# Probation Referral Outcome:

## Inventory of Scientific Findings

The future of probationer assessment is recidivism prediction. Which we believe dovetails with treatment outcome or effectiveness. Specifically designed for probationer screening, the **Probation Referral Outcome (PRO)**. The **PRO** is available for inclusion in recidivism research. Contact us at **research@online-testing.com**.

Professional Online Testing Solutions, Inc.
**www.online-testing.com**

More in-depth Probation Referral Outcome (PRO) information is available at:
**www.probation-referral-outcome.com**

# TABLE OF CONTENTS

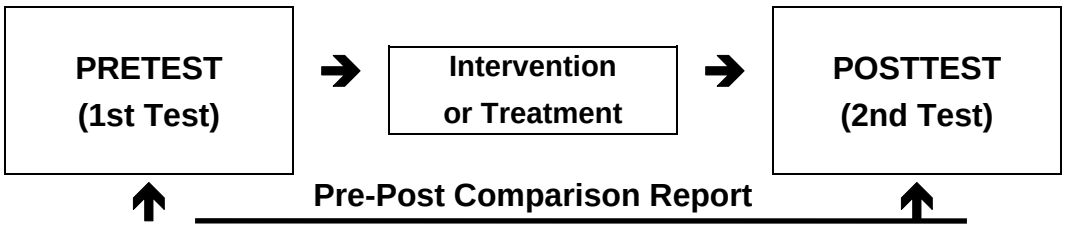
## INTRODUCTION

## PROBATION REFERRAL OUTCOME

In the past, probationers were assumed to have been rehabilitated, cured or made well by virtue of having completed an assigned or mandated treatment (intervention, counseling or psychotherapy) program. However, this is not necessarily the case, particularly when treatment is court ordered or mandated by a probation officer.

Judges, probation officers and others make counseling and treatment referrals. And when treatment is completed, the question invariably arises, "Was treatment effective?" The Probation Referral Outcome (PRO) helps answer that question.

The Probation Referral Outcome (PRO) is a probationer (offender) treatment outcome or treatment effectiveness test. The PRO is administered twice: at treatment intake (pretest) screening and again at treatment completion (posttest). The pretest serves as the baseline for posttest comparison.




The Probation Referral Outcome (PRO) consists of 165 items and takes 25 to 30 minutes to complete. The PRO has eight scales (measures): Truthfulness, Violence (Lethality), Alcohol, Drugs, Depression, Anxiety, Self-Esteem and Stress Management. The PRO assesses counseling, treatment and psychotherapy-related change. Some call this treatment outcome, whereas others refer to treatment effectiveness. Regardless of what it is called, the PRO assesses counseling and treatment **change**. Pretest-posttest comparisons embody change. And it is these positive and negative changes that represent treatment effects or change.

Research studies are presented chronologically within this document in the order of completion. **Recent studies are most representative of the PRO.** No attempt has been made to incorporate all PRO research into this document. However, it is representative of the reliability, validity and accuracy of the PRO.

The Probation Referral Outcome (PRO) is an automated computerized assessment instrument. It enables comparison of probationer status prior to, during and upon treatment completion. The proprietary PRO database ensures continued research and development. It includes true/false and multiple choice items and can be completed in 25 to 30 minutes. The PRO contains eight empirically based scales: Truthfulness, Alcohol, Drugs, Violence, Anxiety, Depression, Self-Esteem and Stress Management.

## PRO MEASURES (SCALES)

A description of each PRO scale follows.

### EIGHT PRO SCALES (MEASURES)

1. **Truthfulness Scale:** measures the truthfulness of the probationer while they were completing the PRO. This scale identifies self-protective, defensive or guarded people who minimize their problems or attempt to portray themselves in an overly favorable light. This type of scale is considered necessary, if not essential, in any objective assessment instrument. It would be very naïve to believe that all clients answer all assessment questions truthfully, especially when treatment is court-ordered. All interview and self-report test information is subject to the dangers of untrue answers due to defensiveness, guardedness, or deliberate falsification. The Truthfulness Scale also identifies reading-impaired probationers.

2. **Alcohol Scale:** The Alcohol Scale measures the probationer's alcohol proneness and alcohol-related problems. *Alcohol* refers to beer, wine and other liquor. This scale was developed with the assistance of experienced chemical dependency program staff. The Alcohol Scale is also in alignment with DSM-IV criteria for substance abuse and dependency. This scale is independent of the Drugs Scale.

3. **Drugs Scale:** The Drugs Scale measures drug use and the severity of abuse. *Drugs* refer to marijuana, crack, cocaine, amphetamines, barbiturates, heroin, ecstasy, etc. This scale also incorporates prescription drug abuse. As with the Alcohol Scale, the Drugs Scale was developed with the assistance of experienced chemical dependency program staff and also incorporates DSM-IV criteria for substance abuse and dependency. The Drugs Scale is an independent measure of the client's drug-related problems.

4. **Violence Scale:** Measures the client's use of physical force to injure, damage, or destroy. It identifies individuals that are dangerous to themselves and others.

An ever-present concern when evaluating probationers is lethality or violence potential. Violence is a significant problem in our society. The harm associated with violence (mental, emotional, and physical) is often under-reported by victims and family. And, there are some people who are "violence prone." Violent tendencies can be exacerbated by substance abuse and other co-occurring disorders.

5. **Anxiety Scale:** Anxiety is an unpleasant emotional experience characterized by non-directed fear. Most definitions of anxiety include a sympathetically induced feeling associated with a sense of threat. General symptoms such as nervousness, apprehension and tenseness are included in this definition, as are panic, terror, and somatic correlates of anxiety.

The Anxiety Scale provides a quantitative score that varies directly with client's self-reported symptoms. The presence, severity and magnitude of these symptoms is measured by client's multiple-choice answers, i.e., "rare or never", "sometimes", "often" or "very often".

Two symptom clusters - anxiety and depression - are clinically significant and consistently related in clinical literature. Anxiety and depression represent the most commonly reported symptoms of distress in clinical and counseling settings. The interaction or blending of these symptom clusters is evident in the definition of dysphoria, i.e., a generalized feeling of anxiety, restlessness and depression.

Perceived distress, whether by self or others, represents the major reason people seek help or are referred for counseling and assistance. Estimates of the prevalence of anxiety and depression in general medical practice are very high.

**6. Depression Scale:** Depression is a dejected emotional state that varies from normal to pathological proportions. General symptoms such as depressed mood are included in this definition, as are impaired social-vocational functioning and loss of interest in usual activities. In addition, thoughts of suicide and other cognitive and somatic correlates of depression are included.

Anxiety and depression are not mutually exclusive. Any given case may represent both symptom clusters. For this reason, separate anxiety and depression scales were included in the PRO. It is important to assess both anxiety and depression due to their prevalence in treatment, counseling, intervention and outcome.

**7. Self-Esteem Scale:** reflects a client's explicit valuing and appraisal of self. Self-esteem incorporates an attitude of acceptance-approval versus rejection-disapproval. Self-esteem refers to a person's perception of self. We learn to approve, praise or pardon our own actions, just as we learn to disapprove, feel guilt or condemn other actions. In summary, individuals react to themselves and evaluate their own behavior.

This scale consists of terms which are rated on a point scale to describe the probationer's self-esteem. This procedure is a rapid means of self-rating wherein the inmate describes his own self-esteem in words commonly used in everyday life. Self-esteem incorporates an attitude of acceptance-approval versus rejection-disapproval of oneself. Included are the probationer's attitudes, feelings, beliefs and perceptions of self. **The Self-Esteem Scale is descriptive of the person one believes oneself to be.**

**8. Stress Management Scale:** establishes how well the client copes with or manages stress. It is based on the Stress Quotient (SQ) findings presented in the "PRO Research" section that follows.

The Stress Management Scale is much more than just a measure of stress. It is a measure of how well the respondent copes with stress. Two people can be in the same stressful situation, however, one person is overwhelmed and the other person handles it well. The Stress Management Scale can account for these different reactions to stress.

The following studies summarize research conducted on a variety of clients, e.g., substance abuse inpatients/outpatients, probationers, etc.

Probation Referral Outcome (PRO) research is presented chronologically in the order it was conducted. Chronological presentation enables the reader to follow the evolution of the PRO into a state-of-the-art automated (computerized) screening instrument. More recent studies (toward the end of this document) are most representative of current PRO statistics.

## PRO RESEARCH

### STRESS QUOTIENT

The **Stress Quotient** (SQ) or **Stress Management Scale** is based upon the following mathematical equation:

$$SQ = CS/S \times k$$

The Stress Quotient (SQ) is a numerical value representing a person's ability to handle or cope with stress relative to their amount of experienced stress. CS (Coping Skill) refers to a person's ability to cope with stress. S (Stress) refers to experienced stress. k (Constant) represents a constant value in the SQ equation to establish SQ score ranges. The SQ includes measures of both stress and coping skills in the derivation of the Stress Quotient (SQ) score. The better an individual's coping skills, compared to the amount of experienced stress, the higher the SQ score.

The Stress Quotient (SQ) scale equation represents empirically verifiable relationships. The SQ scale (and its individual components) lends itself to research. Nine studies were conducted to investigate the validity and reliability of the **Stress Quotient** or **Stress Management Scale**.

**Validation Study 1**: This study was conducted (1980) to compare SQ scores between High Stress and Low Stress groups. The High Stress group (N=10) was comprised of 5 males and 5 females. Their average age was 39. Subjects for the High Stress group were randomly selected from outpatients seeking treatment for stress. The Low Stress group (N=10) was comprised of 5 males and 5 females (average age 38.7) randomly selected from persons not involved in treatment for stress. High Stress group SQ scores ranged from 32 to 97, with a mean of 64.2. Low Stress group SQ scores ranged from 82 to 156, with a mean of 115.7. The t-test statistical analysis of the difference between the means of the two groups indicated that the High Stress group had significantly higher SQ scores than the Low Stress group ($t = 4.9$, $p < .001$). This study shows that the SQ or Stress Management Scale is a valid measure of stress coping. The Stress Management Scale significantly discriminates between high stress individuals and low stress individuals.

**Validation Study 2**: This study (1980) evaluated the relationship between the SQ scale and two criterion measures: Taylor Manifest Anxiety Scale and Cornell Index. These two measures have been shown to be valid measures of anxiety and neuroticism, respectively. If the SQ or Stress Management Scale is correlated with these measures it would indicate that the SQ or Stress Management Scale is a valid measure. In the Taylor Manifest Anxiety Scale, high scores indicate a high level of anxiety. Similarly, in the Cornell Index high scores indicate neuroticism. Negative correlation coefficients between the two measures and the SQ were expected because high SQ scores indicate good stress management. The three tests were administered to forty-three (43) subjects selected from the general population. There were 21 males and 22 females ranging in age from 15 to 64 years. Utilizing a product-moment correlation, SQ scores attained a correlation coefficient of -.70 with the Taylor Manifest Anxiety Scale and -.75 with the Cornell Index. Both correlations were significant, in the predicted direction, at the $p < .01$ level. These results support the finding that the Stress Management Scale is a valid measure of stress management. The reliability of the SQ was investigated in ten subjects (5 male and 5 female) randomly chosen from this study. A split-half correlation analysis was conducted on the SQ items. The product-moment correlation coefficient (r) was .85, significant at the $p < .01$ level. This correlation indicates that the SQ or Stress Management Scale is a reliable measure. These results support the Stress Management Scale as a reliable and valid measure.

**Validation Study 3**: In this study (1981) the relationship between the SQ Scale and the Holmes Rahe Social Readjustment Rating Scale (SRRS) was investigated. The SRRS, which is comprised of a self-rating of stressful life events, has been shown to be a valid measure of stress. Three correlation analyses were done. SRRS scores were correlated with SQ scores and separately with two components of the SQ scale: Coping Skill (CS) scores and Stress (S) scores. It was hypothesized that the SQ and SRRS correlation would be negative, since subjects with lower SQ scores would be more likely to either encounter less stressful life events or experience less stress in their lives. It was also predicted that subjects with a higher CS would be less likely to encounter stressful life events, hence a negative correlation was hypothesized. A positive correlation was predicted between S and SRRS, since subjects

experiencing more frequent stressful life events would reflect more experienced stress. The participants in this study consisted of 30 outpatient psychotherapy patients. There were 14 males and 16 females. The average age was 35. The SQ and the SRRS were administered in counterbalanced order. The results showed there was a significant positive correlation (product-moment correlation coefficient) between SQ and SRRS ($r = .4006$, $p<.01$). The correlation results between CS and SRRS was not significant ($r = .1355$, n.s.). There was a significant positive correlation between S and SRRS ($r = .6183$, $p<.001$). The correlations were in predicted directions. The significant correlations between SQ and SRRS as well as S and SRRS support the construct validity of the SQ or Stress Management Scale.

**Validation Study 4**: This validation study (1982) evaluated the relationship between factor C (Ego Strength) in the 16 PF Test as a criterion measure and the SQ in a sample of adults. High scores on factor C indicate high ego strength and emotional stability, whereas high SQ scores reflect good coping skills. A positive correlation was predicted because emotional stability and coping skills reflect similar attributes. The participants were 34 adjudicated delinquent adolescents. They ranged in age from 15 to 18 years with an average age of 16.2. There were 30 males and 4 females. The Cattell 16 PF Test and the SQ scale were administered in counterbalanced order. All subjects had at least a 6.0 grade equivalent reading level. The correlation (product-moment correlation coefficient) results indicated that Factor C scores were significantly correlated with SQ scores ($r = .695$, $p<.01$). Results were significant and in the predicted direction. These results support the SQ or Stress Management Scale as a valid measure of stress management in adult offenders.

In a subsequent study, the relationship between factor Q4 (Free Floating Anxiety) on the 16 PF Test and S (Stress) on the SQ scale was investigated. High Q4 scores reflect free floating anxiety and tension, whereas high S scores measure experienced stress. A high positive correlation between Q4 and S was predicted. There were 22 of the original 34 subjects included in this analysis because the remaining original files were unavailable. All 22 subjects were male. The results indicated that Factor Q4 scores were significantly correlated (product-moment correlation coefficient) with S scores ($r = .584$, $p<.05$). Results were significant and in predicted directions. The significant correlations between factor C and SQ scores as well as factor Q4 and S scores support the construct validity of the SQ scale.

**Validation Study 5**: Psychotherapy outpatient clients were used in this validation study (1982) that evaluated the relationship between selected Wiggins MMPI (Minnesota Multiphasic Personality Inventory) supplementary content scales (ES & MAS) as criterion measures and the SQ scale. ES measures ego strength and MAS measures manifest anxiety. It was predicted that the ES and SC correlation would be positive, since people with high ego strength would be more likely to possess good coping skills. Similarly, it was predicted that MAS and S correlations would be positive, since people experiencing high levels of manifest anxiety would also likely experience high levels of stress. The subjects were 51 psychotherapy outpatients ranging in age from 22 to 56 years with an average age of 34. There were 23 males and 28 females. The MMPI and the SQ were administered in counterbalanced order. The correlation (product-moment correlation coefficient) results indicated that ES and CS were positively significantly correlated ($r = .29$, $p<.001$). MAS and S comparisons resulted in an r of .54, significant at the $p < .001$ level. All results were significant and in predicted directions.

In a related study (1982) utilizing the same population data (N=51) the relationship between the Psychasthenia (Pt) scale in the MMPI and the S component of the SQ scale was evaluated. The Pt scale in the MMPI reflects neurotic anxiety, whereas the S component of the SQ scale measures stress. Positive Pt and S correlations were predicted. The correlation (product-moment correlation coefficient) results indicated that the Pt scale and the S component of the SQ scale were significantly correlated ($r = .58$, $p<.001$). Results were significant and in the predicted direction. The significant correlations between MMPI scales (ES, MAS, Pt) and the SQ scale components (CS, S) support the construct validity of the SQ or Stress Management Scale.

**Reliability Study 6**: The reliability of the Stress Quotient (SQ) or Stress Management Scale was investigated (1984) in a population of outpatient psychotherapy patients. There were 100 participants, 41 males and 59 females. The average age was 37. The SQ was administered soon after intake. The most common procedure for reporting inter-item (within test) reliability is with Coefficient Alpha. The reliability analysis indicated that the Coefficient Alpha of 0.81 was highly significant ($F = 46.74$, $p<.001$). Highly significant inter-item scale consistency was demonstrated.

**Reliability Study 7**: (1985) The reliability of the Stress Quotient (SQ) or Stress Management Scale was investigated in a sample of 189 job applicants. There were 120 males and 69 females with an average age of 31. The SQ was administered at the time of pre-employment screening. The reliability analysis indicated that the Coefficient Alpha of 0.73 was highly significant ($F = 195.86$, $p<.001$). Highly significant Cronbach's Alpha reveals that all SQ scale items are significantly ($p<.001$) related and measure one factor or trait.

**Validation Study 8**: Chemical dependency inpatients were used in a validation study (1985) to determine the relation between MMPI scales as criterion measures and the Stress Quotient (SQ) Scale or Stress Management Scale. The SQ is inversely related to other MMPI scales, consequently, negative correlations were predicted. The participants were 100 chemical dependency inpatients. There were 62 males and 38 females with an average age of 41. The SQ and the MMPI were administered in counterbalanced order. The reliability analysis results indicated that the Coefficient Alpha of 0.84 was highly significant ($F = 16.20$, $p<001$). Highly significant inter-item scale consistency was demonstrated.

The correlation (product-moment correlation coefficient) results between the Stress Quotient (SQ) and selected MMPI scales were significant at the $p < .001$ level and in predicted directions. The SQ correlation results were as follows: Psychopathic Deviate (-0.59), Psychasthenia (-.068), Social Maladjustment (-0.54), Authority Conflict (-0.46), Taylor Manifest Anxiety Scale (-0.78), Authority Problems (-0.22), and Social Alienation (-0.67). The most significant SQ correlation was with the Taylor Manifest Anxiety Scale. As discussed earlier, stress exacerbates symptoms of impaired adjustment as well as emotional and attitudinal problems. These results support the Stress Quotient or Stress Management Scale as a valid measure of stress management.

**Validation Study 9**: In a replication of earlier research, a study (1986) was conducted to further evaluate the reliability and validity of the Stress Quotient (SQ). The participants were 212 inpatients in chemical dependency programs. There were 122 males and 90 females with an average age of 44. The SQ and MMPI were administered in counterbalanced order. Reliability analysis of the SQ scale resulted in a Coefficient Alpha of 0.986 ($F = 27.77$, $p<.001$). Highly significant inter-item scale consistency was again demonstrated. Rounded off, the **Coefficient Alpha for the SQ was 0.99**.

In the same study (1986, inpatients), product-moment correlations were calculated between the Stress Quotient (SQ) and selected MMPI scales. The SQ correlated significantly (.001 level) with the following MMPI scales: Psychopathic Deviate (Pd), Psychasthenia (Pt), Anxiety (A), Manifest Anxiety (MAS), Ego Strength (ES), Social Responsibility (RE), Social Alienation (PD4A), Social Alienation (SC1A), Social Maladjustment (SOC), Authority Conflict (AUT), Manifest Hostility (HOS), Suspiciousness/Mistrust (TSC-II), Resentment/Aggression (TSC-V) and Tension/Worry (TSC-VII). **All SQ correlations with selected MMPI scales were significant (at the .001 level of significance) and in predicted directions.** These results support the SQ scale or Stress Management Scale as a valid measure of stress management.

The studies cited above demonstrate empirical relationships between the SQ scale (Stress Management Scale) and other established measures of stress, anxiety and coping skills. This research demonstrates

that the Stress Quotient (SQ) or Stress Management Scale is a reliable and valid measure of stress management. The SQ has high inter-item scale reliability. The SQ also has high concurrent (criterion-related) validity with other recognized and accepted tests. The SQ scale permits objective (rather than subjective) analysis of the interaction of these important variables. In the research that follows, the **Stress Quotient** or **SQ** is also referred to as the **Stress Management Scale**.

## PROBATION REFERRAL OUTCOME (PRO) RESEARCH

Probation Referral Outcome (PRO) is designed for intake assessment as well as pre-treatment and post-treatment (or intervention) comparison. Probation departments need an objective, accurate, reliable, valid and impartial assessment instrument to augment decision making. The PRO scales evolved from scale items represented in other established assessment instruments. For example, the Truthfulness, Self-Esteem and Stress Management items largely evolved from the Pre-Post Inventory, which is an established clinical or counseling screening instrument. The Alcohol and Drugs Scale items evolved from the SAQ Adult Probation III, which is an established probationer screening instrument. The PRO has a long history of research and development, much of which is contained in the following summary. **PRO research is reported in a chronological format, reporting studies as they occurred.** For current information refer to the more recent studies near the end of this research section.

Initially, a large item pool was rationally developed for PRO scale consideration. Consensual agreement among three Ph.D. level psychologists and other experienced chemical dependency counselors familiar with PRO scale definitions reduced the initial item pool markedly. Final item selection was empirical - comparing statistically related item configurations to known substance abuse groups. Items chosen had acceptable inter-item reliability coefficients and correlated highest with their respective scales. Final item selection was based on each item's statistical properties. Items with the best statistical properties were retained. The PRO was then objectively standardized and normed on inpatient and outpatient chemical dependency clients, probationers and a variety of other counseling clients.

### 1. A Study of PRO Test-Retest Reliability

Any approach to detection, assessment, or measurement must meet the criteria of reliability and validity. Reliability refers to an instrument's consistency of results regardless of who uses it. This means that the outcome must be objective, verifiable, and reproducible. Ideally, the instrument or test must also be practical, economical, and accessible. Psychometric principles and computer technology help to ensure PRO accuracy, objectivity, cost-effectiveness and accessibility.

Reliability is a measure of the consistency of a test in obtaining similar results upon re-administration of the test. One measure of test reliability, over time, is the test-retest correlation coefficient. In this type of study, the test is administered to a group and then the same test is re-administered to the same group at a later date.

Method

College students at two different colleges enrolled in introductory psychology classes participated in this study (1984). A total of 115 students participated and received class credit for their participation. The students were administered the PRO in a paper-pencil test format. One week later they were re-tested with the PRO again.

Results

The results of this study revealed a significant test-retest product-moment correlation coefficient of $r = 0.71$, $p<.01$. These results support the reliability of the PRO. Test-retest consistency was very high

and indicates that the PRO scores are reproducible and reliable over a one week interval.

## 2. Validation of the Truthfulness Scale

The Truthfulness Scale in the PRO is an important psychometric scale as these scores establish how truthful the respondent is while completing the PRO. Truthfulness Scale scores determine whether or not PRO profiles are accurate and are integral to the calculation of Truth-Corrected PRO scale scores.

The Truthfulness Scale identifies respondents who are self-protective, recalcitrant and guarded, as well as those who minimized or even concealed information while completing the test. Truthfulness Scale items are designed to detect respondents who try to fake good or put themselves into an overly favorable light. These scale items are statements about oneself that most people would agree to. The following statement is an example of a Truthfulness Scale item, “Sometimes I worry about what others think or say about me.”

This preliminary study used the Truthfulness Scale items in the Probation Referral Outcome (PRO) to determine if these Truthfulness Scale items could differentiate between respondents who were honest from those that were trying to fake good. It was hypothesized that the group trying to fake good would score higher on the Truthfulness Scale than the group instructed to be honest.

Method

Seventy-eight Arizona State University college students (1985) enrolled in an introductory psychology class were randomly assigned to one of two groups. Group 1 comprised the “Honest” group and Group 2 comprised the “Fakers” group. Group 1 was instructed to be honest and truthful while completing the test. Group 2 was instructed to "fake good" while completing the test, but to respond "in such a manner that their faking good would not be detected." The Truthfulness Scale was embedded in the test as one of the five scales. Truthfulness Scale scores were made up of the number of deviant answers given to the 21 Truthfulness Scale items.

Results

The mean Truthfulness Scale score for the Honest group was 2.71 and the mean Truthfulness Scale score for Fakers was 15.77. The results of the correlation (product-moment correlation coefficient) between the Honest group and the Fakers showed that the Fakers scored significantly higher on the Truthfulness Scale than the Honest group ($r = 0.27$, $p < .05$).

The Truthfulness Scale successfully measured how truthful the respondents were while completing the test. The results of this study reveal that the Truthfulness Scale accurately detects "Fakers" from those students that took the test honestly.

## 3. Validation of Four Probation Referral Outcome Scales Using Criterion Measures

In general terms, a test is valid if it measures what it is supposed to measure. The process of confirming this statement is called validating a test. A common practice when validating a test is to compute a correlation between it and another (criterion) test that purports to measure the same thing and that has been previously validated. For the purpose of this study, the four Probation Referral Outcome scales (Truthfulness, Alcohol, Drugs and Stress Management) were validated with comparable scales on the Minnesota Multiphasic Personality Inventory (MMPI). The MMPI was selected for this validity study because it is the most researched, validated and widely used objective personality test in the United States. The PRO scales were validated with MMPI scales as follows. The Truthfulness Scale was validated with the L Scale. The Alcohol Scale was validated with the MacAndrews Scale. The Drugs Scale was validated with the MacAndrews and Psychopathic Deviant scales. The Anxiety Scale was

validated with the Taylor Manifest Hostility and Authority Conflict. The Stress Management Scale was validated with the Taylor Manifest Anxiety, Psychasthenia, Social Maladjustment and Social Alienation scales.

Method

One hundred (100) chemical dependency inpatients (1985) were administered both the PRO and the MMPI. Tests were counterbalanced for order effects -- half were given the PRO first and half were administered the MMPI first.

Results and Discussion

Product-moment correlation coefficients were calculated between PRO scales and MMPI scales. These results are summarized in Table 1. Correlation results presented in Table 1 show that all PRO scales significantly correlated (.001 level of significance) with all represented MMPI scales.

The **Truthfulness Scale** correlates significantly with all of the represented MMPI scales in Table 1. Of particular interest is this scale's highly significant positive correlation with the MMPI Lie (L) Scale. A high L Scale score on the MMPI invalidates other MMPI scale scores due to untruthfulness. This helps in understanding why the Truthfulness Scale is significantly, but negatively, correlated with the other represented MMPI scales. Similarly, the MMPI L Scale correlates significantly, but negatively, with the other PRO scales.

**Table 1. (1985) Product-moment correlations between MMPI scales and PRO scales**

| (MEASURES) | Truthful-ness | Alcohol | Drugs | Stress Mgmt |
|---|---|---|---|---|
| **L (Lie) Scale** | 0.72 | -0.38 | -0.41 | 0.53 |
| **Psychopathic Deviant** | -0.37 | 0.52 | 0.54 | -0.59 |
| **Psychasthenia** | -0.34 | 0.38 | 0.41 | -0.68 |
| **Social Maladjustment** | -0.25 | 0.34 | 0.26 | -0.54 |
| **Authority Conflict** | -0.43 | 0.31 | 0.47 | -0.46 |
| **Manifest Hostility** | -0.45 | 0.34 | 0.47 | -0.58 |
| **Taylor Manifest Anxiety** | -0.58 | 0.47 | 0.46 | -0.78 |
| **MacAndrews** | -0.40 | 0.58 | 0.62 | -0.33 |
| **Social Alienation** | -0.47 | 0.35 | 0.45 | -0.67 |

The **Alcohol Scale** correlates significantly with all represented MMPI scales. This is consistent with the conceptual definition of the Alcohol Scale and previous research that has found that alcohol abuse is associated with mental, emotional and physical problems. Of particular interest are the highly significant correlations with the MacAndrews (r = 0.58) Scale and the Psychopathic Deviant (r = 0.52) Scale. High MacAndrews and Psychopathic Deviant scorers on the MMPI are often found to be associated with substance abuse. Similarly, the **Drugs Scale** correlates significantly with the MacAndrews (r = 0.62) Scale and the Psychopathic Deviant (r = 0.54) Scale.

The **Stress Management Scale** is inversely related to MMPI scales which accounts for the negative correlations shown in Table 1. The positive correlation with the L scale on the MMPI was discussed

earlier, i.e., Truthfulness Scale. It should be noted that stress exacerbates symptoms of impaired adjustment and even psychopathology. The Stress coping Ability Scale correlates most significantly with the Taylor Manifest Anxiety (r = -0.78) Scale, the Psychasthenia (r = -0.68) Scale and the Social Alienation (r = -0.67) Scale.

These findings strongly support the validity of Probation Referral Outcome (PRO) scales. All of the PRO scales were highly correlated with the MMPI criterion scales they were tested against. The large correlation coefficients support the validity of the PRO. All product-moment correlation coefficients testing the relation between PRO scales and MMPI scales were significant at the $p < .001$ level.

**4. Inter-item Reliability of the Probation Referral Outcome**

Within-test reliability measures to what extent a test with multiple scales measuring different factors, measures each factor independent of the other factors (scales) in the test. It also measures to what extent items in each scale consistently measures the particular trait (or factor) that scale was designed to measure. Within-test reliability measures are referred to as inter-item reliability. The most common method of reporting within-test (scale) inter-item reliability is with Coefficient alpha.

Method

This study (2005) included three separate groups of subjects: 100 outpatients in private practice, 100 substance abuse inpatients, and 189 job applicants -- totaling 389 subjects. Separate inter-item reliability analyses were conducted to compare results across the three groups.

Results and Discussion

The inter-item reliability coefficient alpha and within-test reliability statistics are presented in Tables 2 and 3, respectively. All inter-item reliability coefficient alphas and within-test reliability F-values are significant at $p<.001$. These results support the reliability of the Probation Referral Outcome (PRO). The PRO is a highly reliable instrument.

These results (Table 2 and 3) demonstrate the impressive reliability of the PRO. Reliability was demonstrated with three different groups of people (outpatients, inpatients and job applicants) taking the PRO. In each of these subject samples, all PRO scales (measures) were found to be significantly independent of the other PRO scales as shown by the highly significant within-test F statistics. The F statistic is obtained in within-subjects between measures ANOVA performed on each individual PRO scale in each of the samples.

**Table 2. Inter-item reliability, coefficient alpha. (2005)**
**Outpatients, Substance Abuse Inpatients and Job Applicants (N = 389)**

| PRO SCALES MEASURES | Outpatients (N = 100) | Inpatients (N = 100) | Job Applicants (N = 189) |
|---|---|---|---|
| **Truthfulness Scale** | 0.81 | 0.79 | 0.81 |
| **Alcohol Scale** | 0.86 | 0.93 | 0.83 |
| **Drugs Scale** | 0.80 | 0.85 | 0.79 |
| **Violence Scale** | 0.84 | 0.79 | 0.78 |
| **Anxiety Scale** | 0.85 | 0.81 | 0.79 |
| **Depression Scale** | 0.83 | 0.84 | 0.78 |
| **Self-Esteem Scale** | 0.88 | 0.89 | 0.85 |
| **Stress Management** | 0.81 | 0.84 | 0.85 |

**Table 3. Within-test reliability, F statistic.**

| PRO SCALES MEASURES | Outpatients (N = 100) | Inpatients (N = 100) | Job Applicants (N = 189) |
|---|---|---|---|
| **Truthfulness Scale** | 21.73 | 53.15 | 45.91 |
| **Alcohol Scale** | 9.29 | 31.46 | 47.75 |
| **Drugs Scale** | 27.19 | 16.34 | 58.18 |

| | | | |
|---|---|---|---|
| **Violence Scale** | 14.63 | 10.81 | 20.12 |
| **Anxiety Scale** | 15.97 | 19.21 | 23.67 |
| **Depression Scale** | 12.64 | 29.27 | 22.69 |
| **Self-Esteem Scale** | 39.54 | 48.42 | 34.81 |
| **Stress Management** | 46.74 | 16.20 | 195.86 |

All F statistics are significant at $p<.001$.

The F statistics show that each Probation Referral Outcome (PRO) scale measures essentially one factor (or trait). In addition, all PRO scales show high inter-item reliability. This is demonstrated by the Standardized Cronbach's Coefficient alpha - a widely used test of inter-item reliability when using parallel models. This measure reveals that all items in each PRO scale are significantly related and measure just one factor. In other words, each PRO scale measures one factor, yet the factor being measured is different from scale to scale.

The inter-item reliability coefficients show very similar results across the three subject samples. The Truthfulness Scale, Alcohol Scale and Drugs Scale are in close agreement. The Stress Management Scale shows similar results for the chemical dependency groups but the job applicant group had a slightly lower coefficient alpha. This difference might be accounted for by the fact that individuals applying for a job would not want to show themselves in a bad light by indicating they have an emotional, substance abuse, stress-related or mental health problem.

**5. Relationships of Selected PRO Scales and Polygraph Examination**

A measure that has often been used in business or industry for employee selection is the Polygraph examination. The polygraph exam is most often used to determine the truthfulness or honesty of an individual while being tested. The Polygraph examination is more accurate as the area of inquiry is more "situation" specific. Conversely, the less specific the area of inquiry, the less reliable the Polygraph examination becomes.

Three Probation Referral Outcome scales were chosen for this study; Truthfulness Scale, Alcohol Scale and Drugs Scale. The Truthfulness Scale was chosen because it is used in the PRO to measure the truthfulness or honesty of the respondent while completing the PRO. The Alcohol and Drugs Scales are well suited for comparison with the polygraph exam because of the situation specific nature of the scales. Alcohol and drug items are direct and relate specifically to alcohol and drug use. The comparison with the Truthfulness Scale is less direct because of the subtle nature of the Truthfulness Scale items as used in the PRO. The respondent's attitude, emotional stability and tendencies to fake good affect the Truthfulness Scale. It was expected that the Alcohol and Drugs Scales would be highly correlated with the polygraph results and the Truthfulness Scale would show a somewhat less but nonetheless significant correlation.

Method

One hundred and forty-eight (148) job applicants (2005) were administered both the PRO scales and the

Polygraph examination. Tests were given in a counterbalanced order, half of the applicants were given the PRO scales first and the other half of the applicants were administered the polygraph first. The subjects were administered the PRO scales and polygraph exam in the same room in the same session with the examiner present for both tests.

Results

The product-moment correlation results between the Polygraph exam and PRO scales indicated there was a significant positive correlation between the Truthfulness Scale and Polygraph exam ($r = 0.23$, $p<.001$). Similarly, significant positive relationships were observed between the Polygraph exam and the Alcohol Scale ($r = 0.54$, $p<.001$) and the Drugs Scale ($r = 0.56$, $p<.001$).

In summary, this study supports the validity of the Probation Referral Outcome (PRO) Truthfulness Scale, Alcohol Scale and Drugs Scale. There were strong positive relationships between the selected PRO scales and the Polygraph examination. The highly significant product-moment correlations between PRO scales and Polygraph examinations demonstrate the validity of the PRO Truthfulness, Alcohol and Drugs measures.

These results are important because the Polygraph exam is a direct measure obtained from the individual being tested rather than a rating by someone else. This is similar to self-report such as utilized in the PRO. The fact that there was a very strong relationship between Polygraph results and PRO scales shows that this type of information can be obtained accurately in self-report instruments.

These results indicate that the PRO Truthfulness Scale is an accurate measure of the respondent's truthfulness or honesty while completing the PRO. The Truthfulness Scale is an essential measure in self-report instruments. There must be a means to determine the honesty or "correctness" of the respondent's answers and there must be a means to adjust scores when the respondent is less than honest. The PRO Truthfulness Scale addresses both of these issues. The Truthfulness Scale measures truthfulness and then applies a correction to other scales based on the Truthfulness Scale score. The Truthfulness Scale ensures accurate assessment. The results of this study show that the PRO is a valid assessment instrument.

## 6. Replication of PRO Reliability in a Sample of Inpatient Clients

In a replication of earlier Probation Referral Outcome (PRO) research, chemical dependency inpatients (2005) were used to evaluate the reliability of the PRO scales.

Method and Results

The PRO scales were administered to 162 inpatients in a chemical dependency facility. The inter-item coefficient alpha statistics are presented in Table 4. These results are in close agreement to reliability results obtained in an earlier study using chemical dependency inpatient clients. The results of the present study support the reliability of the PRO.

Within the subject samples studied, the PRO scales were demonstrated to be independent measures. This mutual exclusivity (significant at $p<.001$) was demonstrated by a within-subjects measures ANOVA performed on each PRO scale. These analyses demonstrate that each PRO scale measures one factor or trait. All PRO scales demonstrate high inter-item congruency, as reflected in the standardized Cronbach's Alpha.

The items on each PRO scale are significantly related to the factor or trait each scale was designed to measure. In other words, each PRO scale measures one factor, and the factor (or trait) being measured differs from scale to scale.

**Table 4. Inter-item reliability, coefficient alpha. Chemical dependency inpatients (2005, N = 162).**

| PRO SCALES | COEFFICIENT | F | P VALUE |
|---|---|---|---|

| MEASURES | ALPHA | VALUE | P< |
|---|---|---|---|
| **Truthfulness Scale** | 0.79 | 13.28 | 0.001 |
| **Alcohol Scale** | 0.92 | 24.39 | 0.001 |
| **Drugs Scale** | 0.87 | 22.23 | 0.001 |
| **Violence Scale** | 0.86 | 21.48 | 0.001 |
| **Anxiety Scale** | 0.81 | 10.92 | 0.001 |
| **Depression Scale** | 0.82 | 12.35 | 0.001 |
| **Self-Esteem Scale** | 0.88 | 24.06 | 0.001 |
| **Stress Management** | 0.99 | 27.77 | 0.001 |

**PRO scales (measures) have been shown to be both mutually exclusive and have high inter-item scale consistency. The PRO has acceptable and empirically demonstrated reliability. In addition, inter-item reliability studies have shown that each PRO scale is an independent measure of the trait (factor) it was designed to measure.**

### 7. Validation of PRO Scales Using DWI Evaluator Ratings

This study (2006) was designed to demonstrate the relationship between PRO scales and DWI evaluator ratings, i.e., concurrent validity. Participating DWI evaluators had over six years' expertise in DWI offender assessment. Evaluators were instructed to complete their normal and usual screening procedures "prior to rating" clients on the scales incorporated into the PRO, i.e., the Alcohol and Drugs Scales. Evaluators were "blind" in the sense that they did not have any knowledge of scale scores at the time of their ratings.

Method and Results

There were 652 DWI offenders included in this study (2006). The participants completed the PRO as part of normal DWI screening and evaluation procedures. Results of staff (evaluator) ratings and scale scores (Alcohol and Drugs Scales) are presented in Table 5. As shown in the table below, the product-moment correlation coefficients between staff ratings and scale scores are highly statistically significant at p<.001.

**Table 5. Agreement Coefficients between Evaluator Ratings and PRO Scale Scores (2006, N=652)**

| PRO SCALES | AGREEMENT COEFFICIENT | SIGNIFICANCE LEVEL |
|---|---|---|
| Alcohol Scale | .66 | P<.001 |
| Drugs Scale | .58 | P<.001 |

It should be noted that these experienced evaluators invested considerable time in reviewing available records and interviewing each client. In contrast, scale scores were arrived at after 10 to 15 minutes of testing time. These results strongly support the validity of the Alcohol and Drugs Scales. Concurrent (criterion related) validity is demonstrated.

In addition, product-moment correlations were computed between these scales and the MAST, Sandler and Court Screening Procedures used by these experienced evaluators. These results are represented in Table 6.

**Table 6. Product-moment correlations (2006, N=652)**
**Mast, Sandler, and Court Procedures**

| PRO SCALES | MAST | SANDLER | COURT PROCEDURE |
|---|---|---|---|
| Alcohol Scale | .67 | .46 | .81 |
| Drugs Scale | .39 | .11 | .36 |

These results support the validity (criterion) of the Probation Referral Outcome (PRO) scales (Alcohol and Drugs Scales). The highest coefficient is between the Alcohol Scale and Court Procedure, indicating that both procedures are essentially reflecting the same information. The Court Procedure involved a review of court records (DUI priors, BAC level, substance abuse-related convictions, MAST results and Sandler scores). These findings support the validity of the Alcohol and Drugs Scales. A positive correlation indicates that predictions from the test will be more accurate than guesses. Whether a validity coefficient is high enough to permit use of the test as a predictor, depends upon numerous factors, such as the importance of prediction and evaluation cost. And, any statistic varies from one sample to another. Even if subjects are drawn randomly from the same population, criterion coefficients will differ from sample to sample.

Using a large sample makes the correlation more dependable. Correlations between a test and criterion are called validity coefficients, coefficients of productivity and concurrent validity. Concurrent validity procedures involve administering a test and comparing test results with identifiable criterion of performance.

**8. Validation of PRO Scales Using the Mortimer-Filkins Test**

In this study (2006), PRO Alcohol and Drugs Scale scores were validated with Mortimer-Filkins total scores. The product-moment correlations are presented in Table 7. There were 1,299 participants included in the study.

**Table 7. Product-moment correlations. (2006, N = 1,299)**
**Mortimer-Filkins versus PRO Alcohol And Drugs Scales**

| PRO Measures | First Sample Coefficients | Second Sample Coefficients |
|---|---|---|
| Alcohol Scale | .469 | .344 |
| Drugs Scale | .260 | .289 |

The Mortimer-Filkins total score correlate highly significantly ($p<.001$) with the PRO Alcohol Scale and Drugs Scale. These high correlations support the validity of the Alcohol and Drugs Scales.

**9. Validation of PRO Scales Using the MacAndrews Scale**

This study (2006) evaluated relationships between the MacAndrews Scale (in the Minnesota Multiphasic Personality Inventory) and the PRO Alcohol Scale and Drugs Scale. Product-moment correlations are reported in Table 8. There were 1,181 participants included in the study.

**Table 8. Product-moment correlations. (2006, N = 1,181)**
**MacAndrews Scale versus PRO Alcohol and Drugs Scales**

| | Significance |
|---|---|

| PRO Measures | MacAndrews | Level |
|---|---|---|
| Alcohol Scale | .1680 | P<.02 |
| Drugs Scale | .1696 | P<.02 |

A positive correlation is demonstrated between the MacAndrews Scale and the Probation Referral Outcome (PRO) Alcohol Scale and Drugs Scale. These results support the concurrent validity of the PRO Alcohol Scale and the Drugs Scale.

**10. Validation of PRO Scales Using DRI Scales as Criterion Measures**

This study (2007) compared the Driver Risk Inventory (DRI) with the Probation Referral Outcome (PRO). The DRI has been demonstrated to be a valid, reliable and accurate DWI offender assessment instrument. The PRO is designed for treatment intake assessment and pretest-posttest comparisons. It contains seven measures or scales: Truthfulness, Alcohol, Drugs, Distress, Resistance, Self-Esteem and Stress management. Four of these seven PRO scales are analogous (although independent) and directly comparable to DRI measures or scales. The DRI is designed for DWI offender evaluation.
The DRI contains five measures or scales: Truthfulness, Alcohol, Drugs, Driver Risk and Stress Management. Although the scales designated Truthfulness, Alcohol, Drugs and Stress Management are independent and differ in the PRO and DRI, they were designed to measure similar behaviors or traits. Thus, although essentially composed of different test questions in the PRO and DRI test booklets, these comparable measures or scales do have similarity.

Method
The PRO and DRI were administered in group settings to 154 adult offenders, in counter balanced order. All of the subjects in this study were male inmates. The demographic composition was as follows. There were 98 Caucasians, 25 Hispanics, 13 American Indians, 12 Blacks and six other ethnicities'. Five age categories were represented: 16-25 years (N = 26), 26-35 years (N = 74), 36-55 years (N = 38), 46-55 years (N = 11) and 56 or older (N = 5). Six educational levels were represented: Eighth grade or less (N = 7), Partially completed high school (N = 50), High school graduates (N = 70), Partially completed college (N = 16), College graduates (N = 9), and Professional/graduate school (N = 2). Each participant completed both the PRO and the DRI. Although all inmates volunteered to participate in this study, inmate motivation varied.

Results and Discussion
The results of this study are presented in Table 9. The results demonstrate highly significant relationships between the analogous PRO and DRI scales. The DRI has been shown to be a valid measure of substance abuse in DWI offenders; hence, the following correlation results support the validity of the PRO.

It was noted that inmate motivation varied widely. This is evident in the Stress Management correlation coefficient of .7642. Even though this is a significant correlation ($p<.001$), the Agreement Coefficient could be expected to be even higher because these scales were nearly identical and only differed by the number of test items. It is reasonable to conclude that low motivation on the part of many inmate volunteers contributed to

lower Agreement Coefficients. Inmate volunteers were serving DWI-related sentences and these tests had no bearing on their incarcerated status or sentences. However, in spite of widely varied inmate motivation, Agreement Coefficients for all five sets of scale comparisons were highly significant.

**Table 9. Product-moment correlations 2007 study of male inmates (N = 154).**

| DRI versus PRO Scales | Agreement Coefficients |
|---|---|
| Truthfulness Scale | .6405 |
| Alcohol Scale | .3483 |
| Drugs Scale | .3383 |
| Stress Management | .7642 |

All product-moment correlations are significant at p<.001.

These results support the relationships between independent, but analogous DRI and PRO scales. Correlation coefficients for this study are presented in Table 10. And, these concurrent validity findings support the accuracy of the PRO Truthfulness Scale, Alcohol Scale, Drugs Scale, and Stress Management Scale. These PRO scales measure what they were intended to measure.

**11. Validation of the PRO Self-Esteem Scale**

This study (2007) evaluated ratings of experienced counselors and the Probation Referral Outcome (PRO) Self-Esteem Scale. These counselors had at least 10 years' experience and an MA degree in counseling.

Two counselors independently rated each client's self-esteem. They reviewed client outpatient files containing court histories, progress notes, diagnoses, MMPI and Incomplete Sentence materials. Each patient was interviewed for a minimum of 30 minutes. Product-moment correlation coefficients were calculated for each rater and are presented in Table 10.

| Table 10. Staff Ratings and PRO Self-Esteem Scale (2007, N=89) | | |
|---|---|---|
| **Product-moment correlation coefficients significant at p<.05.** | | |
| **PRO Scale** | **First Rater** | **Second Rater** |
| Self-Esteem | .12 | .19 |

The results of this study show that staff ratings of client's self-esteem and the PRO Self-Esteem Scale are statistically significantly correlated. These results support the accuracy of the PRO Self-Esteem Scale. Even though this study was completed over a six month period, all comparisons were significant.

**12. Validation of the PRO with MMPI Scales as Criterion Measures**

This study (2007) validated PRO scales using analogous scales from the MMPI. The PRO Truthfulness Scale was correlated with the MMPI L (Lie) Scale. The PRO Alcohol Scale and Drugs Scale were correlated with the MMPI MacAndrews Scale and Psychopathic Deviate Scale. The PRO Stress Management Scale was correlated with the Hypomania (Mam) and Taylor Manifest Anxiety (MAS) Scales. The PRO Self-Esteem Scale was correlated with the Psychasthenia (PT) and the Social Alienation (SOA) Scales.

Method and Results

The participants in this study (2007) were 118 chemical dependency inpatients. Tests were administered in counterbalanced order. Product-moment correlation coefficients between analogous Probation Referral Outcome (PRO) and MMPI scale scores are discussed individually.

The **Truthfulness Scale** (L, r=0.72) correlates highly significantly with the MMPI Lie (L) Scale.

Although independent of each other, the MMPI - L Scale and the PRO - Truthfulness Scale are conceptually similar. Each consists of items that most people agree or disagree with. And, they both determine client honesty. The **Alcohol Scale** correlates significantly with the MacAndrews Alcohol (ALC, r=0.58) Scale and the Psychopathic Deviate (PD, r=0.52) Scale. The **Drugs Scale** correlates significantly with the MacAndrews (ALC, r=0.62) Scale and the Psychopathic Deviate (PD, r=0.54) Scale. High PD and ALC scores on the MMPI are often associated with substance abuse. The **Stress Management Scale** correlates significantly with the Hypomania (Mam r=0.37) and Taylor Manifiest Anxiety (MAS, r=0.78) Scales. The **Self-Esteem Scale** correlates significantly with the Psychasthenia (PT, r=0.34) and the Social Alienation (SOA, r=0.36) Scale.

All correlations were highly statistically significant. These results strongly support the validity of the Probation Referral Outcome (PRO). Validity refers to a test measuring what it is purported to measure. The PRO is an accurate assessment instrument. The PRO measures what it is designed to measure.

### 13. A Study of PRO Reliability in a Sample of Inpatients

The present study (2008) was conducted to investigate reliability of PRO scales in a sample of outpatient participants.

Method and Results

There were 227 adult outpatient participants included in the study. This sample is summarized as follows: Gender (149 males, 65.9% and 78 females, 34.4%). Age: 18 or less (10, 4.4%); 19 through 29 (77, 33.9%); 30 through 39 (97, 42.7%); 40 through 49 (33, 14.5%); 50 through 59 (6, 2.6%) and 60 + (4, 1.8%). Ethnicity: Caucasian (151, 66.5%); Black (27, 11.9%); Hispanic (44, 19.4%); Native American (4, 1.8%); and Other (1, 0.4%). Education: 8th grade or less (20, 8.8%); Partially Completed High School (67, 29.5); G.E.D. (16, 7.0%); High School Graduate (78, 34.4%); Partially Completed College (33, 14.5%); Technical/Business School (3, 1.3%); College Graduate (9, 4.0%) and Professional/Graduate School (1, 0.4%). Marital Status: Single (126, 55.5%); Married (61, 26.9%); Divorced (30, 13.2%); Separated (6, 2.6%) and Widowed (4, 1.8%). Reliability coefficient alphas are presented in the Table 14.

**Table 11. Reliability coefficient alphas. Outpatients (2008, N=227)**

| PRO Scales | Coefficient Alpha | Significance Level |
|---|---|---|
| Truthfulness Scale | .87 | P<.001 |
| Alcohol Scale | .90 | P<.001 |
| Drugs Scale | .89 | P<.001 |
| Violence Scale | .88 | P<.001 |
| Anxiety Scale | .89 | P<.001 |
| Depression Scale | .91 | P<.001 |
| Self-Esteem Scale | .95 | P<.001 |
| Stress Management Scale | .92 | P<.001 |

These results are in close agreement with reliability coefficient alphas found in previous PRO studies. These results again demonstrate the internal consistency of the Probation Referral Outcome.

### 14. Reliability of the PRO in a Large Sample of Outpatients

The purpose of the present study (2008) was to test the reliability of PRO scales in a large sample of

outpatients.

Method and Results

The PRO was administered to 887 adult outpatient participants as part of routine evaluation programs. Subjects were administered PRO scales individually in paper-pencil test format. There were 663 males and 224 females. The demographic composition of this sample is summarized as follows. Age: 18 or less (65, 7.3%); 19 to 29 (335, 37.8%); 30 to 39 (321, 36.2%); 40 to 49 (113, 12.8%); 50 to 59 (34, 3.8%) and 60 + (18, 2.0%). Ethnicity: Caucasian (615, 69.4%); Black (181, 20.4%); Hispanic (66, 7.4%); Asian (7, 0.8%); Native American (13, 1.5%) and Other (4, 0.5%). Education: 8th grade or less (40, 4.5%); Partially Completed High School (201, 25.0%); G.E.D. (7, 8.2%); High School Graduate (255, 27.4%); Partially Completed College (204, 23.1%); Technical/Business School (13, 1.5%); College Graduate (46, 5.2%); Professional/Graduate School (45, 5.1%). Marital Status: Single (488, 55.1%); Married (217, 24.4%); Divorced (102, 11.5%); Separated (63, 7.1%); Widowed (15, 1.7%).

Reliability coefficient alphas are presented in Table 12.

This study supports the reliability of the Probation Referral Outcome (PRO). The PRO produces similar results upon repetition. The PRO is a reliable outcome assessment instrument.

**Table 12. Reliability coefficient alphas. Outpatients (2008, N=887)**

| PRO Scales | Coefficient Alpha | Significance Level |
|---|---|---|
| Truthfulness Scale | .89 | P<.001 |
| Alcohol Scale | .90 | P<.001 |
| Drugs Scale | .91 | P<.001 |
| Violence Scale | .87 | P<.001 |
| Anxiety Scale | .90 | P<.001 |
| Depression Scale | .87 | P<.001 |
| Self-Esteem Scale | .91 | P<.001 |
| Stress Management Scale | .92 | P<.001 |

## 15. PRO Reliability, Scale Risk Range Accuracy and Gender Differences

This study (2008) was conducted to examine the reliability, gender differences and accuracy of PRO scales in a sample of adult participants. The participants completed the PRO pretest at intake prior to beginning their counseling programs. Reliability of the PRO, gender differences in client scale scores and risk range percentile score accuracy was investigated in the present study.

Method and Results

**The subjects in this study consisted of 174 adult counseling clients.** Demographic composition of these participants is as follows: Males: 140 (80.5%); Females: 34 (19.5). Age: 19 & under (3%); 20-29 (35%); 30-39 (33%); 40-49 (20%); 50-59 (7%) and 60 & over (2%). Ethnicity: Caucasian (90%); Black (6%); Hispanic (2%) and Other (2%). Education: Eighth grade or less (3%); Some H.S. (15%); H.S. graduate (67%) and Some college (14%). Marital Status: Single (53%); Married (26%); Divorced (14%); Separated (5%) and Widowed (2%).

Accuracy of the PRO

Risk range percentile scores are calculated for each PRO scale. These risk range percentile scores are derived from scoring equations based on responses to scale items and Truth-Corrections, then converted to percentile scores. There are four risk range categories: **Low Risk** (zero to 39th percentile), **Medium Risk** (40 to 69th percentile), **Problem Risk** (70 to 89th percentile) and **Severe Problem or Maximum Risk** (90 to 100th percentile). Risk range percentile scores represent degree of severity.

Analysis of the accuracy of PRO risk range percentile scores involves comparing the risk range percentile scores obtained from PRO test results to the predicted risk range percentages as defined above. The percentages of participants expected to fall into each risk range are the following: Low Risk **(39%)**, Medium Risk **(30%)**, Problem Risk **(20%)** and Severe Problem or Maximum Risk **(11%)**. The actual percentage of individuals falling in each of the four risk ranges, based on their risk range percentile scores, was compared to these predicted percentages. The risk range percentile score results for the 174 participants administered the PRO are presented in Table 13. The obtained risk range scores can be compared to the predicted risk range scores that are shown in the right-hand column of the table.

**Table 13. Risk Range Percentile Scores, N = 174 adult clients (2008).**

| Risk Range | Truthful-ness | Alcohol | Drugs | Violence | Depression | Anxiety | Self-Esteem | Stress Mgmnt | Predicted |
|---|---|---|---|---|---|---|---|---|---|
| Low | 39.7 | 40.8 | 37.4 | 39.9 | 38.5 | 38.9 | 39.7 | 38.5 | **39%** |
| Medium | 30.4 | 29.9 | 29.6 | 30.8 | 30.7 | 31.4 | 29.8 | 30.5 | **30%** |
| Problem | 19.6 | 19.0 | 21.8 | 18.1 | 20.5 | 20.7 | 19.6 | 20.1 | **20%** |
| Maximum | 10.3 | 10.3 | 11.2 | 11.2 | 10.3 | 9.0 | 10.9 | 10.9 | **11%** |

These results show that obtained risk range percentile scores closely approximated the predicted risk range percentile scores for each of the seven PRO scales presented in Table 13. **These results indicate that the PRO is a very accurate risk assessment instrument.** The results of the comparisons between obtained risk percentages and predicted percentages show that all obtained scale risk range percentile scores were within 1.9 percent of predicted.

Gender Differences

T-tests were calculated to assess possible sex differences. These results are presented in Table 14.

**Table 14. T-test comparisons of sex differences. (2008, N=174)**
**PRO Adult Client Sex Differences**

| PRO Scales | Males (N=140) Mean | Females (N=34) Mean | T-Test Comparisons |
|---|---|---|---|
| Truthfulness Scale | 37.74 | 40.47 | n.s. |
| Alcohol Scale | 17.81 | 15.65 | n.s. |
| Drugs Scale | 11.39 | 10.97 | n.s. |
| Violence Scale | 10.10 | 9.65 | n.s. |
| Anxiety Scale | 11.79 | 12.21 | n.s. |
| Depression Scale | 9.45 | 10.67 | n.s. |
| Self-Esteem Scale* | 26.41 | 30.50 | n.s. |
| Stress Management* | 148.71 | 149.06 | n.s. |

*Note: the Self-Esteem and Stress Management Scales are reversed in that higher scores denote lower risk.

Significant sex differences were not demonstrated on any of the seven PRO scales. Males and females in this sample did not score significantly differently on the PRO scales. This is an important consideration and gender differences will continue to be investigated in the PRO.

Reliability of the PRO

Reliability coefficient alphas are presented in Table 15. The results of this study support the statistical reliability of the PRO. All coefficient alphas are significant at $p<.001$. Most scale reliability coefficients are well above the professionally accepted .80 level for assessment instruments. These results show that the

PRO is a reliable risk assessment instrument.

| Table 15. Reliability coefficient alphas (2008, N = 174). | |
|---|---|
| **PRO Scales** | **Coefficient Alphas** |
| Truthfulness Scale | .92 |
| Alcohol Scale | .90 |
| Drugs Scale | .83 |
| Violence Scale | .86 |
| Anxiety Scale | .89 |
| Depression Scale | .88 |
| Self-Esteem Scale | .94 |
| Stress Management | .92 |

All coefficient alphas are significant at p<.001.

## 16. A Replication Study of Reliability and Accuracy of the PRO-Pretest

This study (2008) continued research of the PRO to investigate the reliability, validity and accuracy of the PRO. Only PRO-Pretest results are summarized in this study. Adult counseling clients were included in this study from different testing settings.

Method and Results

**The subjects in this study consisted of 668 adult counseling clients.**

Demographic composition of these participants is as follows: Males: 565 (84.6%); Females: 103 (15.4). Age: 19 & under (18%); 20-29 (30%); 30-39 (29%); 40-49 (17%); 50-59 (5%) and 60 & over (2%). Ethnicity: Caucasian (81%); Black (8%); Hispanic (7%); Native American (1%) and Other (1%). Education: Eighth grade or less (16%); Some H.S. (19%); H.S. graduate (55%) and Some college (10%). Marital Status: Single (61%); Married (20%); Divorced (13%); Separated (5%) and Widowed (1%).

Accuracy

Client scale scores are classified according to the risk (degree of severity) they represent. Four categories of risk are assigned: Low risk (zero to 39$^{th}$ percentile), Medium risk (40 to 69$^{th}$ percentile), Problem risk (70 to 89$^{th}$ percentile), and Severe Problem (90 to 100$^{th}$ percentile).

By definition, the expected percentage of clients assigned to each risk category is, 39% in Low risk, 30% in Medium risk, 20% in Problem risk and 11% in Severe Problem. The actual percentages of clients placed in the four risk categories based on their scale scores are compared to these expected percentages. Table 16 presents these comparisons.

**Table 16. Risk Range Percentile Scores, PRO-Pretest (2008, N = 668).**

| Scale | Low Risk (39%) | Medium Risk (30%) | Problem Risk (20%) | Severe Problem (11%) |
|---|---|---|---|---|
| Truthfulness Scale | 39.7 (0.7) | 28.4 (1.6) | 20.7 (0.7) | 11.2 (0.2) |
| Alcohol Scale | 39.8 (0.8) | 29.7 (0.3) | 19.3 (0.7) | 11.2 (0.2) |
| Drugs Scale | 39.5 (0.5) | 28.9 (1.1) | 20.1 (0.1) | 11.5 (0.5) |
| Violence Scale | 37.7 (1.3) | 31.3 (1.3) | 19.8 (0.2) | 11.2 (0.2) |
| Anxiety Scale | 39.5 (0.5) | 29.7 (0.3) | 18.8 (1.2) | 12.0 (1.0) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Depression Scale | 41.3 | (2.3) | 28.1 | (1.9) | 20.0 | (0.0) | 10.6 | (0.4) |
| Self-Esteem Scale | 38.8 | (0.2) | 29.9 | (0.1) | 20.8 | (0.8) | 10.5 | (0.5) |
| Stress Management | 38.2 | (0.8) | 29.9 | (0.1) | 20.8 | (0.8) | 11.1 | (0.1) |

As shown in the graph and table above, the PRO-Pretest scale scores are accurate. The objectively obtained percentages of clients falling in each risk range are very close to the expected percentages for each risk category. All of the obtained risk range percentages were within 1.6 percentage points of the expected percentages.

These results demonstrate that the PRO scale scores accurately identify client risk.

Reliability of the PRO

Reliability coefficient alphas are presented in Table 17.

**Table 17. Reliability coefficient alphas. PRO-Pretest (2008, N = 668).**

| PRO Scales | Coefficient Alphas |
|---|---|
| Truthfulness Scale | .89 |
| Alcohol Scale | .91 |
| Drugs Scale | .90 |
| Violence Scale | .88 |
| Anxiety Scale | .86 |
| Depression Scale | .87 |
| Self-Esteem Scale | .94 |
| Stress Management | .93 |

All coefficient alphas are significant at $p<.001$.

The results of this study support the statistical reliability of the PRO-Pretest. All coefficient alphas are significant at $p<.001$. Reliability coefficients are well above the professionally accepted .80 level. These results show that the PRO-Pretest is a highly statistically reliable risk assessment instrument.

## 17. PRO Pretest-Posttest Comparison Study

This study (2008) compared pretest and posttest results in a sample of treatment program adults. Statistical analyses of the pretest data were conducted to study reliability, validity and accuracy of the PRO at pretest assessment. PRO Posttest reliability was also investigated. PRO Pretest and Posttest data was analyzed for all clients who participated in the study. Not all participants who completed the pretest also completed the posttest. A distinction is made between comparisons involving all pretest and posttest, and those comparisons that involve pre-post comparisons for the same participant.

Method and Results

**Pretest: There were 506 participants that completed the PRO at Pretest.** Demographic composition of these participants is as follows: Males: 443 (87.5%); Females: 63 (12.5). Age: 12 & under (2%); 13 (6%); 14 (15%); 15 (28%); 16 (41%) and 17 (8%). Ethnicity: Caucasian (72%); Black (23%); Hispanic (3%); Native American (1%) and Other (2%). Education: Eighth grade or less (60%); Some H.S. (40%) and H.S. graduate (1%).

**Posttest: There were 209 participants that completed the PRO at Posttest. Of these 209, 122 individuals had both pretest and posttest data.** Demographic composition of these participants is as follows: Males: 197 (94.5%); Females: 12 (5.7). Age: 12 & under (0%); 13 (2%); 14 (9%); 15 (21%); 16

(40%) and 17 (29%). Ethnicity: Caucasian (71%); Black (26%); Hispanic (2%); Native American (0%) and Other (1%). Education: Eighth grade or less (42%); Some H.S. (56%) and H.S. graduate (2%).

Reliability of the PRO

Reliability coefficient alphas are presented in Table 18. All alpha coefficients for all of the Probation Referral Outcome (PRO) scales are at or above the .80 level.

**Table 18. Reliability coefficient alphas. (2008, N = 506 Pretest, 209 Posttest).**

| PRE-POST SCALES | Pretest Alphas | Posttest Alphas |
|---|---|---|
| Truthfulness Scale | .85 | .86 |
| Alcohol Scale | .86 | .80 |
| Drugs Scale | .87 | .81 |
| Violence Scale | .85 | .83 |
| Anxiety Scale | .82 | .84 |
| Depression Scale | .86 | .87 |
| Self-Esteem Scale | .91 | .93 |
| Stress Management | .89 | .89 |

All coefficient alphas are significant at $p<.001$.

These results demonstrate that the PRO is a reliable instrument or test. Posttest results show that the Alcohol and Drugs Scales had slightly lower alphas than did the Pretest results. The Pretest-Posttest interval varied from one to ten months.

As a result of intervention/treatment, clients may vary somewhat in their perceived substance use or abuse problem at Posttest. It is likely that each troubled adult progressed at different rates of understanding, acceptance and, where warranted, recovery. In contrast, clients' Pretest scores reflect consistent substance abuse problems perceptions. These scales are reliable. Intervention/treatment may contribute to the clients' understanding and clarification of his or her problems.

The results of this study support the reliability of the PRO-Posttest. By comparing Pretest reliability coefficients with Posttest reliability coefficients it can be seen that the PRO maintains high test-retest reliability. The PRO can be re-administered because, as these results demonstrate, the retest reliability coefficients vary around pretest reliability coefficients, which are impressive. In these pretest-posttest comparisons the interval varied from one to ten months.

Accuracy of the PRO

The accuracy of the eight PRO measurement (or severity) scales is presented in Table 19 for pretest assessments. Client risk assessment is calculated for the Pretest scores. Posttest results are then compared to these Pretest scores using the Pretest cutoff scores for each risk range category.

The Pretest percentages of clients scoring in the four risk categories (low, medium, problem and severe problem) are compared to predicted percentages for each of the seven measurement scales. The differences between obtained and predicted percentages are shown in parentheses in the table. The closeness of obtained Pretest scale scores and the predicted Pretest scale scores determines accuracy. All of the 209 Posttest results were summarized in the comparison table below the Pretest results.

**Table 19. Pretest Scale Risk Ranges (2008, N = 506)**

| Pretest Scale | Low Risk (39% predicted) | Medium Risk (30% predicted) | Problem Risk (20% predicted) | Severe Problem (11% predicted) |
|---|---|---|---|---|

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Truthfulness | 40.7 | **(1.7)** | 29.3 | **(0.7)** | 19.9 | **(0.1)** | 10.1 | **(0.9)** |
| Alcohol | 38.1 | **(0.9)** | 31.3 | **(1.3)** | 19.9 | **(0.1)** | 10.7 | **(0.3)** |
| Drugs | 39.9 | **(0.9)** | 29.9 | **(0.1)** | 19.9 | **(0.1)** | 10.3 | **(0.7)** |
| Violence | 38.5 | **(0.5)** | 29.5 | **(0.5)** | 20.7 | **(0.7)** | 11.3 | **(0.3)** |
| Anxiety | 37.2 | **(1.8)** | 29.2 | **(0.8)** | 22.1 | **(2.1)** | 11.5 | **(0.5)** |
| Depression | 39.4 | **(0.6)** | 31.1 | **(1.1)** | 18.7 | **(1.3)** | 10.8 | **(0.2)** |
| Self-Esteem | 39.1 | **(0.1)** | 29.1 | **(0.9)** | 20.5 | **(0.5)** | 11.3 | **(0.3)** |
| Stress Management | 39.1 | **(0.1)** | 29.3 | **(0.7)** | 20.9 | **(0.9)** | 10.7 | **(0.3)** |

As shown in the graph and table above, obtained Pretest risk range percentages for all risk categories and all PRO scales were within **2.1** percentage points of the predicted percentages. Of the 28 possible comparisons (7 scales x 4 risk ranges) between attained and predicted percentages, 24 were within one percentage point from the predicted percentage. Only four obtained risk range percentages were greater than 1.0% from the predicted percentage, and these were within 2.1 percent. These results demonstrate the accuracy of the Probation Referral Outcome (PRO) at the pretest or before intervention and/or treatment. The above table demonstrates that the PRO accurately measures client risk for all risk categories and all PRO scales.

Pre-Post Comparisons: Posttest Scale Scores Using Pretest Cutoff Scores

Risk range percentages for the PRO are established using Pretest data. This is because Pretest data serves as the baseline (or comparison standard) of attained test scores prior to intervention/treatment. This allows Posttest risk range percentages to be compared to Pretest percentages. Improvement on the Posttest is indicated by a higher percentage of clients scoring in the low risk range. This sequence is summarized as follows: Pretest – Intervention/Treatment – Posttest. It would be expected that more clients would score in the low risk range on the Posttest because scale scores are obtained after intervention/treatment has occurred. Effective treatment is demonstrated by lower Posttest scale scores. Higher Posttest scores (in comparison to Pretest scores) are often associated with no treatment.

In Table 20, the percentage differences between Pretest and Posttest scores are shown in parentheses. These differences are calculated as Posttest percentage – Pretest percentage or posttest minus pretest scores. The pretest-posttest comparison that is of interest is the "Low Risk" category. Because it is this category that is most affected by intervention and treatment. Effective intervention/treatment results in more people shifting to the Low risk category because clients have worked through their problems that existed at program intake. Positive differences in the Low risk category mean that Posttest percentages are higher than Pretest percentage, which establishes that intervention/treatment was effective. Negative differences between Pretest and Posttest mean that fewer clients score in that category on the Posttest than on the Pretest. In other words, if the number of clients attaining Low risk scores does not increase, then intervention/treatment either wasn't given or didn't result in positive change. Subtracting the Posttest percentages shown in the table below from the Pretest percentages (presented earlier) results in the differences shown in parentheses in the table below. All Pretest data (N=506) and all Posttest data (N=209) are included in these comparisons.

**Table 20. Pretest – Posttest Risk Range Comparisons (2008, N = 506 Pretest, 209 Posttest)**

| **Posttest Scales** | **Low Risk** | | **Medium Risk** | | **Problem Risk** | | **Severe Problem** | |
|---|---|---|---|---|---|---|---|---|
| | Attained Posttest % | Pre-Post Difference | Attained Posttest % | Pre-Post Difference | Attained Posttest % | Pre-Post Difference | Attained Posttest % | Pre-Post Difference |
| Truthfulness | 39.7 | **(-1.0)** | 30.6 | **(1.3)** | 19.2 | **(0.7)** | 10.5 | **(0.4)** |
| Alcohol | 43.1 | **(5.0)** | 36.3 | **(5.0)** | 19.6 | **(-0.3)** | 1.0 | **(-9.7)** |
| Drugs | 70.3 | **(30.4)** | 25.4 | **(-4.5)** | 2.9 | **(-7.0)** | 1.4 | **(-8.9)** |
| Violence | 76.6 | **(38.1)** | 14.8 | **(-14.7)** | 6.2 | **(-14.5)** | 2.4 | **(-8.9)** |
| Anxiety | 63.2 | **(26.0)** | 20.5 | **(-9.3)** | 9.1 | **(-11.0)** | 7.2 | **(-4.3)** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Depression | 67.4 | **(28.4)** | 20.9 | **(-8.9)** | 7.7 | **(-12.5)** | 4.1 | **(-7.0)** |
| Self-Esteem | 71.3 | **(32.2)** | 19.6 | **(-9.5)** | 7.2 | **(-13.3)** | 1.9 | **(-9.4)** |
| Stress Management | 71.8 | **(32.7)** | 20.1 | **(-9.2)** | 6.2 | **(14.7)** | 1.9 | **(-8.8)** |

Lower percentages for Medium, Problem and Severe Problem risk ranges are the result of clients being shifted down into the Low risk range at Posttest. That is why negative percentages are reported in Medium, Problem and Severe Problem categories.

The results shown above demonstrate that there were dramatic client improvements on Posttest scores for nearly all PRO scales. The Truthfulness Scale is an exception. Clients' Posttest and Pretest Truthfulness Scale scores were nearly the same. One theory regarding elevated Truthfulness Scale scores is "positive contagion" or the client's desire to respond as their counselor would like them to. Another interpretation might be that the intervention/treatment programs simply might not have addressed "honesty" in the adult's intervention/treatment program. Some degree of "open-honest" orientation is evident in most, if not all treatment programs. However, "honesty" may simply not have been focused upon as a treatment goal. This Truthfulness Scale outcome indicates that troubled clients were equally honest on posttest and pretest settings. This outcome was unexpected and will be studied further in subsequent Probation Referral Outcome studies.

The Violence Scale showed the largest Posttest improvement (lower scores). Over 38 percent more of the clients scored in the low risk range on Posttest. The Drugs, Self-esteem and Stress Management Scales also demonstrate a large improvement (lower scores) on Posttest. These scales improved by 30 percent or more on Posttest. The Alcohol Scale showed an improvement on Posttest of 5 percent for the low risk range and 5 percent for the medium risk range.

Mean Scale Scores Pre-Post Comparisons

There were 122 adults for whom both Pretest and Posttest scores were available. Comparisons of these clients' Pretest and Posttest scores are presented in Table 21.

T-test comparisons of the means for each PRO scale (the one exception is the Truthfulness Scale) indicate that the differences between Pretest and Posttest scores on all scales were significantly different. This means that Posttest scale scores were, on average, significantly lower than Pretest scale scores for these clients.

**Table 21. Pretest-Posttest Scale Comparisons (2008, N=122)**

| PRO Scales | Pretest Mean Score | Posttest Mean Score | T-value | Level of significance |
|---|---|---|---|---|
| Truthfulness Scale | 20.9 | 20.5 | t = 0.16 | n.s. |
| Alcohol Scale | 15.7 | 13.2 | t = 2.52 | p=.013 |
| Drugs Scale | 19.8 | 12.4 | t = 6.54 | p<.001 |
| Violence Scale | 18.1 | 11.8 | t = 7.49 | p<.001 |
| Anxiety Scale | 10.4 | 8.1 | t = 4.15 | p<.001 |
| Depression Scale | 17.1 | 16.9 | t = 0.14 | n.s. |
| Self-Esteem Scale | 19.5 | 31.7 | t = 8.19 | p<.001 |
| Stress Management | 101.8 | 128.2 | t = 6.83 | p<.001 |

Note: Scores on the Self-Esteem and Stress Management Scales are reversed in that higher scores are associated with better self-esteem and stress management. There were 122 clients included in this analysis.

With the exception of the Truthfulness Scale and the Depression Scale, all PRO Posttest scale scores are lower than Pretest scale scores. That is to say, clients showed improvement on all PRO scales (other than the Truthfulness Scale) after having been in treatment. There were 122 adults included in this study that had taken both the Pretest and Posttest.

These Pre-Post scale comparisons are in agreement with the Pre-Post risk range comparisons. The largest pre-post scale score differences occurred on the Self-esteem, Distress, Stress management and Drugs Scales. The Anxiety Scale also demonstrated a large pre-post scale score difference. The Alcohol Scale also had significantly different pre-post scale score differences. These measures support the view that clients benefited from having been in treatment.

Earlier, while discussing Truthfulness Scale results, we referenced "positive contagion" as a possible explanation of this test data. The theory refers to a transmission of ideas and feelings from person (counselor) to person (client) by suggestion, empathy or sympathy. Perhaps the adults were subconsciously attempting to answer items the way they believed their counselor would want them to at the posttest. Another possible explanation is that these intervention/treatment programs simply did not focus on "honesty" as a treatment objective. In contrast, at the pretest these troubled adult may have answered test items more candidly and spontaneously. Regardless of the theory, Truthfulness Scale answers were essentially the same at pretest testing and posttest testing. And as noted earlier these unexpected results will be studied in subsequent Probation Referral Outcome (PRO) research.

## 18. PRO Pre-Post Outcome Study

This study (2008) examined treatment outcome. Adult clients who were administered both the PRO Pretest and Posttest participated in this study. Pretest scale scores represent the severity of client problems going into treatment, whereas, Posttest scale scores represent clients' level of problem severity after having had treatment or at some time during treatment.

The PRO can be administered again after 30 days or longer, for example, 3 months, 6 months, etc. The 30-day time referent in the PRO enables giving the test to the same client after 30 days. The presented outcome analyses are the scale score comparisons between Pretest and Posttest. PRO Pretest scale scores are expected to be higher than Posttest scale scores because participants are expected to improve after having been in treatment. Outcome analyses help determine positive, neutral, or negative change.

Method and Results

**There were 69 participants that completed the PRO Pretest and Posttest.** Demographic composition of these participants is as follows: Males: 57 (82.6%); Females: 12 (17.4). Age: 19 & under (1.4%); 20-29 (43.5%); 30-39 (33.3%); 40-49 (15.9%) and 50-59 (5.8%). Ethnicity: Caucasian (67.2%); Black (11.9%); Hispanic (1.5%); Native American (13.4%) and Other (6.0%). Education: Some H.S. (19.1%); H.S. graduate (55.9%); Some college (7.3%) and College graduate (17.6%). Marital Status: Single (45.8%); Married (29.0%); Divorced (7.2%); Separated (10.1%) and Widowed (1.4%).

Pre-Post Outcomes

Pretest and Posttest scale scores are presented in Table 22. The table presents mean scale scores, maximum score, and t-values for the difference between the means and level of significance for each pre-post comparison.

On average, clients lowered their level of problem severity after having been in treatment. All posttest scale scores were lower than Pretest scale scores. The Alcohol Scale, Anxiety Scale and Depression Scale did not attain statistically significantly differences.

| Table 22. Pretest-Posttest Scale Comparisons (2008, N=69) | | | | | | |
|---|---|---|---|---|---|---|
| | Pretest | | Posttest | | | Level of |
| PRO Scales | Mean Score | Maximum | Mean Score | Maximum | T-value | Significance |
| Truthfulness Scale | 28.3 | 52 | 25.2 | 44 | t = 3.54 | p<.001 |
| Alcohol Scale | 14.8 | 53 | 12.9 | 40 | t = 1.56 | n.s. |
| Drugs Scale | 11.7 | 33 | 8.1 | 30 | t = 3.81 | p<.001 |
| Violence Scale | 12.5 | 38 | 10.0 | 31 | t = 3.07 | p<.003 |
| Anxiety Scale | 8.4 | 23 | 7.4 | 23 | t = 1.68 | n.s. |
| Depression Scale | 9.4 | 25 | 8.9 | 25 | t = 1.89 | n.s. |
| Self-Esteem Scale | 20.5 | 52 | 26.8 | 52 | t = 2.88 | p<.005 |
| Stress Management | 108.7 | 198 | 126.3 | 214 | t = 3.09 | p<.003 |

Note: Scores on the Self-Esteem and Stress Management Scales are reversed in that higher scores are associated with better self-esteem and stress management.

Maximum scale scores also demonstrate that posttest maximum scores were lower than pretest maximum scores, except the Self-esteem Scale which had equal maximum scores. These results further demonstrate that clients improved or positively changed after having been in treatment.

Comparisons of scale scores are a straightforward way of evaluating treatment program effectiveness. And, these comparisons quantify treatment outcome in an objective and standardized way. Not only can it be shown that participants improve after treatment, but the level of improvement is also quantified. Some participants improve more than others. These outcome comparisons are highly individualized, yet important in individual posttest analysis.

PRO Reliability

Reliability coefficient alphas for pretest results are presented in Table 23. All alpha coefficients for all of the Probation Referral Outcome scales are above the .80 level. PRO scales are reliable. The professionally accepted reliability standard is .75.

| Table 23. Reliability coefficient alphas. (2008, N = 69 Pretest). | | |
|---|---|---|
| PRE-POST SCALES | Pretest Alphas | Level of Significance |
| Truthfulness Scale | .89 | p<.001 |
| Alcohol Scale | .85 | p<.001 |
| Drugs Scale | .88 | p<.001 |
| Violence Scale | .87 | p<.001 |
| Anxiety Scale | .81 | p<.001 |
| Depression Scale | .86 | p<.001 |
| Self-Esteem Scale | .92 | p<.001 |
| Stress Management | .92 | p<.001 |

PRO Accuracy

PRO accuracy is based on Pretest scores. The percentages of clients scoring in the four risk categories (low, medium, problem and severe problem) are compared to predicted percentages for each of the seven measurement scales. These results are presented in Table 24. Predicted percentages are shown in the top row of the table. The differences between attained and predicted percentages are shown in

parentheses in the table. Small differences between attained and predicted percentages mean the scale is accurate.

**Table 24. Pretest Scale Risk Ranges (2008, N = 69)**

| Pretest Scale | Low Risk (39% predicted) | | Medium Risk (30% predicted) | | Problem Risk (20% predicted) | | Severe Problem (11% predicted) | |
|---|---|---|---|---|---|---|---|---|
| Truthfulness | 38.4 | **(0.6)** | 29.6 | **(0.4)** | 20.4 | **(0.6)** | 11.6 | **(0.6)** |
| Alcohol | 39.3 | **(0.3)** | 29.3 | **(0.7)** | 20.1 | **(0.1)** | 11.3 | **(0.3)** |
| Drugs | 38.2 | **(0.8)** | 31.6 | **(1.6)** | 20.1 | **(0.1)** | 10.1 | **(0.9)** |
| Violence | 38.7 | **(0.3)** | 30.0 | **(0.0)** | 19.4 | **(0.6)** | 11.9 | **(0.9)** |
| Anxiety | 39.2 | **(0.2)** | 33.1 | **(3.1)** | 17.3 | **(2.7)** | 10.3 | **(0.7)** |
| Depression | 39.5 | **(0.5)** | 30.5 | **(0.5)** | 19.6 | **(0.4)** | 10.4 | **(0.6)** |
| Self-Esteem | 39.0 | **(0.0)** | 29.3 | **(0.7)** | 20.7 | **(0.7)** | 11.0 | **(0.0)** |
| Stress Management | 39.0 | **(0.0)** | 30.2 | **(0.2)** | 20.1 | **(0.1)** | 10.7 | **(0.3)** |

Starting with the Low Risk column, the largest difference between attained and predicted was 0.8 percent. Attained Low Risk PRO scale scores were within 0.8 percent of their predicted 39 percent. This means that Low Risk scores are 99 percent accurate. Medium Risk scores were within 1.6 percent of their predicted 30 percent. This means that Medium Risk scale scores are 98 percent accurate. Problem Risk scores were within 0.7 percent of their predicted 20 percent and are 99 percent accurate. Severe Problem scores were within 0.9 percent of their predicted 11 percent and are 99 percent accurate. These small differences between attained and predicted risk range scores demonstrate the accuracy of the PRO.

The PRO is an objective outcome assessment test. The same test given at pretest or intake is re-administered after treatment or at specified intervals during treatment. The pretest sets the standard or baseline for subsequent comparison after or during treatment. PRO scales assess important client attitudes and behavior that can change after treatment. The amount of change that clients experience after treatment is determined by pretest and posttest scale score comparisons. The amount of change a client experiences (treatment outcome) helps determine the whether the client underwent positive or negative change (or in some cases, no change).

## 19. PRO Pre-Post Outcome Study in a Sample of Adult Clients

This study (2009) examined PRO test results for a sample of adult treatment clients. There were 175 clients that were administered both a pretest and Posttest. These Pretest—Posttest comparisons are presented and discussed. Included in this study are PRO test statistics on the reliability, validity and accuracy of the PRO for these adult clients.

Method and Results

**There were 175 participants that completed the PRO at Pretest and Posttest.** Demographic composition of these participants is as follows: Males: 157 (89.7%); Females: 18 (10.3). Age: 13 & under (2.9%); 14 (10.9%); 15 (24.0%); 16 (50.9%) and 17 (11.4%). Ethnicity: Caucasian (74.3%); Black (22.9%); Hispanic (2.3%) and Other (0.6%). Education: 7$^{th}$ grade or less (12.3%); 8$^{th}$ grade (36.4%); 9$^{th}$ grade (42.0%); 10$^{th}$ grade (6.8%); 11$^{th}$ grade (1.9%) and H.S. graduate (0.6%).

Reliability statistics are presented for both Pretest and Posttest data. Slight reductions in Posttest reliability coefficients indicate that clients changed, to a varying extent, their perception of "problem." They tend to redefine their interpretation of what constitutes a "problem." PRO validity and accuracy

statistics are presented for Pretest data. This was done because Pretest scores set baseline performance upon which to compare Posttest scores. The interval between Pretest and Posttest administrations varied from 1 month to 23 months.

Adult risk is conceptualized as low risk (zero to 39th percentile), medium risk (40 to 69th percentile), problem risk (70 to 89th percentile) and severe problem (90 to 100th percentile) risk. The expected percentage of adults scoring in each risk range (for each PRO scale) is, low risk (39%), medium risk (30%), problem risk (20%) and severe problem risk (11%). A problem is not identified until a adult's scale score is at (or exceeds) the 70th percentile. The scores associated with the 39th, 69th and 89th percentiles are referred to as cut-off scores. Scores above the cut-off score fall in the next higher risk range.

Accurate identification of problems is necessary to make appropriate referral to intervention and treatment. Andrews, Bonta & Hoge concluded that placing low risk offenders in wrong treatment levels can be detrimental to society and the offenders (Andrews, DA, Bonta, J, & Hoge, RD. Classification for Effective Rehabilitation: Rediscovering Psychology. Criminal Justice and Behavior, 1990, 17(1): 19-52.). Thus, it is important to identify offender problems and determine their severity so offenders can be placed in appropriate levels of intervention and treatment. Similar logic is applicable to adult counseling clients. Identification of clients' problems is the first step in intervention and treatment.

PRO risk range percentile scores are obtained by adding test item points and truth correction, if applicable. These raw scores are then converted to percentile scores by using cumulative percentage distributions. Each scale has its own distribution and risk range cut-off scores. Pretest results are summarized in Table 25. Adult obtained Pretest scores are compared to the predicted percentage for each risk range. The predicted percentages are presented in parentheses under the name (low, medium, problem, severe problem) of each risk range. Differences between predicted and obtained scores are presented in parentheses (in bold type). The smaller the difference the more accurate the scale.

Posttest data use risk range cut-off scores established by Pretest data. The percentage of clients that fall in each risk range at Posttest is due entirely by Pretest—Posttest differences. Posttest results are presented in Table 26. Posttest scores are expected to be lower than Pretest scores with the biggest difference being an increase in the Low risk range at Posttest. Differences between Pretest and Posttest risk range percentages are shown in parentheses in Table 34. Positive differences in these percentages means there are more clients in that risk range at Posttest than there were at Pretest. Negative differences signify more clients at Pretest than Posttest.

Table 25 presents the graph and table of adult Pretest risk range percentages. As shown in this graph and related table, obtained risk range percentages are within 1.7 percentage points of the predicted percentages.

These results demonstrate that PRO scale scores are accurate. Placement of clients into appropriate risk ranges is approximately 98 percent accurate.

**Table 25. Pretest Scale Risk Ranges (N = 175, 2009)**

| Pretest Scale | Low Risk (39% predicted) | | Medium Risk (30% predicted) | | Problem Risk (20% predicted) | | Severe Problem (11% predicted) | |
|---|---|---|---|---|---|---|---|---|
| Truthfulness | 37.8 | **(1.2)** | 29.9 | **(0.1)** | 21.7 | **(1.7)** | 10.5 | **(0.5)** |
| Alcohol | 40.6 | **(1.6)** | 28.7 | **(1.3)** | 19.6 | **(0.4)** | 11.1 | **(0.1)** |
| Drugs | 39.8 | **(0.8)** | 28.9 | **(1.1)** | 20.3 | **(0.3)** | 11.1 | **(0.1)** |
| Violence | 38.0 | **(1.0)** | 30.3 | **(0.3)** | 20.9 | **(0.9)** | 10.9 | **(0.1)** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Anxiety | 37.8 | **(1.2)** | 29.4 | **(0.6)** | 21.2 | **(1.2)** | 11.6 | **(0.6)** |
| Depression | 39.5 | **(0.5)** | 29.1 | **(0.9)** | 20.6 | **(0.6)** | 10.8 | **(0.2)** |
| Self-Esteem | 39.4 | **(0.4)** | 29.4 | **(0.4)** | 20.9 | **(0.9)** | 10.3 | **(0.3)** |
| Stress Management | 39.0 | **(0.0)** | 29.8 | **(0.2)** | 20.1 | **(0.1)** | 11.1 | **(0.1)** |

The percentage differences between Pretest and Posttest scores are presented in Table 26. These differences (shown in parentheses) are calculated by subtracting the Posttest percentage from the Pretest percentage. Positive differences in risk range percentages between Pretest and Posttest mean that Posttest percentages are higher than Pretest percentages.

In general, clients' scores on Posttest are lower than on Pretest resulting in risk range percentages shifting toward the lower end. This would be expected in effective treatment/intervention programs. The Low risk range percentage increases at Posttest as a result of clients attaining lower scores at Posttest.

**Table 26. Posttest Scale Risk Ranges (N=175, 2009)**

| **Posttest Scales** | **Low Risk** | | **Medium Risk** | | **Problem Risk** | | **Severe Problem** | |
|---|---|---|---|---|---|---|---|---|
| | Attained Posttest % | Pre-Post Difference | Attained Posttest % | Pre-Post Difference | Attained Posttest % | Pre-Post Difference | Attained Posttest % | Pre-Post Difference |
| Truthfulness | 21.2 | **(-15.6)** | 31.3 | **(1.4)** | 27.6 | **(5.8)** | 19.9 | **(9.4)** |
| Alcohol | 48.4 | **(7.8)** | 32.0 | **(3.3)** | 18.7 | **(-0.9)** | 0.9 | **(-10.2)** |
| Drugs | 71.8 | **(32.0)** | 23.7 | **(-5.2)** | 3.2 | **(-17.1)** | 1.3 | **(-9.8)** |
| Violence | 73.7 | **(35.7)** | 17.4 | **(-12.9)** | 6.6 | **(-14.3)** | 2.2 | **(-8.7)** |
| Anxiety | 57.6 | **(19.8)** | 23.7 | **(-5.7)** | 11.4 | **(-9.8)** | 7.3 | **(-4.3)** |
| Depression | 76.6 | **(38.1)** | 14.8 | **(-14.7)** | 6.2 | **(-14.5)** | 2.4 | **(-8.9)** |
| Self-Esteem | 68.7 | **(29.3)** | 18.7 | **(-10.7)** | 9.8 | **(-11.1)** | 2.8 | **(-7.5)** |
| Stress Management | 67.4 | **(28.4)** | 20.9 | **(-8.9)** | 7.7 | **(-12.5)** | 4.1 | **(-7.0)** |

This result indicates that intervention and treatment were effective. Negative percentages for Medium, Problem and Severe Problem categories are the result of fewer clients scoring in those risk ranges at Posttest compared to Pretest. For simplicity and clarity we refer to the difference between Pretest and Posttest scores. When pretest scale scores are higher than analogous Posttest scale scores, positive change has taken place.

The Truthfulness Scale results show just the opposite. Posttest scores were higher than Pretest scores. One possible explanation for this outcome is that adults fake good at Posttest, they give the response they think the counselor wants them to give. This phenomenon has been called "therapeutic contagion." Consequently, their Truthfulness Scale score goes up in comparison to their Pretest score. Truthfulness Scale scores apply truth-correction to other scale scores, consequently, Posttest scores are being truth-corrected more than are Pretest scores. Pretest-posttest differences could be even greater than what is shown in the table below if truth-correction were not applied.

**Table 27. Pretest-Posttest Scale Comparisons (N=175, 2009)**

| **PRO Scales** | **Pretest Mean Score** | **Posttest Mean Score** | **T-value** | **Level of significance** |
|---|---|---|---|---|
| Truthfulness Scale | 22.46 | 26.60 | t = 4.31 | **p<.001** |
| Alcohol Scale | 15.17 | 12.97 | t = 2.77 | **p=.006** |
| Drugs Scale | 19.54 | 11.99 | t = 8.13 | **p<.001** |
| Violence Scale | 17.53 | 11.75 | t = 8.41 | **p<.001** |
| Anxiety Scale | 10.35 | 8.25 | t = 4.45 | **p<.001** |

| | | | | |
|---|---|---|---|---|
| Depression Scale | 9.74 | 7.72 | t = 3.96 | **p<.001** |
| Self-Esteem Scale | 20.52 | 30.41 | t = 7.60 | **p<.001** |
| Stress Management | 101.70 | 124.77 | t = 7.11 | **p<.001** |

The results in Table 27 show that there were dramatic client improvements on Posttest scores for all PRO scales. The Truthfulness Scale is an exception. The Violence Scale showed the largest Posttest improvement (lower scores). Nearly 36 percent more of the adults scored in the low risk range at Posttest. The Drugs, Self-Esteem and Stress Management Scales also demonstrate a large improvement (lower scores) at Posttest. The Anxiety Scale showed about a 20 percent improvement.

The Alcohol Scale showed an improvement at Posttest of 8 percent (increase) for the low risk range and 10 percent (decrease) for the severe problem risk range. Of the 11 percent of adults who had scored in the severe problem range on the Alcohol Scale at Pretest, only 1 percent remained in the severe problem range at Posttest. There were 175 adults for whom both Pretest and Posttest data were available. Mean or average scale score for each PRO scale for these clients' is presented in Table 35. These results indicate that all scales were statistically significantly different. Posttest scale scores were, on average, significantly lower (the one exception is the Truthfulness Scale) than Pretest scale scores for these clients, which represents positive change

With the exception of the Truthfulness Scale all PRO scale comparisons demonstrate that Posttest scale scores are lower than Pretest scale scores. These clients showed improvement on all PRO treatment scales after having been in treatment. However, the Pretest-Posttest intervals were not the same for all clients. It is likely that higher Pretest-Posttest intervals would result in higher or greater positive change between Pretest and Posttest scores. Significant pre-post score differences occurred on the Self-Esteem, Violence, Stress Management and Drugs Scales. The Anxiety, Depression and Alcohol Scales also demonstrated significant pre-post scale score differences. These treatment measures demonstrate positive change, which means that clients benefited from having been in treatment.

Truthfulness Scale results present an interesting phenomenon. Clients scored significantly higher at Posttest than at Pretest. "Therapeutic contagion" is a possible explanation of this test data. The theory refers to a transmission of ideas and feelings from person (counselor) to person (troubled adult) by suggestion, identification or transference. Perhaps the adults were subconsciously attempting to answer items the way they believed their counselors would want them to at posttest. In contrast, at pretest these clients may have answered test items more defensively. Regardless of the theory, Truthfulness Scale answers were significantly different at pretest and posttest testing. These results will be studied in subsequent Probation Referral Outcome (PRO) research. Within-test reliability, or inter-item reliability coefficient alphas for the Probation Referral Outcome (PRO) are presented in Table 28. As demonstrated in the table, Alpha coefficients for all PRO scales are well above the professionally accepted standard of .75. All of the PRO scales attained reliability coefficients at or above .81. These results show that the PRO is a reliable assessment.

**Table 28. PRO Reliability (2009)**

| PRE-POST SCALES | Pretest Alphas | Posttest Alphas |
|---|---|---|
| Truthfulness Scale | **.86** | **.86** |
| Alcohol Scale | **.86** | **.84** |
| Drugs Scale | **.87** | **.84** |
| Violence Scale | **.85** | **.82** |
| Anxiety Scale | **.83** | **.81** |
| Depression Scale | **.84** | **.85** |
| Self-Esteem Scale | **.91** | **.93** |

| | | |
|---|---|---|
| Stress Management | **.88** | **.90** |

All coefficient alphas are significant at p<.001. Pretest-posttest reliability coefficients demonstrate that the PRO maintains high test-retest reliability. The PRO can be re-administered because the Posttest reliability coefficients are just as high as Pretest reliability coefficients.

Predictive validity is shown by nearly 100% correct identification of adults who have problems. The Alcohol and Drugs Scales accurately identified adults who admitted to drinking and drug problems. The PRO Alcohol Scale identified nearly all (98.1%) of the adults who admitted having an alcohol problem These adults are classified as problem drinkers and 98.1 percent of them had Alcohol Scale scores at or above the 70th percentile. The Alcohol Scale correctly identified almost all of the adults categorized as problem drinkers. The Drugs Scale identified nearly all (97.6%) of the adults who admitted to a drug problem. These adults had Drugs Scale scores at or above the 70th percentile. These results substantiate the accuracy of the Drugs Scale.

The PRO correctly identified nearly all adults who had substance abuse problems. PRO scale scores at or above the 70$^{th}$ percentile identifies problems. These results support the accuracy of the Alcohol Scale and the Drugs Scale. The higher the scale score, the more severe the problem. PRO scale scores do not identify a problem until a score is at or above the 70$^{th}$ percentile. With this problem identification threshold and scores, nearly 100 percent of problem clients were identified. The seventy percent problem threshold is a clear indication that a problem exists. These results support using this risk range percentile cutoff for problem identification.

In summary, the PRO accurately identified clients that had identifiable (serious) problems. Validity analyses clearly demonstrate that the PRO impressively meets these criteria. PRO Alcohol and Drugs Scales identify almost all adults who have alcohol or drugs problems. The PRO measures what it purports to measure. Furthermore, these statistics demonstrate that the PRO is a reliable test.

## 20. PRO: Assessing Treatment Outcome

Assessing treatment **outcome** involves answering the question: Has the client improved, stayed the same or gotten worse? Many practitioners, referral sources and treatment agencies have wanted an accurate way to objectively assess counseling and treatment effectiveness or outcome. On the surface, this outcome question appears straightforward. But, what should be used as the criteria for treatment program effectiveness? The issue of outcome criteria will likely always be controversial.

The following study (2009) demonstrates effect by comparing participants' Pretest and Posttest scores.

Method and Results

**There were 232 participants that completed the PRO Pretest and Posttest.** Demographic composition of these participants is as follows: Males: 210 (90.5%); Females: 22 (9.5). Age: 13 & under (2.6%); 14 (12.5%); 15 (23.7%); 16 (51.7%) and 17 (9.5%). Ethnicity: Caucasian (72.0%); Black (25.9%); Hispanic (1.7%) and Other (0.4%). Education: 7$^{th}$ grade or less (17.2%); 8$^{th}$ grade (32.3%); 9$^{th}$ grade (37.9%); 10$^{th}$ grade (10.3%); 11$^{th}$ grade (1.7%) and High School graduate (0.4%).

The primary measure of treatment outcome in the Probation Referral Outcome (PRO) is the Comparison Index. This index compares pretest (first test administration) scale scores with posttest (second or subsequent test administration) scale scores. All PRO scales are represented in the Comparison Index.

For each scale, the index is gotten by subtracting the posttest scale score from the pretest scale score (pretest minus posttest). A positive difference represents client improvement, that is, their scale score was lower at posttest than it was at pretest.

If the difference between pretest and posttest scale scores is zero, the adult stayed the same. And, a negative difference means that negative change occurred, i.e., the posttest scale score was higher than its analogous pretest score.

The pretest-posttest Comparison Index is presented in the following table. For each PRO scale the mean or average scale score is presented for pretest and posttest scores along with the difference (pretest-posttest) presented in the right-hand column. There are 232 adults included in this analysis. These adults had both pretest and posttest data.

**Table 29. Pre-Post Comparison Index (N=232, 2009)**

| PRO Scales | Pretest Mean Score | Posttest Mean Score | Pretest-Posttest Difference |
|---|---|---|---|
| Truthfulness Scale | 23.22 | 21.47 | 1.75 |
| Alcohol Scale | 15.27 | 12.60 | 2.69 |
| Drugs Scale | 19.65 | 11.70 | 7.95 |
| Violence Scale | 17.56 | 11.75 | 5.81 |
| Anxiety Scale | 10.63 | 8.37 | 2.26 |
| Depression Scale | 12.67 | 9.71 | 2.96 |
| Self-Esteem Scale | 20.60 | 30.68 | 10.08 |
| Stress Management | 98.78 | 124.45 | 25.67 |

Note: Scores on the Self-esteem and Stress Management Scales are reversed in that higher scores are associated with better self-esteem and stress management.

For all Probation Referral Outcome (PRO) scales, posttest scores were lower than pretest scores, which characterizes positive change. Posttest scores were significantly lower than pretest scores at the $p < 0.001$ level of significance. Lower scale scores at posttest means that treatment programs were positively effective.

Truthfulness Scale score pre-post comparison demonstrates that the adults became significantly more open and honest while completing the PRO at posttest. The adults were less inclined to deny, minimize problems or attempt to fake good. Clients' alcohol and drug problem severity was positively changed after treatment. Lower posttest Alcohol Scale scores shows that these adults significantly reduced their alcohol problem severity after being in treatment. Drugs Scale scores were lower, by a wide margin, at posttest compared to pretest scores. Treatment helped to significantly lower clients' severity of drug abuse. Results of the Violence Scale score comparisons show that after treatment the adults had significantly less distress, anxiety and depression. Treatment helped the adults re-establish their emotional well-being. Positive treatment experience is demonstrated by Anxiety Scale score comparisons. Adults became significantly more open and cooperative, and, less resistant. The Self-Esteem Scale pre-post comparison demonstrates that the adults significantly improved their perceived self-worth and value. Treatment helped clients positively change their self-esteem. Stress Management Scale pre-post comparison indicates that clients were better able to cope with stress after having been in treatment. All of these pre-post scale comparisons demonstrated statistically significant differences (at the $p<.001$ level) between pretest and posttest scale scores.

Probation Referral Outcome (PRO) scale scores are objective and accurate measures. The 30-day time referent in the PRO enables the same test to be administered again to the same adult at 30 day or longer intervals. Comparisons between pretest and posttest scores provide an objective and accurate way to compare scores. PRO scale comparisons represent outcome criteria. Pretest scores are the standard or baseline for comparison. Prior history is eliminated from scale scores and the 30-day time referent enables us to use the same test at posttest. This procedure holds testing (and outcome) variables constant so that change in adult responses can be attributed to treatment. The Pre-Post Comparison Index table demonstrates that positive change occurred.

## 21. Probation Referral Outcome (PRO) Reliability and Accuracy in a Large Sample of Adults

Reliability and accuracy of the Probation Referral Outcome (PRO) were examined in a large sample of adult counseling clients (2009).

Method and Results

**There were 600 participants that completed the PRO Pretest and/or Posttest.** Demographic composition of these participants is as follows: Males: 478 (79.7%); Females: 122 (20.3%). Age: 20 & under (8.3%); 21-29 (33.0%); 30-39 (31.2%); 40-49 (19.8%); 50-59 (6.2%) and 60 and over (1.5%). Ethnicity: Caucasian (77.9%); Black (6.5%); Hispanic (4.0%); Asian (0.5%); Native American (5.5%) and Other (5.5%).

Risk range accuracy of the Probation Referral Outcome (PRO) was examined by determining the differences between predicted and attained risk range percentages. Small differences between predicted and attained scale scores represent high accuracy. Table 31 provides accuracy calculations for each Probation Referral Outcome scale for this sample of adult respondents.

As shown in Table 31, Pre-Post scale scores are highly accurate. The objectively obtained percentages of adults falling into each risk range are very close to the expected percentages for each risk category. All attained risk range percentages were within 3.1 percentage points of the predicted percentages.

**Table 31. Probation Referral Outcome Risk Range Accuracy (N=600, 2009)**

| Scale | *Low Risk (39%)* | | *Medium Risk (30%)* | | *Problem Risk (20%)* | | *Severe Problem (11%)* | |
|---|---|---|---|---|---|---|---|---|
| Truthfulness Scale | 40.7 | **(1.7)** | 31.8 | **(1.8)** | 17.6 | **(2.4)** | 9.9 | **(1.1)** |
| Alcohol Scale | 42.0 | **(3.0)** | 28.2 | **(1.8)** | 19.9 | **(0.1)** | 9.9 | **(1.1)** |
| Drugs Scale | 40.5 | **(1.5)** | 30.9 | **(0.9)** | 18.6 | **(1.4)** | 10.0 | **(1.0)** |
| Violence Scale | 39.8 | **(0.8)** | 32.3 | **(2.3)** | 18.0 | **(2.0)** | 9.9 | **(1.1)** |
| Anxiety Scale | 41.3 | **(2.3)** | 30.4 | **(0.4)** | 17.9 | **(2.1)** | 10.4 | **(0.6)** |
| Depression Scale | 39.4 | **(0.6)** | 31.1 | **(1.1)** | 18.7 | **(1.3)** | 10.8 | **(0.2)** |
| Self-Esteem Scale | 42.2 | **(3.2)** | 26.9 | **(3.1)** | 21.4 | **(1.4)** | 9.5 | **(1.5)** |
| Stress Management | 40.1 | **(1.1)** | 29.6 | **(0.4)** | 19.7 | **(0.3)** | 10.6 | **(0.4)** |

Inter-item reliability was calculated for the eight PRO scales. Cronbach's alpha (α) coefficients are presented in Table 30. All attained reliability coefficients exceed the professionally accepted standard for reliability (.75) by a considerable margin.

**Table 30. Reliability coefficient alphas. (N =600, 2009).**

| PRE-POST SCALES | PRO Alphas | Level of Significance |
|---|---|---|
| Truthfulness Scale | .91 | p<.001 |
| Alcohol Scale | .86 | p<.001 |
| Drugs Scale | .86 | p<.001 |
| Violence Scale | .86 | p<.001 |
| Anxiety Scale | .85 | p<.001 |
| Depression Scale | .87 | p<.001 |
| Self-Esteem Scale | .94 | p<.001 |
| Stress Management | .93 | p<.001 |

## 22. Establishing Treatment Outcome with PRO Pretest and Posttest Score Comparisons

The Probation Referral Outcome (PRO) Pretest and Posttest were administered to a group of adults before and after treatment (2009). The same adults were administered the Pretest prior to undergoing treatment and the Posttest after completing treatment.

Method and Results

**There were 464 participants that completed both the PRO Pretest and the PRO Posttest.** Demographic composition of these participants is as follows: Males: 405 (87.3%); Females: 59 (12.7%). Age: 13 & under (9.9%); 14 (17.0%); 15 (27.6%); 16 (37.6%) and 17 (8.0%). Ethnicity: Caucasian (66.7%); Black (30.4%); Hispanic (0.2%) and Other (2.7%).

Mean Scale Scores Pre-Post Comparisons

There were 464 adults for which both Pretest and Posttest scores were available. Pretest and Posttest score comparisons are presented in Table 32.

*T*-tests results, comparing the average Pretest and Posttest scores of each PRO scale, indicate that the score differences found for all scales (excepting the Self-Esteem Scale) were statistically significant. The Posttest scale scores were, on average, significantly lower than Pretest scale scores. Lower scores at Posttest represent decreased problem severity, which is interpreted as positive treatment outcome or positive change.

As shown in Table 32, with the exception of the Self-Esteem Scale, for which average Pretest and Posttest scores were nearly identical, all mean PRO Posttest scale scores are lower than mean Pretest scale scores. This means that clients showed improvement in all areas measured by PRO scales (other than the Self-Esteem Scale) after completing treatment. A lower score upon Posttest (after treatment) represents positive change.

**Table 32. Pretest-Posttest Scale Score Comparisons (2009, N=464)**

| PRO Scales | Pretest Mean Score | Posttest Mean Score | T-value | Level of significance |
|---|---|---|---|---|
| Truthfulness Scale | 29.54 | 18.99 | 12.54 | p<.001 |
| Alcohol Scale | 51.94 | 44.31 | 10.19 | p<.001 |
| Drugs Scale | 52.27 | 37.12 | 18.10 | p<.001 |
| Violence Scale | 44.34 | 15.60 | 31.75 | p<.001 |
| Anxiety Scale | 10.81 | 8.32 | 3.73 | p<.001 |
| Depression Scale | 9.65 | 9.43 | 3.91 | p<.001 |
| Self-Esteem Scale | 24.80 | 24.34 | 4.38 | n.s. |
| Stress Management | 48.17 | 31.27 | 19.49 | p<.001 |

Note: Scores on the Self-Esteem and Stress Management Scales are reversed in that higher scores are associated with better self-esteem and stress management. There were 464 clients included in this analysis.

Correlation analyses were also performed for Pretest and Posttest scores for each scale (N=424). The Pearson's *r* coefficients attained for each scale (all significant at p<.001) are as follows: *Truthfulness Scale*, *r*=.107; *Alcohol Scale*, *r*=.438; *Drugs Scale*, *r*=.381; *Violence Scale*, *r*=.317; *Anxiety Scale*, *r*=.268; *Self-Esteem Scale*, *r*=.326 and *Stress Management Scale*, *r*=.419.

These strong and significant correlations mean that the Probation Referral Outcome (PRO) administered at Pretest effectively measures the same constructs that are measured at Posttest. In other words, the PRO Scales hold to what they are designed to measure both before and after treatment. This is important because as client mindset and risk levels change (prior to, during, and after treatment), the assessment must be able to account for these changes while still effectively measuring what it is purported to measure.

## 23. Gender Differences in the Probation Referral Outcome (PRO)

Probationer assessments should account for possible gender differences. Crime-related research has shown there are differences between male and female probationers. The PRO has been standardized on both male and female respondents.

Gender Differences

T-tests were calculated for all PRO scales to assess possible sex differences (2010). These results are presented in Table 33. Differences in scores were significant at the p<.001. The two exceptions were the Alcohol Scale and Drugs Scale scores, for which the average scores of males and females were comparable.

**Table 33. T-test comparisons of score differences by gender. (2010, N=308)**
**PRO Adult Client Sex Differences**

| PRO Scales | Males (N=140) Mean | Females (N=34) Mean | T-Value |
|---|---|---|---|
| Truthfulness Scale | 26.76 | 16.35 | 3.78 |
| Alcohol Scale | 21.06 | 20.71 | n.s. |
| Drugs Scale | 23.65 | 21.89 | n.s. |
| Violence Scale | 19.50 | 30.24 | -4.52 |
| Anxiety Scale | 11.13 | 6.71 | 2.68 |
| Depression Scale | 10.87 | 8.21 | 2.76 |
| Self-Esteem Scale* | 23.50 | 35.82 | -2.44 |
| Stress Management Scale* | 99.51 | 74.18 | 2.35 |

Significant sex differences were seen on the PRO Truthfulness, Distress, Resistance, Self-Esteem and Stress Management Scale scores. Males averaged higher Truthfulness Scale scores, which means that in the case of this sample, males were more likely to deny or minimize problems than their female counterparts.

For both the Violence Scale and the Stress Management Scales, females' average scores were more problematic (severe) in than the average scores of males. Females in this sample had more pronounced distress and less ability to effectively manage stress.

In regards to the Resistance and Self-Esteem Scales, males averaged more problematic scale scores than females. Males in this sample may have had more impaired self-esteem and were more resistant to receiving help or treatment. Sex differences have been incorporated into PRO gender standardization and sex differences will continue to be explored in future PRO research.

## SUMMARY

In conclusion, this document is not intended as an exhaustive compilation of Probation Referral Outcome (PRO) research. Yet, it does summarize many studies and statistics that support the reliability and validity of the PRO. Based on this research, the PRO presents an increasingly accurate picture of counseling clients and the risk they represent. The PRO provides an empirical foundation for responsible decision making. The PRO is research-based.

Summarized research indicates that the PRO is a reliable, valid and accurate instrument for client outcome assessment. It is reasonable to conclude that the PRO does what it purports to do. The PRO acquires a vast amount of relevant information for staff review prior to decision making. Empirically based scales are objective and accurate. Assessment has shifted from subjective opinions to objective accountability.

The PRO is a research-based assessment instrument or test. Its pre-post design is uniquely advantageous to counseling/treatment research. Using the Pretest as a baseline for Posttest comparison ensures accurate outcome or counseling and/or treatment effectiveness measures. The same test is administered twice; once before treatment (Pretest) and once after treatment (Posttest).

The PRO cumulative database was built with ongoing research in mind. Each PRO that is administered is included (via test answers) in its cumulative database, so each test users' client population is included in annual standardization research. As always, test data is utilized in a confidential (no names) manner. This proprietary database provides a large and continually expanding amount of data, ideal for research purposes. Ongoing database research and test standardization ensures optimum PRO accuracy and performance. And we believe treatment effectiveness dovetails with coming recidivism research. We would like the Probation Referral Outcome (PRO) to be included in recidivism research.

Areas for future PRO research are varied and complex. Professional Online Testing Solutions, Inc. will continue its research and development efforts. Database research is a primary focus. Consistent with the foregoing, Professional Online Testing Solutions, Inc. encourages other scientists to participate in PRO research. Treatment effectiveness is a key area of inquiry for referral sources like probation officers, counselors, therapists, and other professionals working with clients in substance abuse and mental health treatment settings.

Parties interested in using the Probation Referral Outcome (PRO) in future research should contact Professional Online Testing Solutions, Inc., P.O. Box 44256, Phoenix, Arizona 85064-4256. Our email address is research@online-testing.com and our toll-free telephone number is 1 (800) 231-2401. Our office hours are 7:30AM to 4:30PM Mountain Standard Time.

Rahnuma Khandaker
Senior Research Analyst

Professional Online Testing Solutions, Inc.
P.O. Box 32974, Phoenix, Arizona 85064-2974


* * *