, whereas content tapped by many of the other scales often changes over a period of time as a result of treatment intervention.

Behavioral Dysfunction (BD)

This scale assesses change in the child's/adolescent's ability to organize tasks, complete assignments, concentrate, and handle frustration, including times of inattention, hyperactivity, and

impulsivity. Although many of the items on this scale tap features of specific disorders (e.g., Attention Deficit Hyperactivity Disorder) the scale is not intended to be diagnostic but rather to track change suggested by the literature, focus groups and hospital records.

Y-OQ™ Total

The total score is simply a summation of items from all six scales. It reflects total distress in a child's/adolescent's life. Like the OQ® 45.2 total (Lambert & Burlingame, 1996; Lambert et al., 2000), this value tends to be the best index to track global change and has the highest reliability and validity.

Overall the Y-OQ™ is proposed as a brief screening and outcome assessment tool to track the behavioral and subjective experience of a child or adolescent, as well as their ability to function in society. A copy of the instrument and license agreement can be found in Appendix G.

ADMINISTRATION

The Y-OQ™ requires no instructions beyond those printed on the answer sheet. A parent or a significant adult¹ figure rates each of the 64 items on a 5 point Likert scale.

It should be mentioned that parents taking this test can be affected by the attitudes of those administering the test. It is very important that the test administrator encourage the subject to fill out the scale in an honest and conscientious manner. Negative attitudes about the test or assessment communicated by clinicians or others who administer the instrument might significantly affect the results.

Time

Under usual circumstances subjects will complete the test in about seven minutes. Some especially careful individuals may require as much as 18-20 minutes, while others may complete the test in four to five minutes.

Scoring

Scoring the Y-OQ™ is a straightforward procedure, involving simple addition of item values. For example, if item #1 is endorsed at a 3, the weight given #1 for both subscale and total Y-OQ™ score is 3. Total and subscale scores are arrived at by adding items that load on each.

It should be noted that there are eight negatively scored items that occur in several subscales of the Y-OQ™. These items are #7, 16, 24, 32, 45, 47, 53, and 60. The negatively scored items tap "healthy" behaviors that might be positively impacted by mental health care services. That is, treatment might not only attenuate negative symptoms, but also increase positive behaviors in the child or adolescent. Accordingly, the scoring of these "healthy" behaviors is different from the rest of the items. The following weights are associated with scale values for items 7, 16, 24, 32, 45, 47, 53 and 60.

<u>Y-OQ™ Response</u>	<u>Appropriate Weight</u>
4	-2
3	-1

¹ To be referred to as Aparent@ for the rest of the text.

2	0
1	1
0	2

The *Total* score (TOT) is calculated by summing the patient's ratings across all 64 items. This yields a TOT range from -16 to 240. The higher the score, the more disturbed the individual.

The Intrapersonal Distress score (ID) is calculated by summing the patient's ratings on items 1, 3, 5, 9, 15, 17, 25, 32, 33, 34, 41, 49, 53, 57, 61, 62, 63, and 64. The ID has a score range of -4 to 68.

The Somatic score (S) is calculated by summing the patient's ratings on items 2, 10, 18, 26, 35, 42, 50, and 54. The S has a score range of 0 to 32.

The Interpersonal Relations score (IR) is calculated by summing the patient's ratings on items 4, 7, 11, 16, 19, 24, 27, 36, 37, and 43. The IR has a score range of -6 to 34.

The Social Problems score (SP) is calculated by summing the patient's ratings on items 6, 13, 22, 29, 31, 39, 47, and 55. The SP has a score range of -2 to 30.

The Behavioral Dysfunction score (BD) is calculated by summing the patient's ratings on items 8, 14, 23, 30, 40, 45, 48, 52, 56, 59, and 60. The BD has a range score of -4 to 40.

The Critical Items score (CI) is calculated by summing the patient's ratings on items 12, 20, 21, 28, 38, 44, 46, 51, and 58. The CI has a range score of 0 to 36.

Computer Scoring

A computer administration and scoring program (OQ®-Analyst) is available for clinical and research purposes. In addition to administration and scoring, it provides predictive algorithms that identify outpatient cases that may be at risk for treatment failure (i.e., symptom deterioration & premature termination). Further description of these algorithms and their underlying research is found in the Basic Interpretive Strategy section of this manual.

(Details on the software can be obtained by contacting 1-888-647-2673 or www.oqmeasures.com.)

Missing Data

If subjects omit answers to items, substitute values can be prepared by computing the mean of the remaining items for the respective subscales and rounding to the nearest whole number. This value can then be inserted into the test in the place of the missing value. However, this procedure will have an unknown effect on the reliability and validity of the scale scores. Therefore, it is highly recommended that protocols with missing data be viewed with considerable caution.

NORMATIVE DATA

Normative data for the Y-OQ™ have been collected from a wide variety of sites ranging from community normals to severe inpatient and residential treatment settings. Each site has contributed to the depth and breadth of knowledge available regarding the psychometric properties and uses of the Y-OQ™. Normative data from different treatment settings provides a guide to practitioners,

researchers, and treatment monitors regarding the severity of symptoms experienced by a child or adolescent and the patients of which setting the child's level of symptomatology most resembles.

Two separate inpatient and outpatient samples were collected through the intake offices of a large multi-state western health care corporation (approximately 600). Specifically, all children and adolescents who presented for treatment at outpatient and inpatient facilities located in Utah and Idaho owned by this corporation completed the instrument as part of their initial screening. By collecting data at intake from all new patients, initial level of disturbance for these populations was captured. Over 500 protocols were collected from four inpatient and outpatient sites.

Two community mental health centers (CMHC) also contributed to the outpatient data. These CMHCs are operated by state funded public mental health systems in the western United States (one services a city of approximately 1 million, and the other covers a diverse western county). These two samples contributed approximately 600 protocols to the present analyses.

Another outpatient clinical sample came from a western state's youth protection and reform system. This state system typically requires psychosocial interventions, as well as detention or probation. A total of 719 adolescents (64% male, 36% female) from this state system were part of the present analyses.

Additional outpatient samples were obtained from two managed care companies, one servicing clients throughout the United States, and another along the Eastern seaboard. These clinical samples consisted of 577 subjects. Protocols were simply given to parents of adolescents presenting for treatment within the care system at intake and pre-determined points throughout treatment.

A residential treatment group, collected from a private residential treatment facility also in the western U.S. was collected ($n = 242$) provides a preliminary comparison group for this special population. However, a cautionary note is sounded with this sample since admission policies for residential programs can dramatically vary changes the overall severity estimates of this population. Nonetheless, Gillman (1998) conducted research investigating the need for separate norms for the residential treatment population. Using a sample ($n = 242$) of adolescents at a residential treatment facility, he found the mean score for his residential patients to be significantly higher [$F(3, 3096) = 879.85, p < .001$] than the means for all previous normative groups, including the inpatient sample. Gillman found that all subscale scores were significantly different from the subscale scores of the community and outpatient samples. Additionally, four of the six subscale scores for the residential sample differed significantly from those of the inpatient sample. The ID and CI subscales were the only subscales that failed to reach significance. These findings suggest the need to develop residential norms to accompany the other treatment settings already included in this manual. Although replication is essential, Gillman's norms are included in Table 1 to assist the interested reader.

Collectively, across the three samples (community normal, outpatient, and inpatient), males slightly outnumber females (approximately 60% to 40%, respectively) with approximately 29% of the total sample coming from the community normal sample, 60% coming from outpatients, and 11% coming from inpatients.

The means and standard errors for the four normative groupings are depicted in Table 1A. As expected, there are very large differences on the total scores between the three samples, $F(2, 3822) = 1,137.3, p < .000$. Community sample total means were substantially below those taken from clinical settings (inpatient and outpatient). Reliable differences also exist between the two patient

populations with inpatients demonstrating more overall symptoms than outpatients. Furthermore, residential patients exhibit the highest levels of symptom severity.

Table 1A: Normative Group Data for the Y-OQ™ Total Score			
Sample	N	Mean	SE
Residential	242	115.4	2.13
Inpatient	435	110.4	1.70
Outpatient	2,297	68.4	.82
Community	1,091	21.4	.80

Analysis of domain scores (Table 1B) strongly support statistically significant differences between the community normal and the three clinical samples on nearly every subscale. Using a Bonferroni correction at alpha of .05, all subscales reliably distinguished between all three groups.

Table 1B: Normative Group Raw Score Data for the Y-OQ™ Domains								
	Residential (N = 242)		Inpatient (N = 435)		Outpatient (N = 2,297)		Community (N = 1,091)	
Sample	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Intrapersonal Distress	34.1	0.72	34.6	0.58	22.2	0.27	8.1	0.80
Somatic	8.1	0.33	8.6	0.26	5.9	0.10	3.1	0.10
Interpersonal Relations	18.5	0.40	16.5	0.36	9.4	0.16	0.3	0.15
Social Problems	16.6	0.41	14.3	0.33	6.7	0.12	0.6	0.09
Behavioral Dysfunction	26.6	0.50	24.8	0.41	17.3	0.21	6.5	0.23
Critical Items	11.7	0.39	11.6	0.28	6.9	0.11	2.7	0.09

In an effort to replicate these normative values, a recent sample of 15,614 protocols were obtained from two large, multi-state HMO's covering millions of lives. The samples were obtained through primarily outpatient settings (82%), with 6% of the sample drawn from inpatient settings, 4% from inpatient day treatment, 2% from residential treatment programs, and 6% from community normals. The Y-OQ total score means for each of these samples are presented in Table 2A. In this sample, using a Bonferroni correction at alpha of .05, each treatment condition sample significantly differed from the others, with the exception of the samples from the Residential and Inpatient Day Treatment settings.

Table 2A: Validation Group Data for the Y-OQ™ Total Score			
Sample	N	Mean	SE
Residential	325	89.74	2.19
Inpatient Regular	629	104.99	1.54
Inpatient Day Tx	587	87.08	1.54
Outpatient	12,807	75.35	0.33
Community Normal	881	20.51	0.90

Note the difference between mean scores for the Residential Treatment settings in Table 1A (115.4) versus Table 2A (89.74). The change in means between 1996 and 2005 are striking. These differences may reflect different populations given changes in the cohort groups over the past decade. They may also reflect differences in admission policies or, alternatively, sampling error. Further statistical analysis of domain scores (Table 2B) strongly support theoretical differences between the outpatient samples and the three severe clinical samples on nearly every subscale. Using a Bonferroni correction at alpha of .05, all the subscales reliably distinguished the treatment settings from the community normals. Furthermore, the subscales were tested to investigate whether they reliably distinguished between all four groups. Comparisons identified Inpatient Day Treatment and Residential Treatment settings as statistically similar for subscales ID, S, IR, and BD. The Outpatient sample did not differ significantly from the Inpatient Day Treatment and Residential Treatment samples on subscale S, the Inpatient Regular sample did not differ from the Residential Treatment sample on subscale SP, and the Inpatient sample did not differ significantly from the Residential Treatment and the Inpatient Day Treatment samples on subscale BD.

Table 2B: Validation Group Raw Score Data for the Y-OQ™ Domains								
	Residential (N = 325)		Inpatient Reg. (N = 629)		Inpatient Day Tx (N = 587)		Outpatient (N = 12807)	
Sample	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Intrapersonal Distress	28.4	0.70	36.4	0.48	27.4	0.51	25.8	0.11
Somatic	7.0	0.30	9.4	0.23	7.1	0.21	7.2	0.05
Interpersonal Relations	12.2	0.44	13.9	0.33	12.4	0.31	9.7	0.07
Social Problems	11.3	0.36	10.3	0.28	8.9	0.23	6.4	0.05
Behavioral Dysfunction	20.1	0.51	21.7	0.38	21.6	0.38	18.4	0.08
Critical Items	10.7	0.35	13.2	0.26	9.6	0.23	7.9	0.05

Table 3A presents data from three outpatient treatment settings from diverse geographic locations and demographic characteristics, illustrating the comparability of the Y-OQ across settings. No differences are significant between sites on total or subscale means.

Further replication of these initial values comes from data collected at two large, multi-state HMO's and a community normal group, all from the same locale, and is presented in Table 3B. Data comparisons, using a Bonferroni correction and an Alpha level of .05, reveal similar patterns to those noted in Table 3A. While statistically significant differences were found on the Total score and all subscales between the two HMOs, the absolute magnitude is small, suggesting that differences may be due to large sample sizes.

CUTOFF SCORE

One of the major purposes of an outcome measure is to identify when a client is functioning in the range of normality, or when they are significantly disturbed. This ability to identify the range in which a patient is functioning is useful in both clinical and research arenas. According to Jacobson et al. (1999), a cutoff score calculated as a weighted mean between the means of two separate groups is useful at determining which group a patient belongs to. Therefore, a Cutoff score has been

Table 3A: Total and Subscale Y-OQ™ Raw Scores for Outpatients from Three Health Maintenance Organizations							
Source	Total Y-OQ™	Subscale Scores					
		BD	CI	ID	IR	S	SP
Company A							
Mean	71.2	18.1	7.5	24.1	9.2	6.2	5.5
N	171.0	189.0	190.0	179.0	187.0	189.0	187.0
SD	36.3	9.2	5.1	12.3	7.5	4.9	5.3
Company B							
Mean	84.9	20.2	8.5	26.2	12.8	6.6	9.0
N	226.0	328.0	339.0	297.0	340.0	340.0	330.0
SD	37.9	8.9	5.4	12.4	7.4	5.2	5.9
Company B							
Mean	66.6	19.3	6.4	22.1	8.4	5.5	4.9
N	255.0	255.0	255.0	255.0	255.0	255.0	255.0
SD	32.7	9.1	4.3	11.8	7.0	4.4	4.2
Total							
Mean	74.1	19.5	7.6	24.3	10.5	6.1	6.8
N	652.0	772.0	784.0	731.0	782.0	784.0	772.0
SD	36.4	9.1	5.1	12.3	7.6	4.9	5.6

Note: Company A is a large multi-state HMO located on the Eastern seaboard. It is primarily an outpatient facility, but includes a hospital. Company B is a large community mental health center in the Western United States serving primarily Medicare/Medicaid patients. It serves a catchment area of 750,000 people. Company C is a large multi-state HMO serving patients throughout the United States. The current sample was collected primarily from East and West Coast offices and the Southern States.

Table 3B: Total and Subscale Y-OQ™ Raw Scores for Outpatients from Two Health Maintenance Organizations							
Source	Total Y-OQ™	Subscale Scores					
		BD	CI	ID	IR	S	SP
Company D							
Mean	79.6	19.2	8.5	27.9	10.0	7.7	6.5
N	7,979	7,979	7,979	7,979	7,979	7,979	7,979
SD	36.99	9.2	5.5	12.9	7.8	5.3	6.1
Company E							
Mean	75.3	18.2	8.1	24.7	10.1	6.9	7.3
N	6,754	6,754	6,754	6,754	6,754	6,754	6,754
SD	38.1	9.6	5.5	12.9	7.6	5.2	5.7
Community Normal							
Mean	20.5	6.3	2.7	7.8	0.2	3.0	0.6
N	881.0	881.0	881.0	881.0	881.0	881.0	881.0
SD	26.8	7.4	2.9	9.8	5.2	3.3	2.9

calculated between the community sample and the two clinical samples (inpatient and outpatient combined) since this seems the most logical place to compare individuals for treatment outcome. By identifying whether a client's score is above or below this Cutoff, a researcher or clinician can determine whether the subject's score most resembles that of the community normals, or of the clinical population. Cutoffs for the Y-OQ™ total score and subscale scores are as follows: Total: 46; Intrapersonal Distress: 16; Somatic: 5; Interpersonal Relations: 4; Social Problems: 3; Behavioral Dysfunction: 12; Critical Items: 5. These cutoff values are used in the graphs in an illustration in appendix A for the purpose of monitoring patient progress. The formula (Jacobson & Truax, 1991) used to devise these cutoffs is:

$$c = \frac{(SD_1)(mean_2) + (SD_2)(mean_1)}{SD_1 + SD_2}$$

Using this formula, cutoffs can be derived between any two normative samples for comparative purposes in evaluating treatment outcome. The accuracy of these cutoff scores is examined specifically in the Sensitivity and Specificity Section later in the manual. We recommend that the cutoff scores presented in this manual be used for this general purpose since they are based on large and diverse samples. If special populations are being assessed, however, it may be appropriate to construct new normative samples and compute new cutoff scores for that particular group.

RELIABLE CHANGE INDEX

In addition to a Cutoff score between the community mean and the clinical means, Jacobson et al. (1999) recommend that a reliable change index be established, with the goal of aiding an outcome measure to identify when a client has made changes in their level of symptomatology. The RCI is used to determine if the change exhibited by an individual in treatment is reliable or clinically significant (Jacobson & Truax, 1991). In order for an individual's score to be considered clinically significantly changed, it must cross a cutoff score and have a magnitude greater than the RCI. The RCI value that has been computed is 13, meaning that an individual's score must change by at least 13 points on the Y-OQ™ to be considered clinically significantly changed. The RCIs for each of the subscales are as follows: Intrapersonal Distress 8, Somatic 5, Interpersonal Relations 4, Social Problems 5, Behavior Dysfunction 8, and Critical Items 5. The formula for computing the RCI is given below.

$$RCI = \frac{(pre)-(post-treatment)}{S_{diff}} = 1.96$$

$$S_{diff} = \sqrt{2SE^2}$$

$$SE = SD \sqrt{1 - r_{xx}}$$

The standard error of measure (SE) is computed using the internal consistency of the Y-OQ™, which is 0.97 for the total score (computed using all three normative samples), and a pooled standard deviation value (SD). The resulting SE value is inserted into the standard error of difference formula

(S_{diff}). This value is then multiplied by the z-value of the significance level desired, in this case 1.96 ($p < 0.05$). The resulting value represents the size of the difference needed to achieve reliably significant change given the error of the instrument and the standard deviations of the normative samples.

As with the cutoff score, we recommend the use of the RCI presented here for most general purposes, as it is based on large and diverse normative samples. If specialized or more specific RCI values are desired, appropriate norms can be gathered and new RCI values can be derived using the formulas given above.

GENDER DIFFERENCES

Table 4A depicts the gender differences within each sample using the total score (inpatient, outpatient, and community). Statistically significant gender differences were found when the three samples were combined $F(1,3670) = 8.22$, $p = .004$. Within each sample, however, the gender differences varied. Using $p < .05$ as the significance criteria, gender differences were not significant within the inpatient sample, but were significant within the outpatient and sample community sample. Thus, it appears no correction for gender is necessary when the instrument is used in inpatient settings. However, gender considerations may be important in interpreting scores for community and outpatient populations. Specifically, the mean score obtained by females in outpatient settings is higher, on average, than that obtained by males ($M = 71.2$ and 67.3 , respectively); while females in the community population is lower, on average, than that obtained by males ($M = 19.8$ and 23.3 , respectively).

Table 4A: Comparison of Gender Scores on the Y-OQ™ - Total Score				
Sample	N	Mean	S.E.	Sig. (2-tailed)
Inpatient				
Male	269	109.1	2.12	.32
Female	166	112.6	2.82	
Outpatient				
Male	1,399	67.3	.34	.03
Female	751	71.2	.48	
Community				
Male	554	23.3	1.17	.03
Female	532	19.8	1.10	

These findings were again replicated (see Table 4B) with a more recent sample of patients from two multi-state HMO's. Significant gender differences were noted in the all samples ($\text{sig.} = 0.005$). However, as with the earlier normative data (Table 3B), absolute value of these differences were small ranging from 3-4 points.

In addition to gender differences in total scores, males and females were found to vary reliably on two Y-OQ™ subscales (see Table 5A). Males irrespective of sample have higher behavioral dysfunction scale scores than do females, $F(1, 3670) = 58.13$, $p = .000$. On the other hand, females in both clinical samples (outpatient and inpatient) have higher somatic scale scores than do their male counterparts, $F(1, 2584) = 45.72$, $p = .000$, and higher intrapersonal distress scale scores, $F(1, 2584)$

Table 4B: Comparison of Validation Gender Scores on the Y-OQ™ - Total Score			
Sample	N	Mean	S.E.
<i>Residential</i>			
Male	181	87.8	2.95
Female	144	92.1	3.27
<i>Inpatient Regular</i>			
Male	273	106.7	2.38
Female	356	103.7	2.01
<i>Inpatient Day Tx</i>			
Male	401	95.8	1.82
Female	186	89.9	2.89
<i>Outpatient</i>			
Male	7,350	76.7	0.43
Female	5,457	73.6	0.50
<i>Community Normal</i>			
Male	449	22.1	1.30
Female	432	18.9	1.24

= 30.70, $p = .000$. Inasmuch as the behavior dysfunction scale taps behavior frequently associated with attention deficit disorder, this finding is to be expected, given that three to six times more males than females are identified with ADHD (Whalen, 1989). Similarly, the Werry, Reeves and Elkind's (1987) review finds that females predominate in the incidence of somatization disorders (an internalizing disorder). In addition to the differences stated above, preliminary results from a large Outdoor Behavioral Health Treatment Program indicate the possibility of gender differences in this population (Russell, 2003). Russell (2003) found that males ($n=98$) being treated in an outdoor behavioral setting show significantly lower mean Y-OQ™ total score than their female counterparts ($n=46$) being treated in the same setting. Russell (2003) reports that at intake, female patients' mean scores were 104.72, while male mean scores were 94.05. Although the reported sample is small, these differences indicate a need for further study of gender differences among separate treatment populations.

Again, these findings were replicated (see Table 5B) using the recent data obtained from a large, multi-state HMO. In each of the clinical samples, significant, albeit small, differences were found between the genders on the Somatic subscale. Furthermore, significant gender differences were found on the Interpersonal Distress subscale in the Inpatient Day Treatment and Outpatient samples. Gender differences were also discovered on the Behavioral Dysfunction subscale among the Inpatient Regular and the Outpatient clinical samples. Interestingly, the Outpatient sample was the only one to demonstrate significant gender differences across all subscales and the total score. These findings are likely a statistical anomaly due to the large sample size for both genders in the Outpatient Sample. As with the earlier samples, differences (females scoring significantly higher on the Somatic subscale and significantly lower on the Behavioral Dysfunction subscale) were congruent with the prevailing theory that female adolescents tend to somatize, while male adolescents tend to externalize their difficulties. The ability of the Y-OQ™ to measure significant differences in the Behavioral Dysfunction and Somatic subscales between the genders contributes support to the construct validity of the measure. Finally, the Community Normal sample

demonstrated significant gender differences on the Interpersonal Relations, Social Problems, and Behavior Dysfunction subscales, as well as the Total Score.

Table 5A: Comparison of Gender Scores on the Y-OQ™ - Domain Scores						
Sample	Inpatient		Outpatient		Community	
	N = 269	N = 166	N = 1,399	N = 751	N = 554	N = 532
<i>Interpersonal Distress</i>						
Male	33.7	.74	21.2	.34	8.2	.41
Female	36.1	.91	24.2	.49	8.1	.42
<i>Somatic</i>						
Male	8.0	.32	5.5	.13	3.0	.14
Female	9.4	.42	6.8	.19	3.2	.14
<i>Interpersonal Relations</i>						
Male	16.3	.46	9.5	.21	.7	.22
Female	16.7	.59	9.6	.29	-.08	.21
<i>Social Problems</i>						
Male	14.0	.40	6.7	.15	1.0	.13
Female	14.9	.59	6.7	.22	.3	.11
<i>Behavioral Dysfunction</i>						
Male	25.4	.51	17.8	.27	7.9	.34
Female	23.8	.68	16.5	.35	6.8	.29
<i>Critical Items</i>						
Male	11.7	.36	6.7	.13	2.9	.12
Female	11.6	.45	7.3	.19	2.9	.12
<i>Total Score</i>						
Male	109.0	2.12	67.3	.34	23.3	1.17
Female	112.6	2.82	71.2	.48	19.8	1.10

In addition to the statistical differences in scores due to the gender of the child, recent data from two large Western HMO's indicate that different types of respondents may differ significantly on the Y-OQ™ total score and subscale scores. This data is presented in Table 5C. Using a Bonferroni correction and an Alpha level of .05, Fathers (n=1653) who responded to the Y-OQ™ were found to differ significantly from Mothers (n=6859), Guardians (n=851), Clinicians (n=33), and Residential House Counselors (n=521) on all subscales, as well as the Total Score. Specifically, Fathers rated their children lower than the aforementioned groups. Additionally, on the ID, S, and the BD subscales, Mothers were found to differ significantly from Guardians and Residential House Counselors, rating their children at a higher distress level. On the IR subscale, Clinicians, Residential House Counselors, and Mothers were found to score significantly higher than Guardians. On the SP subscale, Mothers and Guardians scored significantly lower than Residential House Counselors, and on the Y-OQ™ Total Score, Guardians scored significantly lower than Clinicians, Mothers, and House Counselors. Due to the sample sizes, statistical differences were somewhat erratic in these analyses; however, the general trend was for Clinicians to score the highest and fathers to score the lowest across subscales and Total Score, while middle scorers differed according to scales (see Table 5C). These scores underscore the importance of caution when interpreting scores from differing respondents on the Y-OQ™.

Table 5B: Comparison of Validation Gender Scores on the Y-OQ™ - Domain Scores										
Sample	Residential		Inpatient Regular		Inpatient Day Tx		Outpatient		Community Normal	
	N=341	N=330	N=273	N=357	N=531	N=239	N=8,677	N=6,488	N=546	N=521
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
<i>Intrapersonal Distress</i>										
Male	23.0	0.71	35.8	0.74	25.3*	0.55	24.6*	0.14	8.3	0.42
Female	24.7	0.75	36.9	0.62	28.5*	0.82	26.3*	0.16	8.1	0.43
<i>Somatic</i>										
Male	6.2*	0.27	8.9*	0.34	6.6*	0.22	6.7*	0.06	3.0	0.14
Female	7.4*	0.30	9.9*	0.30	8.0*	0.33	7.7*	0.07	3.2	0.15
<i>Interpersonal Relations</i>										
Male	9.5	0.42	14.4	0.50	11.8	0.33	9.9*	0.08	0.7*	0.23
Female	8.7	0.46	13.7	0.43	11.8	0.49	8.9*	0.9	-0.1*	0.22
<i>Social Problems</i>										
Male	9.4	0.34	10.7	0.42	8.5	0.24	6.7*	0.06	1.0*	0.14
Female	9.8	0.40	10.1	0.37	9.0	0.39	5.9*	0.07	0.3*	0.12
<i>Behavioral Dysfunction</i>										
Male	17.1	0.50	23.6*	0.58	21.0	0.40	19.2*	0.10	7.5*	0.34
Female	16.0	0.50	20.2*	0.49	20.0	0.62	16.4*	0.11	5.6*	0.30
<i>Critical Items</i>										
Male	9.1	0.34	13.4	0.41	09.2	0.24	7.8*	0.06	2.8	0.13
Female	9.6	0.36	13.1	0.34	10.0	0.39	8.0*	0.07	2.7	0.13
<i>Total Score</i>										
Male	74.4	2.16	106.7	2.38	82.4	1.63	75.0*	0.40	23.3*	1.18
Female	76.3	2.32	103.8	2.01	87.2	2.55	73.3*	0.46	19.8*	1.12

*Significant differences between genders at Alpha=.05 (2-tailed)

Table 5C: Comparison of Respondent Scores on the Y-OQ™ - Domain Scores											
		<i>Respondent</i>									
		Clinician		Mother		House Counselor		Guardian		Father	
		N=33		N=6,859		N=521		N=851		N=1,653	
		Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Scale	ID	28.9	1.66	27.6	0.16	23.5	0.53	22.3	0.47	18.3	0.36
	S	8.6	0.93	7.7	0.06	5.9	0.20	6.0	0.17	5.0	0.11
	IR	12.0	1.28	9.8	0.09	9.9	0.33	7.7	0.27	5.6	0.20
	SP	9.1	1.22	6.2	0.07	8.6	0.27	6.4	0.19	4.3	0.15
	BD	20.36	1.62	19.0	0.11	17.0	0.42	16.8	0.35	13.0	0.25
	CI	10.2	1.00	8.3	0.07	8.6	0.25	7.9	0.20	5.8	0.13
	Total	89.1	6.34	78.6	0.45	73.5	1.66	67.2	1.40	52.0	1.06

AGE DIFFERENCES

Data from the community normal normative sample was examined by five age groupings that correspond somewhat with those used in other diagnostic instruments (e.g., Connors Parent Rating

Y-OQ™ Administration and Scoring Manual

Scale, Child Behavior Checklist). Slightly higher elevations were found on the total Y-OQ™ score for the two older age groups, creating a 5 to 8 point spread between the younger and older subjects (Table 6A); these differences between age groups were significant; $F(4, 1086) = 2.82, p < .05$. The subscale averages (means) in this same sample further explain the differences noted on the total Y-OQ™ scores (Table 6B). The greatest age difference was on the intrapersonal distress Scale where the two older groups showed higher distress scores than the two younger groups, $F(4, 1086), 6.78, p < .01$. Additionally, the two older groups had higher levels of interpersonal relation distress than the three younger groups; $F(4, 1083), 4.69, p < .01$. The only other age differences were on the social problems subscale where the 15-17 year olds demonstrated more social problem distress than the 9-11 year olds; $F(4, 1086) = 3.34, p < .05$; and on the somatic subscale where the 15-17 year olds demonstrated more somatic distress than the 4-5 year olds; $F(4, 1086) = 3.29, p < .05$.

Table 6A: Y-OQ™ Total Score by Age in Community Normal Sample			
Age – Range	N	Total Score	
		Mean	S.E.
Age = 4-5	207	19.0	1.84
Age = 6-8	359	20.4	1.39
Age = 9-11	284	20.7	1.57
Age = 12-14	132	25.8	2.29
Age = 15-17	105	27.3	2.58

Table 6B: Y-OQ™ Subscale Scores by Age in a Sample of Community Normals										
	Age - Range									
	Age = 4-5		Age = 6-8		Age = 9-11		Age = 12-14		Age = 15-17	
	N = 207		N = 359		N = 284		N = 132		N = 105	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Interpersonal Distress	6.1	.67	7.5	.51	8.5	.57	10.4	.84	10.8	.94
Somatic	2.5	.23	3.0	.17	3.3	.20	3.4	.29	3.7	.32
Interpersonal Relations	.04	.35	.10	.26	-.22	.30	1.3	.44	1.8	.49
Social Problems	.84	.20	.73	.15	.24	.17	.47	.25	1.4	.29
Behavioral Dysfunction	7.0	.52	6.4	.40	6.2	.44	6.9	.65	6.4	.73
Critical Items	2.5	.20	2.6	.15	2.6	.17	3.4	.25	3.1	.28

Table 6C: Y-OQ™ Subscale Scores by Age and Clinical Samples

Sample/Age	Residential			Inpatient Reg.			Inpatient Day Tx		
	Mean	N	S.E.	Mean	N	S.E.	Mean	N	S.E.
<i>Interpersonal Distress</i>									
4-5	25.0	3	5.03	34.3	9	3.01	28.6	29	2.10
6-8	24.8	35	2.19	36.6	37	2.05	29.4	144	0.98
9-11	27.8	77	1.56	39.7	79	1.29	26.5	188	0.91
12-14	23.6	232	0.86	36.2	171	0.93	25.2	282	0.76
15-17	23.0	324	0.75	35.7	334	0.65	24.4	127	1.22
<i>Somatic</i>									
4-5	7.7	3	1.20	7.4	9	1.23	7.3	29	0.67
6-8	5.5	35	0.67	8.5	37	0.67	7.7	144	0.42
9-11	7.0	77	0.63	11.1	79	0.70	7.0	188	0.37
12-14	6.6	232	0.34	8.9	171	0.44	6.7	282	0.31
15-17	7.0	324	0.30	9.5	334	0.31	7.0	127	0.47
<i>Interpersonal Relations</i>									
4-5	12.3	3	1.45	20.3	9	2.24	13.8	29	1.36
6-8	11.0	35	1.31	19.2	37	1.01	14.8	144	0.52
9-11	12.5	77	1.02	17.9	79	0.87	12.1	188	0.56
12-14	9.4	232	0.51	13.8	171	0.63	11.0	282	0.46
15-17	7.9	324	0.44	12.3	334	0.43	9.3	127	0.67
<i>Social Problems</i>									
4-5	8.7	3	1.76	13.6	9	1.89	8.6	29	0.80
6-8	9.8	35	0.96	13.0	37	0.77	9.9	144	0.39
9-11	10.3	77	0.76	11.9	79	0.65	8.1	188	0.39
12-14	9.3	232	0.43	9.0	171	0.52	8.3	282	0.36
15-17	9.7	324	0.40	10.3	334	0.40	8.6	127	0.54
<i>Behavioral Dysfunction</i>									
4-5	23.3	3	4.63	30.7	9	3.2	24.7	29	1.57
6-8	21.4	35	1.68	29.9	37	1.18	25.3	144	0.60
9-11	22.0	77	1.20	27.9	79	0.97	21.4	188	0.66
12-14	16.6	232	0.55	21.8	171	0.70	19.3	282	0.57
15-17	14.7	324	0.48	19.0	334	0.48	16.4	127	0.79
<i>Critical Items</i>									
4-5	8.0	3	2.31	14.8	9	1.91	12.0	29	0.92
6-8	9.7	35	1.15	15.3	37	1.06	10.2	144	0.43
9-11	11.0	77	0.74	16.0	79	0.72	9.7	188	0.42
12-14	9.0	232	0.39	13.1	171	0.50	9.1	282	0.35
15-17	9.2	324	0.37	12.3	334	0.35	8.6	127	0.50
<i>Total Score</i>									
4-5	85.0	3	14.47	121.1	9	10.92	95.0	29	5.8
6-8	82.1	35	6.86	122.6	37	5.05	97.3	144	2.67
9-11	90.6	77	4.95	124.6	79	4.08	84.9	188	2.74
12-14	74.5	232	2.52	102.8	171	2.94	79.5	282	2.33
15-17	71.4	324	2.32	99.2	334	2.08	74.3	127	3.58

Table 6C Continued						
Sample/Age	Outpatient			Comm. Norm.		
	Mean	N	Mean	Mean	N	SE
<i>Interpersonal Distress</i>						
4-5	21.0	1,130	0.35	6.1	206	0.54
6-8	24.2	2,933	0.22	7.5	356	0.46
9-11	26.6	3,102	0.22	8.6	276	0.62
12-14	26.6	3,760	0.21	10.4	128	1.04
15-17	25.4	4,240	0.21	10.6	102	1.13
<i>Somatic</i>						
4-5	6.4	1,130	0.14	2.5	206	0.20
6-8	6.7	2,933	0.09	3.0	356	0.17
9-11	7.5	3,102	0.09	3.4	276	0.20
12-14	7.3	3,760	0.09	3.5	128	0.30
15-17	7.3	4,240	0.08	3.7	102	0.38
<i>Interpersonal Relations</i>						
4-5	9.0	1,130	0.21	0.1	206	0.31
6-8	9.4	2,933	0.14	0.1	356	0.26
9-11	9.6	3,102	0.13	-0.2	276	0.31
12-14	10.2	3,760	0.13	1.3	128	0.52
15-17	9.0	4,240	0.12	1.8	102	0.59
<i>Social Problems</i>						
4-5	6.0	1,130	0.14	0.9	206	0.17
6-8	5.9	2,933	0.09	0.7	356	0.15
9-11	5.3	3,102	0.09	0.2	276	0.16
12-14	6.4	3,760	0.10	0.5	128	0.27
15-17	7.4	4,240	0.10	1.3	102	0.42
<i>Behavioral Dysfunction</i>						
4-5	19.3	1,130	0.29	7.1	206	0.52
6-8	20.1	2,933	0.17	6.4	356	0.39
9-11	19.3	3,102	0.16	6.2	276	0.47
12-14	18.0	3,760	0.15	6.9	128	0.64
15-17	15.4	4,240	0.14	6.4	102	0.77
<i>Critical Items</i>						
4-5	7.7	1,130	0.15	2.6	206	0.19
6-8	7.9	2,933	0.09	2.6	356	0.14
9-11	8.0	3,102	0.09	2.6	276	0.18
12-14	8.0	3,760	0.09	3.4	128	0.31
15-17	7.9	4,240	0.09	3.0	102	0.29
<i>Total Score</i>						
4-5	69.5	1,130	1.06	19.1	206	1.65
6-8	74.2	2,933	0.66	20.5	356	1.32
9-11	76.2	3,102	0.63	20.9	276	1.64
12-14	76.4	3,760	0.61	25.9	128	2.66
15-17	72.3	4,240	0.60	21.6	102	0.82

The two clinical samples (outpatient and inpatient) were separately analyzed for age differences. These values can be found in Appendix B. An age difference pattern was found in the outpatient sample on the behavioral dysfunction subscale. The 15-17 year old group reported lower behavioral distress when compared with all the younger age groups; $F(4, 1783) = 17.23, p = .000$. Also in the outpatient sample, the 15-17 age group consistently reported higher scores than the three younger age groups on the social problems scale; $F(4, 1783) = 10.611, p = .000$. The only other age difference found in the outpatient sample was on the intrapersonal distress scale; $F(3,305) = 4.22, p < .05$. The 4-5 age group had lower values than all the older age groups and the 9-11 age group had lower scores than the 15-17 age group. The only age difference in the inpatient sample consisted of the 9-11 age group, reporting lower scores than the 15-17 age group on the Social Problems Subscale; $F(4, 414) = 8.36, p = .000$. Such age group differences, taken as a whole, illustrate the developmental issues associated with the onset of adolescence, wherein major transitions (e.g., puberty, entry to junior high school, etc.) are taking place, which upset self-esteem and coincide with the search for identity (Cf. Block & Robins, 1993, Grotevant & Cooper, 1986).

Using recent data obtained from the aforementioned HMO's, age range differences were again explored for each subscale in each clinical sample, as well as the Community Normal sample. Table 6C presents this data.

In this data, patterns of statistical significance are somewhat complicated, and can be read in Appendix B; however, general trends can be noted. On the ID Subscale, across treatment settings, the 9-11 year-old sample appeared to score the highest, with scores trailing off for younger and older children. In comparison to this, community normal scores were highest for the oldest children and lowest for the youngest. On the S subscale, scores were fairly comparable across age ranges for each treatment setting, with slightly lower scores for younger children in the Outpatient and Inpatient Regular settings. Again, on the S subscale, the Community normal sample showed a general trend of lower scores for younger children. On the IR subscale, the Residential and Inpatient settings showed a trend of higher scores for younger children, whereas the Outpatient showed similar scores across ages. Again, on the IR subscale, the Community Normal sample showed a general trend of higher scores for older children. On the SP subscale, the Residential sample showed highest scores for the 9-11 year old range, with scores tapering off somewhat towards older and younger children. Meanwhile, the Regular Inpatient sample demonstrated a trend of higher scores for younger children, while the Inpatient Day Treatment, Outpatient, and Community Normal samples showed a trend toward higher scores for the youngest and oldest children with scores decreasing towards the ages between the highest and lowest. On the BD subscale, across all treatment settings, samples demonstrated a general trend toward higher scores for younger children, whereas the Community Normal sample demonstrated fairly equivalent scores for all age groups except the 4-5 age range, which scored higher than the others. On the CI subscale, the Residential, Inpatient Regular, Outpatient samples showed a trend of higher scores for the 9-11 year-old groups with scores tapering off towards the younger and older age groups, whereas the Inpatient Day Treatment sample showed higher scores for younger children and the Community Normal showed the reverse: lower scores for younger children. Finally, on the Y-OQ™ Total Score, the Residential, Inpatient Regular, and Outpatient samples showed a pattern of higher scores for the 9-11 year-old age group, with scores decreasing for older and younger children, while the Inpatient Day Treatment sample showed higher scores for younger children and the Community Normal sample showed a trend toward higher scores for older children.

OTHER NORMATIVE GROUPS

The Y-OQ, to date, has been translated into Spanish, French, German, Hmong, and Laotian. Researchers are gathering data for African-American families and Hawaiian Island-Polynesian families. Results of analyses of Lao-speaking families (Mills, Manivanh, Burlingame, Wells, Peterson, & Nuttal, 1997) found the mean Y-OQ™ total scores to be significantly higher (46.58) than the means of the community normal sample (22.40), however, less elevated than the outpatient treatment sample (77.41). These findings are questionable, however, since Southeast Asians have been reported to have less favorable attitudes toward psychology and psychologists. This in turn, may have affected respondent scores (Dinh, Sarason, & Sarason, 1994). Furthermore, investigators reported that some parents were confused about the structure of a Likert scale format although it was explained to them in Laotian. Given accurate findings, Laotian parents may be experiencing somewhat more agitation or perturbation in their children than parents of the English speaking normative community sample. In support of these findings, Dinh et al. (1994) found that Vietnamese students consistently reported more problems in their relationships with their parents than did American-born Asian students.

Using data from two multi-state HMO's, racial differences were examined (see Table 6D). No statistically significant differences were found between races in these samples on the ID, S, SP, and CI subscales, and Y-OQ™ Total Score. Using a Bonferroni correction and an Alpha level of .05, the Asian (n=225) sample was found to be significantly higher than the Caucasian (n=6502) sample on the IR subscale. No other racial differences were found on the IR subscale. Furthermore, on the BD subscale, the Asian and Caucasian samples were found to score significantly higher than the Native American (n=110) sample, and the Asian sample was found to be significantly higher than the Latino (n=707) sample. As a cautionary note, sample sizes are quite small for some samples in these analyses, thus limiting the generalizability of the findings presented above.

Table 6D: Means According to Race

		Race									
		Asian		Caucasian		Latino		African American		Native American	
		N = 225		N = 6,502		N = 707		N = 41		N = 110	
		Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Scale	ID	24.4	0.86	23.5	0.17	22.7	0.50	20.8	1.80	21.2	1.35
	S	6.4	0.34	6.8	0.07	6.5	0.21	5.9	0.59	5.7	0.48
	IR	11.0	0.49	9.2	0.10	9.7	0.29	7.6	1.10	8.5	0.66
	SP	7.7	0.36	6.5	0.07	6.9	0.21	5.5	0.89	6.1	0.49
	BD	18.8	0.63	17.1	0.12	16.5	0.35	14.9	1.43	14.0	0.81
	CI	8.1	0.36	7.8	0.07	8.1	0.23	9.0	0.93	7.5	0.53
	Total	76.3	2.50	71.0	0.49	70.3	1.49	63.7	5.61	62.9	3.5

PSYCHOMETRIC PROPERTIES

Reliability

Internal Consistency

The reliability of the Y-OQ™ was estimated using Cronbach's alpha with the community ($n = 1091$) and clinical normative samples ($n = 2732$). As seen in Table 7, the total score on the Y-OQ™ had a remarkably high internal consistency estimate of .97 across the samples. The critical item and somatic subscales had the lowest internal consistency estimates of the six.

Table 7: Internal Consistency values for the Y-OQ™ Total and Domain Scores			
	Community	Patient	Total
	N = 1,091	N = 2,732	N = 3,823
Intrapersonal Distress	.90	.90	.93
Somatic	.66	.72	.74
Interpersonal Relations	.77	.85	.88
Social Problems	.63	.81	.84
Behavioral Dysfunction	.84	.89	.91
Critical Items	.63	.75	.77
Total Score	.95	.96	.97

Secondly, internal consistency reliability estimates were calculated for each of the normative samples (Burlingame et al, 2004). Estimates of internal consistency for the Y-OQ™ normative samples included a non-clinical sample drawn from a large elementary school ($N = 423$), the original community normative sample of 681 participants, and the clinical normative samples, consisting of outpatient ($n = 342$) and inpatient ($n = 174$), equaling 516. Estimates were made using Cronbach's alpha (Cronbach, 1970). In these samples, the Y-OQ™ total score demonstrated high internal consistency estimates of .94 across all four settings (Burlingame, Mosier et al., 2001). These estimates are similar to those stated above, and replicate the support given to the internal consistency of the total score. Subsequent studies (Girona, 2000) have supported these findings (i.e. internal consistency estimate = 0.95).

Additionally, separate subscale consistency estimates were calculated, and ranged from .51 to .90. Of the six subscales, the Critical Item and Somatic subscales showed the lowest internal consistency estimates (see Table 8A). These scales cover very broad content areas (e.g., diverse somatic complaints and equally diverse behaviors suggesting a need for inpatient treatment). Girona (2000) reports similar results. He hypothesizes that lower internal consistency for these scales reflects the diverse sets of behaviors tapped by items in these scales, which leads to item heterogeneity and more variability in scale scores.

Table 8A: Internal Consistency Values for the Y-OQ™ Total and Domain Scores					
		Internal Consistency			
		Student	Community	Inpatient	Outpatient
		N = 427	N = 681	N = 174	N = 342
Subscale	Intrapersonal Distress	.84	.90	.87	.88
	Somatic	.72	.68	.66	.70
	Interpersonal Relations	.69	.79	.79	.81
	Social Problems	.51	.71	.74	.76
	Behavioral Dysfunction	.85	.86	.85	.85
	Critical Items	.61	.63	.64	.70
Total Score		.93	.95	.94	.94

Finally, subscale internal consistency values were replicated using data gathered from two multi-state HMO's. Table 8B presents internal consistency statistics by treatment setting.

Table 8B: Internal Consistency Values for the Y-OQ™ Total and Domain Scores Across Treatment Settings						
		Internal Consistency				
		Community Normal	Outpatient	Inpatient Day Tx	Inpatient Regular	Residential
		N = 881	N = 12,807	N = 587	N = 629	N = 325
Subscale	Intrapersonal Distress	0.90	0.89	0.89	0.85	0.90
	Somatic	0.66	0.70	0.71	0.70	0.77
	Interpersonal Relations	0.78	0.82	0.83	0.82	0.86
	Social Problems	0.64	0.76	0.74	0.77	0.77
	Behavioral Dysfunction	0.84	0.86	0.89	0.87	0.90
	Critical Items	0.64	0.71	0.74	0.71	0.81
Total Score		0.95	0.95	0.95	0.94	0.96

Overall, the high reliability estimate of the total Y-OQ™ suggests a strong single factor underlying the six subscales of the instrument. These findings are particularly meaningful since the Total score is that most frequently used to track patient change. The presence of a strong single factor was also found on the OQ@-45.2 (Lambert et al., 2004).

Test-retest Reliability

Burlingame, Mosier et al. (2001) calculated test-retest reliability correlation coefficients from two separate community sub-samples. These samples were drawn from local elementary schools, in which retests were completed following two and four weeks. Table 9 illustrates a strong relationship between the first administration of the Y-OQ™ and scores taken at two-week ($r=.84$, $n=56$) and four-week ($r=.81$, $n=93$) intervals. Aggregated interval scores produced an average test-retest reliability coefficient of .83. Furthermore, all subscale test-retest correlations were significant at $p < .001$, ranging from .56 to .82. Both the total score and the separate subscales appear to demonstrate good to excellent test-retest reliability (Burlingame, Mosier et al., 2001). Notably, the lowest values

were associated with the CI scale. These results may be due to the low endorsement of these items (i.e., restricted range), or the small number of items.

Table 9: Reliability Estimates for the Y-OQ™ Using Normal Sample								
Time Elapsed Between Administrations		Subscale Scores						
		ID	S	IR	SP	BD	CI	Total
Two Weeks	N = 56	.82	.70	.75	.78	.82	.56	.84
Four Weeks	N = 53	.79	.67	.57	.71	.78	.65	.81
Total	N = 149	.78	.69	.66	.75	.79	.61	.83

Note. Scores are significant at $p < .001$

In general, internal consistency and test-retest reliability findings indicate that the Y-OQ™ has excellent reliability.

Frequency of Administration

Recent psychotherapy outcome research has indicated value and scientific justification for increasing the number of administrations of outcome measures (Bryk & Raudenbush, 1988; Willet, 1989). Willet asserts that increasing the number of administrations (“waves”) increases the reliability of growth rate data although repeated administrations may result in retest effects whereby raters respond with increasing carelessness (mechanical responding) or with social desirability motives.

Durham et al. (2002) investigated these issues employing a randomized block design. 172 subjects completed the Y-OQ™ in one of four experimental conditions (weekly, biweekly, monthly, pre-post) using an eight-week overall interval. For each of the experimental conditions, the slopes of the growth curves were plotted and found to show reliable decline (weekly: $= -1.49$, $t(32) = -4.5$, $p < .05$; biweekly: $= -1.16$, $t(25) = -3.95$, $p < .05$; monthly: $= -.88$, $t(31) = -2.67$, $p < .05$; and pre/post: $= -.62$, $t(80) = -3.04$, $p < .05$). These findings implicate a retest artifact with the Y-OQ™. Furthermore, the magnitude of the retest artifact increased with the administration schedule, as demonstrated by the growth curve slopes of the pre/post condition differing significantly from those of the weekly ($t(212) = -2.24$, $p < .05$) and biweekly ($t(212) = -2.16$, $p < .05$) conditions. Interestingly, Durham et al. report no evidence for social desirability or mechanical responding affecting Total Y-OQ™ scores. They note that the average retest effect approximated half the value of the RCI. Therefore, despite the statistical significance of the retest artifact value, it is not clinically significant. However, it should be remembered as one increases repeated administering. In summary, repeated administrations of the Y-OQ™ provide the most sensitive tracking process, but may also be affected by retest artifacts. We recommend that the Y-OQ™ be administered on a weekly or biweekly schedule, according to the discretion of the clinician, recognizing the influence of the retest artifact on patients’ scores.

Validity

Criterion-Related Validity

In establishing the validity of an outcome measure, it is necessary to demonstrate that the instrument measures what it purports to measure. One indication of a measure’s ability to report useful findings is how well the scores on the measure in question correlate with an established

measure of a similar construct. This is referred to as criterion-related validity. Two studies have compared Y-OQ™ total and subscale scores with parallel total and subscales scores from the Child Behavior Checklist (Achenbach, 1991) and the Connors Parent Rating Scale (Connors, 1990), two measures of behavioral difficulties in children and adolescents. Both studies included a community normal sample and a patient sample.

In the first study, the parents of a sample of 41 elementary children (age 6 to 12) were used as the community normal sample. In every case, the highest correlation found between subscales were those expected to converge (Table 10), indicating adequate convergent validity. Moreover, small and insignificant coefficients can be noted between dissimilar scales (e.g., somatic and aggressive behavior, antisocial and critical items) suggesting adequate divergent validity. Examples of convergent validity include the intrapersonal distress scale intended to tap anxiety and depression, which correlated highly with the anxious/depressed scales of both criteria. In addition, the highest correlation of the CBCL Withdrawn Scale was found with the Y-OQ™ intrapersonal distress scale. A modest relationship was found between the somatic scale and the criteria somatic complaint scales. Moderately high correlations were found between the interpersonal relations scale and criteria measures of aggressive or conduct disordered behavior. As expected, the highest correlations were found between the SP scale and measures of delinquent and antisocial behavior. While the behavioral dysfunction scale correlated highest with the criteria scales tapping attention problems. However, unexpected highly correlated values for Connors Conduct Disorder scale were also found on the Y-OQ™ subscales of intrapersonal distress, social problems, behavioral dysfunction, and critical items. Finally, a moderately high correlation was found between the Y-OQ™ and CBCL total scores.

Table 10: Criterion-Related Validity Estimates for the Y-OQ™ Using A Normal Elementary School Sample (N = 41)

Criterion	Y-OQ™ Subscales						
	ID**	S	IR	SP	BD	CI	Total
<i>Child Behavior Checklist</i>							
Anxious/Depressed	.71*	.31	.43*	.47*	.45*	.52*	.63*
Withdrawn	.48*	.20	.32*	.14	.21	.21	.37*
Somatic Complaints	.29	.40*	.27	.10	.28	.13	.34*
Aggressive Behavior	.44*	-.04	.61*	.52*	.56*	.42*	.55*
Delinquent Behavior	.31	-.04	.50*	.65*	.48*	.27	.45*
Attention Problems	.41*	.02	.54*	.44*	.54*	.36*	.51*
Total Problems	.67*	.34*	.67*	.52*	.66*	.51*	.75*
<i>Connors Parent Rating Scale</i>							
Anxious-Shy	.61*	.50*	.24	.25	.48*	.62*	.60*
Psychosomatic	.36*	.66*	.01	-.12	.14	.32*	.31
Conduct Disorder	.65*	.27	.61*	.62*	.71*	.64*	.76*
Antisocial	.18	.04	.20	.46*	.19	.18	.23
Hyperactive-Immature	.49*	.28	.29	.42*	.47*	.50*	.53*

**Note: ID = Intrapersonal Distress

CI = Critical Items

S = Somatic

BD = Behavioral Dysfunction

IR = Interpersonal Relations

SP = Social Problems

Tot. = Total Problems

$p < .05$

An inpatient sample of 80 was collected to examine the relationship between the Child Behavior Checklist and relevant Y-OQ™ subscales. As evidenced in Table 11, somewhat higher coefficients resulted in this study when compared with a sample of normal elementary aged children, suggesting that the normal school sample may have lower values due to restricted range. Several examples of convergent validity exist between the two instruments (Table 11). The highest relationships (mid to high 80's) exist between aggressive and delinquent behavior subscales of the CBCL and the interpersonal relations and social problems subscales of the Y-OQ™. More moderate relationships (.40-.48) exist between somatic and withdrawn scales of the CBCL. Finally, a strong relationship (.84) was found between the Total Y-OQ™ score and its counterpart on the CBCL. Overall, the findings from this study suggested that the relationships between the Y-OQJ and established criteria are very promising.

Table 11: Criterion-Related Validity Data from Patient Populations (N = 80)							
Criterion	Y-OQ™ Subscales						
	ID**	S	IR	SP	BD	CI	Tot.
<i>Child Behavior Checklist</i>							
Anxious/Depressed	.70*	.37*	.31*	.20	.51*	.57*	.64*
Withdrawn	.48*	.34*	.26*	.25*	.31*	.29*	.45*
Somatic Complaints	.33*	.64*	.04	.13	.12	.31*	.27*
Aggressive Behavior	.33*	.01	.79*	.62*	.73*	.36*	.64*
Delinquent Behavior	.22*	.13	.66*	.83*	.59*	.26*	.53*
Attention Problems	.50*	.25*	.39*	.36*	.70*	.50*	.62*
Total Problems	.64*	.34*	.67*	.57*	.78*	.61*	.82*

**Note: ID = Intrapersonal Distress

CI = Critical Items

S = Somatic

Tot. = Total Problems

BD = Behavioral Dysfunction

IR = Interpersonal Relations

SP = Social Problems

* $p < .05$

The results from the first study were replicated in the second study. In this study, the parents of a sample of 423 elementary children (age 5 to 12) were used as the community normal sample (see Table 12). The clinical sample ($n = 61$) consisted of children (ages 4-17) receiving treatment from a mental health agency in the western United States (see Table 13). As with the first study of criterion-related validity, correlations between measures were strongest for conceptually similar subscales. Lower correlations were found on dissimilar subscales. A moderately high correlation between the Y-OQ™ and CBCL total scores was also found with this study. Similar findings between this study and the first criterion-related validity study add further support to the convergent validity of the Y-OQ™. However, Burlingame, Mosier et al. (2001) caution that this data provides somewhat weak divergent validity for the Y-OQ™.

Table 12: Criterion-Related Validity Estimates for the Y-OQ™ Using A Normal Elementary School Sample (N = 423)						
Criterion	Y-OQ™ Subscales					
	ID**	S	IR	SP	BD	Total
<i>Child Behavior Checklist</i>						
Anxious/Depressed	<u>.70*</u>	.41	.42*	.19*	.42*	.60*
Withdrawn	<u>.64*</u>	.38*	.39*	.25*	.39	.56*
Somatic Complaints	<u>.44*</u>	<u>.61*</u>	.24*	.20*	.32*	.44*
Aggressive Behavior	.55*	.36*	<u>.63*</u>	.49*	.62*	.67*
Delinquent Behavior	.48*	.37*	.49*	<u>.58*</u>	.53*	.60*
Attention Problems	.48*	.36*	.41*	.32*	<u>.64*</u>	.58*
Total Problems	.73*	.57*	.61*	.42*	.69*	<u>.78*</u>
<i>Conners Parent Rating Scale</i>						
Anxious-Shy	<u>.67*</u>	.48*	.44*	.25*	.46*	.62*
Psychosomatic	.38*	<u>.62*</u>	.22*	.15*	.32*	.41*
Conduct Disorder	.60*	.39*	<u>.60*</u>	.44*	.61*	.66*
Antisocial	.35*	.25*	.42*	<u>.48*</u>	.41*	.41*
Hyperactive-Immature	.52*	.40*	.48*	.38**	<u>.67*</u>	.64*
Restless-disorganized	.60*	.46*	.48*	.31*	<u>.59*</u>	.64*

*p < .01

Table 13: Criterion-Related Validity Data From Patient Populations (N = 61)						
Criterion	Y-OQ™ Subscales					
	ID**	S	IR	SP	BD	Total
<i>Child Behavior Checklist</i>						
Anxious/Depressed	<u>.78*</u>	.48*	.45*	.27	.38*	.67*
Withdrawn	<u>.55*</u>	.35*	.37*	.25	.16	.45*
Somatic Complaints	.33	<u>.63*</u>	.21	.20	.19	.38*
Aggressive Behavior	.42*	.26	<u>.71*</u>	.62*	.57*	.65*
Delinquent Behavior	.41*	.25	.57*	<u>.60*</u>	.37*	.55*
Attention Problems	.43*	.25	.35*	.32*	<u>.57*</u>	.50*
Total Problems	.69*	.50*	.65*	.56*	.61*	.78*

*p < .01

Construct Validity

To establish the validity of a measure, researchers must demonstrate that scores reflect meaningful information regarding the construct the measure was originally designed to measure. This is referred to as construct validity. The construct validity of the Y-OQ™ is supported by a study comparing inpatient and outpatient scores on the Y-OQ™ with those of the community samples. Burlingame, Mosier et al. (2001) assumed that scores would be ordered from most pathological to least pathological. They demonstrated the community sample to be the least disturbed, followed by the community mental health outpatient, the private outpatient, with the

Y-OQ™ Administration and Scoring Manual

inpatient sample demonstrating the most disturbance. As predicted, these samples were found to be reliably different ($F(3,4080) = 680.9, p < .001$). However, post-hoc comparison indicated no significant difference between the average symptoms reported by the community mental health and private sector outpatients. Table 14 demonstrates the evident differences between clinical and non-clinical groups.

Table 14: Comparison of Level of Psychopathology as Measured by the Total Y-OQ™ Score Across Patient and Non-patient Samples			
Comparison Group	N	Mean	S.E.
Community Normal	683	23.9	6.77
Community Mental Health Outpatient	2881	76.6	9.21
Private Care Outpatient	342	76.3	8.89
Inpatient	174	110.8	8.74

Note: $F(3, 4080) = 680.9$ (significant, $p < .001$)

Children in the community sample were, on average, the most healthy, with inpatient children, on average, being the most severely disturbed. A fourth normative group, adolescents in residential treatment settings, was the focus of two different studies. In the first study, Gillman (1998) found that a sample of youth at a residential treatment center demonstrated significantly higher scores ($F(3, 3096) = 879.85, p < .001$) on the Y-OQ™ than other normative groups. Furthermore, Mosier et al. (2001) used the Y-OQ to evaluate a new in-home treatment program intended to replace inpatient treatment for children and adolescents. They reported that the mean Y-OQ™ score of initial severity for this sample to be significantly higher ($t(276) = 1.99, p < .05$) than the Y-OQ™ inpatient normative sample. Taken together, these studies indicate that residential treatment youth constitute a distinct treatment group in terms of severity and pattern of Y-OQ™ scores. Furthermore, the Y-OQ™ appears to discriminate between symptom levels of patients found in disparate treatment settings. Therefore, the Y-OQ™ appears to accurately measure a wide range of symptom levels, which is one aspect of the construct originally envisioned in the creation of the measure.

Sensitivity to Change

Sensitivity to change, specifically the change occurring as a result of treatment, is perhaps the most important property of an outcome measure. Establishing the Y-OQ™'s sensitivity to change helps further establish its construct validity. Since the Y-OQ™ was designed as an assessment tool to track patient progress and treatment outcome rather than diagnosing clients, its utility rests upon its sensitivity to change during and following participation in therapy or treatment. A logical criterion against which to compare the Y-OQ™'s sensitivity to change is subsequent administrations of the Y-OQ™. The Y-OQ™ has proven to be a stable measure over short time periods, which suggests that over brief periods of time (i.e. two and four week intervals) *without* treatment intervention, Y-OQ™ scores should remain relatively constant. Therefore, observed changes in Y-OQ™ scores over relatively brief time periods which exceed the magnitude expected by known instability coefficients (test-retest estimates), suggest meaningful changes in the pattern and level of a subject's symptomatology. To a great extent, the Y-OQ™'s construct validity depends on its ability to detect change following interventions such as psychotherapy. Therefore, it is expected that the scores of patients receiving psychological or psychopharmacological interventions should decrease over time. Unfortunately, there are several different means of operationalizing "sensitivity

to meaningful change.” The simplest is to calculate a pre-to-post-treatment difference score. The reliability of this difference score, calculated according to the method described by Allen and Yen (1979), provides a frame of reference to compare difference scores. This method essentially estimates the inflation in difference scores due to shared variance in pre-treatment and post-treatment observations (i.e., correlation between the two scores).

A second means of measuring sensitivity to change is by employing methods proposed by Jacobson and Truax (1991). By applying two statistical indices to the pre-post change, it is possible to classify clients as “recovered” or “significantly changed.” The first index assumes that following clinical intervention, a client’s outcome post-test score should be in the normal population range of functioning as opposed to the dysfunctional population range. The cut-off score of 46, described earlier, establishes the threshold between the functional and dysfunctional populations. The second statistical index, the Reliable Change Index (RCI), controls for variation in pre-to-post-treatment scores attributable to measurement error inherent in the test (known test-retest variability). When both of these indices are met, a client is considered to have demonstrated clinically significant change. Specifically, if the pre-to-post-treatment difference score is greater than the RCI value (13), *and* the post-test score has crossed the threshold between a dysfunctional and functional population, the client is said to have demonstrated clinically significant change.

In contrast to the RCI, Speer (1992) describes the Edwards-Nunnally (EN) method, involving centering confidence intervals of 2 standard errors of measurement on the client’s unbiased estimated or true initial score, as a third method of classifying client change. If the client’s post-score falls outside the confidence interval, it is considered significantly different from the pre-score at $p < .05$. The EN method is considered to be somewhat more conservative than the RCI method, due to its adjustment for regression to the mean by using the estimated true score. The formulas used in calculating each of these methods are listed below in Table 15.

Table 15: Formulas Used in Measuring Sensitivity to Change	
Method	Formula
Allen and Yen	$r_{DD'} = \frac{1}{2} (r_{xx} + r_{yy}) - r_{yx} / 1 - r_{yx}$
Jacobson and Truax	$RCI = x_A - x_L / S_{diff}$
Edwards-Nunnally	$x_L > \text{or} < [r_{xx} (x_A - M) + M] \pm 2 SD(1-r_{xx})^{1/2}$

Note. r_{xx} = internal reliability estimate of time one Y-OQ total scores; r_{yy} = internal reliability estimate of time two Y-OQ total scores; r_{yx} = the correlation between time two Y-OQ total score and time one Y-OQ total score; x_L = individual’s raw last assessment score; x_A = individual’s raw assessment score; S_{diff} = standard error of the differences between x_A and x_L ; M = population mean of total Y-OQ scores; SD = population standard deviation of Y-OQ total scores; r_{xx} = test-retest reliability of the Y-OQ.

With the purpose of directly comparing these three methods of detecting and interpreting change, Mosier, Burlingame, Nebeker, and Wells (1997) gathered data from an outpatient sample of 185 on the Y-OQ™. The sample included 103 males and 83 females ranging in age from 4-18 years ($M = 12.24$, $SD = 3.40$). The majority ($n=124$) were collected from an outpatient setting; however, the sample also included children and adolescents receiving day or residential treatment. Pre-to-post-treatment change scores followed a normal distribution ranging from -151 to 82. The average participant change score was 16.6 ($SD=37.8$), indicating overall improvement in the sample meeting the RCI criteria.

Mosier et al.'s (1997) pre- to posttest difference scores, when estimated using the method suggested by Allen and Yen (1979), yielded a reliability estimate of .78. This value is within the conventional range of acceptability, leading to greater confidence in drawing comparative conclusions using difference scores. While high difference score reliability is expected (given the aforementioned reliability of the Y-OQ™), it cannot be guaranteed. The importance of generating empirical estimates of such cannot be underestimated.

Although all methods support the Y-OQ™ with regard to sensitivity to change in the Mosier et al. (1997) sample, each produces a different profile of overall patient change. The Edwards-Nunnally method indicated that, of the 185 patients, 85 were improved, 22 unchanged, and 78 deteriorated. Using the Reliable Change Index, 91 subjects were classified as significantly improved, 59 as unchanged, and 35 as deteriorated. Actual frequency totals and percentage of hits and misses are shown in Table 16. Although there is a high rate of agreement between the two methods when analyzing which subjects are improving (93%), more discrepancy exists among those subjects who are classified as either remaining the same or deteriorating (45% and 45%, respectively).

Table 16: Cross-Tabulation of EN and RCI Classification Methods				
	Reliable Change Index:			
Edwards-Nunnally	Improved	Unchanged	Deteriorated	Total
Improved	(87%) 79 (93%)	(10%) 6 (7%)	**	85
Unchanged	(13%) 12 (55%)	(17%) 10 (45%)	**	22
Deteriorated	**	(73%) 43 (55%)	(100%) 35 (45%)	78
Total Table	91	59	35	185

Note. Percentages for EN values are on the left side and are read from top to bottom. Percentage values of classification for RCI method are to the right of the actual frequency count and are read horizontally.

The EN method classified 46% of Mosier et al.'s (1997) subjects as improved while the RCI method classified 49% improved. This finding conforms to literature-based expectations, with the majority of subjects exhibiting improvement while a portion of presumably more difficult or recalcitrant cases remain the same or deteriorate. The EN method classifies more subjects as deteriorators (i.e., worse off than at admission or initial administration) and fewer subjects as unchanged than the RCI method. However, this is primarily due to the asymmetrical confidence interval that is generated via the EN method and should not be interpreted to mean that the EN method is necessarily superior to the RCI method. Moreover, the findings of this study are in agreement with those reported by Speer (1992), giving greater confidence in both (i.e., replication).

Burlingame, Mosier et al. (2001) then used the RCI method to calculate the sensitivity to change of the Y-OQ™ in a larger sample of outpatient and inpatient youth, compared with a community normal sample. The samples consisted of 81 student normals, 675 outpatients and 174 inpatients. The sample ranged from age 4 to 17 ($M=11.17$, $SD=3.43$). Y-OQ™'s were administered at the third, fifth, and tenth sessions of treatment. Using the Cutoff of 46 and the RCI of 13, Burlingame, Mosier et al. (2001) designated 17% of their clinical sample as recovered (falling below the Cutoff while changing more than 13), 37% as improved (symptom scores improving by more than 13), 31% as unchanged, and 15% as deteriorated (symptom scores increasing by more than 13). The reported an average change between pre- and posttest scores for the clinical sample of 17.7 points ($SD=34.8$), while the student normal sample difference score mean was 4.34 ($SD=2.48$). This difference was significant ($t(927)=3.44$, $p<0.001$), indicating that the Y-OQ™ is able to reliably detect changes in

symptom levels across short periods of treatment, and supporting the conclusion that the Y-OQ™ demonstrates excellent construct validity.

An additional study supports the sensitivity to change of the Y-OQ™ with a wilderness therapy treatment population. Hoag, Savicki, and Burlingame (August 2001) investigated the effectiveness of a wilderness therapy program employing the Y-OQ™ as a dependent measure of symptomatic levels in adolescents enrolled in the program. Phase one tracked 277 adolescents (201 male, 76 female) who completed wilderness therapy and a one year follow-up period. At intake, the average Y-OQ™ score for their sample was 101. Results indicated that adolescents' symptom levels significantly decreased during the program, and continued to decrease over the one-year follow-up. Furthermore, they discovered that males and females changed at similar rates, and age was not an influential factor in change. This preliminary study indicates the similarity between wilderness therapy populations and residential treatment populations, as well as the usefulness of the Y-OQ™ at tracking clinical change over the course of treatment.

Additionally, Berrett (2000) further investigated the Y-OQ™'s construct validity by examining sensitivity to change at the item level. This analysis was based upon a large sample of clinical children and adolescents (N = 729) gathered from four different sites (ranging from outpatient to residential and inpatient settings) as well as a community sample of non-treated children and adolescents. These subjects all completed the 64-item Y-OQ™ (Berrett, 2000).

Individual item slopes, estimating the amount of change occurring over time, were generated via hierarchical linear modeling and compared within and between the various populations described above (i.e., community versus different clinical groups). In contrasting the combined clinical samples and the community normal sample, 27 of the 64 Y-OQ™ items were found to have slopes where the clinical sample demonstrated significantly more change (in the direction of improvement) than the non-clinical sample. The majority of these items are now found in a shortened version of the Y-OQ™ (Y-OQ™- 30). When the individual item slopes are regressed to form an average slope (or an average change estimate), this average item slope is also greater for the clinical than the community sample (Berrett, 2000).

Item slopes were also compared between different levels of clinical treatments with the hypothesis being that more intensive levels of intervention would yield differential change patterns with more change being seen with more intensive interventions. This hypothesis was partially supported as 45 of the 64 original items had larger slopes in residential than outpatients (i.e., more intensive treatment produced greater change). There were not significant differences between the patterns of change in outpatients and day-treatment patients (Berrett, 2000).

Sensitivity to change is the most important characteristic of a treatment outcome instrument. Ideally, any demonstrated change in scores would represent actual behavioral change. In actuality, measurement error, regression to the mean, and other possible confounding variables require statistical techniques to determine how much change in recorded scores is necessary to validly demonstrate that meaningful change has occurred in behavior. Overall, the findings of the Mosier et al. (1997), Burlingame, Mosier et al. (2001), and the Berrett (2000) studies indicate that change is occurring in the hypothesized direction. These findings support the Y-OQ™ as a measure that is sensitive to the changes occurring in the pattern and level of subjects' symptomatology measured over the course of treatment, contributing to the construct validity of the measure.

Sensitivity to Psychopathology

Support for the construct validity of the Y-OQ™ was also sought by comparing inpatient and outpatient scores on the Y-OQ™ with those of the community samples (see Table 17). It was assumed that statistically significant differences between the means of the clinical and normative samples would suggest that the Y-OQ™ can distinguish between these groups. Further, it was expected that the mean scores for the groups would be ordered from the most pathological to least pathological. We expected the inpatient sample to be most disturbed, followed by the outpatient, and then the community sample. A one-way ANOVA was conducted to determine the difference between sample means. Results comparing the three samples were significant at the .001 level. T-tests were conducted following the ANOVA for post-hoc comparisons as well as to quantify the differences between the three samples.

Table 17: Comparison of Level of Psychopathology as Measured by the Total Y-OQ™ Score Across Patient and Non-patient Samples			
F (2,1196) = 591.4 (significant, $p < .001$)			
<i>Comparison Group</i>	N	Mean (S. D.)	t-Value (D.F.)
Community	683	23.2 (26.6)	t = 23.9*
Outpatient	342	78.6 (36.4)	
Outpatient	342	78.6 (36.4)	t = 4.16*
Inpatient	174	100.0 (40.37)	

* $p < .05$

The data in Table 17 clearly suggest that the Y-OQ™ reflects pathology in line with expectations. That is, there were clear differences between clinical and non-clinical groups. The children in the community sample were, on average, the healthiest. The inpatients on average were most severely disturbed.

Furthermore, Gray et al. (2004) report that the Y-OQ™ can be useful in identifying potential for juvenile offenders in the justice system of Utah. In a sample of 719 adolescents from 12 to 17 years of age, who had been entered into the Juvenile Justice system of Utah, the Cutoff scores of the Y-OQ™ classified 49% of the adolescents above the Outpatient Cutoff. Furthermore, 17% of this sample was similar to the inpatient normative group. Additionally, the CI, ID, and SP subscales for these youth were significantly different than the community normal scores (CI=5.6 vs. 3.3, ID=17.9 vs. 10.8, SP=8.31 vs. 1.1). Finally, Gray et al. (2004) indicate that recidivism in the Juvenile Justice sample was directly correlated with Y-OQ™ symptom severity, indicating that the higher the Y-OQ score was, the more likely the subject was to have repeat offenses. Replication of these findings is underway in other states across the country. These findings indicate that the Y-OQ™ total and subscale scores are useful tools at evaluating the amount of distress a child or adolescent is experiencing, which contributes to the conclusion that the measure demonstrates construct validity.

Sensitivity and Specificity

Sensitivity and specificity further establish the construct validity of a measure. Sensitivity refers to the proportion the “true positives,” members of the clinical groups, correctly identified. Burlingame, Mosier et al. (2001) reported the sensitivity of the Y-OQ™ to be .82. Approximately 82% of the true members of clinical groups were properly classified as clinical, while 18% were placed in the normal group (misclassified) using a cutoff score of 46.

Specificity refers to the proportion of the “true negatives,” or members of the community normal group, correctly identified. Burlingame Mosier et al. (2001) discovered the specificity of the Y-OQ™ to be .89, indicating that 89% percent of the members of the normal group were correctly placed with a cutoff score of 46.

Other operating characteristics include Positive Predictive Power (PPP) =.77 and Negative Predictive Power (NPP) =.81. The sensitivity, specificity, and PPP values were further supported in a study conducted by Atkin (2000). Performing a receiver operating characteristic (ROC) analysis, Atkin determined the sensitivity (true positives) and the false positives produced by a measure on the Y-OQ™. ROC analyses plot curves which indicate the sensitivity of measures. The area under the curve (AUC) demonstrates a measure’s predictive ability. Atkin (2000) indicated that, with AUC values up to .94, the Y-OQ™ accurately discriminated between normal and clinical samples. Furthermore, with a population of children and adolescents sampled at their primary care physician’s office, Anthony Tzoumas (2001) found that the Y-OQ™ total score correctly classified 85% of his sample into their correct treatment categories. Finally, Burlingame, Mosier et al. (2001) performed ROC analyses for each of the subscales of the Y-OQ™. Their results are presented in Table 18.

Table 18: AUC Values for Y-OQ™ Total and Subscale Scores			
Scale	AUC	SE	95% CI
Total	0.94	0.01	0.91-0.96
BD	0.94	0.01	0.91-0.96
ID	0.91	0.02	0.88-0.95
CI	0.86	0.03	0.80-0.91
IR	0.85	0.03	0.79-0.90
SP	0.84	0.03	0.78-0.90
S	0.81	0.03	0.74-0.87

With AUC values ranging from .81 to .94, the subscales of the Y-OQ™ demonstrate excellent sensitivity and specificity for the subscales of the measure as well, contributing further support to the construct validity of the measure.

These sensitivity and specificity values are similar to those obtained by the OQ@45.2 (Lambert, et al., 2004) and provide an index of the accuracy of the Y-OQ™ as a screening tool.

It is important to note that, while the sensitivity and specificity are values are clearly useful at the levels obtained, some consideration of the measurement process itself may indicate why they are not even higher. By definition, we have considered the children of the community sample as normal and the children from the patient samples as abnormal. Yet, parents have individually defined thresholds of sufficient concern to seek treatment for their child; therefore, some children who, in the opinion of most professionals should be receiving treatment, are not receiving help (the “false negative”

child) because their parents have not chosen to refer them. Other parents may refer their child for much less serious difficulties, even prophylactic reasons such as divorce adjustment (the “false positive” child).

BASIC INTERPRETIVE STRATEGY

The Y-OQ™ was designed for two general purposes: first, to measure the outcomes of mental health treatment by repeated administrations, and second, to screen potential clients with the purpose of establishing need for treatment or alert mental health professionals to the need for more careful evaluation.

Parents complete the Y-OQ™ at intake. Scores may then be employed at three different levels to determine the need for treatment. The most conservative procedure is to first examine the Y-OQ™ Total severity elevation. Cronbach’s alpha coefficients for the Y-OQ™ indicated that the test can best be understood as having one main factor underlying all the subscales. By comparing the individual’s score to the current normative groups, the clinician may determine whether the child’s scores suggest that he or she is in need of treatment. Clinicians may further discover what level of treatment will most likely be warranted--outpatient or a more restrictive environment.

As noted above, recent studies have shown that the Y-OQ™ total score can be used to correctly classify patients into their appropriate treatment settings (Burlingame, Mosier et al., 2001; Tzoumas, A., 2001; Atkin, 2000; Gillman, 1998). Moreover, a recent modification of the Y-OQ™ (Y-OQ™-12, see below) shows promise as a screening tool for primary care settings. It must again be emphasized, however, that the Y-OQ™ should not be used as a diagnostic tool, but rather as a screening measure.

The second and third levels of interpretation speak to the issues of “. . .how wrong and where to look.” The clinician considering the Y-OQ™ scores may next examine any subscale elevations, particularly the Critical Items (CI) subscale. CI questions were written specifically to assess concerns critical for increased protective actions, such as inpatient hospitalization (e.g., “Sees, hears, or believes things that are not real,” “Has times of unusual happiness or excessive energy,” etc.).

Further evaluation of both subscale elevations and individual items aids the clinician at the third level of evaluation in generating hypotheses about the most effective treatment protocol. For example, frequent endorsement of Critical Items, such as “Believes that others can hear her/his thoughts, or that s/he can hear the thoughts of others” may indicate the need for a psychiatric evaluation for medication as opposed to parent training.

The Y-OQ™ is designed to be brief enough for repeated administrations to parents. Scores can be plotted on a “Y-OQ™ Tracking Sheet” to monitor the parent’s report of perceived changes over the course of treatment. When the Total score falls below the community normal cutoff score of 46, the clinician knows that the parent sees the child as behaving “within normal limits” and can adjust treatment accordingly (e.g., work to consolidate gains, prepare for termination, etc.). Erratic scores suggest to the clinician the need for careful evaluation of the treatment protocol, the introduction of second-level treatment, the influence of risk factors that mitigate against rapid treatment response, etc. Brown et al. (2001) demonstrated that clinicians can employ regression analysis using the patient’s intake Y-OQ™ score and a change score from repeated administrations to identify those patients most at-risk of premature termination. These patients may then receive more intensive treatment, increasing their chance of positive outcomes.

Bishop et al. (2005) further advanced the ability of clinicians to identify potential treatment failures early by testing rationally derived algorithms based upon the RCI, cut scores and normative

change data. Change scores on the Y-OQ™ may be used to identify trajectories of change in clients with different levels of distress and who are treated in different settings. Bishop et al. report that the algorithm correctly classifying 77.3% of deteriorators and 81.7% of improvers from therapy using an archival data set (N=300). The computational intensity of this method is eliminated by the using of the aforementioned computer software solution—OQ®-A. By identifying which clients are in danger of treatment failure, the clinician can then adjust treatment or employ different intervention strategies to intercede before the client withdraws from therapy or deteriorates further (Bishop et al., 2005).

USE OF THE INSTRUMENT FOR TREATMENT PLANNING

General Treatment Planning Issues

Given that published literature on the Y-OQ™ has not addressed many issues related to treatment planning, the following should be seen as possibilities for clinical practice. It is equally clear, however, that the same statement could be applied to the entire topic of treatment planning in child and adolescent psychotherapy. Treatment planning for child and adolescent psychotherapy is a topic without robust literature support, and the Y-OQ™ literature reflects this lack. While significant advances have been made, Kazdin (1995) asserted that, in child and adolescent treatment, “. . . research is at an early stage in relation to the range of clinical dysfunctions that has been systematically studied and the types of research questions that are addressed” (p. 125). Fonagy (1996) similarly noted that, at this point, research should direct efforts toward studies of “ecological” relevance, with somewhat less emphasis on those designed for maximal internal validity. Therefore, the following should be viewed as possibilities for clinical practice.

The Y-OQ™ may be employed as a means of tracking parents’ perceptions of their child’s change. Similarly, watchful clinicians or treatment teams may, by repeated administrations of the Y-OQ™, observe the effects of additional interventions or environmental impacts, such as psychotropic medications.

Clinically, Y-OQ™ administrations may be seen as a structured anecdotal report from a parent (M. Latkowski, personal communication, October, 1997), which decreases the time the clinician must take to ask the parent about the child’s progress. The first administration provides a baseline level from which treatment effects can be monitored. When scores fluctuate in unpredictable directions, the therapist can often learn about other external stressors which wouldn’t otherwise be discovered but may need to be addressed therapeutically. Latkowski has also found that clinicians who carefully review the Y-OQ™ tracking sheet tend to shift toward therapeutic impact questions as opposed to diagnostic questions. In other words, with the parents providing the clinician with treatment targets, therapists spend less time in evaluation, and move more directly to the question, “What needs to be done to address this difficulty?”

In the managed health care corporation that assisted with the development of the Y-OQ™, child psychiatrists have employed it as a “need-for-treatment monitoring device,” (R. Ferre, personal communication, March, 1997). In the practice of child psychiatry, once a child has reached a stable, therapeutic dose of medication, a portion of clinicians’ efforts is devoted to quarterly check-ups for medication evaluation. These physicians have clinic secretaries regularly mail out Y-OQ™ protocols. When Y-OQ™’s are received back, they are scored and the results are entered on the Y-OQ™ tracking sheet. This practice of mailing Y-OQ™’s to patients’ caretakers does not represent an ecological validity problem since the Community Normative samples were obtained through the

same method. Most Clinical Normative samples for the Y-OQ™ were obtained in the waiting room; however, it is not known if there are any differences in clinical level scores obtained on the measure when it is mailed compared to waiting rooms. Future research may focus on this area. A computerized administration and scoring system is available and is designed to streamline the process of entering and scoring data from the Y-OQ™, significantly decreasing the time therapists would require to receive feedback from an administered measure, as well as facilitating the transfer of data to health care corporation headquarters. In the capitated health care system, the physician's timely review of tracking sheets may inform him or her about the continuing efficacy of treatment or alert him or her to the need for a new evaluation, intensive focus on patient retention, resumption of more intensive treatment, or further treatment modifications (Brown et al., 2001).

Applications to Treatment Planning Issues

Identification of Primary and Secondary Problems

As indicated above, the Y-OQ™ was not created as a diagnostic instrument. Existing standardized diagnostic instruments such as the Child Behavior Checklist (Achenbach, 1991), and the Personality Inventory for Children (Lachar, 1982) are more appropriate for that purpose. Conceptually, however, the clinician may examine subscale patterns of the instrument to glean major presenting problems expressed by parents. So-called "broad-band" difficulties with externalizing or internalizing behavior or potential psychotic spectrum difficulties may become apparent. Individual item responses could spark further questioning or evaluation in such arenas as eating disorders (e.g., "Has lost significant amounts of weight without medical reason") or somatization (e.g., "Complains of stomach pain or feeling sick more than other children of the same age"). Some clinicians have used extreme item or subscale scores as treatment targets. This practice, however, should not be relied upon, since the nature of the symptom(s) expressed and the child's parents' concern may not match the measure. Diagnosis of a primary clinical problem may not be congruent with the problem most important to treatment. It is possible, for instance, that in terms of long-term success of the ADHD child, his or her social skill deficits are much more deleterious than poor study and homework skills, the more usual focus of concern.

Appropriate Level of Care

As indicated previously, Total score elevations differ significantly between inpatient and outpatient clinical samples on the Y-OQ™. In a rough clinical sense, the clinician employing the Y-OQ™ can say that a Total score of 110, for instance, is most comparable to children being treated on an inpatient basis. Similarly, high scores on the Critical Items subscale indicate possible need for a more protected or restricted treatment environment, such as inpatient or residential treatment. However, recent studies (Gillman, 1998; Mosier, 2001) provide cautionary notes to this practice, demonstrating that Y-OQ™ subscale scores from youth in residential treatment or in-home settings differed significantly from community and outpatient norms, and showed significantly elevated subscale scores compared to inpatient samples.

Potential Use or Limits for Treatment Planning in a Managed Care Setting

Many managed care organizations are employing outcomes assessment to a greater degree than ever before. Brown et al. (2001) illustrate the practical CQI application of the Y-OQ™ and its related measures in a large managed care setting. PacifiCare Behavioral Health, Inc. (PBH) invites each of its providers to regularly administer a shorter form of the Y-OQ™, the Youth Outcome Questionnaire 30 (Y-OQ™-30), to a parent of each patient (See Related Instrument section below for description of Y-OQ™-30). The instrument is completed at regular intervals, thus tracking

therapeutic progress. In the office, the measure is faxed to a central PBH location, where it is scored and entered into PBH's database. PBH then creates a feedback loop with the patient's therapist, informing him or her of the client's progress. The Y-OQ™ score can then be entered into the client's chart. The therapist employs the measure as: (a) an intake measure of initial severity of symptoms and index of risk factors which moderate expectations for rapid improvement; (b) a tracking device of therapeutic change; (c) a warning system to alert the therapist of the danger of premature termination and the need for more intensive intervention; and (d) a potential summary source for demonstration of effectiveness of therapeutic interventions. The OQ®-A software (www.oqmeasures.com) provides similar automated functionality making scores and algorithmic predictions immediately available to a clinician or practice group.

Clinicians whose practice is independent of third party payers may also find this process effective. For example, one therapist created a data set of Y-OQ™ scores for the patients in his private practice (Asay, Lambert, Gregersen, & Goates, 2002). Examining the individual recovery curves, plotted from repeated measures of the Y-OQ™ for each child and adolescent ($N = 40$), he tracked the effectiveness of treatment. Data showed that 75% of his sample met requirements for reliable change following 12 sessions. Asay et al. assert that clinicians could compare individual improvement to the clinician's average client recovery, as well as to national norms, by creating a database of their patients' scores, thus assessing client change and therapeutic effectiveness.

Corporations can further employ analyses of repeated Y-OQ™ and Y-OQ™-30 scores for: (a) reporting therapeutic effectiveness to subscriber companies and for profiling individual providers; and (b) answering further research questions, such as evaluating the efficacy of innovative treatment approaches. The Y-OQ™ has been shown to be effective in this area. Hoag, Lambert, Jenkins, Hyde, Lindsay, and Harvey (1998) used the Y-OQ™ to assess the effectiveness of an experimental treatment program for delinquent and ungovernable youth. In their study, 87% of the experimental group, compared to 56% of controls, demonstrated recovery (scores below the community normal cut-off of 46) or improvement (scores lower than pre-test by 13 or more points) on Y-OQ™ post-test scores. They showed that the Youth Reclamation Incorporated treatment protocol produced more clinically significant change and was more cost-effective than the standard court-ordered residential treatment for ungovernable youth. Similarly, studies mentioned earlier indicate usefulness of the measure as a treatment evaluation instrument in various settings such as an in-home service program designed to be an alternative to inpatient treatment for severely disturbed children and adolescents (Mosier et al., 2001), outdoor behavioral healthcare programs (Russell, 2003, 2002), and a partial day treatment program for children with severe emotional disturbances (Robinson, 2000). Each study supports the appropriateness of using the Y-OQ™ to evaluate the usefulness of clinical programs.

USE OF THE INSTRUMENT FOR TREATMENT MONITORING

Purpose of Treatment Monitoring

Treatment monitoring provides ecological relevance for researchers and even clinicians. Furthermore, Treatment monitoring answers in standardized format the question, "How is the patient doing at this stage of treatment?" Fonagy (1996) asserts that monitoring is essential to bridge the gap between academic research and actual clinical practice:

Monitoring the process of a service goes hand in hand with routine monitoring of outcomes of clinical practice.... Where outcomes are poorer than anticipated on the

bases of research findings, and the discrepancy is not accounted for by the deviations from laid down standards of performance, the monitoring process has provided further research questions about the essential components of treatment or patient characteristics which place limits on treatment effectiveness. In an ideal world further theoretical and clinical development follow, leading to research which in turn addresses the shortcomings of the treatment protocol. (p. 39)

Administrators of health care organizations find treatment monitoring equally essential to assist decision making regarding whether specific patients benefit from particular services, as well as what the duration and intensity of treatment should entail. Research such as that conducted by Hoag, et al. (1998), Mosier et al. (2001), Robinson (2000), Russell (2002; 2003), and Brown et al. (2001) demonstrating the effectiveness of the Y-OQ™ in assessing the usefulness of clinical programs; as well as studies demonstrating the ability of the Y-OQ™ to correctly classify patients into their proper treatment settings (Gillman, 1998; Tzoumas, A., 2001) lay the groundwork for further applications of the Y-OQ™ to treatment monitoring efforts. Echoing Fonagy's vision of treatment monitoring, Burlingame et al. (1995) urge administrators to see treatment monitoring as an informational process through which the organization, as well as providers, can calibrate the treatment process and, therefore, its "products."

Families may also benefit from treatment monitoring. They, too, are asking, "How is my child doing?" In the absence of more formalized information, parents often rely on their impressions of the recent past. This may limit their ability to track gradual but substantial changes that may have occurred in their child. Clinicians report increased consumer satisfaction in a child psychiatry practice when they routinely share the Y-OQ™ tracking charts with the parents of clients.

How to Use the Instrument for Treatment Monitoring

Practitioners who have integrated the Y-OQ™ into their practices usually ask parents to complete the questionnaire at each session. As described above, they reportedly interpret the scores at three levels: First, the overall severity and/or response to Critical Items indicates the parents' sense of crisis or agitation with their child. Even during ongoing treatment, the occurrence of three data points that suggest increasing behavioral problems should incite the clinician to query the parents in an effort to determine the source of negative change. Second, initial perusal of subscales frequently alludes to the general therapeutic thrust required; i.e., higher levels of externalizing behaviors require a different therapeutic protocol than higher levels of internalizing behaviors. And third, following significant progress with the child, the clinician is alerted to consider less restrictive treatment alternatives.

USE OF THE Y-OQ™ FOR TREATMENT OUTCOMES ASSESSMENT

As indicated previously, the Y-OQ™ was designed specifically for the purpose of tracking outcomes. Below, we discuss some advantages and limitations of the Y-OQ™ for this purpose. It is important to note that the Y-OQ™ in its present form addresses outcomes only from the point of view of the patient's primary care taker. Children are not considered reliable informants of their behavioral and mental states. Furthermore, adolescents are commonly coerced into treatment, and are notoriously poor informants. As noted earlier, such Y-OQ™ forms have been developed. For example, the Y-OQ™-30 is a shortened version of the Y-OQ™ developed by Burlingame, Jasper et al. (2001) to provide a quick assessment of the symptomatic status of an adolescent, from their own

perspective. This sampling of the client's own experience is a rich potential data source for clinicians. For example, adolescent reported change on the Y-OQ™-30 co-varies highly with parent-reported change on the same instrument, despite absolute values of symptom distress differing (Russell, 2003).

Naturally, employing only the original parent-completed Y-OQ™ may limit the generalizability and usefulness of outcome data since these data come from a single source (i.e., the parent or guardian). Additional factors affect the value of outcome data generated by significant others' ratings as well. For example, outcome data collected on a weekly from basis different informants (e.g., mother, father, an older sibling, etc.) bringing the child to weekly sessions may show large differences between scores from informants because of the amount of information they possess, or their response bias toward the child. Therefore, the outcome data collected may not be meaningfully interpreted. Future research could focus on differences in scores found by different caretakers completing the measure. This source of variance has not been systematically studied yet. Similar problems exist within inpatient treatment centers. In these settings, parents may not have enough contact to adequately rate the child, and no one particular inpatient staff member may have spent enough time with the child to accurately rate the full range of behaviors sampled by the Y-OQ™. Thus, the development of other versions of the Y-OQ™, such as the Y-OQ™-30 self-report form, significantly increase the availability of data collected from different sources addressing the efficacy of treatment. This solution directly addresses the problems associated with outcome research with youth.

These problems as methodological issues are inherent in all data that are collected from significant others, and therefore not unique to the Y-OQ™. However, it is important to remember that these, and related short-comings, will influence conclusions regarding the effectiveness of treatment interventions as measured by the Y-OQ™.

Use of Findings from the Y-OQ™ with Other Evaluation Data

To date, the Y-OQ™ has rarely been used simultaneously with other measures of treatment outcome. It is therefore difficult to compare outcomes across measures for the purpose of identifying measures that produce larger or smaller effect sizes. However, one study described earlier (Gironda, 2000) employed the Y-OQ™ as the reference criterion against which a newer outcome measure was validated. Gironda calculated the concurrent validity for the Y-OQ™ and the OQ® to be .76 ($p < .05$). As stated previously, similarly high correlations have been found between the Y-OQ™ and other outcome measures such as the CBCL and Connors (Burlingame, Mosier et al., 2001; Atkin, 2000). Future research should compare differential sensitivity to change among these outcome measures, explicating whether one measure is more capable of detecting change in particular populations.

Provision of Feedback Regarding Outcomes Assessment Findings

Feedback information based upon scores from Y-OQ™ administrations can be utilized in a wide range of applications. Frequently, parents ask about the purpose of the measure and their personal results. Clinicians are left to determine their own course of action in these circumstances. They may choose to include a full disclosure of the results. Such an inquiry is essentially the equivalent of a patient's parents asking the question, "How is my son doing...is he getting better?" These should be handled on an individual basis. Charting the progress of a specific client may also be quite informative to a clinician and even provide validating feedback as to therapeutic setbacks, stagnation, or rate and pattern of progress (Brown et al., 2001; Asay et al., 2002).

The most meaningful feedback for a clinician and third party providers is typically the aggregate scores of clients and sessions. Once Y-OQ™ results have been accumulated across multiple clients and sessions, the resulting data may provide critical feedback on the progress of patients, typical patterns of improvement for the patients of different clinicians, and the effectiveness of treatments found in various, hospitals, and regions. The computerized administration and scoring program provides one method of accessing this data. Clients can take the Y-OQ™, and the software provides tabled results describing clinician or clinic efficacy in terms of percentage of clients improved and/or recovered across the number or sessions used.

Brown et al. (2001) opines that outcomes are difficult to compare and interpret without a valid statistical method of accounting for variations in severity and difficulty of cases. Case-mix adjustments provide equivalent findings when comparisons across hospitals, clinics, programs or providers are made. Case mixing is accomplished through a variety of methods on any number of relevant variables (e.g., initial severity scores, patient diagnosis, chronicity of disorders). The goal of each method is to balance or match comparisons based on the patient's disorder or severity of illness. Brown et al. (2001) demonstrated that employing a patient's initial score on the Y-OQ™-30 (see Related Instruments section below) as one effective method of case-mix adjustment. In their study, a large health maintenance organization used this method to compare treatment outcomes and provide therapists with feedback regarding patient progress and outcome.

Limitations in the Use of the Y-OQ™ for Outcomes Assessment

Perhaps the biggest limitation of the Y-OQ™ remains the area of implementation. As with all outcome measures, the most difficult challenge with the Y-OQ™ is the systematic collection of data from informants across clinical sessions with the goal of interpreting the overall pattern of change. It is possible that the problem of missing data will be resolved to some extent as more and more HMOs require such data from their health care provider panels. As mentioned above, Brown et al. (2001) illustrated the use and efficacy of one such corporate application in a large, western behavioral healthcare corporation. Their study demonstrates initial efforts to implement treatment planning issues using outcome assessment. Furthermore, Brown et al. report that PBH plans to apply the program throughout their organization in the near future, thus confirming the usefulness of the program.

Psychometric limitations of the Y-OQ™ include evidence of a retest artifact (Durham, et al., 2002), as well as an absence of factor analyses for support for the rationally derived subscales (Hughes, Lambert, Burlingame, & Wells, 2000). These limitations are offset by strong support for reliability and validity as well as the instrument's demonstrated sensitivity to change.

ADDITIONAL VERSIONS AND RELATED PRODUCTS OF THE Y-OQ™

Self-report Version

An adolescent self-report version of the Y-OQ™ has been developed called the Youth-Outcome Questionnaire™ Self-report (Y-OQ™SR; Wells et al., 1996).

Shortened Version

Clinicians often request a shorter form of the Y-OQ™, which can quickly and accurately gauge the current symptom level of a child or adolescent. Based upon this feedback, Berrett (1999)

investigated items on the Y-OQ™ that were differentially sensitive to change. This research laid the foundation for the 30-item instrument (Y-OQ™-30) initially developed under the proprietary label Youth Life Status Questionnaire (Y-LSQ) (Burlingame, Jasper et al., 2001; Brown et al., 2001). The Y-OQ™-30 provides a single total score without the original Y-OQ™ subscales. The preliminary parent and adolescent self-report forms were collapsed into a single omnibus instrument eliminating the use of two forms. Furthermore, when used by parents of adolescents, the initial psychometric properties of this omnibus form are promising. Internal consistency estimates, based upon parent- and self-report, approximate that demonstrated by the longer version (.96 and .93 respectively). Moreover, sensitivity and specificity for both parent- (.80 and .75 respectively) and adolescent self-report (.70 and .63. respectively) are good. Early research supports the ability of the measure to capture patient change in a manner that is comparable to the parent instrument – Y-OQ™.

Y-OQ™-12

The Youth Outcome Questionnaire-12 (Y-OQ™-12), another modification of the Y-OQ™, has shown success as a screening tool in primary care settings. A. Tzoumas (2001) hypothesized that the Y-OQ™'s high positive predictive power could assist pediatricians in identifying children in need of mental health treatment. With this goal in mind, Tzoumas reported that total Y-OQ™ scores correlated with high utilization of primary care. He also identified twelve items from the original Y-OQ™ that most accurately discriminated between community and clinical samples. These twelve items formed the Y-OQ™-12, a quick, psychometrically sound, screening tool for use in physicians' offices. In subsequent research, J. Tzoumas (2001) reported that the Y-OQ™-12 demonstrated adequate internal consistency (.87). The new measure showed good sensitivity, ranging from .77 to .99, and adequate specificity (.62 to .77), with values varying by clinical setting. These initial statistics indicate that the Y-OQ™-12 exhibits promising psychometric properties, and may therefore aid clinicians in quickly assessing mental health treatment needs in a primary care setting. However, the Y-OQ™-12 has not shown sensitivity to change, and is not applicable as a short measure of treatment progress and outcome.

Software

The OQ®-Analyst (OQ®-A) is a software program that provides automated administration and scoring for the Y-OQ™ (www.oqmeasures.com). Automation of outcomes assessment has shown to significantly reduce staff time (i.e., administering and scoring the tests) and increase the use of outcomes information by clinicians. The OQ®-A calculates patient change and compares progress with expected change trajectories alerting clinicians with real time alerts for clients whose outcome is in doubt. The monitoring system was built through studying the progress of clients who had weekly sessions of therapy. This progress information was then used to predict the kind and amount of change that is typical of those who eventually failed to respond to treatments. Research clearly indicates that providing this kind of information to therapists will improve therapy outcomes.

CONCLUSION

Although further research is essential and continues, to date, evidence suggests that the Y-OQ™ holds considerable promise. The measure requires little time to complete, is easy to administer, and is available at nominal cost. It has been translated into Spanish, French, German, Hmong, and Laotian; and ethnic group comparisons are underway. Thus far, it has shown acceptable-to-excellent

reliability and validity, demonstrated sensitivity to change, and been found to be useful both by clinicians and health care administrators.

Currently, analyses indicate that, like the OQ®-45, interpretation of the Total score rather than subscale scores, is the most justifiable tracking procedure. However, subscales and individual item responses may be employed for hypothesis generation as with any other objective measure. Recent measure refinements include confirmatory factor analyses, the establishment of norms for more specific treatment populations (residential treatment), parallel self-report versions for adolescents, and a primary care physician screening version. Each of these developments has increased the support and utility of the Y-OQ™ as a clinical, research, and administrative tool. We anticipate revisions of the Y-OQ™ and this manual as data becomes available. In the meantime, inasmuch as our goal is to have created a measure that is well standardized, brief, and inexpensive, we invite clinicians, agencies, and other researchers to share their data with us as they utilize the instrument. We hope to create a data repository from sites throughout the country, which will permit continuing development and refinement.

REFERENCES

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Allen, J. S., Jr., Tarnowski, K. J., Simonian, S. J., Elliot, D., & Drabman, R. S. (1991). The generalization gap revisited: Assessment of generalized treatment effects in child and adolescent behavior therapy. *Behavior Therapy*, 22, 393-405.
- Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Asay, T. P., Lambert, M. J., Gregersen, A. T., & Goates, M. K. (2002). Using patient-focused research in evaluating treatment outcome in private practice. *Journal of Clinical Psychology*, 58, 1213-1225.
- Atkin, G. Q. (2000). Concurrent-criterion validity of the Youth Outcome Questionnaire. (Doctoral dissertation, Brigham Young University, 2000). *Dissertation Abstracts International*, 61, 582.
- Atkin, Q., Whoolery, M.L., Peterson, G., Burlingame, G.M., Wells, M.G., & Nebeker, R.S. (1997, April). *Reliability and validity of the Youth Outcome Questionnaire*. Paper presented at annual meeting of the Western Psychological Association convention, Seattle, WA.
- Baer, R.A., & Nietzel, M.T. (1991). Cognitive and behavioral treatment of impulsivity in children: A meta-analytic review of the outcome literature. *Journal of Clinical Child Psychology*, 20, 400-412.
- Barnett, R.J., Docherty, J.P., & Frommelt, G.M. (1991). A review of psychotherapy research since 1963. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30, 1- 14.
- Berrett, K. M. S. (1999). Youth Outcome Questionnaire: Item sensitivity to change. (Doctoral dissertation, Brigham Young University, 1999/2000). *Dissertation Abstracts International*, 60, 4876.
- Bickman, L., Guthrie, P. R., Foster, E. M., Lambert, E. W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. A. (1995). *Evaluating managed mental health services: The Fort Bragg experiment*. New York, NY: Plenum Press.
- 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, 207-222.
- Block, J.H., & Robins, R.W. (1993). A longitudinal study of consistency and change in self-esteem from early adolescence to early adulthood. *Child Development*, 64, 909-923.
- Brown, G. S., Burlingame, G. M., Lambert, M. J., Jones, E., & Vaccaro, J. (2001). Pushing the quality envelope: A new outcomes management system. *Psychiatric Services*, 52, 925-934.
- Brown, J. (1987). A review of meta-analyses conducted on psychotherapy outcome research. *Clinical Psychology Review*, 7, 1-23.
- Bryk, A.S., & Raudenbush, S.W. (1988). Toward a more appropriate conceptualization of research on school effects: A three level hierarchical linear model. *American Journal of Education*, 21, 65-108.
- Burlingame, G. M., Jasper, B. W., Peterson, G., Wells, M. G., Lambert, M. J., Reisinger, C. W., & Brown, G. S. (2001). *Youth-Life Status Questionnaire*. Wilmington, DE.: 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.

- Burlingame, G. M., Wells, M. G., Hoag, M. J., Hope, C. A., Nebeker, R. S., Konkel, K., McCollam, P., Peterson, G., Lambert, M. J., Latkowski, M., Ferre, R., & Reisinger, C. W. (2001). *Youth Outcome questionnaire*. Wilmington, DE: American Professional Credentialing Services. .
- Burlingame, G. M., Wells, M.G., & Lambert, M. J. (1996). *Youth Outcome Questionnaire*. Stevenson, MD: American Professional Credentialing Services.
- Burlingame, G. M., Wells, M.G., Lambert, M. J., & Cox, J. C. (2004). Youth Outcome Questionnaire (Y-OQ). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment*. (3rd ed, Volume 2). Mahwah, NJ: Lawrence Erlbaum Associates; pp 235-274.
- Burlingame, G.M., Lambert, M.J., Reisinger, C.W., Neff, W.L., & Mosier, J.I. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration*, 22, 226-236.
- Casey, R.J., & Berman, J.S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin*, 98, 388-400.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Connors, C. K. (1989). *Manual for Connors Rating Scales*. N. Tonawanda, NY: Multi- Health Systems.
- Connors, C.K. (1990). *Connors' Rating Scales Manual*. North Towanda, NY: Multi-Health Systems. Credentialing Services: Wilmington, DE.
- Cronbach, L.J. (1970). *Essentials of psychological testing* (3rd ed.) New York: Harper & Row.
- Cummings, N. A. (1987). The future of psychotherapy: One psychologist's perspective. *American Journal of Psychotherapy*, 41, 349-360.
- Department of Health and Human Services, (1999). *Mental health: A report of the Surgeon General*. Retrieved 28, December, 2002 from <http://www.surgeongeneral.gov/Library/MentalHealth/home.html>
- Derogatis, L. R., & DellaPietra, L. (1994). Psychological tests in screening for psychiatric disorder. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 22-54). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dinh, K., Sarason, B., & Sarason, L. (1994). Parent-child relationships in Vietnamese immigrant families. *Journal of Family Psychology*, 8, 471-488.
- Durham, C. J., McGrath, L. D., Burlingame, G. M., Schaalje, G. B., Lambert, M. J., & Davies, D. R. (2002). The effects of repeated administrations of self-report and parent-report scales. *Journal of Psychoeducational Assessment*, 20, 240-257.
- Durlak, J. A., Wells, A. M., Cotten, J. K., & Johnson, S. (1995). Analysis of selected methodological issues in child psychotherapy research. *Journal of Clinical Child Psychology*, 24, 141-148.
- Fonagy, P. (1996, October). *Evaluating the effectiveness of interventions in child psychiatry: The state of the art*. Invited address at the Kansas Conference on Clinical Child Psychology, October, 1996.
- Gillman, K. D. (1998). Symptom status and risk factors of adolescents in residential treatment: Developing norms for the Youth Outcome Questionnaire (Y-OQ) and determining the utility of the Y-OQ and the Y-OQ Prognostic Assessment (Y-OQ PA) in predicting treatment setting. (Doctoral dissertation, Brigham Young University, 1998). *Dissertation Abstracts International*, 59, 0416.

- Gironda, M. A. (2000). A validity study of the Youth-Outcome Questionnaire and the Ohio Scales. (Doctoral dissertation, Pace University, 2000). *Dissertation Abstracts International*, 61, 2813.
- Gray, D., Konkel, K., Achilles, J., Burlingame, G. M., Haggard, L., Norman, J., & McMahon, W. (2000) *Juvenile Offenders: Suicide Risk and Psychiatric Symptoms*. Poster session presented at the American Academy of Child & Adolescent Psychiatry Convention, New York.
- Grossman, P.G., & Hughes, J.N. (1992). Self-control interventions with internalizing disorders: A review and analysis. *School Psychology Review*, 21, 229-245.
- Grotevant, H.D., & Cooper, C.R. (1986). Individuation in family relationships. *Human Development*, 29, 82-100.
- Hinshaw, S.P., Lahey, B.B., & Hart, E.L. (1993). Issues of taxonomy and comorbidity in the development of conduct disorder. *Development and Psychopathology*, 5, 31-49.
- Hoag, M.J., & Burlingame, G.M. (1997). Evaluating the effectiveness of child and adolescent group treatment: A meta-analytic review. *Journal of Clinical Child Psychology*, 26, 234-246.
- Hoag, M.J., Lambert, M.J., Jenkins, P.H., Hyde, R.F., Lindsey, S.B., & Harvey, Q. (October, 1998). An intensive family-centered early warning intervention pilot project program for juvenile delinquents and ungovernable youth: An outcome analysis of the youth reclamation program. Paper presented at the annual meeting of the Kansas Conference on Clinical Child Psychology. Lawrence, KA.
- Hoag, M. J., Savicki, K., & Burlingame, G. M. (August 2001) *Efficacy of wilderness therapy: Individual and familial aspects of change*. Paper presented as part of a symposium at the annual conference for the American Psychological Association. San Francisco, CA.
- Hughes, L. N., Lambert, M. J., Burlingame, G. M., & Wells, M. G. (2000, April). Factor analysis of the Youth Outcome Questionnaire: Further evidence of construct validity. In G. M. Burlingame (Chair), *Innovations in tracking outcome: The Youth Outcome Questionnaire – construct validity, outcome predictions, psycho-social screening, and in-home treatment evaluation*. Symposium conducted at the 80th Annual Convention of the Western Psychological Association, Portland, OR.
- Jacobson, N.S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Kazdin, A. E. (1995). Scope of child and adolescent psychotherapy research: Limited sampling of dysfunctions, treatments, and client characteristics. *Journal of Clinical Child Psychology*, 24, 125-140.
- Kazdin, A. E. (2003) Psychotherapy for children and adolescents. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*, (5th ed., pp. 543-589). New York: John Wiley.
- Kazdin, A.E. (1990). Premature termination from treatment among children referred for antisocial behavior. *Journal of Child Psychology and Psychiatry*, 31, 415-425.
- Kazdin, A.E. (1991). Treatment research: The investigation and evaluation of psychotherapy. In M. Hersen, A. E. Kazdin, & A. S. Bellack (Eds.), *The clinical psychology handbook* (2nd ed., pp. 293-312). New York: Pergamon.
- Kazdin, A.E. (1993). Psychotherapy for children and adolescent psychotherapy research: Limited sampling of dysfunctions, treatments, and client characteristics. *Journal of Consulting and Clinical Psychology*, 58, 729-740.
- Kazdin, A.E. (1988). *Child psychotherapy: Developing and identifying effective treatments*. Elmsford, NY: Pergamon.

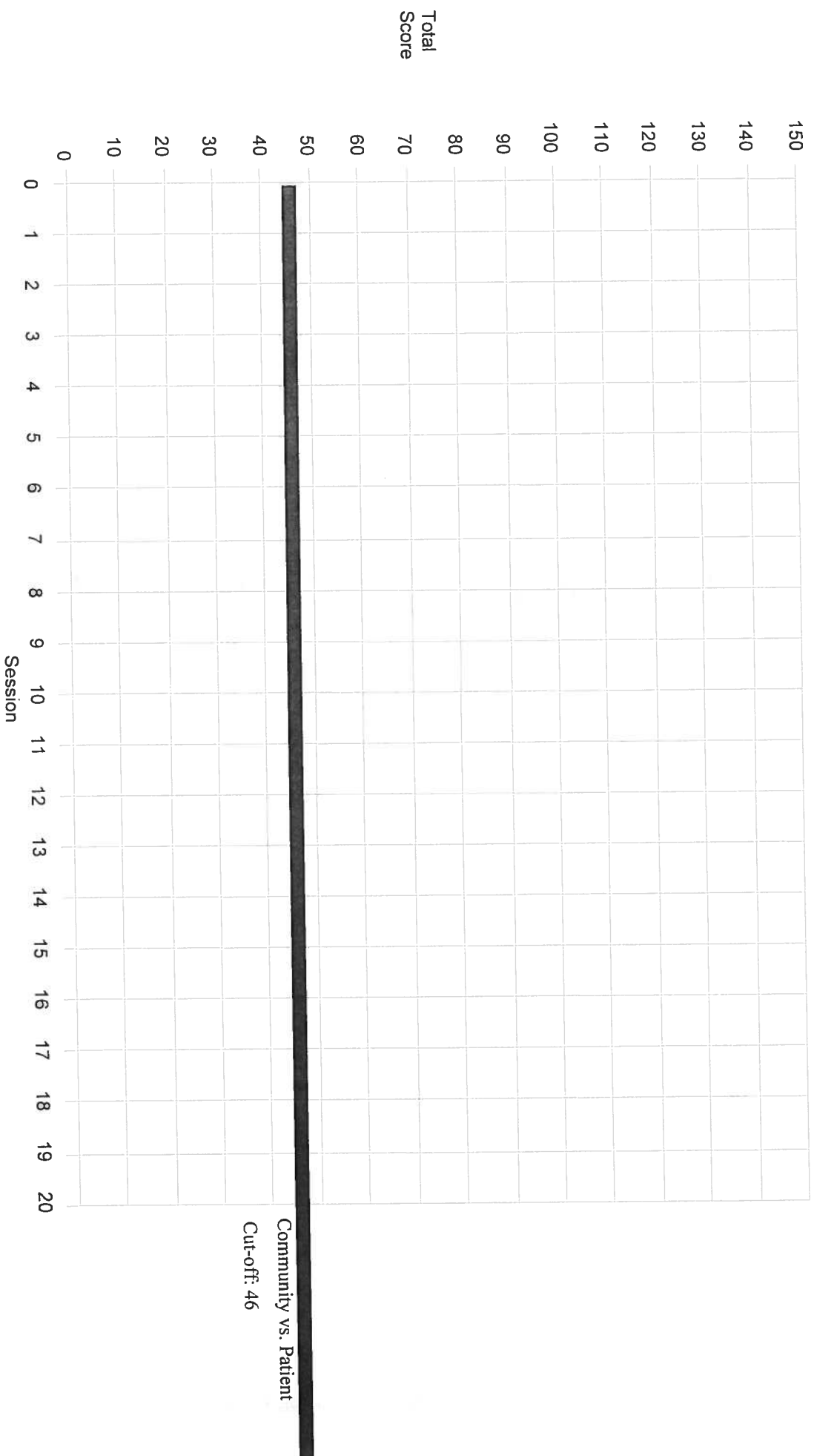
- Kazdin, A.E., Bass, D., Ayers, W.A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology*, 58, 729-740.
- Kendall, P.C., & Morris, R.J. (1991). Child therapy: Issues and recommendations. *Journal of Consulting and Clinical Psychology*, 59, 777-784.
- Kovacs, M., & Paulaskas, S. (1986). The traditional psychotherapies. In H. D. Qual & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (3rd Ed., pp. 496-522). New York: Wiley.
- Lachar, D. (1982). *Personality Inventory for Children (PIC) revised format manual supplement* Los Angeles: Western Psychological Services.
- Lambert, M. J., Morton, J.J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., Shimokawa, K., Christopherson, C., and Burlingame, G. M. (2004). *Administration and Scoring Manual for the Outcome Questionnaire -45*. Orem, UT: American Professional Credentialing Services.
- Lambert, M. J., & Burlingame, G.M. (1996). *Outcome Questionnaire 45.2*. American Professional Credentialing Services: Wilmington, DE.
- Lambert, M.J., & Gregersen, A., Burlingame, G. (2003). Outcome Questionnaire - 45: Chapter 35. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed). Mahwah NJ: Lawrence Erlbaum Associates.
- Lambert, M.J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed., pp. 139-193). New York: John Wiley.
- Lambert, M.J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., & Burlingame, G.M. (1996). *Outcome Questionnaire 45.2*. Wilmington, DE: American Professional Credentialing Services.
- Lambert, M.J., Thompson, K.C., Nebeker, R.S., & Andrews, A. (June, 1996). *The re-test artifact and its implications for establishing dose-effect estimates of patient change following psychotherapy*. Paper presented at the annual meetings of the Society for Psychotherapy Research. Amelia Island, Florida.
- McCollam, P.M., Burlingame, G.M., Vanderwal, G., Hardinger, C., & Wells, M.G., (1997, April). *The Youth Outcome Questionnaire: Characteristics of a juvenile delinquent population*. Paper presented at annual meeting of the Western Psychological Association convention, Seattle, WA.
- McCollam, P. & Burlingame, G. M. (2000, April). Optimal recovery curves: Using the Youth Outcome Questionnaire in treatment settings. In G. M. Burlingame (Chair), *Innovations in tracking outcome: The Youth Outcome Questionnaire -- construct validity, outcome predictions, psycho-social screening, and in-home treatment evaluation*. Symposium conducted at the 80th Annual Convention of the Western Psychological Association, Portland, OR.
- Mills, J. R., Manivanh, T., Burlingame, G.M., Wells, M.G., Peterson, G., & Nuttall, M. (1997, April). *Developing the Youth Outcome Questionnaire (Y-OQ) for Lao refugee children and adolescents*. Paper presented at the annual meetings of the Western Psychological Association Convention, Seattle, WA.
- Mosier, J. (1998). *The predictive validity of the Youth Outcome Questionnaire: Prognostic Assessment*. Unpublished masters thesis, University of Utah, Salt Lake City.
- Mosier, J. (2001). Predicting treatment outcome using psychosocial characteristics: A re-examination of the Youth Outcome Questionnaire Prognostic Assessment. (Doctoral dissertation, Brigham Young University, 2001). *Dissertation Abstracts International*, 62, 2070.

- Mosier, J., Burlingame, M. G., Wells, M. G., Ferre, R., Lakowski, M., Johansen, J., Peterson, G., & Walton, E. (2001). In-home, family-centered psychiatric treatment for high-risk children and youth. *Children's Services: Social Policy, Research, and Practice*, 4, 51-68.
- Mosier, J.I., Burlingame, G.M., Nebeker, R.S., & Wells, M.G. (1997, April). *Correlates of change in treatment in a clinical sample of children and adolescents*. Paper presented at annual meeting of the Western Psychological Association convention, Seattle, WA.
- Neeleman, L. (2001). Factor analysis of the Youth-Outcome Questionnaire: Further evidence of construct validity. (Doctoral dissertation, Brigham Young University, 2001). *Dissertation Abstracts International*, 62, 2110.
- Newman, F.L., Ciarlo, J. A., & Carpenter (1999). Guidelines for selecting psychological instruments for treatment planning and outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (2nd ed.), (pp. 153-170). Mahwah, NJ: Lawrence Erlbaum Associates.
- Prout, H.T., & DeMartino, R.A. (1986). A meta-analysis of school based studies of psychotherapy. *Journal of School Psychology*, 24, 285-292.
- Roberts, A. R., & Camasso, M. J. (1991). The effects of juvenile offender treatment programs on recidivism: A meta-analysis of 46 studies. *Notre Dame Journal of Law, Ethics, & Public Policy*, 5, 421-441.
- Robinson, K. E. (2000). Outcomes of a partial-day treatment program for referred children. *Child & Youth Care Forum*, 29, 127-137.
- Russell, K. (2003). An Assessment of Outcomes in Outdoor Behavioral Healthcare Treatment. *Child & Youth Care Forum*, 32, 355-381.
- Russell, K. C. (2002). Outdoor treatment for adolescents with problem behaviors: An outcomes study. *Behavioral Health Management*, 22, 14-18.
- Russell, R.L., Greenwald, S., & Shirk, S.R. (1991). Language change in child psychotherapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 51, 42-53.
- Shirk, S.R., & Russell, R.L. (1992). A reevaluation of estimated of child therapy effectiveness. *Journal of American Academy of Child and Adolescent Psychiatry*, 31, 703-709.
- Speer, D.C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology*, 60, 402-408.
- Tingey, R., Lambert, M., Burlingame, G., & Hansen, N. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, 6, 109-123.
- Tramonta, M.G. (1980). Critical review of research on psychotherapy outcome with adolescents: 1967- 1977. *Psychological Bulletin*, 88, 429-450.
- Tzoumas, A. C. (2001). The Youth-Outcome Questionnaire-Primary care medicine screening instrument. (Doctoral dissertation, Brigham Young University, 2001/2002). *Dissertation Abstracts International*, 62, 3817.
- Tzoumas, J. L. (2001). *Validation study of the Youth Outcome Questionnaire-Primary care medicine (YOQ-PCM) screening instrument*. Unpublished doctoral dissertation, Brigham Young University, Provo, Utah.
- Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist*, 47, 1578-1585.
- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117, 450-468.

- Weisz, J.R., Weiss, B., Alicke, M.D., & Klotz, M.L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, 55, 542-549.
- Wells, M. G., Burlingame, G. M., & Lambert, M. J. (1999). Youth Outcome Questionnaire. In M.E. Maruish (Ed.), The use of psychological testing for treatment planning and outcome assessment (2nd ed.), (pp. 497-453). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wells, M. G., Burlingame, G. M., Rose, P. M. (1996). Youth-Outcome Questionnaire Self-report. American Professional Credentialing Services: Wilmington, DE.
- Wells, M. G., Mohlman, J., Trecarico, S., Gandhi, P., Turner, L., Stark, C., Peterson, S., Burlingame, G. M. (1997, April). *Measuring therapeutic change among child and adolescent inpatients using the Youth Outcome Questionnaire*. Paper presented at the annual meetings of the Western Psychological Association, Seattle, WA.
- Wells, M.G., Burlingame, G.M., Lambert, M.J., Hoag, M. & Hope, C. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Youth and Adult Outcome Questionnaires. *Psychotherapy: Theory, Research, and Practice*, 33, 275-283.
- Werry, J. S. (1986). Physical illness, symptoms and allied disorders. In H. C. Quay and J. S. Werry (Eds.), Psychopathological disorders of childhood, 3d ed. New York: Wiley, 232-293.
- Whalen, C. K., Henker, B., Hinshaw, S. P., & Granger, D. A. (1989). Externalizing behavior disorders, situational generality, and the type A behavior pattern. *Child Development*, 60, 1453-1462.
- Whalen, C.K. (1989). Attention deficit and hyperactivity disorders. In T. H. Ollendick, and M. Hersen (Eds.). Handbook of child psychopathology. New York: Plenum, 131-169.
- Willet, J.B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, 49, 587-602.

Appendix A: Y-OQ™ Tracking Sheet

Y-OQ™ TRACKING SHEET



Date _____

Name: _____ Age: _____ Sex: _____ Therapist: _____ Agency: _____

Y-OQ™ Administration and Scoring Manual

Appendix B: Outpatient and Inpatient Age Differences

Y-OQ™ Total Score by Age in an Outpatient Sample			
Age – Range	N	Total Score	
		Mean	S.E.
Age = 4-5	120	70.50	3.5
Age = 6-8	356	72.32	2.0
Age = 9-11	449	76.20	1.8
Age = 12-14	412	75.66	1.9
Age = 15-17	447	70.45	1.8

Y-OQ™ Subscale Scores by Age in a Sample of Outpatients										
	Age - Range									
	Age = 4-5		Age = 6-8		Age = 9-11		Age = 12-14		Age = 15-17	
	N = 120		N = 356		N = 449		N = 412		N = 447	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Intrapersonal Distress	19.3	1.18	23.6	.68	25.2	.61	24.9	.63	23.0	.61
Somatic	6.3	.46	5.8	.27	6.8	.24	6.8	.25	6.1	.24
Interpersonal Relations	9.8	.70	9.7	.41	10.2	.36	10.9	.38	10.1	.37
Social Problems	6.2	.53	5.8	.31	6.2	.28	7.2	.29	8.17	.28
Behavioral Dysfunction	21.1	.86	19.9	.50	20.2	.45	18.5	.46	15.7	.45
Critical Items	7.8	.47	7.3	.27	7.6	.24	7.4	.26	7.3	.25

Y-OQ™ Total Score by Age in an Inpatient Sample			
Age - Range	N	Total Score	
		Mean	S.E.
Age = 4-5	NA	NA	NA
Age = 6-8	NA	NA	NA
Age = 9-11	36	111.69	5.9
Age = 12-14	144	112.15	2.9
Age = 15-17	235	110.27	2.3

Appendix C: Application for a Y-OQ™ License

HOW TO OBTAIN AN OQ®-PAPER & PENCIL PRODUCT LICENSE

(Prices good through July 31, 2006)

20051102 MAIL OR FAX Form

STEP 1: Read the **OQ® PAPER & PENCIL PRODUCT BINDING LICENSE AGREEMENT** and the accompanying, **PAPER & PENCIL PRODUCT LICENSE APPLICATION & ORDER FORM, CAREFULLY.** *(PLEASE NOTE: A separate license fee must be paid for each OQ® Product applied for.)*

STEP 2: Review the following table to select the type of license that best suits your needs. Enter the appropriate license fee for your size Practice/Organization on the **PAPER & PENCIL PRODUCT LICENSE APPLICATION & ORDER FORM.**

The fee shown is a **ONE TIME** fee.

LICENSE TYPE:	WHO SHOULD ORDER:	FEE:
STUDENT	One individual. License is valid while student is working on their first terminal graduate degree. License expires upon issuing of degree.	\$30.00
INDIVIDUAL	One person in a solo, group practice or otherwise	\$75.00
SMALL GROUP	Two to five persons in a given group practice or legal entity employed by or bound by contract with this group practice or entity. Group/entity must be located in one state or providence.	\$250.00
MEDIUM GROUP	Six to fourteen persons in a given group practice or legal entity employed by or bound by a contract with this group practice or entity. Group/entity must be located in one state or providence.	\$900.00
LARGE GROUP	Fifteen to fifty persons in a given group practice or legal entity employed by or bound by a contract with this group practice or entity. Group/entity must be located in one state or providence.	\$2400.00
FACILITY	Any number of persons over fifty in number, employed by or bound by a contract with one given facility at one distinct location.	\$3500.00
SINGLE STATE	Any number of persons, over fifty in number, employed by or bound by a contract with a legal entity operating within the confines of a single state or providence with more than one distinct location.	\$7000.00
MULTI-STATE	Any number of persons employed by, or bound by a contract with a legal entity operating within the confines of up to ten states or providence's.	Call
NATIONAL	Any number of persons employed by, or bound by a contract, with a legal entity operating within the confines of a single nation.	Call
INTERNATIONAL	Any number of persons employed by or bound by contract with a legal entity operating in more than one nation.	Call

STEP 3: DETERMINE IF YOU WILL NEED MORE THAN THE ONE MANUAL SUPPLIED WITH EACH LICENSE. Please be aware that the Manual is copyrighted and **may not** be duplicated. Enter any extra manual charge (\$25.00 per manual) on the **PAPER & PENCIL PRODUCT LICENSE APPLICATION & ORDER FORM.** Most products (instruments) are also available in other languages. Contact OQ Measures for availability. Please note: manuals are available only in English.

STEP 4: DECIDE HOW YOU WANT YOUR ORDER SHIPPED TO YOU. Determine shipping costs and enter the amount on the **PAPER & PENCIL PRODUCT LICENSE APPLICATION & ORDER FORM.** All orders are payable in U S Dollars.

SERVICE	CHARGE	
Shipping and Handling for <i>all</i> Orders	\$25.00	\$ 25.00
Priority Mail for U. S. Orders add	\$5.00	\$
Express Mail for U. S Orders add	\$20.00	\$
International Orders Global Priority (Additional Mail Charge) add	\$25.00	\$
Please Enter This Total Amount On Line D of The LICENSE-APPLICATION & ORDER FORM:	TOTAL:	\$

STEP 5: Enclose a check, money order or credit card payment in U S dollars for the Total of the License Fee, Extra Manuals and Total of the Shipping and Handling Charges. Return it with the enclosed **PAPER & PENCIL PRODUCT LICENSE APPLICATION & ORDER FORM.**

What You Will Receive With Your Order:

- One (1) copy of each instrument you order. Additional copies may be purchased upon request; please specify language desired.
 - You have a lifetime license to use this instrument (unless otherwise noted, as with student licenses), and may make as many copies of this instrument as necessary for the purpose of administration.
- One (1) copy of the manual corresponding to the instrument you ordered. This applies to all levels of licenses. These are available only in English. Additional copies may be purchased upon request.

OQ[®]-PAPER & PENCIL PRODUCT LICENSE APPLICATION & ORDER FORM

(2005-11-02 Mail or Fax Form)

TERMS AND CONDITIONS:

Applicant is hereby applying for a/an:

Please check type of product(s):	
<input type="checkbox"/> OQ [®] -45.2	<input type="checkbox"/> Y-OQ [®] 2.01
<input type="checkbox"/> OQ [®] -30.1	<input type="checkbox"/> Y-OQ [®] -30.1
<input type="checkbox"/> OQ [®] -10.1	<input type="checkbox"/> Y-OQ [®] -SR 2.0

Please check one level of license:			
<input type="checkbox"/> Student	<input type="checkbox"/> Small Group	<input type="checkbox"/> Facility	<input type="checkbox"/> National
<input type="checkbox"/> Individual	<input type="checkbox"/> Medium Group	<input type="checkbox"/> Single State	<input type="checkbox"/> International
<input type="checkbox"/>	<input type="checkbox"/> Large Group	<input type="checkbox"/> Multi-State	<input type="checkbox"/>

License as is defined on the attached "OQ[®] Paper & Pencil Product Binding License Agreement." Applicant is enclosing the corresponding License and order processing Fees (Please enter these fees below). Applicant has read the Binding Paper & Pencil License Agreement on the accompanying this Order & Application Form and accepts and agrees to the Paper & Pencil Binding License Agreement. OQ[®] Measures, LLC ("OQ MEASURES") accepts this Application and grants this License to Applicant subject to the Binding Paper & Pencil Binding License Agreement. The License is not effective or granted unless this Application is approved by an authorized officer or representative of OQ MEASURES. Applicants who are not accepted for licensure will have their license fee returned to them.

AGREED TO BY:

Applicant Name (Individual, Corporation, etc.)

Date

Authorized Signature (Applicant)

Printed Name and Title of Authorizing Authority for Group or Organization.

ORDER INFORMATION:	ENTER FEE:
A) License Fee (Taken from STEP 2 Table):	\$
B) Extra Manuals: Number of Extra Manual Wanted = _____ for \$25.00 each.	+ \$
C) Additional Instrument (please specify language): _____ for \$25 each.	+ \$
D) Total of Shipping, Handling, and Additional Postage Charges (Taken from STEP 4 Table):	+ \$
E) Sub Total (Add Lines A, B, C, and D).	= \$
F) Utah Residents must include UT State Sales Tax with their order (6.60% of line D)	+ \$
G) Grand Total Enclosed With This Order Total of Lines D and E	= \$

Payment Information:	<input type="checkbox"/> Check or Money Order Enclosed or <input type="checkbox"/> Credit Card Payment (Details Below)
Credit Card Information: Visa <input type="checkbox"/> MasterCard <input type="checkbox"/> Expiration Date (MM / YY) (_____ / _____)	
Name as it appears on Credit Card (Please Print Carefully) _____	
Billing address and zip code (Please Print Carefully) _____	
Card Number: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	

Contact Information:	Shipping Information: Please Print or Type Clearly
Phone # 1:	Name:
Phone # 2:	Organization:
Fax #:	Address:
e-mail:	Address:
Other:	City State ZIP Country

Return This Form (Order & Application and License) Along With Your Total Payment To:

OQ[®] Measures, LLC

P.O. Box 521047

Salt Lake City, UT 84152-1047

Make Checks Payable To: OQ Measures, LLC

PHONE: 1-888-MH-SCORE (1-888-647-2673)

FAX: (801) 990-4236

EMAIL: info@OQMeasures.com

WEB: <http://www.OQMeasures.com>

OQ®-PAPER & PENCIL PRODUCT BINDING LICENSE AGREEMENT

(2005-11-C2 Mail or Fax Form)

1. **Licensee.** If the OQ® Measures, LLC (Hereafter "OQ® Measures") or its designee has approved of the Application of the Applicant by the act of returning to the Applicant correspondence indicating this fact, then the Applicant is the "Licensee" under this License Agreement.
2. **OQ® Product.** "OQ® Product" means the paper and pencil version of the health care protocol, outcome screening, progress tracking or outcome prognostic measure, and work of authorship for which the Applicant is applying for on the accompanying "OQ® - PRODUCT LICENSE APPLICATION & ORDER FORM."
3. **License.** Subject to the terms and conditions of this Agreement, OQ® Measures grants to the Licensee a license to use, copy and distribute the specific OQ® product accompanying an Administration & Scoring Manual, but only in connection with Licensee's bona fide health care practice (the "License") as the Applicant has applied for and been approved for. This Administration & Scoring Manual may *NOT* be duplicated. Student licenses expire upon issuing of the student's first terminal degree or five years after the issue of the Student License, whichever comes first. The Licensee is granted a license only to the specific OQ® Product being applied for on the Application & Order Form.
4. **Modifications.** Licensees may not modify, translate into other languages or change the content, wording or organization of OQ® product or create any derivative work based on OQ® Product. Licensees may put the OQ® Product into other written, non-electronic, non-computerized, non-automated formats, provided that the content, wording and organization are not modified or changed.
5. **Copies, Notices and Credits.** Any and all copies of the OQ® Product made by Licensee must include the copyright notice, trademarks and other notices and credits in the OQ® Product. Such notices may not be deleted, omitted, obscured or changed by Licensee.
6. **Use, Distribution and Charges.** The OQ® Product may only be used and distributed by Licensee in connection with Licensee's stated bona fide health care practice and may not be used or distributed for any other purpose. Without limiting the generality of the foregoing, Licensee may not distribute copies of the OQ® Product beyond the scope of the applied for license or to other persons for use by other persons. Such other persons should apply to OQ® Measures for a license to use the OQ® Product. Licensee may not charge any client, patient, organization or other entity for use of the OQ® Product.
7. **Responsibility.** BEFORE USING OR RELYING UPON THE OQ® PRODUCT IT IS THE RESPONSIBILITY OF LICENSEE TO ASCERTAIN THE SUITABILITY OF THE OQ® PRODUCT FOR ANY AND ALL USES MADE BY LICENSEE. THE OQ® PRODUCT IS NOT A DIAGNOSTIC TOOL AND SHOULD NOT BE USED AS SUCH. THE OQ® PRODUCT IS NOT A SUBSTITUTE FOR AN INDEPENDENT MEDICAL OR OTHER APPROPRIATE PROFESSIONAL EVALUATION. ANY AND ALL USE OF AND RELIANCE ON THE OQ® PRODUCT BY LICENSEE IS AT LICENSEE'S SOLE RISK AND IS LICENSEE'S SOLE RESPONSIBILITY. LICENSEE SHALL INDEMNIFY OQ® MEASURES AND ITS OFFICERS, DIRECTORS, EMPLOYEES AND REPRESENTATIVES, AND THE AUTHORS OF THE OQ® PRODUCT AGAINST, AND HOLD THEM HARMLESS FROM, ANY AND ALL CLAIMS AND LAW SUITS ARISING FROM OR RELATING TO ANY USE OF OR RELIANCE ON THE OQ® PRODUCT PROVIDED BY OQ® MEASURES TO LICENSEE. THIS OBLIGATION TO INDEMNIFY AND HOLD HARMLESS INCLUDES A PROMISE TO PAY ANY AND ALL JUDGMENTS, DAMAGES, ATTORNEYS' FEES, COSTS AND EXPENSES ARISING FROM ANY SUCH CLAIM OR LAW SUIT.
8. **Disclaimer.** LICENSEE ACCEPTS THE OQ® PRODUCT "AS IS" WITHOUT WARRANTY OF ANY KIND. OQ® MEASURES DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT. OQ® MEASURES DOES NOT WARRANT THAT THE OQ® PRODUCT IS WITHOUT ERROR OR DEFECT. OQ® MEASURES SHALL NOT BE LIABLE FOR ANY CONSEQUENTIAL, INDIRECT, SPECIAL, INCIDENTAL OR PUNITIVE DAMAGES. THE AGGREGATE LIABILITY OF OQ® MEASURES FOR ANY AND ALL CAUSES OF ACTION (INCLUDING THOSE BASED ON CONTRACT, WARRANTY, TORT, NEGLIGENCE, STRICT LIABILITY, FRAUD, MALPRACTICE, OR OTHERWISE) SHALL NOT EXCEED THE FEE PAID BY LICENSEE TO OQ® MEASURES. THIS LICENSE AGREEMENT, AND SECTIONS 7 AND 8 IN PARTICULAR, DEFINES A MUTUALLY AGREED UPON ALLOCATION OF RISK. THE FEE REFLECTS SUCH ALLOCATION OF RISK.
9. **Construction.** The language used in this Agreement is the language chosen by the Parties to express their mutual intent, and no rule of strict construction shall be applied against any Party.
10. **Entire Agreement.** This Agreement is the entire agreement of the Parties relating to the OQ® Product.
11. **Governing Law.** This Agreement is made and entered into in the State of Utah and shall be governed by the laws of the State of Utah. In the event of any litigation or arbitration between the Parties, such litigation or arbitration shall be conducted in Utah and the Parties hereby agree and submit to such jurisdiction and venue. Notice to commence any litigation or arbitration should be directed to: OQ MEASURES LLC, 2150 S 1300 E, Ste 529, Salt Lake City, Utah 84106.
12. **Modification.** This Agreement may not be modified or amended.
13. **Transferability.** This Agreement may not be transferred, bartered, loaned, assigned, leased or sold by the licensee.
14. **Violations.** Violations of any provision or stipulation of this Agreement will result in immediate revocation of this license. Punitive damages may be assessed.

