

The Test-Retest Reliability and Internal Consistency
of the Youth Outcome Questionnaire

A Thesis

Presented to the

Department of Educational Psychology

Brigham Young University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Matthew Lawrence Whoolery

June 1997

Table of Contents

| | Page |
|--|------|
| Table of Contents..... | ii |
| List of Appendices..... | iii |
| List of Tables..... | iv |
| List of Figures..... | v |
| Thesis..... | 1 |
| Chapter 1: Introduction and Literature Review..... | 1 |
| Chapter 2: Methods..... | 15 |
| Chapter 3: Results..... | 22 |
| Chapter 4: Discussion..... | 29 |
| Chapter 5: Journal Ready Article..... | 41 |
| References..... | 56 |
| Abstract | |

List of Appendices

| Appendix | Page |
|--|------|
| A. The Youth Outcome Questionnaire™ 2.0..... | 62 |
| B. Consent, Confidentiality, Information, and Questionnaire Forms..... | 64 |
| Consent and Confidentiality..... | 64 |
| Information Sheet..... | 65 |
| Questionnaire for YOQ Second Administration..... | 66 |
| C. Test-Retest Correlation Coefficients Tables..... | 67 |
| D. Retest Effect and Sample Distribution Descriptives..... | 70 |

List of Tables

| Table | Page |
|---|------|
| 1. Demographic Characteristics of Sample..... | 16 |
| 2. Test-Retest Correlation Coefficients..... | 23 |
| 3. Comparison of Internal Consistency Estimates From Present Study and YOQ Manual..... | 27 |
| 4. Test-Retest Correlation Coefficients for 2-Weeks Elapsed Time..... | 67 |
| 5. Test-Retest Correlation Coefficients for 4-Weeks Elapsed Time..... | 68 |
| 6. Test-Retest Correlation Coefficients for 6-Weeks Elapsed Time..... | 69 |
| 7. Range, Skewness, and Kurtosis of YOQ-Total and Subscale Score Distributions..... | 72 |
| 8. Retest Effect For YOQ Total Score and Subscales..... | 73 |

List of Figures

| Figures | Page |
|-----------------------------------|------|
| 1. Histogram for CI scores..... | 70 |
| 2. Scatterplot for CI scores..... | 70 |
| 3. Histogram for IR scores..... | 71 |
| 4. Scatterplot for IR scores..... | 71 |

THE TEST-RETEST RELIABILITY AND INTERNAL CONSISTENCY OF THE YOUTH OUTCOME QUESTIONNAIRE

CHAPTER ONE

Recent changes in this nation's mental health care system have led to the need for instruments which effectively track psychotherapy outcome. The increasing costs (U.S. Department of Commerce, 1992) and utilization of health care services have been markedly escalating since the early 1980's (Cummings, 1987). These changes have led to a shift in power from suppliers (doctors, psychologists) to the consumers -- patients, employers, and third-party providers (Linder, 1991). As a result, psychology must begin to demonstrate its usefulness and cost-effectiveness to consumers (Leviton, 1996).

Outcome assessment tools are used to assess the effectiveness of services provided. To demonstrate the effectiveness of psychotherapy with children and adolescents, a reliable assessment instrument, which shows sensitivity to patient change over short periods of time, is essential (Burlingame, Lambert, Reisinger, Neff, & Mosier, 1995).

While a great deal of research dealing with outcomes of psychotherapy has been conducted for the adult population, few studies are directed toward children and adolescents (Kazdin, 1991). It has been suggested that children compose approximately 30-35% of the total psychotherapy population (Latkowski, personal communication, February, 1996). However, several researchers report only 6% (Weisz, Weiss, Alicke, & Klotz, 1987) and 7% (Kazdin, Bass, Ayers, & Rodgers, 1990) of outcome research concentrates on child psychotherapy outcome. In addition, a significant amount of money is spent in the treatment of

child/adolescent patients within managed care settings (Kelleher & Long, 1994). This magnifies the need for more child outcome research and effective, cost efficient outcome measures for children.

An instrument recently developed that holds promise in meeting these needs is the Youth Outcome Questionnaire (YOQ, Burlingame, Wells, & Lambert, 1996). It was developed to track progress in therapy for children and adolescents between the ages of four and seventeen years old.

One of the indicators for the need of the YOQ has been the local response to its appearance. Though the YOQ was originally developed for Intermountain Health Care and Primary Children's Medical Center, it has recently been adopted by the State of Utah and a number of mental health care providers in the area.

Additional research is needed on the YOQ to establish its reliability. Preliminary research has been done on the internal consistency (Burlingame, McCollam, Nebeker, Wells, Hoag, Hope, & Mosier, 1996), but further research is needed to establish the test-retest reliability of the YOQ. According to Kaplan & Saccuzzo (1982), "Reliability is one of the basic foundations of behavioral research. If a test is not reliable, it will not be possible to demonstrate that it has any meaning" (p. 114). For the YOQ to be helpful in tracking change in children through therapy, it must reliably measure their psychosocial functioning. Two main types of statistical analyses to establish reliability are stability over time (test/retest) and internal consistency.

Test/retest reliability measures how consistent a measure is over time. The same test is administered twice and the scores are correlated, yielding a coefficient of stability. High stability

over time shows that the test will reliably measure the same thing on separate occasions.

Test/retest assessment "takes into account errors produced by differences in conditions associated with the two occasions on which the test is administered" (Aiken, 1994, p. 85) but does not measure error due to different sample of test items.

The internal consistency is used to measure the homogeneity of scale items. It also allows the researcher to determine redundant items and decide if the measure can or ought to be shortened. Using both of these methods of assessing reliability would estimate both the stability over time and the consistency of test items in measuring the same domain.

Importance of This Study

The YOQ is presently used extensively in clinical settings but further research is needed to determine its reliability. It is important that a measure that claims to be tracking progress in therapy be both stable and sensitive. If there is any change from one administration of the YOQ to the next, the change should be due to meaningful attention in a young person's life and not random fluctuation (error). This reliability study for the YOQ is necessary to ensure that it is a practical, consistent, and sensitive measure of children's change in therapy. Its widespread use not only communicates the need for it, but also the importance of knowing its basic psychometric properties.

Purpose

The purpose of the present study was two-fold. First to make a preliminary estimate of the test-retest reliability of the total YOQ and the individual subscales, and second to replicate the research on the internal or inter-item consistency of the entire YOQ and each subscale. A

sample (N = 423) was drawn from a Salt Lake County, Utah elementary school during the months of September and October 1996.

This study was a preliminary study to assess the basic properties of the YOQ with regards to test-retest reliability. Future research is aimed at continuing to gather normative samples and assessing the YOQ's sensitivity to change between several sessions. If the YOQ proves to be reliable, its further use as a psychotherapy outcome measure will be supported.

Research Questions

This study deals with two basic analyses of reliability, test-retest and internal consistency. The research questions address each of these analyses for the YOQ and its subscales:

- 1) Do the subscales and the total YOQ remain stable over time for a normal population with an acceptable degree of statistical reliability?
- 2) Does the YOQ Total Score measure a single factor? Do the YOQ subscales measure a homogenous set of symptoms?

Literature Review

The literature review is organized into five main parts. First is the issues surrounding research of the outcomes of psychotherapy for children and adolescents. Next the availability and effectiveness of current measures will be addressed. Third, the need for integration of research and clinical practice will be reviewed. Fourth, statistical analyses and methods for assessing reliability will be discussed. Finally, issues regarding managed health care and accountability will be examined.

Research Issues in Child Psychotherapy Outcome

Recent reviews of psychotherapy outcome research demonstrate the effectiveness of psychotherapy. In a review of 1,080 studies Bergin & Garfield (1994) reported an average effect size of .82, meaning that 79.4% of persons who went through psychotherapy were better off than a no-treatment group. In a recent article on psychotherapy effectiveness, Seligman (1995) concluded that patients benefited very substantially from psychotherapy. Nevertheless, the research on outcome of psychotherapy for children and adolescents lags far behind the adult research. (Kazdin, 1993).

In an analysis of more than 15,000 studies, Allen, Tarnowski., Simonian, Elliott, & Drabman (1991) found that only six percent of psychotherapy outcome studies concentrated on children and adolescents. Tramontana (1980) found this earlier and out of the twenty studies included in his review, only five were methodologically sound. Along with a lack of studies of child psychotherapy outcome, Kendall & Morris (1991) emphasized that many of the methodological improvements in adult outcome research have not been incorporated into the studies of child psychotherapy outcome.

On the other hand, Durlak, Wells, Cotten, and Johnson (1995) report that research in child psychotherapy is more methodologically sound and adequate in amount than is typically assumed. However, they also state that very important variables are ignored in the research, such as: communication skills, academic achievement, family variables, and assessing clinical significance of outcomes. Kazdin et al. (1990) agrees that any conclusions about the effectiveness of therapy for children and adolescents must be heavily qualified because of the lack of quality and quantity in child psychotherapy outcome research.

Despite these weaknesses, Kazdin (1993) makes the conclusion that psychotherapy for children and adolescents is effective. From his meta-analysis, children treated with psychotherapy were found to have improved more than 76% of a comparable control sample. This estimate agrees with Weisz, Weiss, Alicke, & Klotz (1987) who reported an effect size of 0.79 for treated children (75% of the treated children were better off than those in the untreated control group).

In a recent article, Kazdin (1995) outlined some of the important methodological issues facing child and adolescent psychotherapy research. One of the problems he mentioned was the need to separate dysfunction into smaller domains so that research can more readily identify whether specific treatments are effective for more narrow types of presenting problems. For example, anxiety and depression are frequent client presenting problems but are not frequently studied in outcome research. Kazdin also noted that a standard outcome assessment practice would help to create a common knowledge base. Standardized measures could provide a consistent profile of children with a variety of presenting problems, allowing more accurate integration of research findings.

Another important point made by Kazdin in this article was the need to understand contextual issues that influence outcome. Family variables, parent psychopathology, developmental level, race, ethnicity, and culture are all likely to have an impact upon any therapeutic intervention. These factors must be taken into account for both in treatment and in assessing psychotherapy outcome.

Hendren (1993) also called for the need to take contextual factors into consideration, stating that "the family and the environment have much more of an impact on the outcome of a

particular disorder in a young person than in an adult" (p. 341). Factors that he suggested be included in measures of outcome include family participation and constellation, socioeconomic status, and child's previous exposure to therapy. The need to include contextual factors in research on child psychotherapy outcome is a common thread through much of the recent research on methodology. Other calls for improvements in methodology include measuring outcome at multiple points, rather than just a pretest/post-test design (Stoolmiller, Duncan, Bank, & Patterson, 1993) and follow-up assessment (Kazdin, 1993). Barnett, Docherty, and Frommelt (1991) also emphasized the need for adequate measurements that are standardized, valid, and reliable. Many of the current measures lack sophistication.

One of the major issues that seem to be called for in the literature reviewed here is the need to improve the quality and quantity of outcome research for children and adolescents. A reliable, valid measure of child psychotherapy outcome would improve the ability to measure outcome across different domains. It would also improve clinicians' ability to determine the effectiveness of certain treatments for specific symptoms. A measure would need to be stable over time to be a useful outcome tool. For it to be effective in measuring outcome across different variables, the scales of the instrument would need to be homogenous, or internally consistent.

Current Measures

This section outlines characteristics of good outcome measures as well as some of the more commonly used measures currently used for child psychotherapy outcome assessment. Though each has its strengths, there are some specific weaknesses that are important to note.

Characteristics of mental health outcome measures. Vermillion and Pfeiffer (1993)

identify four important characteristics which outcome measures should possess. First, the measures should have *known reliability and validity estimates*. Second, the measures should be *practical* (brief, easily administered and scored). Third, they should be *suitable for the population* with whom they are used (i.e., they measure what is relevant and informative to the researcher). Finally, outcome measures should demonstrate *sensitivity to meaningful change*. This is particularly relevant to psychotherapy outcome because a measure is needed which can be employed frequently to detect slight changes following participation in therapy (Lambert & Hill, 1994).

Shortcomings of current child outcome measures. Although numerous child assessment

instruments are currently available, few have been designed for the sole purpose of child psychotherapy outcome measurement. Most have been developed for psychodiagnostic purposes. Though they may be used for outcome assessment, these instruments may lack some of the aforementioned characteristics (Hope, Atkin, Hoag, Mosier, Burlingame, & Wells, 1996) As a result, their effectiveness as outcome measures for children is somewhat questionable.

The Child Behavior Checklist . The Child Behavior Checklist (CBCL; Achenbach,

1991) is probably the most commonly used measure of child psychological functioning. The CBCL is a parent-report measure for children ages four to eighteen. It has 118 behavioral and emotional problems. It also contains a Teacher Report Form (TRF) and a Youth Self Report form (YSR) which allows for cross-assessment between the measures. The data is rated on a 0-1-2 scale based upon how true an item is about their child. The CBCL has many advantages.

One of its main strengths is the extent to which it has been researched. The CBCL is also easily scored, and relatively inexpensive to administer.

The CBCL also has some weaknesses, including its length and sensitivity. Although 118 items is not extremely excessive, it still becomes a fairly lengthy process that does not provide the convenience that many desire. Length also becomes a factor if it is to be used at multiple scoring points during the process of therapy. A recent study (Drotar, Stein, & Perrin, 1995) also questioned the sensitivity of the CBCL to psychotherapeutic change, especially when the child has less serious behavioral problems. These authors also mentioned difficulties in interpreting the data from the CBCL to ethnically, culturally, or economically diverse populations.

The main weakness of the CBCL as well as the other current measures is that they were developed for psychodiagnostic purposes, not for outcome assessment. Therefore they are not made for detecting change over short periods of time (Hope, et al., 1996).

The Revised Behavior Problem Checklist . The Revised Behavior Problem Checklist (RPBC; Quay, 1983) is an 89-item questionnaire concentrating on children between the ages of six and eighteen. A parent or teacher of the child fills it out, using a three-point scale. There are six subscale scores and a general score for general psychosocial functioning. The strengths of the RPBC are its ease of scoring and administration and its use of multiple domains. The weaknesses of the RPBC are a limited amount of data on its validity and reliability, lack of information on sensitivity to treatment change, and its lack of specific time interval suggested between administrations.

The Conner's Parent Rating Scale. The Conner's Parent Rating Scale (CPRS; Conners, 1990) is a widely used measure of children's behavioral problems. The ages for the CPRS are between three and seventeen. There are two versions, one with 93 items yielding eight subscales and one with 48 items yielding five factors. It is a parent-rated questionnaire on a four-point scale. Though the longer version has research supporting its reliability and validity, the shorter (and more used) has little research support. The CPRS's strengths are that it is easily administered and scored and focuses on several areas of functioning.

General weaknesses among the commonly used measures of children's psychosocial functioning are a lack of sound psychometric properties, a limited focus, and utility. The main deficiency though is a lack of sensitivity to change over time. The purpose of assessment/diagnostic measures is to assess a child's current psychosocial functioning. Outcome measures, on the other hand, are for tracking change. Assessment measures and outcome measures can be used concurrently, one as a pretest/posttest, and the other for tracking change through the process of therapy.

Researcher and Clinician Issues

Researchers and clinicians often stand on different ends of the same problem, both discounting the validity of the other's pursuits. For example, the types of cases studied by researchers are often different from those seen in therapy (Weisz, et al., 1987). Concerning this gap, Kazdin (1993) stated that "a critical issue for the field is to address the chasm between research and practice" (p.652). He suggests three ways that this gap can begin to be filled.

Clinical samples are often referred from schools, courts, etc. while research samples are most often taken from volunteers in schools. Instead of taking their samples from different

places, a more rigorous investigation of clients referred for treatment would increase the relevancy of research findings to clinical practice.

Second, he recommends that clinicians increase their use of single case (N=1) research designs to test the effectiveness of interventions and test hypotheses. Though formal designs may not always be used, some systematic evaluation can increase the ability to draw conclusions about treatment.

Third, many interventions can be evaluated by using quasi-experimental designs that permit inferences about treatment effectiveness. Such research can benefit the researcher by generating new ideas, problems, and hypotheses.

Durlak et al. (1995) addressed another point that contributes to this distance between research and clinical practice. These authors noted some concerns expressed by practitioners that research may often not relate to the kinds of cases they see. For example, much research has been conducted on white children in the middle class, yet many therapists deal with a vastly different population. Some of these concerns were addressed in their review of 516 studies. First, there are very few studies of non-behavioral interventions. Secondly, little research has been done on minority subjects. Thirdly, long-term psychotherapeutic interventions have not been studied. And last, few treatments have been assessed, leaving out important factors like family constellation, parental psychopathology, and availability of social support. All of these factors are dealt with in a clinical setting, but often left unstudied by researchers.

An outcome measure that is stable and internally consistent can help bridge this gap between researchers and practitioners (Burlingame, et al., 1995). It can improve the ability to assess the effectiveness of psychotherapy for youth and adolescents of different ages and ethnic

backgrounds. A valid, reliable outcome measure can help to generate the new ideas and assess their effectiveness. The gap between research and clinical practice could be also shortened by providing a practical, effective way to research child psychotherapy outcome.

Reliability

The reliability of a measure is possibly the single most important characteristic. Without reliability, there can be no validity. The degree of reliability also sets the upper limit of the validity of an instrument. Test-retest and internal consistency are the most common measures of reliability.

A measure can be split into two components, a true score and measurement error. The test-retest correlation coefficient, r , measures how stable a score is over time. High test-retest reliability is “indicated by a correlation coefficient that approaches +1.00” (Cozby, 1985, p. 130). An acceptable range for the reliability coefficient of a psychological measure is considered to be approximately .80 or higher (Aiken, 1994). The fraction left from +1.00 is a measure of the “random fluctuations of performance from one test session to another” (Anastasi, 1976, p. 110). Therefore, the higher the correlation coefficient, the more reliable the test score.

Retest reliability "shows the extent to which scores on a test can be generalized over different occasions" (Anastasi, 1988, p. 117). For a measure that claims to track change, the less receptive it is to the daily fluctuations (error) in a child's life, the more reliable it will be. For this measure, it is measuring specifically the amount of error in the parent's reporting of the child's symptomatology.

For a measure that has subscales to measure different domains of functioning, internal consistency is especially important. The internal consistency coefficient is “the mean of all

possible split-half reliability estimates. It reflects the degree to which the item content is similar” (Lemke & Wiersma, 1976).

Internal consistency will assess whether the subscales are homogeneous, or are measuring the same domain. If the subscales are measuring a single domain, the "test scores will be less ambiguous" (Anastasi, 1988, p. 122) and provide accurate information concerning the various areas of psychosocial functioning. The range of acceptable scores for the internal consistency coefficient is similar to that for the test-retest coefficient. A low internal consistency score indicates that the total measure or individual sub-tests are measuring a heterogeneous set of symptoms.

Reliability estimates are affected by the distribution of scores. Restricted ranges of scores produce spuriously low test-retest coefficients (McCall, 1994; Anastasi, 1988; Kenney, 1987). The correlation score is “greatly influenced by the heterogeneity of the sample; the more homogenous the sample, the lower the value of the correlation coefficient” (Hopkins, Glass, & Hopkins, 1987, p. 87).

Also, reliability estimates are heavily dependent upon the sample from which they are drawn. It is preferable to evaluate the reliability using a population similar to that for which a measure is developed (Anastasi, 1988).

Managed Health Care and Accountability

One of the major changes that faces contemporary psychology is the advent of managed health care. As monetary costs for children's mental health care skyrocket, the industry has begun to demand accountability and to require psychologists to produce positive outcomes. According to Leviton (1996), the challenge for psychology at present is to "assure

coherent integration of relevant theories, knowledge bases, and public health practice" (p.42).

The crunch created by the changes in our health care system is creating a need for the union of practice, theory, and outcome. The call for accountability also introduces the need for ways to accurately measure outcome.

Burlingame, Lambert, Reisinger, Neff, and Mosier (1995) state that "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).

Standardized measurements of outcome are vital to the accountability demanded by the new system. The use of outcome measures along with assessment measures allows the practitioner and mental health care provider to track change through the process of therapy along with a pretest/posttest assessment of a child's psychosocial functioning.

CHAPTER TWO: METHODS

The basic purpose of this study was to assess the test-retest reliability and internal consistency of the Youth Outcome Questionnaire (YOQ). This chapter will present the methodology used in this study. This study used a basic correlational design to assess the 1) internal consistency and 2) test/retest reliability of the YOQ. The chapter will be organized in the following sections: population, sample, instruments, procedures, and limitations.

Population

The population chosen for this project included parents of children from the ages of 4-17 years old. The YOQ was specifically developed to measure parent-reported psychosocial functioning of children and adolescents in this age group. The YOQ is used in outpatient and inpatient clinical settings.

Sample

Subjects included a community sample obtained from an elementary school in Salt Lake County, Utah. This sample contained parents of children, kindergarten through fifth grade (ages ranged from five through twelve), totaling 423 subjects. The school is located in a semi-rural area about twenty miles from a major city.

Demographic information on Table 1 shows relatively equal amounts of female/male participants in each age group. Also, it appears that a majority of children were Caucasian, with parents who had completed High School. The mean family income for this sample was approximately \$47,000 per year.

Table 1

Demographic Characteristics of Sample

| Age (N = 427) | | Gender (N = 427) | | Ethnicity (N = 358) | | Income (N = 353) | | Education (N = 348) |
|------------------|----|---------------------|-----|------------------------|-------|---------------------|--------|------------------------|
| mean | SD | M | F | White | Other | mean | SD | H.S. Grad |
| 8 | 2 | 196 | 231 | 97% | 3% | 47,141 | 17,484 | 99% |

Instruments

The instrument used was the Youth Outcome Questionnaire (Burlingame, Wells, and Lambert, 1996). The YOQ is a 64-item parent-report measure of child and adolescent psychosocial functioning. It was developed to remedy the limitations of present measures being used for psychotherapy outcome with children and adolescents.

The YOQ was designed to be a measure of child psychotherapy outcome that could meet the needs of practitioners, researchers, and managed health care systems. It will help increase practitioner accountability by providing a continuous set of points at which change in therapy can be measured. Therefore, the changes, if any, that take place in therapy can be evaluated during the process of therapy, not simply after its completion. Also, its brevity and low-cost will reduce the cost and time taken in administering and scoring tests. The YOQ was not developed to replace the existing assessment measures, but ideally to be used to complement them as a tracking measure.

The YOQ consists of six subscales: intrapersonal distress (ID), somatic (S), interpersonal problems (IP), behavioral dysfunction (BD), social problems (SP), and critical items (CI). A brief description of each of these scales follows.

The *intrapersonal distress* (ID) scale assesses the amount of emotional stress in the child's life. It contains items measuring anxiety, energy level, depression, hopelessness, and fearfulness.

The *somatic* (S) subscale assesses somatic symptoms that the child is facing. Some items are: dizziness or headaches, bowel difficulties, pain or weakness in joints, and sleep disturbances.

The *interpersonal problems* (IR) scale measures the amount of difficulty the child has in relationships with friends, family, and other adults. Assessment is made regarding the child's attitude toward adults and authority, relationships with family, and quality of communication with others.

The *social problems* (SP) scale's purpose is to assess problem behaviors that are socially related. These behaviors include running away from home, truancy from school, substance abuse, and sexual problems.

The *behavioral dysfunction* (BD) subscale assesses attention deficit symptoms including completion of tasks, hyperactivity, organizational skills, and ability to deal with boredom.

The *critical items* (CI) subscale is used to identify children with more severe clinical problems such as significant weight loss, hallucinations, self-harming behaviors, and delusions.

The YOQ was developed to fulfill the limitations of current measures. Therefore, it is brief, easy to administer and score, and available at a low cost. It is also hoped that the YOQ will be sensitive to change over short periods of time. The subscales of the YOQ should be helpful in assessing functioning across various domains.

The YOQ questionnaire items are interval level data on a five-point Likert scale. The YOQ is completed by parents or guardians. The parents/guardians are asked to decide how well each item describes their child from never (0) to almost always true (4).

Preliminary research on the YOQ is promising, yet incomplete. A sample has been collected which consisted of 148 community members, 43 outpatients, and 55 inpatients. An internal-consistency estimate revealed a Cronbach's alpha of .96 (Burlingame, Wells, Hoag, Hope, Nebeker, Konkel, McCollam, Peterson, Lambert, Latkowi, Ferre, & Reisinger, 1996). This high alpha coefficient suggests that the YOQ items are probably very related, and probably measuring one similar construct. Test-retest reliability has not been conducted on the YOQ. The present study was conducted to determine the test-retest reliability of the YOQ.

Preliminary assessment of the validity of the YOQ has shown that the YOQ has good concurrent criterion validity when compared to the Child Behavior Checklist (CBCL, Achenbach, 1991) and Conner's Parent Rating Scale (Conners, 1990). The YOQ-Total Score and CBCL Total Score have a high degree of correlation, suggesting high validity between the tests (Atkin, Whoolery, Peterson, Burlingame, Wells, & Nebekar, 1997).

Continuing research aims at continuing to gather normative samples, assessing the YOQ's sensitivity to change between two sessions (Wells, Burlingame, & Latkowski, 1996),

and establishing test-retest reliability. This study was conducted to assess the test-retest reliability and replicate the internal consistency studies of the YOQ.

Procedures

The sample was parents of an elementary school in Salt Lake County, Utah. A newsletter, which briefly described the study, informed parents that a packet will soon be sent home with each student. Two days later, each student in the school was sent home with a packet, to be completed by a consenting parent or guardian (N = 650). The packet contained a cover letter from the principal, a Consent and Confidentiality form (see Appendix B), the three test measures--the Youth Outcome Questionnaire (YOQ), the Child Behavior Checklist (CBCL), and the Conner's Parent Rating Scale-93 (CPRS)--and an information sheet (see Appendix B). The CBCL and CPRS were included for a project assessing the concurrent criterion validity of the YOQ.

On the information sheet, parents who wished to continue participation (by filling out another YOQ in two or four weeks) gave their name, address, and phone number (if desired). The students then returned the completed packets to the school and were rewarded with a small candy bar. Of the 650 that were sent out originally, 451 were returned, yielding an original return rate of 69.4%. An attempt to encourage parents who had not returned packets was made, yielding an additional 20 packets.

After the initial collection process was complete, parents who agreed to continue participation (82.5% of those returned, N=372) were sent a YOQ, a brief questionnaire (see Appendix B), and self-addressed stamped envelope. The first YOQ administrations were assigned an ID number. This number was printed on the second YOQ sent to that individual.

The test-retest subsample was divided in half and sent YOQ's for the two-week and four-week retest. The subjects then mailed back the completed YOQ and questionnaire. Parents who indicated that we could call them were telephoned and reminded to return the YOQ. This yielded over 30 more completed YOQ test-retest sets. Questionnaires of parents with children receiving psychological treatment, medications, or who had a significant trauma were not used for this study, since a normal sample is needed.

The number of complete test-retest YOQ sets was $N=242$, yielding a return rate of 65.1%. ID numbers were used to match the first and second administrations of the YOQ. The researcher checked the age and gender to double-check that the YOQ's were filled out for the same child. Incorrect, incomplete, or unusable YOQ's not used totaled 48. The total usable sets of YOQ's were 194.

The first administration of the YOQ was used for the internal consistency analysis. Incomplete or blank YOQ's were not used. The total usable first administrations of the YOQ equaled 427.

The usable YOQ's were then entered into a computer and scored. Money was subsequently donated to the school and a newsletter was sent home thanking parents and notifying them of the amount of money donated to the school.

Research Design and Data Analysis

The research design was a basic correlational study to determine the internal consistency and test/retest reliability of the YOQ. For the internal consistency, Cronbach's (1951) coefficient alpha statistical method was used. The results obtained in this study ($N = 427$) were also combined with a larger sample that included clinical subjects ($N = 1249$,

Burlingame, Wells, Hoag, Hope, Nebeker, Konkel, McCollam, Peterson, Lambert, Latkowski, Ferre, & Reisinger, in press). The Pearson Product Moment Correlation was used to assess test/retest reliability. The data was analyzed using Microsoft Excel and SPSS.

To assure accuracy of the data entry into the computer, a double-entry audit was implemented. The data was entered in two times by different people. The results were then checked for discrepancies. These discrepancies were then checked for accuracy by looking once again at the raw data. For the few errors that were found, corrections were made from the original data before the data analyses were conducted.

CHAPTER THREE: RESULTS

This research concentrated on two questions: 1) Do the subscales and the total YOQ remain stable over time for a normal population?, and 2) Are the separate subscales and the entire YOQ internally consistent? This chapter will present the results of the research aimed at answering these questions.

Test-Retest Reliability

The Pearson Product Moment Correlation (r) was used to assess the test/retest reliability of the YOQ Total score and each of the six subscales. Table 2 summarizes the results.

The YOQ Total score's test-retest reliability is 0.81 ($N = 194$). The subscale scores range from 0.57 for the CI (Critical Items) to 0.78 for the BD (Behavioral Dysfunction). All test/retest correlation coefficients were significant at the $p < .001$ level. The correlation scores for each individual administration (2, 4, and 6 weeks) can be seen in Appendix B. A Test of Difference Between Two Independent Correlation Coefficients found that the coefficients were not significantly different. Therefore, they were combined into one score representing all elapsed times.

YOQ-Total score. The YOQ Total score's reliability estimate is .81 ($N = 194$). The range of scores for this scale was limited. Out of a possible range of 256, the scores range only 131 and 124 for the time one and time two administrations. The distribution of scores was positively skewed and leptokurtic (See Appendix D, Table 7). The standard deviation statistics were small (22.02 and 22.18) and the means were very low for both administrations (16.20 and

14.35).

Table 2

Test-Retest Correlation Coefficients

| | Mean | N | SD | SE | r | Sig. |
|-------------|-------|-----|-------|------|-----|------|
| Total YOQ | 16.70 | 194 | 22.02 | 1.58 | .81 | .000 |
| Total YOQ-2 | 14.35 | 194 | 22.18 | 1.59 | | |
| BD | 5.78 | 194 | 6.88 | .49 | .78 | .000 |
| BD-2 | 5.06 | 194 | 6.56 | .47 | | |
| CI | 2.09 | 194 | 2.35 | .17 | .57 | .000 |
| CI-2 | 1.86 | 194 | 2.16 | .16 | | |
| ID | 6.17 | 194 | 7.94 | .57 | .76 | .000 |
| ID-2 | 5.36 | 194 | 7.68 | .55 | | |
| IR | -.49 | 194 | 4.23 | .30 | .62 | .000 |
| IR-2 | -.80 | 194 | 4.61 | .33 | | |
| S | 2.80 | 194 | 3.15 | .23 | .72 | .000 |
| S-2 | 2.70 | 194 | 2.85 | .20 | | |
| SP | .35 | 194 | 2.74 | .20 | .75 | .000 |
| SP-2 | .18 | 194 | 2.60 | .19 | | |

BD subscale. This sample produced a test-retest reliability coefficient of .78 for the BD subscale. This score is the highest of the subscale test-retest coefficients. The BD subscale showed a fairly good range of scores. The range for the BD subscale is from -4 to 40. The range for the first administration scores was from -4 to 31. The second administration scores ranged from -4 to 25. There is some restriction in range, but not as major as some of the other subscales. The standard deviation of the scores was 6.88 and 6.56 for the first and second administrations. The means were very low—out of a possible 44 points, the means were 5.29 and 5.06 for the first and second administrations. The second administration scores for this subscale were less skewed and more resembling the kurtosis of a normal curve than any of the other scales (see Appendix D, Table 7).

CI subscale. The test-retest correlation coefficient for the CI subscale produced by this sample is 0.57. As can be seen in Figures 1 and 2 (see Appendix D), the range for this subscale was extremely limited. The range for this subscale is from 0 to 36, but the scores produced by this normal sample only ranged from 0 to 14 for the first administration and 0 to 12 for the second administration. Specifically, 98% of the respondents' scores were from 0 to 8. The means (2.17 and 1.86 out of a possible 36) were extremely low. In addition, the standard deviation for this scale was only 2.34. This subscale's score distribution was positively skewed and leptokurtic.

ID subscale. The ID subscale r is estimated at .76. The distribution of scores is somewhat limited in range. The ID scores have a possible range of -4 to 68. The scores for the first administration range from -4 to 36 and the second from -4 to 40. So out of a possible range of 72 points, the normal sample used for this study only produced scores ranging only 40

- 44 points. The distribution for these scores was positively skewed and leptokurtic. The standard deviation is relatively high at 7.94 and 7.68 for the first and second administrations.

IR subscale. . The possible range of scores for the IR subscale is from -6 to 68, while the first and second administrations of the YOQ produced ranges from -6 to 16 and -6 to 15, respectively. In addition, 98% of the respondents' scores were below 10. So out of a possible range of 74, this population only produced ranges from 22 and 21. A scatterplot and histogram of scores can be seen in Appendix D, Figures 3 and 4. The IR subscale score distribution was positively skewed and leptokurtic. The means and standard deviations were also low (means = -0.45, -0.80; SD = 4.61, 4.62).

S subscale. The S subscale test-retest reliability estimate was .72 (N = 194). The normal sample used for this study also showed an extremely restricted range for this subscale. The range of scores for the S subscale is from 0 to 32. The respondents' scores only ranged from 0 to 17 for each administration. This accounts for just less than half of the possible range. The standard deviations and means were low for both administrations. Also, the distribution of scores was positively skewed and leptokurtic.

SP subscale. The SP subscale's r estimate was .75. Once again, the normal population used for this study produced a restricted range of responses. The scores for the SP subscale can range from -2 to 30. However the two administrations of the YOQ produced ranges from -2 to 13 and -2 to 9. Out of a possible range of 32 points, this population produced ranges of only 15 and 11. This extremely limited range samples less than half of the possible scores on the subscale. The S subscale's score distribution was positively skewed and leptokurtic.

Retest Effect

A look at the mean scores in Table 2 for the first and second administrations of the YOQ shows a systematic temporal trend (see Appendix D, Table 8 for mean changes). Without exception, the mean scores for the YOQ Total score and all the subscales go down from the first to second administration. The trend decreases over time for the YOQ-Total score from -3.29 to -2.50 to .087 for the 2, 4, and 6 week test-retest administrations. So the degree of the effect seems to decrease as the time between administrations increases.

Internal Consistency

To answer question 2, Cronbach's (1951) coefficient alpha statistical method was used to assess the internal consistency of the YOQ Total score and its subscales.). The second set of data (N = 1676) combining the data gathered in this study with the data from Burlingame, et al. (in press) will be referred to as the 1997 combined data set to distinguish it from the combined sample from the 1996 version of the YOQ manual. The results are summarized in Table 3, compared with estimates from the YOQ Manual (Burlingame, et al., 1996).

Table 3

Comparison of Internal Consistency Estimates From Present Study and YOQ Manual

| | Present Study Sample (N = 427) | School Sample From YOQ Manual (1996) (N = 41) | Combined Sample (1997) (N = 1676) | YOQ Manual Combined Sample (1996) (N = 1199) |
|-----------|--------------------------------------|--|--|---|
| YOQ Total | .93* | .93* | .97* | .97* |
| BD | .82* | .85* | .90* | .91* |
| CI | .61* | .61* | .80* | .76* |
| ID | .87* | .84* | .93* | .93* |
| IR | .69* | .69* | .85* | .89* |
| S | .62* | .72* | .78* | .76* |
| SP | .50* | .51* | .82* | .84* |

* significant at $p < .01$, two-tailed

The internal consistency estimate for the YOQ Total score is 0.93 (N = 427). When combined with the 1997 combined data set, the coefficient was estimated at .97 (N = 1676). The subscale coefficient alpha scores using results from the present study range from 0.50 for the SP (Social Problems) subscale to 0.87 for the CI (Critical Items) subscale. Similar to the YOQ-Total score, the internal consistency estimate is higher for the subscales when using a sample that includes some clinical subjects. Using the 1997 combined sample, the coefficients range

from .78 for the S subscale to .93 for the ID subscale. The coefficient alpha scores are all significant at the $p < .01$ for the YOQ Total score and all of the subscales.

A good range for the reliability coefficient of a psychological measure is approximately .80 or higher (Aiken, 1994). For the sample obtained in this study, the YOQ Total score and the subscales BD and ID have good internal consistency. The IR subscale is in the low-adequate range at .69 ($N = 427$). The remainder of the subscales are still significantly different than zero, but do not meet the commonly accepted range of .70 to .90 or higher. When combined with a larger sample (Burlingame, et al., in press), the YOQ Total score and all subscales show an adequate to good degree of internal consistency.

The results of the data gathered in this study are very similar to those found in the YOQ Manual (Burlingame, et al., 1996) Student sample ($N = 41$). The 1997 combined sample comprising data gathered in this study with data used in the Revised YOQ Manual (Burlingame, et al., in press) is also similar to the combined sample alpha coefficients reported in the YOQ Manual (Burlingame, et al., 1996). A Test of Difference Between Two Independent Correlation Coefficients comparing the school samples showed no significant differences between the coefficients. There was also not a significant difference found between the combined samples alpha coefficients.

CHAPTER FOUR: DISCUSSION

The research questions about 1) the test-retest reliability and 2) the internal consistency of the YOQ will be discussed for each scale. Following the discussion of the results, the practical and theoretical implications of the results of this study will be considered. Finally there will be a section addressing directions for future research.

YOQ Total Score

The YOQ Total score has a test-retest reliability coefficient of 0.81 (N = 194). This finding shows that the Total YOQ has good stability. The coefficients for the different time-elapsed groups ranges from 0.84 (N = 56) for the two-week administration to 0.76 (N = 45) for the six-week group. There were no significant differences in the reliability coefficients between the 2, 4, and 6 week administrations of the YOQ. The YOQ was developed to track change on a weekly or bi-weekly basis. The data in this study supports the stability of the Total YOQ score over the time-periods it was developed to be used.

Other reliable parent-report measures have similar test-retest reliability coefficients. The Child Behavior Checklist (CBC, Achenbach, 1991) is reported to have a two-week test-retest coefficient of 0.89 (Kramer & Conoley, 1992). The Conner's Rating Scales (CRS, Conners, 1990) are reported to have a range of 0.91-0.33 for the different subscales (Kramer & Conoley, 1992). The YOQ has scores that are consistent with other measures considered to be adequately reliable.

The YOQ Total score also shows excellent internal consistency. The Cronbach's alpha coefficient for the YOQ Total score is 0.93 (N = 427). This score is identical to the one reported in the 1996 version of the YOQ Manual (Burlingame, et al., 1996). The combined

samples from the 1996 version of the manual and the combined sample for the version in press are identical (.97, N = 1199; .97, N = 1676). This score is consistent with other measures that are considered adequately reliable such as the Conner's Rating Scales which report internal consistency estimates from 0.61 to 0.95.

This high internal consistency estimate indicates that there is a single underlying factor for the YOQ as a whole. The existence of strong single factor was also reported for the OQ-45.2 (Lambert & Burlingame, 1996).

Behavioral Dysfunction (BD) Subscale

The test-retest coefficient r for the BD subscale is 0.78. This shows that the BD subscale has adequate stability over time. The BD subscale scores showed a higher range and less skewness than most of the other subscales. This may account for the relatively higher score.

The internal consistency estimate for the BD scale is also adequate (0.85). The estimate obtained from combining the data obtained in this study with the data for the upcoming revision of the YOQ Manual (Burlingame, et al., in press) is .90. This supports the use of BD as measuring a single factor. The BD is measuring such things as inattention deficit, hyperactivity, and impulsivity. These factors do tend to inter-correlate, as supported by the alpha coefficient for this subscale. Some range restriction may explain the lower coefficient in the elementary school sample used in this sample and the student sample reported in the 1996 version of the YOQ Manual (Burlingame, et al., 1996).

Critical Items (CI) Subscale

The CI subscale's test-retest reliability estimate obtained from the sample used in this study was .57. The CI subscale's low stability estimate is likely due to range restriction. Extremely restricted ranges such as this produce spuriously low test-retest coefficients (Kenney, 1987; Anastasi, 1988). A homogenous sample, like the one used in this sample, produces lower correlation coefficients (Hopkins, et al., 1987).

This kind of range restriction in a normal population is not surprising for this scale. This scale is computed from items 12, 20, 21, 28, 38, 44, 46, 51, and 58 (See Appendix A for a copy of the YOQ™ 2.0). These items include such symptomatology as hallucinations, suicidal behavior, mania, paranoia, and obsessive-compulsive symptoms. Testing a normal population for such intense symptoms, it is expected that the scores would be very low. To get a more accurate test-retest coefficient for this subscale, it would be preferable to reevaluate the reliability using a population with a group more similar to the population for which it will be used (Anastasi, 1988). In this case, a sample from a clinical waiting list may be a better population for testing the test-retest reliability.

The CI subscale also has low internal consistency at 0.61 (N = 427) when using a normal population. This score is the same as the school sample (N = 41) coefficient alpha found in the Administration and Scoring Manual for the Youth Outcome Questionnaire 2.0 (Burlingame, et al., 1996). When the combined data from a normal and clinical population were used, the coefficient alpha is significantly higher at .80 (N = 1676). There are a couple of possible explanations for the lower internal consistency.

While the BD subscale measures symptoms that tend to co-exist, the CI subscale is measuring a heterogeneous set of symptoms. The symptoms are similar in that they tend to be

more extreme symptoms found in inpatient populations, but they do not necessarily inter-correlate. The CI subscale, therefore, may not be measuring a single factor, but a varied set of factors that only share similarity in extremity. This is not necessarily a negative finding since the CI subscale was not developed to be measuring a single, homogenous construct.

The range restriction also has an effect on the internal consistency coefficient. Since internal consistency is measuring the inter-item correlation, low scores will not show much inter-item correlation. In other words, if the parent only marks one of subscale questions, that one item cannot inter-correlate with any other items. Lower scores will make the coefficient smaller. The comparison of the school samples with the combined samples supports this hypothesis (see Table 3).

The distribution of scores obtained in this study was negatively skewed as well as restricted in range (See Table 7, Appendix D for skewness, kurtosis, and range for all scales). Both of these characteristics lower the internal consistency coefficient.

Intrapersonal Distress (ID) Subscale

The ID subscale shows adequate stability with a Pearson's r of 0.76. This score shows a moderate degree of test-retest reliability. The low score may be a factor of range restriction, though not as extreme as the CI subscale. The ID subscale shows adequate stability, but the score may be spuriously low due to range restriction (McCall, 1994).

The ID subscale shows a good degree of internal consistency with an alpha coefficient of 0.87. When combined with a more heterogeneous sample, the internal consistency estimate was .93. This higher score using combined samples supports the hypothesis that the lower reliability obtained in this study is due at least in part to the range restriction.

This scale was developed to measure the emotional distress of the child. The symptoms include depression, anxiety, fearfulness, hopelessness, and self-harm. These symptoms seem to inter-correlate as supported by the high internal consistency coefficient.

Interpersonal Problems (IR) Subscale

The IR subscale has a low degree of test-retest reliability, producing a Pearson r coefficient of 0.58. This is the lowest of all the subscales. Similar to the problems with the CI subscale, the IR subscale has an extremely restricted range. Again, extremely restricted ranges such as this one produce spuriously low test-retest coefficients. The symptoms measured, including relationships with adults and friends, may also be less stable over time than symptoms that are more behavioral in nature such as those measured by the BD subscale which has a higher stability coefficient.

The IR subscale has a low to moderate degree of internal consistency when using a normal/school population, producing a Cronbach's alpha coefficient of 0.69. When a combined sample is used, the coefficient is much higher (.85, $N = 1676$). An even more drastic difference was reported by Burlingame, et al. (1996) between the school sample and a combined sample. They report a .69 ($N = 41$) for the student sample and .89 ($N = 1199$) for the school combined with clinical and community samples. This suggests that the lower coefficient obtained in this study are due to range restriction.

The acceptable range is usually within the .80 to .90 range. The IR subscale consists of questions 4, 7, 11, 16, 19, 24, 27, 36, 37, and 43. These items measure the child's relationship with family, friends, and adults. It includes symptoms such as arguing, cooperation with rules,

aggressiveness toward adults, fighting, and relations with friends. This coefficient alpha suggests that the IR subscale items are measuring a set of symptoms that inter-correlate.

Somatic (S) Subscale

The S subscale had a moderate to low degree of test-retest reliability. The Pearson r coefficient for this subscale is 0.67. As with the other scales, this restricted range may be spuriously low due to the range restriction. The low degree of stability may also be a factor of the symptoms that it is measuring. Since it is measuring somatic complaints such as headaches, vomiting, and stomach aches—all symptoms that may be transitory or short-lived in nature—the low test-retest reliability may be due to the changing nature of the symptoms measured.

The S subscale shows a low to moderate degree of internal consistency with an alpha coefficient of 0.62 ($N = 427$). When combined with a more diverse sample, the coefficient is higher (.78, $N = 1676$). This scale includes items 2, 10, 18, 26, 35, 42, 50, and 54. The symptoms assessed are items such as stomach pains, enuresis/encopresis, vomiting, headaches, weakness in joints, and constipation or diarrhea. The moderate to low internal consistency score suggests that these items are not measuring a single construct. It could be that a child may experience one or more of these symptoms, but that having one of these somatic problems does not predict the existence of another symptom. As with the CI subscale, a low internal consistency is not necessarily a negative finding. The S subscale was not developed to measure a single construct, but to measure the amount of various somatic complaints.

Another explanation for the low alpha coefficient found in this study is the restricted range problem. A restricted range and negatively skewed distribution both lower the correlation coefficient and make the results unclear. Taking into account the higher coefficient with the

more diverse sample, it suggests that the lower internal consistency estimate obtained in this study is due to range restriction.

Social Problems (SP) Subscale

The SP subscale shows a moderate degree of test-retest reliability with a Pearson's r of 0.71. The low stability coefficient may be due to the restricted range. With the amount of skewness and kurtosis of the score distribution, it is likely that this estimate is lower than the actual reliability of the subscale.

The SP subscale's internal consistency coefficient found in this study is just 0.50. This is consistent with the school sample reported in the Administration and Scoring Manual for the Youth Outcome Questionnaire 2.0 (Burlingame, et. al, 1996). They report a Cronbach's alpha coefficient of 0.51 (N = 41). However, when combined with the patient and community samples, Burlingame, et al. (1996) report an alpha coefficient of .84 (N = 1199). The correlation coefficient also improved drastically from .50 to .82 when the data from this study were combined with a more diverse sample.

As with the other subscales, this distribution was both skewed and restricted in range (see Table 7, Appendix D). These two qualities lower the correlation coefficients. The higher coefficient when using a more diverse sample supports the hypothesis that the lower score is due to range restriction.

The items that make up the SP score are 6, 13, 22, 29, 31, 39, 47, and 55 (see Appendix A for a copy of the YOQ™ 2.0). The symptoms include school truancy, stealing, alcohol/drug use, running away from home, deliberate rule-breaking, inappropriate sexual

behavior, and appropriate guilt for wrong-doing. The moderate degree of internal consistency suggests that these behaviors often tend to inter-correlate.

Retest Effect

The trend to have lower scores on the second administrations is consistent across the scales when all elapsed times are combined. The trend also seems to decrease as the time elapsed increases (see Table 8, Appendix D). This retest effect may be due to a number of reasons or confounding factors, but its consistency from such a large sample suggests that there is a systematic retest effect.

The fact that there is a trend in the YOQ-Total score (and some of the subscales) to decrease as time elapses suggests that the memory of the first administration plays a part in how the person answers the test items the second time. As time increases and the person does not have the items in memory, the retest effect seems to decrease.

The retest effect, though interesting and important to consider when interpreting change in scores, does not have an effect on the test-retest correlation coefficient providing the change is systematic.

Practical Implications of Results

The data from this study found that the YOQ-Total score has the highest degree of both test-retest reliability ($r = 0.81$, $N = 194$) and internal consistency (Cronbach's $\alpha = 0.93$, $N = 427$). This suggests that the YOQ Total score is the one that will be the most reliable, stable estimate of a child's psychosocial functioning. The two-week test-retest correlation coefficient ($r = 0.84$, $N = 56$) is similar to the Child Behavior Checklist's (Achenbach, 1991) two-week correlation of 0.89 (Kramer & Conoley, 1992).

The test-retest correlations for the subscales are similar to those for the Conner's Rating Scales (Conners, 1990). The CRS do not produce a total score, but report test-retest correlation coefficients for the individual subscales range from 0.91 to 0.33. The YOQ subscales fall within a similar range from 0.57 (CI) to 0.78 (BD). Therefore, the YOQ is as stable over time as other similar scales that are considered to be adequate in reliability.

The CBCL (Achenbach, 1991) does not report internal consistency coefficients for their measure (Kramer & Conoley, 1992). The CRS does not produce a total score, but the internal consistency coefficients range from 0.61 to 0.95. The YOQ subscales fall within a similar range from 0.50 (SP) to 0.82 (BD). So though some of the YOQ subscales are low to moderate in their internal consistency, the scores are comparable to the widely used CRS.

The YOQ does appear to have adequate test-retest reliability and internal consistency. The YOQ-Total score is highest in stability and internal consistency with the subscales ranging from low to high in both reliability factors. The lower scores for test-retest reliability may be largely based on a restricted range of values from the sample used. Low internal consistency estimates obtained in this study also seems to be due to a restricted range of values and a the negatively skewed scores. This project supports the YOQ use as a brief screening and outcome assessment measure to track the psychosocial health of children and adolescents.

Theoretical Considerations of Results

Contributions of the YOQ. Kazdin (1995) mentioned that there was a need to separate dysfunction into smaller domains so that research can more easily identify the effectiveness of treatment options for specific types of clinical presenting problems. One example given was that anxiety and depression are often seen as problems, but not often studied

in outcome research. This research supports the use of the YOQ Intrapersonal Distress subscale as an internally consistent and reliable subscale measuring symptoms of depression, anxiety, and hopelessness. The YOQ may be used to help develop and evaluate treatment options for dealing with children with different kinds of presenting problems. The YOQ, then, can add strength to our understanding and evaluation of specific treatments for narrower types of presenting problems.

The good reliability of the YOQ also fills a gap in the research as mentioned by Stoolmiller, et al. (1993). One call for improvement in the methodology of outcome assessment is measuring outcome at multiple points instead of only the pretest/post-test design. The YOQ's test-retest reliability, as supported by this study, makes it an ideal tool for assessing psychotherapy outcome at multiple points. As the validity and sensitivity to change research is completed, further evaluation of the YOQ as an effective multiple-point outcome tool will be possible.

Retest effect. The consistent decrease in scores from test to retest suggests that there is a systematic retest effect. As outcome measures become more widely used, this effect should be researched more fully to assure that changes in test scores are not lowering simply due to this retest effect.

Data gathering and statistical considerations. This project is largely an applied study and not one that adds much new to the theory of child outcome assessment. However, some theoretical implications are apparent from the data gathering with a normal sample. This study shows the amount of error that can come from gathering a normal sample to study the reliability of a measure that is made to assess psychosocial problems in a clinical population.

The problem with using a normal sample to gather reliability information is two-fold. The first problem is the restriction of range. The results of this study show the drastic restriction in range that using a normal population can cause. This range restriction causes spuriously low test-retest reliability coefficients.

A second implication is that the reliability coefficients could be spuriously high. This could happen if the YOQ scores are looking stable because they are so low. It could be that at the higher score range, the YOQ scores become less stable. So again, gathering a sample that includes a clinical population would increase the range and provide more information.

However, there are also problems with using a clinical sample.

The main problem with using a clinical sample is the reason why this study selected out children receiving treatment. The measure was developed to track change in a clinical population. Therefore, it would be hoped that the YOQ scores would change over a short period of time in a clinical population and in that way be less stable than in a normal population. So using a clinical population also poses some problems.

Recommendations for Further Research

One recommendation for further research would be gathering further test-retest reliability data from a more ethnically diverse sample. This would be helpful in assessing more accurately the test-retest and internal consistency reliability for the YOQ and the generalizability of the results.

Another direction for future reliability assessment would be to gather a sample including parents of children from the ages of 12 to 17. This age group may provide scores that are quite different from the present study's sample. The YOQ Manual (Burlingame, et al., 1996) reports

differences in scores for different age groups. Specifically, the mean for the YOQ-Total scores increases with age. To be able to extend the results of the test-retest reliability findings to all the age groups involved, an age group including these other ages would be essential.

Conclusion

Based on the results of this research, it is concluded that 1) the YOQ is a stable measure of psychotherapy outcome, and 2) the YOQ is internally consistent. This study recommends the use of the Youth Outcome Questionnaire for the assessment of psychotherapy outcome for children and adolescents. The YOQ is psychometrically sound and will be a valuable addition to the research of psychotherapy outcome for children. The stability of the YOQ supports its use in tracking change during psychotherapy.

CHAPTER 5: JOURNAL READY ARTICLE

Running head: YOQ RELIABILITY

The Reliability of the Youth Outcome Questionnaire

Matthew L. Whoolery

Brigham Young University

The Reliability of the Youth Outcome Questionnaire

Recent changes in this nation's mental health care system have led to the need for instruments which effectively track psychotherapy outcome. The increasing costs (U.S. Department of Commerce, 1992) and utilization of health care services have been markedly escalating since the early 1980's (Cummings, 1987). This had led to a shift in power from suppliers to consumers -- patients, employers, and third-party providers (Linder, 1991). As a result, psychology must begin to demonstrate its usefulness and cost-effectiveness to consumers (Leviton, 1996). In addition to these concerns, there is a need for assessment tools which can be used for continual monitoring and endpoint results of psychotherapy. Thus, reliable, valid assessment instruments, which show sensitivity to patient change over short periods of time, are essential (Burlingame, Lambert, Reisinger, Neff, & Mosier, 1995).

While a great deal of psychotherapy outcome research has been conducted for the adult population, few studies are directed toward children and adolescents (Kazdin, 1991). It has been suggested that children compose approximately 30-35% of the total psychotherapy population (Latkowski, personal communication, February, 1996). However, several researchers report only 6% (Weisz, Weiss, Alicke, & Klotz, 1987) and 7% (Kazdin, Bass, Ayers, & Rodgers, 1990) of outcome research concentrates on child psychotherapy outcome. In addition, a significant amount of money is spent in the treatment of child/adolescent patients within managed care settings (Kelleher & Long, 1994). This magnifies the need for more child outcome research and effective, cost-efficient outcome measures for children.

Characteristics of Mental Health Outcome Measures

Vermillion and Pfeiffer (1993) identify four important characteristics which outcome measures should possess. First, the measures should have *known reliability and validity estimates*. Second, the measures should be *practical* (brief, easily administered and scored). Third, they should be *suitable for the population* with whom they are used (i.e., they measure what is relevant and informative to the researcher). Finally, outcome measures should demonstrate *sensitivity to meaningful change*. This is particularly relevant to psychotherapy outcome because a measure is needed which can be employed frequently to detect slight changes following participation in therapy (Lambert & Hill, 1994).

Shortcomings of Current Child Outcome Measures.

Although numerous child assessment instruments are currently available, few have been designed for the sole purpose of child psychotherapy outcome measurement. Most have been developed for psychodiagnostic purposes. Thus, even though they may be used for outcome assessment, these instruments may lack some of the aforementioned characteristics (Hope, Atkin, Hoag, Mosier, Burlingame, & Wells, 1996) As a result, their effectiveness as outcome measures for children is somewhat questionable.

The Youth Outcome Questionnaire.

The Youth Outcome Questionnaire (YOQ, Burlingame, Wells, & Lambert, 1996) was developed to remedy present limitations of current measures being used for outcome assessment with children. It is brief, easy to administer and score, and is available at a low cost. Research is currently being conducted to assess reliability, validity, and sensitivity to change.

The YOQ is a 64-item parent-report questionnaire that requires parents to describe their child's recent behavioral and emotional status (see Appendix A). The parent reads each

statement and rates how well it describes their child on a five-point Likert-type scale ranging from zero to four. It is designed to assess client progress in therapy across five domains of psychosocial functioning. The five domains include: Intrapersonal Distress (depressed, anxious, withdrawn behavior), Somatization (physical problems which suggest internalizing behavior), Interpersonal Relations (aggressive and argumentative behavior with peers and adults), Social Problems (behavior which suggests lack of regard for social norms), and Behavior Disturbances (behavior characterized by hyperactivity, impulsiveness, attention deficits).

Research on the YOQ is promising, yet incomplete. The internal consistency of the YOQ Total score, using Cronbach's alpha (Cronbach, 1951) is reported at .96 (N= 421; Wells, et al., 1996). This high alpha coefficient suggests that there is one strong single factor underlying the six subscales of the YOQ. Discriminant Analysis has found evidence for discriminant validity. The YOQ Total Score highly differentiated between each distinct outpatient, inpatient, and community samples (Wells, et al., 1996).

Future research is aimed at continuing to gather normative samples and assessing the YOQ's sensitivity to change between several sessions. This study was conducted to assess the internal consistency and test-retest reliability of the YOQ.

Rationale for Reliability Study

In reviewing the diversity of outcome measures which have been employed for outcome assessment, it seems clear that many lack the sound psychometric properties necessary to produce reliable and valid data. Furthermore, many instruments have not been studied to establish psychometric properties (Froyd, Lambert, & Froyd, 1996). The ability of an instrument to *consistently measure what it purports to measure* is what we are referring to

when we speak of reliability. Without this essential psychometric component, one can not be sure that the outcome instrument has consistently and accurately measured patient change.

Test-retest reliability measures how consistent a measure is over time. The same test is administered twice and the scores are correlated, yielding a coefficient of stability. High stability over time shows that the test will reliably measure the same thing on separate occasions. Kaplan & Saccuzzo (1982) say "Reliability is one of the basic foundations of behavioral research. If a test is not reliable, it will not be possible to demonstrate that it has any meaning" (p. 114). Establishing reliability (stability) with an instrument which tries to be sensitive to clinical change must subsequently use a population which remains stable over time (i.e., a normal population).

For a measure that has subscales to measure different domains of functioning, internal consistency is especially important. The internal consistency coefficient is "the mean of all possible split-half reliability estimates. It reflects the degree to which the item content is similar" (Lemke & Wiersma, 1976).

Internal consistency will assess whether the subscales are homogeneous, or are measuring the same domain. If the subscales are measuring a single domain, the "test scores will be less ambiguous" (Anastasi, 1988, p. 122) and provide accurate information concerning the various areas of psychosocial functioning. The range of acceptable scores for the internal consistency coefficient is similar to that for the test-retest coefficient. A low internal consistency score indicates that the total measure or individual sub-tests are measuring a heterogeneous set of symptoms.

As test developers strive to develop a psychometrically sound instrument, proof of reliability is crucial if the instrument is to become an effective measure utilized in outcome

assessment. This study will assess the test-retest reliability by administering the YOQ to the same normal population at two separate times. The first administration of the YOQ will be used to assess the internal consistency of the YOQ.

Methods

Subjects

Subjects included a normal community sample obtained from an elementary school. This sample contained parents of children, kindergarten through fifth grade (ages five through twelve), totaling 423 subjects. Demographic information on Table 1 shows relatively equal amounts of female/male participants in each age group. Also, it appears that a majority of children were Caucasian, with parents who had completed High School, with incomes in the upper-middle category.

Table 1

Demographic Characteristics of Sample

| Age (N = 427) | | Gender (N = 427) | | Ethnicity (N = 358) | | Income (N = 353) | | Education (N = 348) |
|------------------|----|---------------------|-----|------------------------|-------|---------------------|--------|------------------------|
| mean | SD | M | F | White | Other | mean | SD | H.S. Grad |
| 8 | 2 | 196 | 231 | 97% | 3% | 47,141 | 17,484 | 99% |

Procedures

The community sample was obtained from parents of an elementary school sample. A newsletter which briefly described the study, informed parents that a packet will soon be sent

home with each student. Two days later, each student in the school was sent home with a packet, to be completed by a consenting parent or guardian. The packet contained a cover letter from the principal, a Consent and Confidentiality form, the three test measures (i.e., the YOQ, the CBCL, and the CPRS-93), and an information sheet. On the information sheet, parents who wished to continue participation (by filling out another YOQ in two to four weeks) gave their name, address, and phone number (if desired). The students then returned the completed packets to the school and were rewarded with a small candy bar. After the collection process was complete, parents who wished to continue participation were sent a YOQ and self-addressed stamped envelope. The test-retest sub-sample was divided in half and sent YOQ's for the two-week and four-week retest. Then they mailed back the completed YOQ. Questionnaires of parents with children receiving psychological treatment were not used for this study, since a normal sample is needed. The questionnaires were then entered into a computer and scored, according to the instructions for each questionnaires. Money was subsequently donated to the school and a newsletter was sent home thanking parents and notifying them of the amount of money donated to the school.

Results

This research concentrated on two questions: 1) Do the subscales and the total YOQ remain stable over time for a normal population?, and 2) Are the separate subscales and the entire YOQ internally consistent?

Test-Retest Reliability

The Pearson Product Moment Correlation (r) was used to assess the test/retest reliability of the YOQ Total score and each of the six subscales. Table 2 summarizes the results.

The YOQ Total score's test-retest reliability is 0.81 ($N = 194$). The subscale scores range from 0.57 for the CI (Critical Items) to 0.78 for the BD (Behavioral Dysfunction). All test/retest correlation coefficients were significant at the $p < .001$ level. The correlation scores for each individual administration (2, 4, and 6 weeks) can be seen in Appendix B. A Test of Difference Between Two Independent Correlation Coefficients found that the coefficients for the 2, 4, and 6 weeks administrations were not significantly different. Therefore, they were combined into one score representing all elapsed times.

An acceptable range for the test-retest reliability coefficient of a psychological measure is approximately .80 or higher (Aiken, 1994). The YOQ-Total score shows a good degree of stability. The subscales vary from low to moderate in their test-retest reliability.

Table 2

Test-Retest Correlation Coefficients For All Elapsed Times

| | Mean | N | SD | SE | r | Sig. |
|-------------|-------|-----|-------|------|-----|------|
| Total YOQ | 16.70 | 194 | 22.02 | 1.58 | .81 | .000 |
| Total YOQ-2 | 14.35 | 194 | 22.18 | 1.59 | | |
| BD | 5.78 | 194 | 6.88 | .49 | .78 | .000 |
| BD-2 | 5.06 | 194 | 6.56 | .47 | | |
| CI | 2.09 | 194 | 2.35 | .17 | .57 | .000 |
| CI-2 | 1.86 | 194 | 2.16 | .16 | | |
| ID | 6.17 | 194 | 7.94 | .57 | .76 | .000 |
| ID-2 | 5.36 | 194 | 7.68 | .55 | | |
| IR | -.49 | 194 | 4.23 | .30 | .62 | .000 |
| IR-2 | -.80 | 194 | 4.61 | .33 | | |
| S | 2.80 | 194 | 3.15 | .23 | .72 | .000 |
| S-2 | 2.70 | 194 | 2.85 | .20 | | |
| SP | .35 | 194 | 2.74 | .20 | .75 | .000 |
| SP-2 | .18 | 194 | 2.60 | .19 | | |

Internal Consistency

To answer question 2, Cronbach's (1951) coefficient alpha statistical method was used to assess the internal consistency of the YOQ Total score and its subscales. The results are summarized in Table 3.

Table 3

Internal Consistency Coefficients for YOQ Total score and Subscales

| | N of Cases | N of items | Coefficient Alpha |
|-----------|------------|------------|-------------------|
| YOQ Total | 427 | 64 | .93* |
| BD | 427 | 11 | .82* |
| CI | 427 | 9 | .61* |
| ID | 427 | 18 | .87* |
| IR | 427 | 10 | .69* |
| S | 427 | 8 | .62* |
| SP | 427 | 8 | .50* |

* significant at $p < .01$, two-tailed

The internal consistency estimate for the YOQ Total score is 0.93 (N = 427). The coefficient alpha scores for the subscales ranges from 0.50 for the SP (Social Problems) subscale to 0.87 for the CI (Critical Items) subscale. The coefficient alpha scores are all significant at the $p < .01$ for the YOQ Total score and all of the subscales. A good range for the reliability coefficient of

a psychological measure is approximately .80 or higher (Aiken, 1994). The YOQ Total score and the subscales BD and ID have a good amount of internal consistency. The IR subscale is in the low-adequate range at .69 (N = 427). The remainder of the subscales are still significantly different than zero, but do not meet the commonly accepted range of .70 to .90 or higher.

Discussion

This research reports on the development of a brief questionnaire that can be used to assess change in children receiving psychological treatment. The goal of the creators of the YOQ was to create a cost-effective, parent-report symptomatic measure of psychosocial functioning in children that is psychometrically sound.

Results demonstrated that the overall questionnaire and its subscales possess good stability. The lower test-retest coefficients were likely due in part to the range restriction as a result of using a normal population. Extremely restricted ranges such as this one produce spuriously low test-retest coefficients (Kenney, 1987; Anastasi, 1988).

The YOQ Total had the highest of the internal consistency coefficients (.93, N = 427). These results suggest that the test as a whole measures a similar, common variable. The subscales ID, BD, and IR have moderate to good internal consistency, supporting the use of these subscales as measuring a discreet variable. The subscales CI, S, and SP have low internal consistency estimates. These subscales likely measure a more heterogeneous set of symptoms.

Future Directions

Future research is aimed at getting in-patient and out-patient samples containing children with a wider variety of scores and ages (i.e., including children over 12 years). It is anticipated

that the test-retest correlation coefficients will be stronger in a sample with a wider variety of scores and ages.

Conclusions

Based on the results of this research, it is concluded that 1) the YOQ is a stable measure of psychotherapy outcome, and 2) the YOQ is internally consistent. This study recommends the use of the Youth Outcome Questionnaire for the assessment of psychotherapy outcome for children and adolescents. The YOQ is psychometrically sound and will be a valuable addition to the research of psychotherapy outcome for children. The stability of the YOQ supports its use in tracking change during psychotherapy

References: Chapter 5

- Anastasi, A. (1988). Psychological Testing (6th Ed.). New York: Macmillan Publishing Company.
- Burlingame, G. B., Lambert, M. J., Reisinger, C. W., Neff, W. L., Mosier, J. (1995). Pragmatics of Tracking Mental Health Outcomes in Managed Care Setting. The Journal of Mental Health Administration, 22, 226-236.
- Burlingame, G.M, Wells, M.G., & Lambert, M.J. (1996). Youth Outcome Questionnaire. American Professional Credentialing Services: Stevenson, MD.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Cummings, N. A. (1987). The future of psychotherapy: One psychologist's perspective. American Journal of Psychotherapy, 41, 349-360.
- Froyd, J. E., Lambert, M. J., & Froyd, J. D. (1996). A review of practices of psychotherapy outcome measurement. Journal of Mental Health, 5, 12-15.
- Hope, C. A., Atkin, Q., Hoag, M. J., Mosier, J. I., Burlingame, G. M., Wells, M. G. (1996). Outcome Assessment for Children and Adolescents: Review and Recommendations. Paper presented at the annual Western Psychological Association Conference, San Jose, California, in April.
- Kaplan, R.M. & Saccuzzo, D.P. (1982). Psychological Testing. Monterey, CA: Brooks/Cole Publishing Company.
- Kazdin, A. E. (1991). Effectiveness of Psychotherapy With Children and Adolescents. Journal of Consulting And Clinical Psychology, 59, 785-798.

Kelleher, K. & Long, N. (1994). Barriers and new directions in mental health service research in the primary care setting. Journal of Clinical Child Psychology, 23(2), 133-142.

Kenney, D.A. (1987). Statistics for the social and behavioral sciences. Boston: Little, Brown, and Company.

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), Handbook of psychotherapy and behavior change, (4th ed.), (pp. 72-113). New York: John Wiley.

Lemke, E. & Wiersma, W. (1976). Principles of psychological measurement. Chicago: Rand McNally College Publishing Company.

Leviton, L. C. (1996). Integrating psychology and mental health: challenges and opportunities. American Psychologist, 51(1), 42-51.

Linder, J. C. (1991). Outcome measurement: Compliance tool or strategic initiative? Health Care Management Review, 16, 21-33.

U.S. Department of Commerce (1992). Survey of Current Business. Washington, DC.

Vermillion, J., & Pfeiffer, S. (1993). Treatment outcome and continuous quality improvement: Two aspects of program evaluation. Psychiatric Hospital, 24, 9-14.

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., Hoag, M., & Hope, C. (1996).
Conceptualization and measurement of patient change during psychotherapy: Development of
the Youth Outcome Questionnaire (YOQ). Psychotherapy: Theory, Research, and Practice.

References

- Achenbach, T. M. (1991). Manual for the Child Behavior Checklist and 1991 Profile. Burlington: University of Vermont, Department of Psychiatry.
- Aiken, L.R. (1994). Psychological testing and assessment. (8th ed.). Massachusetts: Simon & Schuster, Inc.
- Allen, J.S.Jr., Tarnowski, K.J., Simonian, S.J., Elliott, D., & Drabman, R.S. (1991). The generalization map revisited: assessment of generalized treatment effects in child and adolescent behavior therapy. Behavior Therapy, *22*, 393-405.
- Anastasi, A. (1976). Psychological testing (4th ed.). New York: Macmillan Publishing Co., Inc.
- Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan Publishing Co., Inc.
- Atkin, Q., Whoolery, M.L., Peterson, G., Burlingame, G.M., Wells, M.G., & Nebekar, R.S. (1997, April). Reliability and validity of the Youth Outcome Questionnaire. Paper presented at annual convention of the Western Psychological Association, Seattle, WA.
- Barnett, R.J., Docherty, J.P., & Frommelt, G.M. (1991). A review of child psychotherapy since 1963. Journal of American Academy of Child and Adolescent Psychiatry, *30*(1), 1-14.
- Bergin, A.E. & Garfield, S.L. (1994). Handbook of psychotherapy and behavior change (4th Edition). New York: Wiley & Son.

Burlingame, G. B., Lambert, M. J., Reisinger, C. W., Neff, W. L., Mosier, J. (1995). Pragmatics of Tracking Mental Health Outcomes in Managed Care Setting. The Journal of Mental Health Administration, 22, 226-236.

Burlingame, G.M., McCollam, P., Nebekar, S., Wells, M.G., Hoag, M.J., Hope, C.A., & Mosier, J. (1996, April). The reliability and validity of the Youth Outcome Questionnaire. Paper presented at the annual convention of the Western Psychological Association. San Jose, CA.

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., Ferr, R., & Reisinger, C. (1996). Administration and scoring manual for the Youth Outcome Questionnaire 2.0. American Professional Credentialing Services: Stevenson, MD.

Burlingame, G.M, Wells, M.G., & Lambert, M.J. (1996). Youth Outcome Questionnaire. American Professional Credentialing Services: Stevenson, MD.

Conners, C.K. (1990). Conners' Rating Scales Manual. North Towands, NY: Multi-Health Systems.

Cozby, P.C. (1985). Methods in behavioral research. (3rd ed.). London: Mayfield Publishing Company.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Cummings, N. A. (1987). The future of psychotherapy: One psychologist's perspective. American Journal of Psychotherapy, 41, 349-360.

Drotar, D., Stein, R.E.K., & Perrin, E.C. (1995). Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research.

Journal of Clinical Child Psychology, 24(2), 184-192.

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(2), 141-148.

Hendren, R.L. (1993). Adolescent psychotherapy research: A practical review.

American Journal of Psychotherapy, 47(3), 334-343.

Hope, C. A., Atkin, Q., Hoag, M. J., Mosier, J. I., Burlingame, G. M., Wells, M. G.

(1996). Outcome Assessment for Children and Adolescents: Review and Recommendations.

Paper presented at the annual Western Psychological Association Conference, San Jose, California, in April.

Hopkins, K.D., Glass, G.V., & Hopkins, B.R. (1987). Basic statistics for the

behavioral sciences, (2nd ed.). New Jersey: Prentice Hall, Inc.

Kaplan, R.M. & Saccuzzo, D.P. (1982). Psychological Testing. Monterey, CA:

Brooks/Cole Publishing Company.

Kazdin, A. E. (1991). Effectiveness of Psychotherapy With Children and Adolescents.

Journal of Consulting And Clinical Psychology, 59, 785-798.

Kazdin, A.E. (1993). Psychotherapy for children and adolescents: current progress

and future research directions. American Psychologist, 48(6), 644-657.

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(2), 125-140.

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(6), 729-740.

Kelleher, K. & Long, N. (1994). Barriers and new directions in mental health service research in the primary care setting. Journal of Clinical Child Psychology, 23(2), 133-142.

Kendall, P.C. & Morris, R.J. (1991). Child therapy: issues and recommendations. Journal of Consulting and Clinical Psychology, 39(6), 777-784.

Kenney, D.A. (1987). Statistics for the social and behavioral sciences. Boston: Little, Brown, and Company.

Kramer, J.J. & Conoley, J.C. (1992). The eleventh mental measurements yearbook. Lincoln, Nebraska: The University of Nebraska Press.

Lambert, M.J., & Burlingame, G.M. (1996). Outcome Questionnaire 45.2. American Professional Credentialing Services: Wilmington, DE.

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), Handbook of psychotherapy and behavior change, (4th ed.), (pp. 72-113). New York: John Wiley.

Lemke, E. & Wiersma, W. (1976). Principles of psychological measurement. Chicago: Rand McNally College Publishing Company.

Leviton, L. C. (1996). Integrating psychology and mental health: challenges and opportunities. *American Psychologist*, *51*(1), 42-51.

Linder, J. C. (1991). Outcome measurement: Compliance tool or strategic initiative? *Health Care Management Review*, *16*, 21-33.

McCall, R.B. (1994). Fundamental statistics for behavior sciences. (6th ed.). New York: Harcourt Brace College Publishers.

Quay, H.C. (1983). A dimensional approach to behavior disorder: The revised Behavior Problem Checklist. *School Psychology Review*, *12*, 244-249.

Seligman, M.E.P. (1995). The effectiveness of psychotherapy: the Consumer Reports study. *American Psychologist*, *50*(12), 965-974.

Stoolmiller, M., Duncan, T., Bank, L., & Patterson, G.R. (1993). Some problems and solutions in the study of change: significant patterns in client resistance. *Journal of Consulting and Clinical Psychology*, *61*(6), 920-928.

Tramontana, M.G. (1980). Critical review of research on psychotherapy outcome with adolescents: 1967-1977. *Psychological Bulletin*, *88*(2), 429-450.

U.S. Department of Commerce (1992). Survey of Current Business. Washington, DC.

Vermillion, J., & Pfeiffer, S. (1993). Treatment outcome and continuous quality improvement: Two aspects of program evaluation. *Psychiatric Hospital*, *24*, 9-14.

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. & Latkowski, M. (1996, October). Tracking child and adolescent psychotherapy outcome with the Youth Outcome Questionnaire. Paper presented at the Kansas Conference in Clinical Child Psychology, Lawrence, KS.

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 Outcome Questionnaire (YOQ). Psychotherapy: Theory, Research, and Practice.

Appendix A: Youth Outcome Questionnaire

YOQ™ 2.0**PURPOSE:**

The YOQ.2 is designed to describe a wide range of troublesome situations, behaviors, and moods that are common in children and adolescents. You may discover that some of the items do not apply to your child's current situation. If so, please do not leave these items blank but circle the "Never or almost never" category. When you begin to complete the YOQ.2 you will see that you can easily make your child look as healthy or unhealthy as you wish. Please do not do that. If you are as accurate as possible it is more likely that you will be able to receive the help that you are seeking for your child.

DIRECTIONS:

- Read each statement carefully.
- Decide how true this statement is for your child during the past **7 days**.
- Circle the number that most accurately describes your child during the past week.
- Circle only one answer for each statement and erase unwanted marks clearly.

| 0 | 1 | 2 | 3 | 4 |
|-------------------------------|-------------|----------------|-----------------|---------------------------------|
| Never or almost never true | Rarely true | Sometimes true | Frequently true | Almost always or always true |

PLEASE COMPLETE BOTH SIDES.

My Child:

- | | | | | | |
|---|---|---|---|---|--|
| 0 | 1 | 2 | 3 | 4 | 1. Wants to be alone more than other children of the same age. |
| 0 | 1 | 2 | 3 | 4 | 2. Complains of dizziness or headaches. |
| 0 | 1 | 2 | 3 | 4 | 3. Doesn't participate in activities that were previously enjoyable. |
| 0 | 1 | 2 | 3 | 4 | 4. Argues or is verbally disrespectful. |
| 0 | 1 | 2 | 3 | 4 | 5. Is more fearful than other children of the same age. |
| 0 | 1 | 2 | 3 | 4 | 6. Cuts school or is truant. |
| 0 | 1 | 2 | 3 | 4 | 7. Cooperates with rules and expectations. |
| 0 | 1 | 2 | 3 | 4 | 8. Has difficulty completing assignments, or completes them carelessly. |
| 0 | 1 | 2 | 3 | 4 | 9. Complains or whines about things being unfair. |
| 0 | 1 | 2 | 3 | 4 | 10. Experiences trouble with her/his bowels, such as constipation or diarrhea. |
| 0 | 1 | 2 | 3 | 4 | 11. Gets into physical fights with peers or family members. |
| 0 | 1 | 2 | 3 | 4 | 12. Worries and can't get certain ideas off his/her mind. |
| 0 | 1 | 2 | 3 | 4 | 13. Steals or lies. |
| 0 | 1 | 2 | 3 | 4 | 14. Is fidgety, restless, or hyperactive. |
| 0 | 1 | 2 | 3 | 4 | 15. Seems anxious or nervous. |
| 0 | 1 | 2 | 3 | 4 | 16. Communicates in a congenial and appropriate manner. |
| 0 | 1 | 2 | 3 | 4 | 17. Seems tense, easily startled. |
| 0 | 1 | 2 | 3 | 4 | 18. Soils or wets self. |
| 0 | 1 | 2 | 3 | 4 | 19. Is aggressive toward adults. |
| 0 | 1 | 2 | 3 | 4 | 20. Sees, hears, or believes things that are not real. |
| 0 | 1 | 2 | 3 | 4 | 21. Has participated in self-harm (e.g. cutting or scratching self, attempting suicide). |
| 0 | 1 | 2 | 3 | 4 | 22. Uses alcohol or drugs. |
| 0 | 1 | 2 | 3 | 4 | 23. Seems unable to get organized. |
| 0 | 1 | 2 | 3 | 4 | 24. Enjoys relationships with family and friends. |

| 0 | 1 | 2 | 3 | 4 |
|-------------------------------|-------------|----------------|-----------------|---------------------------------|
| Never or almost never true | Rarely true | Sometimes true | Frequently true | Almost always or always true |

My Child:

- | | |
|-----------|---|
| 0 1 2 3 4 | 25. Appears sad or unhappy. |
| 0 1 2 3 4 | 26. Experiences pain or weakness in muscles or joints. |
| 0 1 2 3 4 | 27. Has a negative, distrustful attitude toward friends, family members, or other adults. |
| 0 1 2 3 4 | 28. Believes that others are trying to hurt him/her even when they are not. |
| 0 1 2 3 4 | 29. Threatens to, or has run away from home. |
| 0 1 2 3 4 | 30. Experiences rapidly changing and strong emotions. |
| 0 1 2 3 4 | 31. Deliberately breaks rules, laws, or expectations. |
| 0 1 2 3 4 | 32. Appears happy with her/himself. |
| 0 1 2 3 4 | 33. Sulks, pouts, or cries more than other children of the same age. |
| 0 1 2 3 4 | 34. Pulls away from family or friends. |
| 0 1 2 3 4 | 35. Complains of stomach pain or feeling sick more than other children of the same age. |
| 0 1 2 3 4 | 36. Doesn't have or keep friends. |
| 0 1 2 3 4 | 37. Has friends of whom I don't approve. |
| 0 1 2 3 4 | 38. Believes that others can hear her/his thoughts, or that s/he can hear the thoughts of others. |
| 0 1 2 3 4 | 39. Engages in inappropriate sexual behavior (e.g. sexually active, exhibits self, sexual abuse towards family members or others). |
| 0 1 2 3 4 | 40. Has difficulty waiting his/her turn in activities or conversations. |
| 0 1 2 3 4 | 41. Thinks about suicide, says s/he would be better off if s/he were dead. |
| 0 1 2 3 4 | 42. Complains of nightmares, difficulty getting to sleep, oversleeping, or waking up from sleep too early. |
| 0 1 2 3 4 | 43. Complains about or challenges rules, expectations, or responsibilities. |
| 0 1 2 3 4 | 44. Has times of unusual happiness or excessive energy. |
| 0 1 2 3 4 | 45. Handles frustration or boredom appropriately. |
| 0 1 2 3 4 | 46. Has fears of going crazy. |
| 0 1 2 3 4 | 47. Feels appropriate guilt for wrongdoing. |
| 0 1 2 3 4 | 48. Is unusually demanding. |
| 0 1 2 3 4 | 49. Is irritable. |
| 0 1 2 3 4 | 50. Vomits or is nauseous more than other children of the same age. |
| 0 1 2 3 4 | 51. Becomes angry enough to be threatening to others. |
| 0 1 2 3 4 | 52. Seems to stir up trouble when bored. |
| 0 1 2 3 4 | 53. Is appropriately hopeful and optimistic. |
| 0 1 2 3 4 | 54. Experiences twitching muscles or jerking movement in face, arms, or body. |
| 0 1 2 3 4 | 55. Has deliberately destroyed property. |
| 0 1 2 3 4 | 56. Has difficulty concentrating, thinking clearly, or attending to tasks. |
| 0 1 2 3 4 | 57. Talks negatively, as though bad things are all his/her fault. |
| 0 1 2 3 4 | 58. Has lost significant amounts of weight without medical reason. |
| 0 1 2 3 4 | 59. Acts impulsively, without thinking of consequences. |
| 0 1 2 3 4 | 60. Is usually calm. |
| 0 1 2 3 4 | 61. Will not forgive her/himself for past mistakes. |
| 0 1 2 3 4 | 62. Lacks energy. |
| 0 1 2 3 4 | 63. Feels that he/she doesn't have any friends, or that no one likes him/her. |
| 0 1 2 3 4 | 64. Gets frustrated and gives up, or gets upset easily. |

Appendix B: Consent, Confidentiality, Information, and Questionnaire

CONSENT AND CONFIDENTIALITY

The purpose of this research is to evaluate a new psychological instrument which will be used to assess mental health and distress in children and adolescents. Your scores will contribute significantly to the development of this instrument.

This procedure will include completing three mental health questionnaires and an information sheet. These can usually all be completed in about thirty minutes. This can be done in your own home, and at your convenience. Then, parents who wish to continue participation will complete one more questionnaire in three weeks. This questionnaire will be sent to your home with a self-addressed stamped envelope. The benefits of participation include the chance to see and complete a questionnaire used by psychologists to assess youth's psychological status and functioning, as well as your contribution that will aid future psychotherapy research and assessment. The perceived risks of participation in this study are minimal, if any.

Participation in this study is voluntary. As a participant in this research study you have the following rights: (1) freedom from any pressure; (2) the right to refuse to participate at anytime; and (3) the right to see the results of the completed study.

The information in this questionnaire is personal. A number (rather than your name) will be associated with the questionnaires you complete. For parents who continue participation, and agree to be sent one more questionnaire - your names, phone numbers, and addresses will be crossed out once the completed questionnaire is sent back.

If you have any questions regarding this research, you may contact Dr. Gary Burlingame at 378-7557 or Dr. Gawain Wells at 378-6125. Or, if you have any further questions regarding your rights as a participant in a research project, you may contact Dr. Larry Wood, Chair of the Institutional Review Board, 1122 SWKT, Brigham Young University, Provo, UT 84602. Phone: 378-3405.

Your completed, returned packet is your consent to participate in Part I of the study. Completion of the second portion of the information sheet is your consent continue to participate in Part II of the study.

INFORMATION SHEET

PART I:

Instructions: Please read through the following pages, answering all questions to the best of your knowledge. Please note there may be material printed on both sides of the enclosed material. Also, if you are describing more than one of your elementary-school children, please use a separate packet for each child. This is very important. After you have completed filling out all the contents of the packet, please put them inside the same packet, and have your child return it to his/her teacher. **THANK YOU VERY MUCH** for your participation in this important research.

PLEASE FILL IN ALL THE BLANKS

Child's Age: ___ Child's Grade: ___

Child's Gender: ___ Child's Race: _____

Your Relationship to the Child (circle):

Mother Father Other: (please describe) _____

How long have you known this child (circle):

Less than 3 months More than 3 months

Does your child have any learning disabilities (circle)?

No Yes (please describe, if possible): _____

Estimated Yearly Family Income: _____

Number of Children: ___ Number of Children living at home: ___

Please circle the highest level of education you have received:

HIGH SCHOOL COLLEGE: 1 2 3 4 5 6 7 8 9 years Degree: _____

PART II:

If you would be willing to complete one more questionnaire (YOQ) in approx. three weeks, please indicate your name (not your child=s name), address, and phone number. This information will only be used to send you a YOQ. The YOQ should only take 7 min to complete. We will also send a self-addressed stamped envelope so it can be easily returned. Your personal information will subsequently be destroyed.

NAME OF PARENT/GUARDIAN: _____

Address: _____

Phone Number: _____

Is it okay to call and give you a reminder to return the YOQ if we have not

received it in a week (circle): YES NO

Dear Parent:

Thank you for agreeing to help us with this project. Enclosed should be a copy of the YOQ, and a self-addressed stamped envelope.

Please fill out the YOQ and PA entirely and put them in the envelope provided. There is no need to add postage.

Also, please answer the following questions and RETURN THIS PAPER with the others.

1. Is the child presently receiving counseling or therapy?

yes no

2. Has a significant trauma (i.e. death in the family, personal injury, etc) taken place over the past two weeks that you feel would affect the child's YOQ score?

yes no

3. Has your child started any medications during the past three weeks?

yes no

For those who indicated that we may call you, we will contact you in about a week if we have not yet received your YOQ.

THANKS AGAIN FOR YOUR HELP.

Appendix C

Table 4

Test-Retest Correlation Coefficients for 2-Weeks Elapsed Time

| | Mean | N | SD | SE | r | Sig. |
|-------------|-------|----|-------|------|-----|------|
| Total YOQ | 20.27 | 56 | 25.26 | 3.38 | .84 | .000 |
| Total YOQ-2 | 16.98 | 56 | 24.28 | 3.24 | | |
| BD | 6.93 | 56 | 6.63 | .89 | .82 | .000 |
| BD-2 | 5.13 | 56 | 6.67 | .89 | | |
| CI | 2.50 | 56 | 2.91 | .39 | .56 | .000 |
| CI-2 | 1.96 | 56 | 2.49 | .33 | | |
| ID | 6.73 | 56 | 8.80 | 1.18 | .76 | .000 |
| ID-2 | 6.14 | 56 | 8.31 | 1.11 | | |
| IR | -.14 | 56 | 4.76 | .64 | .75 | .000 |
| IR-2 | -.36 | 56 | 4.96 | .66 | | |
| S | 3.52 | 56 | 4.15 | .55 | .69 | .000 |
| S-2 | 3.30 | 56 | 3.32 | .44 | | |
| SP | .73 | 56 | 3.27 | .44 | .78 | .000 |
| SP-2 | .80 | 56 | 2.84 | .38 | | |

Table 5

Test-Retest Correlation Coefficients for 4-Weeks Elapsed Time

| | Mean | N | SD | SE | r | Sig. |
|-------------|-------|----|-------|------|-----|------|
| Total YOQ | 14.20 | 93 | 20.70 | 2.15 | .81 | .000 |
| Total YOQ-2 | 11.70 | 93 | 20.17 | 2.09 | | |
| BD | 5.01 | 93 | 6.81 | .71 | .77 | .000 |
| BD-2 | 4.67 | 93 | 6.13 | .64 | | |
| CI | 2.00 | 93 | 2.20 | .20 | .65 | .000 |
| CI-2 | 1.67 | 93 | 1.86 | .19 | | |
| ID | 5.51 | 93 | 7.75 | .80 | .79 | .000 |
| ID-2 | 4.47 | 93 | 7.29 | .76 | | |
| IR | -.73 | 93 | 4.00 | .42 | .58 | .000 |
| IR-2 | -1.24 | 93 | 4.17 | .43 | | |
| S | 2.40 | 93 | 2.43 | .25 | .67 | .000 |
| S-2 | 2.37 | 93 | 2.34 | .24 | | |
| SP | -.02 | 93 | 2.42 | .25 | .71 | .000 |
| SP-2 | -.25 | 93 | 2.39 | .25 | | |

Table 6
Correlation Coefficients for 6-Weeks Elapsed Time

| | Mean | N | SD | SE | r | Sig. |
|-------------|-------|----|-------|------|-----|------|
| Total YOQ | 17.4 | 45 | 20.09 | 3.00 | .76 | .000 |
| Total YOQ-2 | 16.53 | 45 | 23.28 | 3.47 | | |
| BD | 5.96 | 45 | 7.26 | 1.08 | .77 | .000 |
| BD-2 | 5.78 | 45 | 7.46 | 1.11 | | |
| CI | 1.78 | 45 | 1.76 | .26 | .49 | .001 |
| CI-2 | 2.07 | 45 | 2.32 | .35 | | |
| ID | 6.84 | 45 | 7.24 | 1.08 | .69 | .000 |
| ID-2 | 6.20 | 45 | 7.61 | 1.14 | | |
| IR | -.42 | 45 | 4.04 | .60 | .52 | .000 |
| IR-2 | -.42 | 45 | 5.04 | .75 | | |
| S | 2.71 | 45 | 2.95 | .44 | .83 | .000 |
| S-2 | 2.64 | 45 | 3.08 | .46 | | |
| SP | .56 | 45 | 2.63 | .39 | .76 | .000 |
| SP-2 | .27 | 45 | 2.61 | .39 | | |

Appendix D

Figure 1

Histogram for CI scores

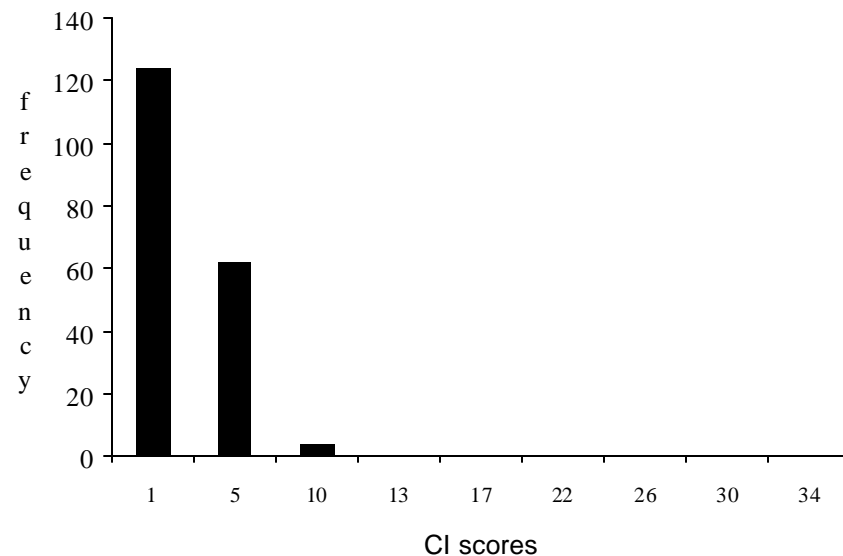


Figure 2

Scatterplot for CI scores

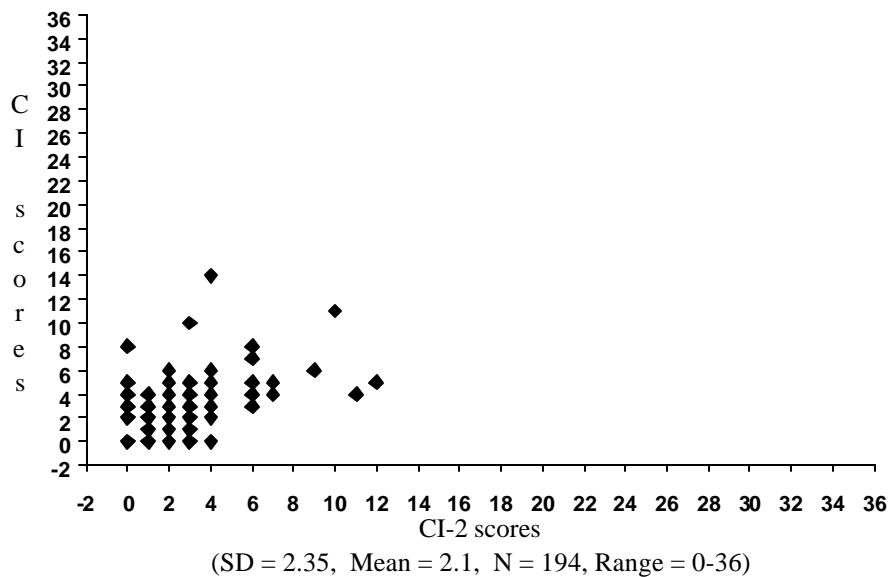


Figure 3

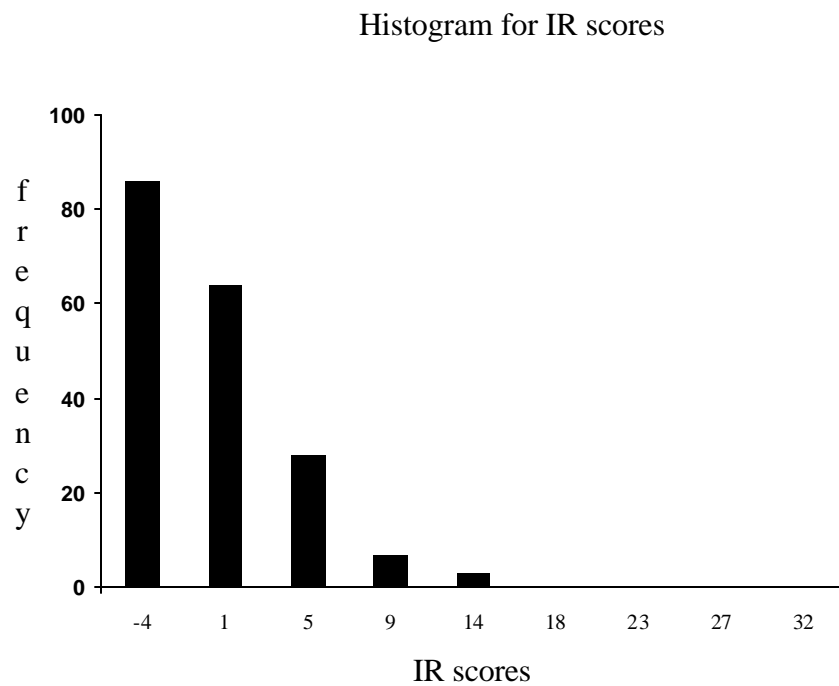


Figure 4

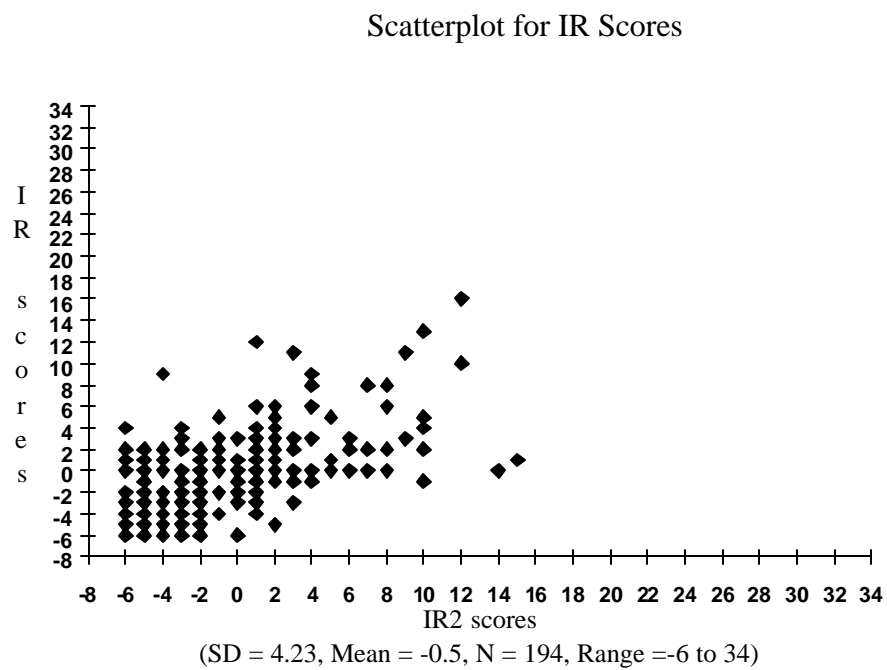


Table 7

Range, Skewness, and Kurtosis of YOQ-Total and Subscale Score Distributions

| | Mean | N | SD | Range (Possible) | Skewness | | Kurtosis | |
|-------------|-------|-----|-------|---------------------|----------|-----|----------|-----|
| | | | | | Stat. | SE | Stat. | SE |
| Total YOQ | 16.20 | 427 | 23.39 | 142 (256) | 1.24 | .12 | 2.10 | .24 |
| Total YOQ-2 | 14.35 | 194 | 22.18 | 124 (256) | 1.05 | .18 | 1.20 | .35 |
| BD | 5.29 | 427 | 6.98 | 42 (44) | 1.05 | .12 | 1.56 | .24 |
| BD-2 | 5.06 | 194 | 6.56 | 29 (44) | .73 | .18 | .07 | .35 |
| CI | 2.17 | 427 | 2.61 | 15 (36) | 1.67 | .12 | 3.38 | .24 |
| CI-2 | 1.86 | 194 | 2.16 | 12 (36) | 1.8 | .18 | 4.35 | .35 |
| ID | 6.16 | 427 | 8.54 | 42 (72) | 1.10 | .12 | .99 | .24 |
| ID-2 | 5.36 | 194 | 7.68 | 21(72) | 1.15 | .18 | 1.7 | .35 |
| IR | -.45 | 427 | 4.62 | 23 (40) | 1.04 | .12 | 1.12 | .24 |
| IR-2 | -.80 | 194 | 4.61 | 21 (40) | 1.09 | .18 | .84 | .35 |
| S | 2.75 | 427 | 3.12 | 17 (32) | 1.50 | .12 | 2.41 | .24 |
| S-2 | 2.70 | 194 | 2.85 | 17 (32) | 1.46 | .18 | 3.15 | .35 |
| SP | .27 | 427 | 2.48 | 15 (32) | 1.55 | .12 | 3.38 | .24 |
| SP-2 | .18 | 194 | 2.60 | 11 (32) | 1.36 | .18 | 1.09 | .35 |

Table 8

Retest Effect for YOQ Total Score and Subscales

| | Mean Change From First To Second Administration | | | |
|-----------------|---|-----------|----------|-------|
| | Two Week | Four Week | Six Week | All |
| YOQ Total Score | -3.29 | -2.50 | -0.87 | -2.35 |
| BD | -1.80 | -0.34 | -0.18 | -0.72 |
| CI | -0.54 | -0.33 | +0.29 | -0.23 |
| ID | -0.59 | -1.04 | -0.64 | -0.81 |
| IR | -0.22 | -0.51 | 0.00 | -0.31 |
| S | -0.22 | -0.03 | -0.07 | -0.10 |
| SP | +0.07 | -0.23 | -0.23 | -0.17 |

The Test-Retest Reliability and Internal Consistency
of the Youth Outcome Questionnaire

Matthew Lawrence Whoolery

Department of Educational Psychology

M.S. Degree, June 1997

ABSTRACT

The purpose of the present study was two-fold. First to assess the test-retest reliability of the Youth Outcome Questionnaire (YOQ)-Total score and the individual subscales, and second to estimate the internal or inter-item consistency of the entire YOQ and each subscale. To determine the reliability of the YOQ, parents of an elementary school (N = 427) in Salt Lake County, Utah were asked to fill out the YOQ for their child. Parents who were willing to continue filled out a second YOQ two or four weeks later (N = 194). The Pearson Product Moment Correlation (r) and Cronbach's coefficient alpha were used to assess the test/retest reliability and internal consistency of the YOQ Total score and its subscales.

This project supports the internal consistency and test-retest reliability of the Youth Outcome Questionnaire. The YOQ Total score is the most reliable score in both aspects ($r = .81$, $\alpha = .93$). The data gathered and analyzed in this study supports the use of the YOQ as an adequately stable and internally consistent parent-report measure of the psychosocial functioning of children and adolescents.

COMMITTEE APPROVAL:

Darwin Gale, Committee Chair

Gary Burlingame, Committee Member

Fred Rowe, Committee Member

Ronald Bingham, Department Chair