

EFFECTS OF COLLIGAN, OSBORNE, SWENSON, AND OFFORD'S
CONTEMPORARY MMPI NORMS ON LACHAR
AUTOMATED INTERPRETATIONS

By

JOYCE LYONS HAWLEY

Bachelor of Arts
University of Tulsa
Tulsa, Oklahoma
1965

Master of Arts
University of Tulsa
Tulsa, Oklahoma
1976

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1991


Theo's
149:12
H39:6e
cup.2

Copyright
by
Joyce Lyons Hawley
May, 1991

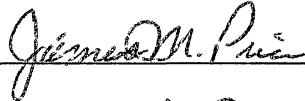
To my mother and father,
with love and thanks

EFFECTS OF COLLIGAN, OSBORNE, SWENSON, AND OFFORD'S
CONTEMPORARY MMPI NORMS ON LACHAR
AUTOMATED INTERPRETATIONS

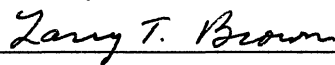
Thesis Approved:

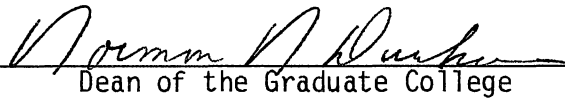


Thesis Adviser









Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to express my deepest thanks to my adviser, Dr. Robert S. Schlottmann. His respect, kindness, and patience have made him an unfailing help and a friend throughout my graduate program.

I would also like to thank Dr. Larry T. Brown, who showed me by precept and example how much fun teaching can be; Dr. Al Carlozzi, who taught me much of what I know about family therapy (and my patients thank him); and Dr. Jim Price, who not only showed me the principles underlying the formulae and the numbers in statistics, but who also gave me a glimpse of the beautiful world of mathematics.

I also extend my thanks to the psychologists at the VA Medical Center, without whose cooperation and generosity in giving their time I could not have completed this project.

I want to express special thanks to my father, Dr. Paul F. Hawley, and my aunt, Dr. Florence Hawley Ellis. Their achievements have inspired me throughout.

Thanks are also due to my wonderful boss, Dr. Camille Lloyd, whose constant encouragement--and her willingness to allow me to take those necessary trips to Stillwater--was amazing.

And finally, thank you Mem, the sister I never had; Fred, Andrea, and Rieka; Anne and Sharon; Corinne, Armin, and Carol; Jeff, Ruth, Lucia and Michael; and Craig, for being the people you are.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.	1
II. LITERATURE REVIEW	4
Development of the MMPI.	4
Clinical Versus Actuarial Prediction	6
Automated and Computer-Assisted Interpretive Systems . .	8
Lachar System of Automated Interpretation for the MMPI .	9
Development of Contemporary Norms.	11
Controversy Regarding Use of Normalizing Procedures. . .	14
Representativeness of the Original Sample.	16
Effects of Use of Contemporary Norms	17
III. STATEMENT OF THE PROBLEM.	19
IV. METHOD.	21
Subjects	21
Procedure.	21
Data Analysis.	23
V. RESULTS	26
VI. DISCUSSION.	34
BIBLIOGRAPHY	40
APPENDICES	45
APPENDIX A - TABLE: MEAN ACCURACY RATINGS.	46
APPENDIX B - TABLE: ONE-WAY REPEATED MEASURES ANOVA.	51
APPENDIX C - TABLE: PAIRWISE COMPARISONS BETWEEN MEANS OF ORIGINAL, COLLIGAN GENERAL, AND COLLIGAN AGE-BY-SEX NORMS	53
APPENDIX D - TABLE: 3 X 4 ANOVA WITH REPEATED MEASURES . . .	55
APPENDIX E - HUBERTY'S PROCEDURES FOR ASSESSING IMPROVEMENT .	57

Chapter	Page
APPENDIX F - MMPI-PREDICTED CLASSIFICATION TABLES	60
APPENDIX G - REDUCTION OF FALSE POSITIVES	63
APPENDIX H - SAMPLE PARAGRAPH AND SUBJECT DATA FORM	66

CHAPTER I

INTRODUCTION

The Minnesota Multiphasic Personality Inventory (MMPI) is one of the most widely used personality instruments in this country (Lubin, Larsen, and Matarazzo, 1984). Developed in the 1940s, it has long demonstrated its clinical and research value. Unlike earlier and generally ineffective objective personality tests, which relied primarily on rational test construction approaches and face validity for item selection, the MMPI was constructed along empirical lines. For each scale, a criterion group made up of persons in a specific diagnostic category was selected. Items were selected for that scale if they discriminated between the criterion group and the normative sample.

It was originally hoped that the individual scales on the MMPI would constitute effective measures of the traits whose names they bore. This did not prove to be the case, and the scales are now more often known by their numbers than by their original names in order to avoid overly simplistic interpretation. However, certain patterns of elevations were found to be associated with certain types of psychopathology. Eventually, the research and clinical lore concerning these relationships were formalized in several systems of MMPI code types, which are elevations on single clinical scales or on combinations of two or three scales. Among these are the systems of Marks and Seeman (1963); Gilberstadt and Duker (1965); Gynther, Altman, and Sletten (1973); and Lachar (1974b).

Since its inception, the MMPI has accumulated an enormous body of research data. At the present time, more than 6,000 articles and books have been published dealing with the MMPI (Buros, 1985). It is, in part, this research base as well as its clinical effectiveness that has gained the MMPI its high standing.

However, the MMPI is now half a century old. The first scale was published in 1940. Since the 1940s, our society has undergone many changes. Various authors suggested that revision of the adult norms would be desirable (Butcher and Owen, 1978), and questioned the adequacy of the normative sample in terms of national representativeness (Anastasi, 1976). In 1989, the MMPI-2 was released, with updated norms and a somewhat revised item pool (Graham, 1990). It is still unclear to what extent these revisions will be accepted. One potential problem with the MMPI-2 norms is that professional, highly-educated individuals are over-represented in the sample (Graham, 1990).

Colligan, Osborne, Swenson, and Offord presented their contemporary norms in 1983. Working with a carefully selected normal sample of 1,408 subjects, which was designed to be geographically comparable to the original sample, they found that mean MMPI scores in their sample were significantly above the mean of the original sample for almost every clinical and validity scale. Colligan et al. (1983) interpreted this finding as indicating changes in social attitudes and behaviors since the original norms were published. They then went on to develop two new sets of norms based on their sample. One set permits comparison of the client's responses to those of the same sex in the general adult population. The other, developed in response to the significant age trends found, provides norms for comparison with seven age- and sex-specific groups.

It is worth noting that the new MMPI-2 norms do not include separate norms for different age groups.

Colligan et al. (1983) were concerned by the known skewness and kurtosis of the distributions of raw scores on the various scales. They chose to normalize the distributions for each scale before preparing the tables for conversion of raw scores to T -scores. In the original norms, simple linear transformations had been used, with the result that the skewness and kurtosis of the raw score distributions were preserved in the distributions of T -scores. Hence, the T -scores were not directly comparable across scales.

Both these changes, the development of new norms using a contemporary normative sample and the use of normalized T -scores for those norms, call for empirical investigation to determine the effect of their use on clinical interpretation. The present study is one such attempt. Using the Lachar (1974b) system of automated interpretation to allow comparison, therapists were asked to rate interpretive paragraphs generated from the three sets of norms: the original Hathaway and Briggs norms, the Colligan et al. general population norms, and the Colligan et al. age-by-sex norms.

CHAPTER II

LITERATURE REVIEW

Development of the MMPI

The test now known as the MMPI was developed over a period of time, with the first scale appearing in 1940 (Dahlstrom, Welsh, and Dahlstrom, 1972). Hathaway and McKinley (1940b) reported that the normal sample was composed of 1,040 subjects. Subjects were less carefully selected than would now be customary. The majority were individual visiting patients at the Mayo Clinic; if the person was not under a doctor's care, the assumption was made that he or she was in good health. Most subjects were married, averaged 35 years of age (ranging from 16 to 65), had an average of eight years of formal schooling, lived in a small town or rural area, and worked at a skilled or semi-skilled trade. In general, the sample was believed to correspond well to the age, sex, and marital status of the general population, according to the 1930 census.

This sample was later revised to consist of 226 males and 315 females (Hathaway and Briggs, 1957). The procedure for selecting protocols for the revised sample was only vaguely described, but apparently included reexamining all records to exclude those that were incomplete or appeared defective.

The 10 clinical and 3 validity scales that make up the standard profile of the MMPI were published in a series of articles as they were completed (Hathaway and McKinley, 1942; McKinley and Hathaway, 1940,

1942, 1944). Scale 0 (Si, Social Introversion-Extroversion) was not originally a part of the MMPI. This was a scale developed by Drake (1946) which was added to the standard group of clinical scales because of its usefulness.

The same basic procedure was used in the development of each of the nine clinical scales put forth by Hathaway and McKinley (1940a) (not including Scale 0), with the exception of Scale 7(Pt). In each case, responses made by a criterion group were compared with those of a normative group. The normative samples used in the derivation of the different scales varied somewhat from scale to scale. Items were selected for a scale based on their differentiating power. Hathaway and McKinley did not make their statistical procedures explicit; however, they stated that they required a significant difference, defined as at least twice the standard error of the proportions, in the percentage frequency of agreement between the criterion group and the normal group. Once items were tentatively selected for a scale, they were subjected to cross validation on a new sample of individuals who fit the criterion (Greene, 1980; Hathaway and McKinley, 1942; McKinley and Hathaway, 1940). The item selection procedure differed for Scale 7(Pt). Since the trait was found to be highly homogeneous, internal consistency methods (tetrachoric correlations) were used for the final selection (McKinley and Hathaway, 1942).

The four validity scales (?, L, F, K) were discussed in a series of publications (Hathaway and McKinley, 1940b; Hathaway and McKinley, 1943; Meehl, 1945; Meehl and Hathaway, 1946). A subsequent article by McKinley, Hathaway, and Meehl (1948) summarized the information concerning these scales and provided normative data for Scale K. These validity scales were employed in response to the problem of deliberate or unconscious efforts to distort the results. It had become evident that

persons who wished to do so could exert considerable influence over their MMPI scores, producing a record that seemed either more normal (defensiveness or faking good) or more abnormal ("plus-getting" or faking bad) than was actually the case (Meehl and Hathaway, 1946). To compensate for such attempts, proportions of K were added to Scales 4(Pd), 7(Pt), 8(Sc), and 9(Ma), and K was substituted for an earlier correction factor incorporated in Scale 1(Hs) when it was developed. The correction factor previously incorporated in Scale 2(D), however, was found not to be surpassed by K; this scale remains the only one with an internal correction factor (McKinley, Hathaway, and Meehl, 1948).

Clinical Versus Actuarial Prediction

In the 1940s, a controversy began to develop which has particular relevance for MMPI interpretation. This is the question of clinical versus actuarial (also referred to as "statistical" or "mechanical") prediction. Meehl (1954) summarized the issues involved in his classical book, Clinical Versus Statistical Prediction. Meehl noted that there are two points at which the approach may be clinical or actuarial: in the process of data gathering, and in the methods used for prediction. Data may be collected through psychometric or nonpsychometric techniques. In prediction, either type of data may be used alone or in combination with either type of method, statistical or clinical. Meehl examined 20 studies comparing the relative efficacy of clinical and actuarial predictions. The predictions in these studies were relatively narrow, involving success in school, prison recidivism, or recovery from major psychosis. Meehl's examination of these studies led him to conclude that in all but one, predictions using the actuarial method were as good as or superior to those made by clinical methods.

Sawyer (1966) examined 45 studies bearing on this issue. He also concluded that the mechanical approach to combination and prediction was always equal or superior to the clinical mode, whether the data were collected mechanically or clinically.

Holt (1970) pointed out methodological problems with a number of studies that seemed to support the actuarial approach, including statistical predictions that used weights derived from the same sample to which they were applied and inadequate criterion measures. He also noted that clinical judgment was inevitably involved in the prediction process at most points. He continued to reject the position that the actuarial methods of prediction are superior.

In 1970, Sines looked at 50 studies comparing statistical and clinical predictions of various sorts of behavioral outcomes, giving particular attention to 14 studies that dealt with central issues in psychopathology, such as prediction of which clients are appropriate for psychotherapeutic attention and the client's response to therapy. Of these 14 studies, Sines found that all but one of them supported the conclusion that actuarial prediction is equal or superior to clinical prediction. He went on to discuss the generally low level of success of both methods, especially in psychopathology. He felt that unreliability of criteria limited both methods, and that for "difficult" clients or rare behaviors, neither method might be capable of high rates of accuracy.

The most recent comprehensive survey is that of Dawes, Faust, and Meehl (1989). They noted that close to 100 studies covering a wide range of judgment tasks concerning diagnosis and prediction have yielded results supporting actuarial approaches, varying from slight to substantial. They also considered that the often modest results of even the

best available methods pointed to the present ceiling in the prediction of human behavior.

At this point, the bulk of evidence favors actuarial approaches to prediction. The actuarial method also lends itself readily to adaptation for use with computers.

Automated and Computer-Assisted Interpretive Systems

With the widespread use of computers, programs have been developed, first for computer scoring of psychological tests, and then for computerized systems of test interpretation. These offer the potential advantage of taking into account much more of the research data on a given test, as well as saving time (Eichman, 1972). However, there are also potential pitfalls in the use of such programs. Eichman pointed out that a computer-assisted interpretation of the MMPI is a blind interpretation, with all the difficulties that presents. He also noted that the documentation of the relationship between the interpretation and the data on which it is based is lacking or insufficient in many of the computerized interpretation systems. Butcher (1978b) voiced concern that material generated by computer may have a spurious aura of authenticity. Recognizing early the future of computer-assisted interpretation of psychological tests, the American Psychological Association Council of Representatives in 1966 adopted a set of standards designed to provide guidelines for such systems (Eichman, 1972). However, not all systems have been equally responsive in following these guidelines (Adair, 1978b; Butcher, 1978b; Eichman, 1972; Sundberg, 1985). In 1985, Sundberg reiterated the warning that computer-based test interpretations are easily perceived as scientifically precise, but are still basically subjective.

He stated that most systems use statements derived from the programmer's awareness of clinical lore. At the time of his article, the relationship of empirical research to the interpretive statements was still not published information for most interpretive systems.

Lachar System of Automated Interpretation for the MMPI

In 1974, Lachar published a system of automated interpretation for the MMPI that received praise from reviewers (Adair, 1978a; Butcher, 1978a). The Lachar system represents, to some extent, a combination of the clinical and actuarial approaches to interpretation (Fowler, 1985; Lachar, 1974a). Individual interpretive paragraphs were derived from clinical experience. The accuracy of each interpretive statement was then investigated in a substantial study (Lachar, 1974b).

First published in 1974, Lachar's system more than met the APA interim standards for automated assessment (Adair, 1978a). Lachar (1974b) intentionally took a conservative approach to interpretation. The definition of high point codes is limited to scales over 69 I. The statements themselves and the rules governing their selection were constructed to interpret the most relevant elements of the variance of a specific profile. The statements are phrased to emphasize caution and minimize the aura of authority of computer-generated material (Lachar, 1974b).

The Lachar system consists of a thorough set of decision rules guiding the selection of specific interpretive paragraphs according to the elevations of the validity and clinical scales (Lachar, 1974b). Three types of narratives can result from this process. A Type I narrative is used when one or more of the clinical scales exceed 69 I. It is found most often in working with psychiatric patients, especially with

inpatients. A Type II narrative is used when no scales exceed 69 I, but the rules for including at least one Significant Elevations statement (see below) are satisfied. It is more commonly found in working with outpatient clients. A Type III narrative is used when neither rules for a Type I narrative nor those for a Type II narrative are met; it is essentially a normal profile (Lachar, 1974b).

A Type I narrative consists of two to five sections. Validity, the first section, contains a paragraph selected according to the elevations on the validity scales, L, F, and K. The Code section consists of interpretive paragraphs selected according to the one or two highest scales above 69 I. If there are more than two scales with elevations above 69 I, the section Significant Elevations will be included; this considers the remaining scales in descending rank order and selects up to three additional interpretive paragraphs, one for each elevated scale. Two additional sections may be included if the appropriate rules are met. Since certain patterns of relationships among scales are known to be informative, the presence of any such pattern is dealt with by using the relevant paragraphs from the Configuration section. A final Comments section may be used for remarks and suggestions, such as "Rule out suicidal ideation" (Lachar, 1974b). A Type II narrative consists of Validity and Significant Elevations sections, with sections for Configuration and Comments if the criteria are met.

One of the concerns about computer-based assessments has been that they often fail to provide documentation indicating the relationship between the test data and the interpretive report (Eichman, 1972). This concern is answered in the case of the Lachar system by the publication of Lachar's (1974b) book, which gives in detail the decision rules which govern inclusion of specific paragraphs in the interpretive narrative.

For each interpretive paragraph, a mean profile based on the number of subjects in Lachar's study whose scores warranted use of that paragraph in their report is also provided.

Lachar (1974b) carried out a large study to provide an initial estimate of the accuracy of the individual paragraphs in his interpretive system, using 1,472 subjects with a wide range of presenting problems. Information about the relative accuracy of each paragraph is made available at the end of each paragraph in Lachar's guide to the automated assessment procedure, along with the frequency with which the paragraph occurred in his study (Lachar, 1974b).

The accuracy of the Lachar system of automated MMPI interpretation has also been investigated in an independent study. Adams and Shore (1976) asked the supervising psychiatrists to rate the overall accuracy of the report and of each individual interpretive paragraph of reports on 98 patients, using Lachar's automated procedure. They found that 87% of the overall reports were rated either "completely accurate" or "mostly accurate."

Development of Contemporary Norms

Colligan et al. (1983), at the Mayo Clinic, became interested in the issue of new norms for the MMPI. They also wanted to determine whether the original norms were outdated by comparing MMPIs from a large sample of men and women who met specifications similar to those used in the original norming process, with those from the standard norm group. Since data for the original group are no longer available, these comparisons were made with the Hathaway and Briggs (1957) refined subsample.

Colligan et al. (1983) selected a population-based sample from an area with a 50-mile radius around the Mayo Clinic. Individuals with

chronic diseases, cancer, rheumatoid or other types of arthritis, chemical dependency, learning disability or mental retardation, and those undergoing psychotherapy were not included in the study. Blood relatives were also excluded, because of the work of Gottesman (1963, 1966) and of Hill and Hill (1973) reporting on the heritability of some of the personality traits considered to be assessed by the MMPI. The complete group of subjects used for developing contemporary norms for the MMPI ultimately consisted of 1,408 final subjects.

To compare their data with that of the original Minnesota normal group, as represented by the 1957 Hathaway-Briggs refined sample, the Colligan group decided to draw a subsample of subjects from their total number of subjects. This census-matched subsample consisted of 335 women and 305 men. To answer the question of whether the original MMPI norms were now outdated for present day use, Colligan et al. (1983) plotted the mean MMPI profile of their census-matched subsample, using the original norms. They also plotted the value which was found to be two standard deviations above the norm in their contemporary sample. On almost every scale, the mean score of the contemporary sample was significantly above a \bar{T} -score of 50 (which consisted of the mean calculated for the original normative group). The profiles reflecting scores two standard deviations above the mean (\bar{T} -score of 70) showed very similar increases in elevation (Colligan et al., 1983).

Thus, the sample representative of the present population of the United States responded to the items of the MMPI in ways which resulted in elevated profiles, which would be interpreted as suggesting more psychological or physical distress than was characteristic of the original normative group. Colligan et al. (1983) concluded that the original

norms were indeed outdated and should be revised for use with today's population.

Colligan et al. (1983) now had the raw scores from their contemporary samples; the next step was to develop the conversions of these raw scores to T-scores. In doing so, Colligan et al. decided to prepare normalized T-scores, rather than using a linear conversion, as had been done in the original norms. This decision was based upon their concern about the considerable skewness and kurtosis of the various MMPI scales. The original norming process used a simple linear transformation to obtain T-scores (except for the validity scores Q, L, and F, for which T-scores were arbitrarily assigned). A T-score of 70 thus does not have the same meaning across scales and does not necessarily represent the 97.3 percentile, although these assumptions are often made. Colligan et al. thought that interpretations and assignment of profile code types might suffer from the variable meaning of a T-score of 70. When they examined the distribution of their sample on each scale using the old norms, the percentage of subjects obtaining T-scores over 70 was, in most instances, significantly different from the expected 2.3%. For women, only clinical scales 1(Hs), 3(Hy), and 7(Pt) were within 1% of the expected value, and none of the scales were this close for men.

In preparing their contemporary norms, Colligan et al. (1983) used a normalization procedure to set equal the areas under the skewed distribution, as defined by percentile points, to the equivalent areas under the normal curve, using the Box-Cox power transformation. The specific transformations used, as well as the regression equations on age, were determined individually for each scale.

In selecting raw score transformations, the Colligan group used two criteria. First, the transformation must result in values which had a

Gaussian distribution. Second, it must permit the expression, as deviations from the mean of the normal sample, of raw scores higher than those found in the normal population, since quite elevated scores may be found in a patient population.

For the general adult norms, Colligan et al. (1983) prepared two sets of tables to convert raw scores to normalized T -scores. There was one table each for male and for female adults in the general population. For the age- and sex-specific norms, Colligan et al. developed separate tables for each age group, also divided by gender. The norms for the general adult population were based on the census-matched subsample. The age-by-sex norms were based on the total sample.

Colligan et al. (1983) speculated that use of their new sets of norms would result in some changes in the typical configuration of the profile and the base-rate frequency of single and multiple high point code types, because of the changes in response frequency among items, the changes in norms, and the new age-specific T -scores. Although a period of adjustment was believed to be needed, their experience led them to believe that the changes would not be drastic ones.

Controversy Regarding Use of Normalizing Procedures

Examination of the literature on the MMPI revealed no thorough evaluation of the effect of using normalized T -scores on profiles, code types, or clinical interpretation. Only one study seems to have touched on this issue prior to the work of Colligan et al. (1983). Weisberger (1965) compared normalized MMPI T -scores with linear MMPI T -scores in predicting which applicants would remain in a religious order. He found very little difference in predictive efficiency of the two types of T -scores; if anything, the normalized profiles were slightly less

effective. Eighty-five percent of the single high point scales were the same if single high points were used. When two and three scale high points were used, agreement dropped to 70% and 52%, respectively.

Since publication of the Colligan et al. (1983) contemporary norms, controversy has developed concerning the use of normalized T -scores with the MMPI. Hsu (1984) criticized this procedure, basing his remarks on a 1980 article by Colligan, Osborne, and Offord. Hsu expressed concern that the underlying distribution of psychopathology might not be normal, in which case the use of a normalizing transformation would not be justified. Colligan, Osborne, and Offord (1984) considered the distribution of psychological traits a theoretical question, and noted that the assumption of normal distribution of personality traits would probably continue to be an accepted one in the social sciences, in the absence of convincing evidence to the contrary.

Several of Hsu's (1984) criticisms were based on the incorrect assumption that Colligan et al. (1983) had used the median split technique of normalizing distributions. Hsu pointed out that the median split method, which uses the percentile ranks of scores rather than the raw scores themselves, can result in loss of information in the transformed scores. He also cited Helmes and Jackson (1982), who had found significant amounts of skewness and kurtosis remaining after use of the median split method of normalizing. Hsu also feared that normalized scores might be distorted by the presence of slight differences in extreme values, such as might result from varying degrees of skewness and kurtosis in the transformed distributions. Colligan, Osborne, and Offord (1984) stated that they had chosen the Box-Cox method of normalization to avoid these difficulties.

Hsu's (1984) last criticism of the contemporary norms was the potential loss of comparability to the large body of research and clinical data on the original MMPI. Colligan, Osborne, and Offord (1984) agreed that this potential existed. They commented that this question would have to be investigated empirically and suggested that, in the interim, clinicians plot profiles using both the original and the contemporary norms to enable them to determine which norms were more useful with various populations.

Representativeness of the Original Sample

The argument by Colligan et al. (1983) that contemporary norms were needed was based on the finding that their present day sample obtained scores significantly above the mean of the original sample. Pancoast and Archer (1989) presented information suggesting that the original MMPI norms may have been inaccurate in terms of representativeness to the general population, even at the time that the test was developed. In nine MMPI studies involving normal adults (683 male, 269 female) which had been conducted between 1948 and 1959, mean elevations greater than a T-score of 50 were found for six to seven clinical scales and for the validity scale K. In 11 studies done between 1965 and 1983 (1,588 males, 1,190 females), two clinical scales and K were consistently elevated above a T-score of 50 for men, while seven clinical scales and K were so elevated for women subjects.

Pancoast and Archer (1989) concluded that the original MMPI norms may have been based on an atypical sample which did not correspond well to mean population values at the time they were collected. They also noted that changes in mean MMPI scores over time seemed minimal. Using a procedure developed by Rosenthal and Rubin (1982) for comparing effect

sizes, they found only one significant change: the score of Scale 7(Pt) increased significantly for women from pre-1960 studies to post-1965 studies.

Effects of Use of Contemporary Norms

Pancoast and Archer (1989) expressed concern that use of new norms might actually render the MMPI less helpful in clinical use, as the resulting changes in scale elevations, and hence in the code types on which interpretations are based, would invalidate the enormous body of clinical interpretation literature. They noted that information on the degree to which old and new norms converge and diverge will be needed, and that a lengthy process of establishing the clinical correlates of new norms may be required.

Thus far, few studies have been done to examine the effects of using the Colligan et al. (1983) contemporary norms. Greene (1985) mentioned that in a sample of 17 MMPI profiles, he found four cases with different 2-point codes and nine cases with the same 2-point codes. In four profiles, a reversal of order of the high point scales occurred between the original and the contemporary norms; Greene stated that in three of these cases, the reversed order of the 2-point codes would cause interpretive differences.

In 1986, Miller and Streiner published an investigation of the differences in the number of elevated scales and the changes in 2- and 3-point code types resulting from use of the Colligan et al. (1983) General norms compared to that of the standard norms, using over 2,000 profiles. They found that the overall elevations of the profiles were reduced, but 64% of the 2-point and 72% of the 3-point code types had no difference in elevation ranks. Twenty-seven percent of the 2-point code types and 50%

of the 3-point code types resulting from use of the Colligan et al. General norms were quite different from those found with the standard norms, showing either considerable rearrangement of the scales or inclusion of a scale which would not have been part of the code type under the original norms. Miller and Streiner recommended using the old norms as well as the contemporary norms until the nature of the differences is thoroughly understood.

Munley and Zarantonello (1989) examined changes in code types when the Colligan et al. (1983) General norms were used to plot distinctive, highly differentiated MMPI profiles that had been derived using the original norms. These profiles were those given in Gilberstadt and Duker (1965) and Marks, Seeman, and Haller (1974) as exemplars of clinical groups. In comparison to the profiles when the standard norms were used, the normalized profiles showed lower average scale elevations and decreased dispersal around mean scores on the various scales. The overall shape of the profiles, however, remained quite consistent. When the 2- or 3-high point code types were compared, 50% remained identical after transformation with the Colligan et al. (1983) General contemporary norms. Forty-four percent showed a change in the order of scale elevations. Only 6% of the transformed profiles yielded code types which included a scale elevation not present in the original code type. Munley and Zarantello also recommended use of both the original contemporary norms until much more work has been done.

CHAPTER III

STATEMENT OF THE PROBLEM

The goal of clinical use of the MMPI is an accurate interpretation of the patient. Once Colligan et al. (1983) had developed their two sets of contemporary norms, the question of the effect on interpretation of using the new norms arose. There was also controversy over the acceptability of normalization of the distributions. The problem to which this study was addressed was the assessment of comparative accuracy of Lachar (1974b) automated interpretations among the Original MMPI norms, the Colligan General norms, and the Colligan Age-by-Sex norms.

Over the years, several widely-used interpretive systems have been developed for use with the MMPI. To a large extent, these are based on clinical judgment, on the accumulated clinical lore of their developers, and of those who have written and taught the use of this test. These systems specify meanings to be assigned to MMPI code types, which are elevations on single scales, or on combinations of two or three scales. If the underlying scores are not an accurate reflection of the individual's standing in relationship to the general population, the correctness of the interpretations based on these scores is called into question. The issue here is whether the use of contemporary norms would succeed in improving accuracy, even with the use of the same set of interpretive rules. It is possible that the interpretive systems somehow compensate for flaws in the original norms or changes in attitudes since the development of the original norms.

The Lachar (1974b) system of automated interpretation was selected for use in this study because Lachar has published all the algorithms used in his system, making it more appropriate for research than those systems for which this information is not available. Lachar's system, like the others, is based primarily on clinical judgment. He also attempted to evaluate the accuracy of individual paragraphs in his system.

Three different results of the new norms on interpretation were considered possible. No significant effect on interpretation might be found. If there were an effect, the accuracy of interpretation could either be enhanced or diminished. To the extent that Lachar's (1974b) interpretations compensate for the non-normal distributions and the relationship between the original \bar{I} -scores and the contemporary population, the use of new norms might have an adverse effect on the accuracy of the interpretations. To the extent that the Lachar interpretations do not take these factors into account, and to the extent that normal distributions more accurately reflect population characteristics, the use of new norms should improve the accuracy of the interpretations. The present study investigated this question by asking therapists to rate Lachar (1974b) automated interpretations based on the original norms as compared to those based on the new norms developed by Colligan et al. (1983). Both the new norms for the general adult population and those subdivided by age group were used.

CHAPTER IV

METHOD

Subjects

Subjects were obtained from the various units of a Veterans' Administration Medical Center in the southwestern United States. Subjects consisted of 78 adult clients currently being seen or recently seen in inpatient or outpatient psychotherapy by eight practicing psychotherapists who possessed a Ph.D. in psychology. Seventy-four subjects were male; four were female. The mean age was 41.22 years; ages ranged from 25 to 74 years. Educational level was reported for 57 subjects. For these, mean educational level was 12.35 years; educational level ranged from 2 to 20 years. Subjects met three criteria: (1) they had been seen by their therapists at least four times, or the therapist believed he/she knew the patient well enough from contact on the unit to rate the interpretive paragraphs; (2) they had completed the MMPI; and (3) they had obtained a score of 69 \bar{I} or more on at least one clinical scale of the MMPI. There were also two exclusion criteria, following Lachar's (1974b) conditions for considering a profile invalid: (1) an F - K raw score difference of 16 or more; and (2) 30 or more items omitted from the record.

Procedure

Each of the eight therapists submitted complete MMPIs from 10

clients. Two male subjects had to be eliminated from the study because of incomplete data. Information on subjects was anonymous; records were kept in the name of the therapist and the subject number assigned by the experimenter when the MMPI profile was received.

Raw scores for each MMPI were supplied with the original record, since patients at this VA medical center take the MMPI by computer. The same computer record also indicated T-scores based on the original norms. These were referred to as "Original T-scores." T-scores based on the contemporary norms were obtained by using the tables in Colligan et al. (1983). Two sets of T-scores were derived from the Colligan et al. norms. One was based on their norms for the general population, divided by sex. The other was based on their norms for the seven specific age- and sex-groups. These scores were referred to subsequently as "Colligan General T-scores," and "Colligan Age-by-Sex T-scores," respectively.

For each of the three profiles obtained through the use of the three sets of norms, the appropriate paragraphs were selected from Lachar's (1974b) automated interpretation procedure. Three sets of automated interpretations were generated, one for each of the three sets of norms. In cases where the same paragraph appeared in more than one interpretation, redundant paragraphs were deleted.

Each paragraph was typed on a separate page. Appropriate identification was given on the back of each page, including therapist's name, age and sex of client, and subject number of client. The therapists also completed a separate page of demographic data on the clients, which was identified on the back of the page in the same way. (A sample paragraph and subject data sheet are shown in Appendix H.) Beneath each interpretive paragraph, a 7-point Likert scale was drawn, with 1 indicating "completely accurate," and 7 signifying "completely inaccurate."

Paragraphs were presented in random order to the therapists. Validity paragraphs were very slightly revised when necessary, to ensure that they made sense standing alone and to avoid biasing the ratings; for example, the phrase "this is a valid profile" was deleted. When the selection criteria for a Code paragraph dictated inclusion of a statement from the Comments section (e.g., the statement, "consider chemotherapy for depression"), that statement was included with the Code paragraph, as it often did not make sense by itself. Clinicians were instructed to rate each paragraph for its accuracy in describing their clients. They were asked to rate each paragraph independently, disregarding the rest of the interpretive paragraphs. The therapists were also asked to assign the subject to one of the four general categories: Neurotic, Psychotic, Characterological, or Indeterminate. This information was included on the page of demographic data.

Data Analysis

An accuracy rating was obtained for each of the four report sections (Validity, Code, Significant Elevations, and Configuration) for each norm group. In cases where a report section included more than one paragraph, the ratings for the paragraphs involved were averaged. Differences in accuracy ratings for the three sets of norms were evaluated using a 3 x 4 (norms used by report section), two-way analysis of variance with repeated measures on both factors. Only data from subjects whose reports contained ratings from all four report sections were included in this analysis to avoid unequal n s. In order to use data from all subjects, an additional one-way, within subjects, ANOVA was performed using only the average rating over all paragraphs included in the report generated by each of the three sets of norms.

At the time the raw data were submitted, therapists classified each client into one of four groups: Neurotic, Psychotic, Characterological, or Indeterminate. Each client was then objectively grouped into one of the same four classifications, according to Lachar's (1974b) predicted classification based on code type. Code types are a commonly used way of referring to MMPI profiles by the most elevated scores above 69 I; for example, a profile whose highest elevations were on scales Pd (Scale 4) and D (Scale 2) would be referred to as a 4-2 code type. The order of the high point scales is usually considered interchangeable. When only one scale is above 69 I, a single high point code is used; for example, 4.

The objective classification into predicted categories was carried out separately for each of the three sets of norms. The three resultant tables of predicted versus actual classes were analyzed using Huberty's (1984) procedure for classification analysis to determine if use of the respective profile improved diagnostic classification beyond chance levels. (See Appendix E for a discussion of Huberty's procedure.)

Percentage changes in code types among the three sets of norms were also calculated. For example, a subject whose highest scales were 4 (Pd) and 8 (Sc) would be a 4-8 code; use of the new norms might result in a code change to a 4-9. Cases in which one scale dropped out in the transition from the Original norm code type to the code type derived from one of the two sets of Colligan contemporary norms were reported as a percentage of the total number of cases. For example, 58% of the total profiles showed changes in code type from the Original MMPI norms to the Colligan General norms. This included 37% of the total profiles in which one scale dropped out between the two norms. Cases in which the same scales were present in the code type in reversed order were treated

similarly. The percentage of code types which changed and which remained different were compared with those found in other studies that had examined shifts in code types between the Original norms and the Colligan General norms.

CHAPTER V

RESULTS

For each paragraph in the Lachar (1974b) system, the mean accuracy rating, standard deviation, and frequency of appearance are reported in Appendix A. The one-way within subjects ANOVA using the overall accuracy rating (N=78) for each of the three norms was statistically significant:

$$F(2,77) = 3.76, p < .05.$$

For means and standard deviations, see Table I. The ANOVA summary table (Table X) is given in Appendix B.

TABLE I
MEANS AND STANDARD DEVIATIONS OF OVERALL REPORT
ACCURACY RATINGS FOR THREE MMPI NORMS

Norm ^a	Mean	Standard Deviation
Original	2.76	0.76
Colligan General	2.63	0.84
Colligan Age-by-Sex	2.61	0.90

^a_{n=78}

Post hoc tests using Tukey's HSD revealed that computerized interpretations based on Colligan Age-by-Sex norms were rated significantly more accurate ($p < .05$) than those based on the original norms. Computerized interpretations based on Colligan General norms were not rated significantly more accurate than those based on original norms, nor was there a significant difference in rated accuracy between interpretations based on the Colligan General norms and those based on the Colligan Age-by-Sex norms. (For results of the post hoc tests, see Appendix C.)

In order to assess the accuracy of the separate sections of the interpretive reports, accuracy ratings were averaged separately for the Validity paragraphs, the Code paragraphs, the Significant Elevations paragraphs, and the Configurations paragraphs. Means and standard deviations for these separate segments are reported in Table II. Subjects whose reports did not include all four sections for each set of norms were deleted, resulting in a reduced N of 29 for this analysis. Differences in accuracy of the three sets of norms were evaluated, using a 3×4 , two-way analysis of variance with repeated measures on both factors. Factors were the three sets of norms (Original, Colligan General, and Colligan Age-by-Sex), and the four report sections (Validity, Code, Configuration, and Significant Elevations). None of the main effects or interactions were significant. The summary table (Table XII) of the 3×4 ANOVA is shown in Appendix D.

Another way of determining the effects of using different norms is by using high point codes to classify accurately people into broad psychiatric categories. The Lachar (1974b) predicted classifications were compared to the actual classification obtained from the patients' therapists.

TABLE II
 MEAN ACCURACY RATINGS AND STANDARD DEVIATIONS
 OF SEPARATE REPORT SEGMENTS FOR
 THREE MMPI NORMS

Norm ^a	Validity	Code	Significant Elevation	Configuration
<u>Original</u>				
Mean	2.93	2.64	2.97	3.07
SD	1.44	1.62	1.12	1.26
<u>Colligan- General</u>				
Mean	2.74	2.82	2.90	2.99
SD	1.06	1.52	1.33	1.46
<u>Colligan Age-by-Sex</u>				
Mean	2.81	2.66	2.87	2.94
SD	1.31	1.55	1.29	1.51

^an=29

At the time the raw data were submitted, therapists classified each client into one of four groups: neurotic, psychotic, characterological, or indeterminate. Each client was then objectively grouped into one of the same four classifications, according to Lachar's (1974b) grouping by profile type, in which profiles are categorized according to the highest scale scores. If all scales were less than 70 T, the subject was classified into a fifth category, Normal. This occurred for 14 subjects with the Colligan General norms and 11 subjects with the Age-by-Sex norms. These subjects had to be omitted from the classification analysis, as

their inclusion would have resulted in unbalanced 4 x 5 tables for these two sets of norms. The three classification tables are shown in Tables XII, XIII, and XIV in Appendix F. It should be noted that use of the two new norms resulted in sufficiently lowered scores for 10 subjects classified as Characterological by their psychologists that these subjects had to be categorized as Normal in the MMPI predicted classifications. The three 4 x 4 tables of predicted versus actual classes were analyzed using Huberty's (1984) procedure. (See Appendix E for a discussion of Huberty's procedure.)

In the population from which the subjects were drawn, the majority of patients receive characterological diagnoses. Hence, the maximum chance criterion was used to evaluate overall hit rates, in accordance with Huberty's (1984) suggestion that this criterion be used when the n 's in the different classifications vary widely. Overall hit rates using the maximum chance criterion and reduction in remaining error percentages (I) are presented in Table III. Significantly lower success was found in using the MMPI for making these classifications than would have occurred by simply categorizing all subjects as Characterological. This held for all three sets of norms.

Although the maximum chance criterion is best for assessing overall classification accuracy, in some situations it is also useful to look at separate group hit rates using the proportional chance criterion. This would occur when it is important to identify accurately classification into certain groups (Huberty, 1984). For example, in this population, identification of psychotic and neurotic patients would be important; in a psychiatric hospital in which most patients were psychotic, it might be needed to detect characterological patients.

TABLE III
 OVERALL HIT RATES IN CLASSIFYING SUBJECTS INTO
 BROAD PSYCHIATRIC GROUPINGS USING
 MAXIMUM CHANCE CRITERION WITH
 THREE MMPI NORMS

Norm ^a	Hits	Hit Rate (%)	% Correct ^b	<u>z</u>	<u>I</u> (%)
Original	43	55.13	67.95	-2.43**	-40.00
Colligan General	33	42.31	67.95	-4.85**	-80.00
Colligan Age-by-Sex	34	43.59	67.95	-4.60**	-76.00

^an=78

^bPercentage of cases correctly classified using the Maximum Chance Criterion.

**p < 0.01

Hit rates and reduction in remaining error percentages (I) for the separate classifications are shown for each of the sets of norms in Tables IV, V, and VI, respectively. It can be seen by looking at these tables that classification of patients into neurotic and psychotic categories was significantly improved by use of the MMPI. This was true for all three sets of norms.

Changes in 1- and 2-point code types were observed among the original norms, the Colligan General norms, and the Colligan Age-by-Sex norms. These results are presented in Table VII. As can be seen, approximately 60% of the code types were changed by transformation to the two new sets of norms.

TABLE IV
SEPARATE GROUP HIT RATES IN CLASSIFYING SUBJECTS
INTO BROAD PSYCHIATRIC GROUPINGS USING
PROPORTIONAL CHANCE CRITERION WITH
ORIGINAL MMPI NORMS

Clinician's Classification	N	Hits	Hit Rate (%)	Chance Frequency of Hits	\underline{z}	$\underline{I}(\%)$
Neurotic	10	7	70	1.28	5.41**	65.59
Psychotic	11	8	73	1.55	5.59**	68.25
Character-ological	53	28	53	36.01	-2.36**	-47.17
Indeterminate	4	0	--	0.21	-0.46	-5.41

**p < .01

TABLE V
SEPARATE GROUP HIT RATES IN CLASSIFYING SUBJECTS
INTO BROAD PSYCHIATRIC GROUPINGS USING
PROPORTIONAL CHANCE CRITERION WITH
COLLIGAN GENERAL MMPI NORMS

Clinician's Classification	N	Hits	Hit Rate (%)	Chance Frequency of Hits	\underline{z}	$\underline{I}(\%)$
Neurotic	10	7	70	1.28	5.41**	65.59
Psychotic	11	7	64	1.55	4.72**	57.67
Character-ological	53	18	34	36.01	-5.30**	-106.04
Indeterminate	4	1	25	0.21	1.80*	20.95

*p < .05

**p < 0.01

TABLE VI
 SEPARATE GROUP HIT RATES IN CLASSIFYING SUBJECTS
 INTO BROAD PSYCHIATRIC GROUPINGS USING
 PROPORTIONAL CHANCE CRITERION WITH
 COLLIGAN AGE-BY-SEX MMPI NORMS

Clinician's Classification	N	Hits	Hit Rate (%)	Chance Fre- quency of Hits	\underline{z}	$\underline{I}(\%)$
Neurotic	10	6	60	1.28	4.46**	54.12
Psychotic	11	8	73	1.55	5.59**	68.15
Character- ological	53	20	38	36.01	-4.71**	-94.26
Indeterminate	4	0	--	0.21	-0.46	-5.41

**p < 0.01

TABLE VII
 PERCENTAGE CHANGES IN 1- AND 2-POINT CODE TYPES
 AMONG THREE MMPI NORMS IN PRESENT STUDY

	Original Norms	Colligan General Norms	Colligan Age-by-Sex Norms
Original Norms	--	58	63
Colligan General Norms	--	--	31

Three studies examined changes in code types between the original norms and the Colligan General norms (Greene, 1985; Miller and Streiner, 1986; Munley and Zarantonello, 1989). Their results are compared with those of the present study in Table VIII. It can be seen from Table VIII that this study found more differences in code types between original and Colligan General norms than any of these three studies. The present study was also the only one to compare code types obtained by use of the Colligan Age-by-Sex norms to those from the Original norms; the other studies used only the Colligan General norms.

TABLE VIII
PERCENTAGE CHANGES IN 1- AND 2-POINT CODE TYPES
BETWEEN ORIGINAL AND COLLIGAN
GENERAL NORMS

	Present Study	Greene (1985)	Miller & Streiner (1986)	Munley & Zarantonello (1989)
Code Types Different From Original to Colligan General Norms	58	24	37	6
One Scale Dropped Out Between Original and Colligan General Norms ^a	37	not given	not given	not given
Codes Same	43	76	48	50
Code Type Same, But Scale Order Reversed ^b	12	24	15	44

^aCases in which one scale dropped out are also included in the total number of code types considered different from the Original norms to the Colligan General norms.

^bCases in which scale order was reversed are also included in the total number of code types considered the same from the Original norms to the Colligan General norms.

CHAPTER VI

DISCUSSION

Earlier studies made comparisons between mean elevations and dispersions of profiles and between code types obtained when scores on the original MMPI norms were transformed to the contemporary norms developed by Colligan et al. (1983). The present study appears to be the first to examine the effects of these new norms on the accuracy of interpretive reports. It also seems to be the only one thus far which has included the age-specific norms also provided by the Colligan group. In the present study, use of the Colligan Age-by-Sex contemporary norms resulted in significant improvement of rated accuracy of overall interpretive reports over those using the original norms, as generated by the Lachar (1974b) system of automated MMPI interpretations. The Colligan General norms did not result in significantly improved accuracy ratings.

The results of the 3 x 4 ANOVA were not significant for either the Norms or the Segments factors, separately, or for the interaction of Norms by Segments. Thus, when looking at the reports section by section, no improvement in accuracy rating was seen. It is difficult to interpret these results since in order to avoid empty cells, the N had to be cut severely. It may be that with a larger sample the separate segments would also have shown differences in accuracy ratings among the norms.

It should be noted that the mean accuracy ratings for overall reports generated by the use of the three norms were actually quite close (see Table I, Chapter V). On a scale of 1 (indicating completely

accurate) to 7 (indicating completely inaccurate), reports based on all three norms were rated fairly accurate. One might speculate whether the differences found, though statistically significant, would be of much clinical utility.

Did the use of the new norms improve classification of patients into broad psychiatric groupings? Although the results of this study are limited to those using the Lachar system, two interesting implications emerged, the first for settings in which the goal is simply to achieve the highest overall hit rate, and the second for situations in which it is important to detect any patients with diagnoses unusual for that site. The population from which the sample for this study was drawn was one in which more than a majority of patients received a characterological diagnosis of one kind or another. Not surprisingly, use of the MMPI did not improve accuracy of classification for such characterological patients. In such a setting, the highest overall chance hit rate is obtained by assigning all patients to the majority classification. This would also be the case in other settings if a predominant number of the diagnoses tended to fall in a particular grouping; for example, "Neurotic," or "Psychotic." However, since there are situations in which it is important to discern any cases which differ from the majority diagnosis, it should be noted that the use of the MMPI significantly improved such separate group hit rates. This was true for all three sets of norms. Also, the \bar{I} values suggest that the Original norms did a better job, with 65% or better reduction in remaining error in detecting both neurotic and psychotic cases.

In analyzing improvement in classifying subjects, one is concerned not only with hit rate and false negatives, but also with false positives. The Huberty analyses described in Appendix E deals with hit

rates and reduction in false negatives by looking at the number of hits compared to the number of subjects actually in that classification.

However, the false positives are also important. Separate group analysis by the Huberty method suggests that use of the MMPI to identify subjects in a majority classification (Characterological in this study), actually reduces hit rate for that classification, as shown in Tables IV, V, and VI (see Chapter V). However, there is another way to look at this issue. One can also compare the number of hits to the number of subjects who were predicted to be in that category, obtaining a percentage indicating the reduction of false positives. Using this method, it may be seen that Characterological subjects are mispredicted very rarely in the MMPI predictions with any of the three sets of norms. These data are shown in Tables XVI, XVII, and XVIII in Appendix G. Between 86% and 88% of those subjects predicted to be Characterological by the MMPI, using the three sets of norms, were actually classified as Characterological by their clinicians. Thus, by reducing the false positives, the MMPI appears to be more useful in these classifications than the Huberty analysis alone would suggest.

What can be learned from the present study with regard to its implications for test construction? The use of normalized scores yielded reports which were slightly improved in accuracy, at least when age differences were also taken into account, over those from the original MMPI norms. The original linear scaled scores of the MMPI were known to be skewed and were not uniform from scale to scale. In the Colligan norms, all the scales are normalized, with the result that the same T-score has the same percentile equivalent from scale to scale. The use of the Colligan norms, however, also results in a generalized lowering of scores, as has also been noted by Greene (1985), Miller and Streiner (1986), and

Munley and Zarantonello (1989). While this lowering was part of the rationale for the development of contemporary norms, it may also account, in part, for the lowered hit rates found in classifying patients into broad categories using the Colligan norms; the lowered \bar{I} -scores resulted in 10-14 patients being classified as normal.

This study did not reveal any information about the distribution of psychiatric traits in the population. If the Colligan General norms had resulted in more accurate interpretations, it would have suggested that the traits are normally distributed in the population. Alternatively, had the Original norms been superior, it would have suggested either that the traits are not normal or that Lachar's (1974b) interpretations are modified by experience to take such peculiarities into account. The question as to whether such traits are distributed normally or in a skewed fashion in the population remains open. What is indicated in this study, however, is the utility of age and sex norms for accurately assessing psychiatric patients.

While many tests have offered separate norms for men and women, it has been less common to use age norms, at least for adults. Yet, this study suggested that having separate norms for age and sex groups appears to make a difference. It seems possible that the answers given by males and females and by persons in different age groups may differ enough to affect report accuracy. The results of this study suggested that development of age-specific norms should also be considered.

Although it had been hoped that individual paragraphs in the Lachar (1974b) system could be evaluated, it was found that the number of subjects was insufficient to determine whether or not individual paragraphs should be modified. Many of the individual paragraphs had zero or very

low frequencies (see Appendix A). A very large number of subjects would be necessary to address this question.

The present study can also be compared with the other three recent studies which examined similarities and differences in 2-point code types based on original norms versus those based on the Colligan General transformed norms. This study found more changes (58%) in 1- or 2-point code types after transformation to Colligan General norms than the Munley and Zarantonello (1989) study (6% different), the Greene (1985) study (24%), or the Miller and Streiner (1986) study (37%). Since the Munley and Zarantonello (1989) study used only the highly differentiated profiles which had been presented in the literature as exemplars of the code types, their much lower rate of differences in code types after transformation is perhaps not surprising. The difference rate of 37% found in the Miller and Streiner (1986) study, which used a very large sample, is closer to the 58% difference rate found in the present study, but still constitutes a considerable variation. These observed changes in code types appear to support Pancoast and Archer's (1989) concern about the lack of correspondence of the new norms to the interpretive literature on the MMPI. Clearly, more studies are needed to establish the parameters of expected similarities and differences in code types between the Original MMPI norms and the Colligan General norms. Until such parameters have been established, it continues to be advisable to use the Original norms in conjunction with either of the Colligan norms.

Neither the Munley and Zarantonello (1989) nor the Miller and Streiner (1986) studies compared code-type changes from the original set of MMPI norms to the Colligan Age-by-Sex norms, nor did they compare the Colligan General norms to the Colligan Age-by-Sex norms. In the present investigation, a majority (63%) of the code types were different when the

Original norms were compared with the Colligan Age-by-Sex norms. This is comparable to the 58% difference rate from Original to Colligan General norms. Code types were similar between the Colligan General and the Colligan Age-by-Sex norms, with 69% remaining the same. Again, further investigations comparing the Original norms with the Colligan Age-by-Sex norms, and comparing the Colligan General and the Colligan Age-by-Sex norms are needed.

A limitation of this study was that the accuracy ratings for each subject were determined by one rater, the subject's psychologist. As these ratings were not necessarily uniform across the 10 clinicians, there was no good way to ascertain the reliability of the ratings.

Further studies to examine the effect on interpretive accuracy of the Colligan contemporary norms are needed. It is strongly recommended that such investigations include the Colligan Age-by-Sex norms, since the present study found significantly improved accuracy ratings only by using the Age-by-Sex norms. Since the publication of the Colligan et al. (1983) contemporary norms, the MMPI-2 has been published. As the classic MMPI is so well established, it is likely that it will continue to be used, at least for some time. Studies comparing accuracy of interpretation of the original MMPI using the Colligan contemporary norms with that of the MMPI-2 are recommended.

BIBLIOGRAPHY

- Adair, F. L. (1978a). Automated psychological assessment (Lachar). In: The Eighth Mental Measurements Yearbook: Vol. I, O. K. Buros, Ed. Highland Park, New Jersey: Gryphon, 945-947.
- Adair, F. L. (1978b). Computerized scoring and interpreting services. In: The Eighth Mental Measurements Yearbook: Vol. I, O. K. Buros, Ed. Highland Park, New Jersey: Gryphon, 940-942.
- Adams, K. M. and Shore, D. L. (1976). The accuracy of an automated MMPI interpretation system in a psychiatric setting. Journal of Clinical Psychology, 32(1), 80-82.
- Alker, H. A. (1978). Minnesota Multiphasic Personality Inventory. In: The Eighth Mental Measurements Yearbook: Vol. I, O. K. Buros, Ed. Highland Park, New Jersey: Gryphon, 931-935.
- Anastasi, A. (1976). Psychological Testing (4th ed.). New York: Macmillan.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society (Series B), 26, 211-243.
- Buros, O. K., Ed. (1985). Ninth Mental Measurements Yearbook. Highland Park, New Jersey: Gryphon.
- Butcher, J. N. (1978a). Automated psychological assessment (Lachar). In: The Eighth Mental Measurements Yearbook: Vol. I, O. K. Buros, Ed. Highland Park, New Jersey: Gryphon, 947-948.
- Butcher, J. N. (1978b). Computerized scoring and interpreting services. In: The Eighth Mental Measurements Yearbook: Vol. I, O. K. Buros, Ed. Highland Park, New Jersey: Gryphon, 942-945.
- Butcher, J. N. and Owen, P. L. (1978). Objective personality inventories: Recent research and some contemporary issues. In: Clinical Diagnosis of Mental Disorders: A Handbook, B. B. Wolman, Ed. New York: Plenum, 475-545.
- Colligan, R. C., Osborne, D., and Offord, K. P. (1980). Linear transformation and the interpretation of MMPI I scores. Journal of Clinical Psychology, 36, 162-165.
- Colligan, R. C., Osborne, D., and Offord, K. P. (1984). Normalized transformations and the interpretations of MMPI I scores: A reply to Hsu. Journal of Consulting and Clinical Psychology, 52(5), 824-826.

- Colligan, R. C., Osborne, D., Swenson, W. M., and Offord, K. P. (1983). The MMPI: A Contemporary Normative Study. New York: Praeger.
- Dahlstrom, W. G. and Welsh, G. S. (1960). An MMPI Handbook: A Guide to Use in Clinical Practice and Research. Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., and Dahlstrom, L. E. (1972). An MMPI Handbook: Vol. I. Clinical Interpretation. Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., and Dahlstrom, L. E. (1975). An MMPI Handbook: Vol. II. Research Applications. Minneapolis: University of Minnesota Press.
- Dawes, R. M., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. Science, 243, 1668-1674.
- Drake, L. E. (1946). A social I.E. scale for the Minnesota Multiphasic Personality Inventory. Journal of Applied Psychology, 30, 51-54.
- Eichman, W. J. (1972). Computerized scoring and interpreting services. In: The Seventh Mental Measurements Yearbook, Vol. 1., O. K. Buros, Ed. Highland Park, New Jersey: Gryphon, 105-106.
- Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. Journal of Consulting and Clinical Psychology, 53(6), 748-759.
- Gilberstadt, H. and Duker, J. (1965). A Handbook for Clinical and Actuarial MMPI Interpretation. Philadelphia: W. B. Saunders.
- Gottesman, I. I. (1966). Genetic variance in adaptive personality traits. Journal of Child Psychology and Psychiatry, 7, 199-208.
- Gottesman, I. I. (1963). Heritability of personality: A demonstration. Psychological Monographs, 77(9), 1-21.
- Graham, J. R. (1990). MMPI-2: Assessing Personality and Psychopathology. New York: Oxford University Press.
- Greene, R. L. (1980). The MMPI: An Interpretive Manual. New York: Grune and Stratton.
- Greene, R. L. (1985). New norms, old norms, what norms for the MMPI? Journal of Personality Assessment, 49, 108-110.
- Gynther, M. D., Altman, H., and Sletten, I. W. (1973). Replicated correlates of MMPI two-point code types: The Missouri actuarial system. Journal of Clinical Psychology, 29, 263-289.
- Hathaway, S. R. (1939). The personality inventory as an aid in the diagnosis of psychopathic inferiors. Journal of Consulting Psychology, 3, 112-117.

- Hathaway, S. R. (1956). Scales 5 (masculinity-femininity), 6 (paranoia), and 8 (schizophrenia). In: Basic Readings on the MMPI in Psychology and Medicine, G. S. Welsh and W. G. Dahlstrom, Eds. Minneapolis: University of Minnesota Press, 104-111.
- Hathaway, S. R. (1976). Foreword. In: The Clinical Use of the Automated MMPI, by R. D. Fowler. Nutley, New Jersey: Hoffmann-La-Roche, iv.
- Hathaway, S. R. and Briggs, P. F. (1957). Some normative data on the new MMPI scales. Journal of Clinical Psychology, 13, 364-368.
- Hathaway, S. R. and McKinley, J. C. (1940a). The measurement of symptomatic depression with the Minnesota Multiphasic Personality Schedule. Psychological Bulletin, 37, 425.
- Hathaway, S. R. and McKinley, J. C. (1940b). A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. Journal of Psychology, 10, 249-254.
- Hathaway, S. R. and McKinley, J. C. (1942). A Multiphasic Personality Schedule (Minnesota): III. The measure of symptomatic depression. Journal of Psychology, 14, 73-84.
- Hathaway, S. R. and McKinley, J. C. (1943). The Minnesota Multiphasic Personality Inventory (rev. ed.). Minneapolis: University of Minnesota Press.
- Helmes, E. and Jackson, D. N. (1982). A comparison of methods of normalizing a discrete distribution. Journal of Clinical Psychology, 38, 581-587.
- Hill, M. S. and Hill, R. N. (1973). Hereditary influence on the normal personality using the MMPI: I. Age-corrected parent-offspring resemblances. Behavior Genetics, 3, 133-144.
- Holt, R. R. (1970). Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? American Psychologist, 25(4), 337-349.
- Hsu, L. M. (1984). MMPI T scores: Linear versus normalized. Journal of Consulting and Clinical Psychology, 52(5), 821-823.
- Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin, 95(1), 156-171.
- Lachar, D. (1974a). Accuracy and generalizability of an automated MMPI interpretation system. Journal of Consulting and Clinical Psychology, 42(2), 267-273.
- Lachar, D. (1974b). The MMPI: Clinical Assessment and Automated Interpretation. Los Angeles: Western Psychological Services.

- Lachar, D. and Alexander, R. S. (1978). Veridicality of self report: Replicated correlates of the Wiggins MMPI Content scales. Journal of Consulting and Clinical Psychology, 46, 1349-1356.
- Lubin, B., Larsen, R. M., and Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States, 1935-1982. American Psychologist, 39, 451-454.
- Marks, P. A. and Seeman, W. (1963). The Actuarial Description of Abnormal Personality. Baltimore: Williams and Wilkins.
- Marks, P. A., Seeman, W., and Haller, D. L. (1974). The Actuarial Use of the MMPI With Adolescents and Adults. Baltimore: Williams and Wilkins.
- McKinley, J. C. and Hathaway, S. R. (1940). A multiphasic personality schedule (Minnesota): II. A differential study of hypochondriasis. Journal of Psychology, 10, 255-268.
- McKinley, J. C. and Hathaway, S. R. (1942). A multiphasic personality schedule (Minnesota): IV. Psychasthenia. Journal of Applied Psychology, 26, 614-624.
- McKinley, J. C. and Hathaway, S. R. (1944). The Minnesota Multiphasic Personality Inventory: V. Hysteria, hypomania, and psychopathic deviate. Journal of Applied Psychology, 28, 153-174.
- McKinley, J. C., Hathaway, S. R., and Meehl, P. E. (1948). The Minnesota Multiphasic Personality Inventory: VI. The K scale. Journal of Consulting Psychology, 12, 20-31.
- Meehl, P. E. (1945). An investigation of a general normality or control factor in personality testing. Psychological Monographs, 59(4), 1-62.
- Meehl, P. E. (1954). Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1965). Seer over sign: The first good example. Journal of Experimental Research in Personality, 1, 27-32.
- Meehl, P. E. and Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. Journal of Applied Psychology, 30, 525-564.
- Miller, H. R. and Streiner, D. L. (1986). Differences in MMPI profiles with the norms of Colligan et al. Journal of Consulting and Clinical Psychology, 54(6), 843-845.
- Munley, P. H. and Zarantonello, M. M. (1989). A comparison of MMPI profile types across standard and contemporary norms. Journal of Clinical Psychology, 45(2), 229-239.

- Newmark, C. S., Ed. (1979). MMPI Clinical and Research Trends. New York: Praeger.
- Pancoast, D. L. and Archer, R. P. (1989). Original adult norms in normal samples: A review with implications for future developments. Journal of Personality Assessment, 53(2), 376-395.
- Rosenthal, R. and Rubin, D. B. (1982). Comparing effect sizes of independent studies. Psychological Bulletin, 92, 500-504.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. Psychological Bulletin, 66(3), 178-200.
- Sines, J. O. (1970). Actuarial versus clinical prediction in psychopathology. British Journal of Psychiatry, 116, 129-144.
- Snedecor, G. W. and Cochran, W. G. (1967). Statistical Methods (6th ed.). Ames, Iowa: Iowa State University Press.
- Sundberg, N. D. (1985). Review of the Western Psychological Services computer-based test report (Lachar) on the Minnesota Multiphasic Personality Inventory. In: The Ninth Mental Measurements Yearbook: Vol. II, J. V. Mitchell, Jr., Ed. Lincoln, Nebraska: Buros Institute of Mental Measurements.
- United States Department of Commerce. Bureau of the Census. (1980). Statistical Abstract of the United States (101st ed.). Washington, D.C.: U.S. Government Printing Office.
- Weisberger, C. A. (1965). Comparison of normalized and linear T-scores in the MMPI. Journal of Clinical Psychology, 21, 412-415.
- Welsh, G. S. (1948). An extension of Hathaway's MMPI profile coding system. Journal of Consulting Psychology, 12, 343-344.
- Welsh, G. S. and Dahlstrom, W. G., Eds. (1956). Basic Readings on the MMPI in Psychology and Medicine. Minneapolis: University of Minnesota Press.
- Wiggins, J. S. (1973). Personality and Prediction: Principles of Personality Assessment. Reading, Massachusetts: Addison-Wesley.

APPENDICES

APPENDIX A

TABLE: MEAN ACCURACY RATINGS

TABLE IX
MEAN ACCURACY RATINGS OF PARAGRAPHS
IN THE LACHAR SYSTEM

Paragraph Number	Original Norms			Colligan General Norms			Colligan Age-by-Sex Norms		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
3	2	2.50	0.71	5	2.20	0.45	5	2.20	0.45
3A	1	2.00	--	4	2.00	0.82	4	2.00	0.82
4	1	3.00	--	2	3.00	0.00	2	3.00	0.00
5	1	2.00	--	1	2.00	--	1	2.00	--
5+5A	2	2.50	0.71	1	3.00	--	1	3.00	--
6	3	2.00	0.00	2	2.00	0.00	2	2.00	0.00
7	4	4.25	1.71	7	3.14	1.68	5	3.60	2.07
8	11	2.82	1.33	13	2.46	0.88	14	2.64	1.28
8+8A	1	5.00	--	6	3.17	1.47	2	3.50	2.12
9	8	2.25	1.83	0	--	--	4	2.25	2.50
9+9A	4	2.75	1.71	0	--	--	4	2.75	1.71
10+10A	1	6.00	--	0	--	--	0	--	--
12	4	2.75	0.96	4	2.75	0.96	4	2.75	0.96
14	31	2.32	1.11	31	2.23	1.02	27	2.26	1.06
16	0	--	--	0	--	--	0	--	--
17	0	--	--	0	--	--	0	--	--
19	0	--	--	0	--	--	0	--	--
22	0	--	--	0	--	--	0	--	--
23	0	--	--	0	--	--	0	--	--
24	0	--	--	1	4.00	--	1	4.00	--
24A	6	4.00	1.90	13	2.69	1.32	13	3.15	1.68
26B	9	3.78	1.48	1	4.00	--	0	--	--
27+27-1	1	3.00	--	5	3.40	1.67	5	2.80	1.92
28	0	--	--	1	1.00	--	1	1.00	--
29+29-1	3	3.00	2.00	7	2.71	0.95	7	2.43	1.34

TABLE IX (Continued)

Paragraph Number	Original Norms			Colligan General Norms			Colligan Age-by-Sex Norms		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
30	0	--	--	0	--	--	0	--	--
31	0	--	--	0	--	--	1	1.00	--
32	0	--	--	1	4.00	--	0	--	--
33	2	4.50	2.12	0	--	--	2	4.50	2.12
34	0	--	--	2	2.00	1.41	1	3.00	--
34A	6	2.50	1.64	0	--	--	1	1.00	--
35	1	1.00	--	0	--	--	1	1.00	--
37+37-1	5	3.00	2.00	2	3.50	3.54	1	6.00	--
38+38-1	5	2.60	1.95	5	2.60	1.95	6	3.33	2.07
38A+38A-1	4	3.50	2.38	4	3.50	2.38	4	3.75	2.22
40	1	3.00	--	1	3.00	--	0	--	--
42+42-1	0	--	--	1	2.00	--	3	1.67	0.58
44B	1	2.00	--	2	2.50	0.71	0	--	--
45+45-1	10	2.50	1.08	4	2.75	1.71	5	2.80	1.30
46+46-1	0	--	--	2	1.5	0.71	0	--	--
47	7	2.71	1.98	5	1.40	0.55	3	1.33	0.58
48+48-1	8	2.63	1.41	2	2.00	1.41	2	2.50	0.71
50	2	2.00	0.00	0	--	--	0	--	--
51	3	1.67	0.58	2	2.00	0.00	2	2.00	0.00
51A	2	1.50	0.71	0	--	--	0	--	--
51B	1	1.00	--	2	4.00	2.83	2	4.50	2.12
53	0	--	--	1	4.00	--	2	1.50	0.71
54+54-1	0	--	--	0	--	--	1	1.00	--
55	1	2.00	--	0	--	--	0	--	--
56B+56B-1	0	--	--	2	2.00	1.41	4	1.50	1.00
57+57-1	2	2.00	0.00	2	2.00	0.00	1	2.00	--
58+58-1	6	2.67	0.52	3	2.67	0.58	3	2.67	0.58
59	7	1.57	0.53	1	1.00	--	1	2.00	--

TABLE IX (Continued)

Paragraph Number	Original Norms			Colligan General Norms			Colligan Age-by-Sex Norms		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
60+60-1	0	--	--	0	--	--	1	2.00	--
61+61-1	4	3.25	0.96	7	2.71	1.11	5	2.80	1.30
61A	1	2.00	--	1	2.00	--	1	2.00	--
61B	4	2.50	0.58	6	2.50	1.05	5	2.20	0.84
61C	4	3.75	2.06	6	3.17	1.94	5	3.20	2.17
63	5	2.00	1.00	3	1.67	1.15	6	2.00	1.10
63A	0	--	--	0	--	--	1	1.00	--
63B+63B-1	4	2.25	0.96	0	--	--	0	--	--
63C+63C-1	1	3.00	--	3	2.00	1.00	5	2.20	0.84
65+65-1	3	4.00	1.73	2	3.50	2.12	1	5.00	--
66	9	2.33	1.41	6	3.33	1.03	7	3.43	1.40
67	10	2.80	1.40	16	3.31	1.45	15	3.40	1.24
68	9	2.56	1.81	9	2.33	1.22	13	2.15	1.14
69	6	2.83	1.33	1	1.00	--	3	1.00	0.00
70	7	3.43	1.51	3	2.33	0.58	7	2.57	1.13
71	7	2.43	1.27	12	2.25	0.97	6	2.67	1.21
72	2	2.00	0.00	0	--	--	3	2.67	0.58
73	12	2.58	1.24	20	2.65	1.63	22	2.95	1.76
74	6	2.67	1.97	6	1.83	0.75	7	2.14	1.77
75	7	1.71	0.49	0	--	--	1	4.00	--
76	6	3.17	1.33	6	3.17	1.33	8	2.63	1.41
77	9	2.11	1.36	12	2.17	0.94	10	2.20	0.92
78	14	2.79	1.25	1	6.00	--	2	5.00	1.41
79	21	2.81	1.47	19	2.68	1.57	18	2.89	1.53
80	4	2.25	1.50	11	2.00	0.89	11	2.00	0.77
81	15	2.13	1.55	2	2.00	1.41	1	1.00	--
82	15	3.40	1.40	12	3.42	1.44	11	3.18	1.33

TABLE IX (Continued)

	Original Norms			Colligan General Norms			Colligan Age-by-Sex Norms		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Paragraph Number									
83	12	3.17	1.64	7	3.29	1.11	11	3.27	1.42
85	2	4.50	2.12	2	4.50	2.12	2	2.50	0.71
86	2	2.00	0.00	7	3.14	1.77	4	3.50	1.73
87	1	2.00	--	0	--	--	0	--	--
89	1	2.00	--	1	2.00	--	1	2.00	--
90	0	--	--	1	1.00	--	2	2.00	1.41
91	0	--	--	3	2.67	2.08	4	2.00	0.82
92	31	3.03	1.52	23	2.87	1.71	24	2.96	1.73
93	25	3.36	1.47	5	3.00	1.41	5	3.00	1.41
94	0	--	--	0	--	--	0	--	--
95	13	3.31	1.32	17	3.00	1.22	15	3.00	1.25
96	17	2.18	1.19	10	1.60	0.84	10	1.40	0.70
97	4	2.75	2.06	3	1.67	0.58	4	2.00	0.82

APPENDIX B

TABLE: ONE-WAY REPEATED MEASURES ANOVA

TABLE X
ONE-WAY REPEATED MEASURES ANOVA
(N=78)

Source	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
Norms	2	0.9924	3.76	<0.025
Ss	77	142.1258		
Ss x Norms	154	20.3282		

APPENDIX C

TABLE: PAIRWISE COMPARISONS BETWEEN MEANS OF
ORIGINAL, COLLIGAN GENERAL, AND COLLIGAN
AGE-BY-SEX NORMS

TABLE XI
 PAIRWISE COMPARISONS BETWEEN MEANS OF
 ORIGINAL, COLLIGAN GENERAL, AND
 COLLIGAN AGE-BY-SEX NORMS,
 USING TUKEY'S HSD
 PROCEDURE

	Original Norms	Colligan General Norms	Colligan Age-by-Sex Norms
Original Norms	--	0.13	0.15*
Colligan General Norms	--	--	-.03

*p < 0.05

APPENDIX D

TABLE: 3 X 4 ANOVA WITH REPEATED MEASURES

TABLE XII
 3 X 4 ANOVA WITH REPEATED MEASURES, USING
 NORMS AND REPORT SEGMENTS AS
 FACTORS
 (N=29)

Source	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
Norms	2	.3890	.45	ns
Ss x Norms	56	24.3100		
Segment	3	4.0233	.34	ns
Ss x Segment	84	331.1387		
Norms x Segment	6	1.0586	.52	ns
Ss x Norms x Segment	168	56.7238		

APPENDIX E

HUBERTY'S PROCEDURES FOR ASSESSING IMPROVEMENT

Huberty's Procedures for Assessing Improvement in
Classification Accuracy Over Chance

Huberty (1984) offered a procedure for assessing whether a particular scheme for classifying individuals into groups improves the accuracy of the classification above that which could be expected by chance. This involves first determining the chance hit rate, then using a standardized test statistic to test for statistical significance. Huberty also presented a reduction in remaining error index, \underline{I} .

Consider \underline{k} populations, such as those made up of persons whose symptoms fall into generally neurotic, psychotic, or characterological patterns. If these populations are of approximately equal size, and if equal samples are drawn by chance from each, we can reasonably assume that $1/\underline{k}$ of the individual will be correctly classified by chance, for the following reasons. If predicted and actual group membership are independent, then the expected frequency in cell $(\underline{g}, \underline{g})$ of a classification table is $\underline{e}_g = \underline{q}_g \underline{n}_g$, where $\underline{q}_g = \underline{n}/\underline{N}$. If $\underline{n}_1 = \underline{n}_2 = \dots = \underline{n}_k = \underline{n}$ and $\underline{q}_g = \underline{n}/\underline{N} = 1/\underline{k}$, then the total chance frequency of hits would be

$$\underline{e} = \sum_{g=1}^{\underline{k}} \underline{e}_g = \sum_{g=1}^{\underline{k}} \underline{q}_g \underline{n}_g = \underline{n}.$$

Therefore, the chance hit rate would be $\underline{n}/\underline{N} = 1/\underline{k}$.

If, as seems possible in the case of pathology, the \underline{k} populations are of different sizes and the sample sizes reflect the differences, it is reasonable to assume that the proportions $\underline{q}_g = \underline{n}_g/\underline{N}$ give good estimates of the probabilities of group membership. Then the chance frequency of hits for Group \underline{g} is

$$\underline{e}_g = \underline{q}_g \underline{n}_g = \underline{n}_g^2 / \underline{N}$$

The chance frequency of hits overall is

$$\underline{e} = \sum_{g=1}^k g_g n_g = 1/N \sum_{g=1}^k n_g^2$$

The standardized normal test statistic for group g , using the above \underline{e}_g is

$$\underline{z} = \frac{(n_{gg} - \underline{e}_g) \sqrt{n_g}}{\sqrt{\underline{e}_g (n_g - \underline{e}_g)}}$$

This statistic is evaluated by the standard table of probabilities for values of \underline{z} . This procedure assesses separate-group classification accuracy.

In some situations it is more appropriate to evaluate any improved prediction against the maximum chance criterion, $\underline{e}_{\max} = (n_g/N)$. Such situations occur when one group contains a large majority of the total cases, allowing the highest chance hit rate by predicting all cases to be in that group, and when the interest is simply in obtaining the highest possible hit rate, with little concern about detecting unusual cases. The \underline{e}_{\max} is substituted for \underline{e}_g in the above formula for \underline{z} .

Huberty also suggested a way of assessing reduction in remaining error obtained by using a classification rule, such as the grouping by code types offered by Lachar. One may use the following reduction in remaining error index:

$$\underline{I} = \frac{\underline{H}_0 - \underline{H}_e}{1 - \underline{H}_e}$$

where \underline{H}_0 is the observed hit rate and \underline{H}_e is the hit rate expected by chance, obtained by the formula for \underline{e}_g above. Using the classification rule, one obtains $100(\underline{I})$ fewer classification errors than would be expected by chance classification.

APPENDIX F

MMPI-PREDICTED CLASSIFICATION TABLES

TABLE XIII
MMPI-PREDICTED CLASSIFICATIONS WITH
ORIGINAL NORMS

Clinicians' Classifications	N	P	C	I	Row Totals
Neurotic	7	2	1	0	10
Psychotic	1	8	2	0	11
Characterological	12	12	28	1	53
Indeterminate	1	2	1	0	4
Column Totals	21	24	32	1	78

TABLE XIV
MMPI-PREDICTED CLASSIFICATIONS WITH
COLLIGAN-GENERAL NORMS

Clinicians' Classifications	N	P	C	I	O	Row Totals
Neurotic	7	1	0	0	2	10
Psychotic	1	7	2	0	1	11
Characterological	13	10	18	2	10	53
Indeterminate	0	1	1	1	1	4
Column Totals	21	19	21	3	14	78

Note: O was used to stand for Normal, as use of the Colligan General norms resulted in some subjects having no clinical scales above a T-score of 70.

TABLE XV
MMPI-PREDICTED CLASSIFICATIONS WITH
COLLIGAN AGE-BY-SEX NORMS

Clinicians' Classifications	N	P	C	I	O	Row Totals
Neurotic	6	3	0	1	0	10
Psychotic	1	8	2	0	0	11
Characterological	11	10	20	2	1	53
Indeterminate	1	1	1	0	1	4
Column Totals	19	24	23	3	11	78

Note: O was used to stand for Normal, as use of the Colligan Age-by-Sex norms resulted in some subjects having no clinical scales above a I-score of 70.

APPENDIX G

REDUCTION OF FALSE POSITIVES

TABLE XVI
 LOWERING OF FALSE POSITIVES WITH ORIGINAL NORMS

Clinician's Classification	N ^a	Hits	Number Predicted to be in Classification	% of Those Predicted Who Were Actually in Classification
Neurotic	10	7	21	33
Psychotic	11	8	24	33
Characterological	53	28	32	88
Indeterminate	4	0	1	0

^aN = 78

TABLE XVII
 PERCENTAGE LOWERING OF FALSE POSITIVES
 WITH COLLIGAN-GENERAL NORMS

Clinician's Classification	N ^a	Hits	Number Predicted to be in Classification	% of Those Predicted Who Were Actually in Classification
Neurotic	10	6	19	32
Psychotic	11	8	24	33
Characterological	53	20	23	87
Indeterminate	4	0	3	0
Normal ^b	0	0	11	0

^aN = 78

^bWhen the Colligan General Norms were used, some subjects had no clinical scales above a T-score of 70 and were therefore classified as Normal.

TABLE XVIII
 PERCENTAGE LOWERING OF FALSE POSITIVES
 WITH COLLIGAN AGE-BY-SEX NORMS

Clinician's Classification	N ^a	Hits	Number Predicted to be in Classification	% of Those Predicted Who Were Actually in Classification
Neurotic	10	7	21	33
Psychotic	11	7	19	37
Characterological	53	18	21	86
Indeterminate	4	1	3	33
Normal ^b	0	0	14	0

^aN = 78

^bWhen the Colligan Age-by-Sex Norms were used, some subjects had no clinical scales above a \bar{I} -score of 70 and were therefore classified as Normal.

APPENDIX H

SAMPLE PARAGRAPH AND SUBJECT DATA FORM

Individuals who obtain similar profiles are often seen as evidencing conflicts which center around impulse control and social conformity. These difficulties are likely chronic and may be seen as more of a problem by others than by the patient.

1	2	3	4	5	6	7
Completely Accurate						Completely Inaccurate

SUBJECT DATA FORM

1. Therapist's name:
2. Subject's code number:
3. Subject's age:
4. Subject's sex:
5. Subject's education:
(number of years completed)
6. Inpatient or outpatient
(please circle one)
7. DSM-III diagnosis:
(all five axes if possible)
8. In your judgment, is this client (please circle one):
 "Neurotic" "Psychotic" "Characterological" "Indeterminate"
 (Please try not to use the indeterminate category unless absolutely
 necessary!)
9. How many therapy sessions have you had with this client?
10. How well do you feel you know this client? (please circle one of
 these choices):

Very well	Well	Fair	Poorly	Very poorly
1	2	3	4	5

VITA

Joyce Lyons Hawley

Candidate for the Degree of

Doctor of Philosophy

Thesis: EFFECTS OF COLLIGAN, OSBORNE, SWENSON, AND OFFORD'S CONTEMPORARY MMPI NORMS ON LACHAR AUTOMATED INTERPRETATIONS

Major Field: Clinical Psychology

Biographical:

Personal Data: Born in Chicago, Illinois, March 6, 1942, the daughter of Dr. Paul F. and Ruama Coit Hawley.

Education: Graduated from Thomas A. Edison High School, Tulsa, Oklahoma, in May, 1960; attended University of California, Santa Barbara, 1960-63; received Bachelor of Arts degree in History from University of Tulsa in January, 1965; received Master of Arts degree in Psychology from University of Tulsa, in May, 1976; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May, 1991.

Professional Experience: Psychological Trainee, Hissom Memorial Center, Sand Springs, Oklahoma, 1972-74; Psychological Assistant (M.A. level), Lloyd E. Rader Children's Diagnostic and Evaluation Center, Sand Springs, Oklahoma, 1977-82; Psychological Associate, Stillwater Community Mental Health Center, Stillwater, Oklahoma, 1982-83; Psychological Associate, Intensive Treatment Center, Juvenile Bureau, Oklahoma Department of Human Services, Sand Springs, Oklahoma, 1983-84; Graduate Teaching Assistant, Oklahoma State University, 1984-85; Psychological Associate, Psychological Services Center, Stillwater, Oklahoma, 1984-86; Psychological Associate, Marriage and Family Therapy Clinic, Stillwater, Oklahoma, 1984-86; Psychological Associate, Tulsa Mental Health Council, Tulsa, Oklahoma, 1985-86; Internship in Clinical Psychology, Veteran's Administration Medical Center, Houston, Texas, 1986-87; Staff Psychologist, Student Counseling Services, University of Texas Health Sciences Center, Houston, Texas, 1987 to present.