A Study of the Validity of the Global Assessment Scale

Daniel J. Dekker
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations
Part of the Psychoanalysis and Psychotherapy Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/2428

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
A STUDY OF THE VALIDITY OF THE
GLOBAL ASSESSMENT SCALE

by

Daniel J. Dekker

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Counseling and Personnel

Western Michigan University
Kalamazoo, Michigan
December 1983

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A STUDY OF THE VALIDITY OF THE
GLOBAL ASSESSMENT SCALE

Daniel J. Dekker, Ed.D.
Western Michigan University, 1983

The study evaluated the concurrent validity of the Global Assessment Scale (GAS), a rating scale which is used to measure overall severity of psychiatric disturbance in program evaluation systems and in research studies. Validity was evaluated by comparing GAS ratings of 60 community mental health outpatients to scores on the MMPI.

The literature review indicated that the GAS has been used in over 90 studies. Thirty-one inter-rater reliability coefficients cited indicated an acceptable median reliability of .80, but studies relevant to the GAS's validity were sparse and marginally supportive. The GAS was utilized primarily to evaluate change due to psychotherapy or medication, or to compare severity of disturbance between experimental groups. Few studies originated from program evaluation settings, but a questionnaire sent to several state departments of mental health indicated that the GAS is used in statewide evaluation systems in five states.

GAS and MMPI data were collected at admission from adult outpatients at a representative community mental health center in a medium-sized city. GAS ratings made by therapists were correlated with four overall severity indexes from the MMPI, as well as with the individual MMPI scales. Correlations with the overall indexes ranged from .29 to .36, indicating minimal shared variance and questioning the GAS's validity as a measure of severity of disturbance. Correlations with the MMPI
clinical scales were also under .36 and indicated that GAS scores are more related to psychotic symptoms than to neurotic ones. Both correlation data and a factor analysis demonstrated that the GAS was most highly related to MMPI scales measuring defensiveness and guardedness. This finding questioned the appropriateness of completing GAS ratings after a 1-hour intake session, and suggested that ratings made later in treatment may be more valid.

It was concluded that the GAS is a questionable measure of overall severity of disturbance because of its low concurrent correlations. Mean GAS scores of patient and normal groups in the literature and in the present study, however, appeared to have accuracy, suggesting that the GAS may have validity for comparing group mean scores. Using GAS scores as cutoffs or criterion measures for individuals, however, is not supported by the study's data and is discouraged.
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Dekker, Daniel Jay

A STUDY OF THE VALIDITY OF THE GLOBAL ASSESSMENT SCALE

Western Michigan University

Copyright 1983 by Dekker, Daniel Jay

All Rights Reserved
ACKNOWLEDGEMENTS

I regret that it is not possible to acknowledge all the individuals without whose help this study could not have been completed. I would like to especially thank my dissertation committee: Dr. Robert Betz, Chairman, for his thoughtful guidance throughout the project and particularly for his involvement during a period when he was not teaching; Dr. Robert Hopkins for his helpful support and encouragement; and Dr. Frederick Gault for his perceptive insights. Dr. Robert Brashear provided appreciated consultation on data processing. I would also like to thank the staff of the Ingham Community Mental Health Center for their cooperation. Dr. Rom Kriauciunas and Dr. Judith Taylor were advocates in obtaining this cooperation and provided appreciated support. The data could not have been collected without the assistance of the referral specialists, receptionists, and secretaries. My thanks to the clients of the center who gave time to participate in the study. I would also like to acknowledge my parents for their support and encouragement during the dissertation period and years of graduate school. Finally, I would like to thank Sharon Beauchamp of Executive Format for her efficient typing of the manuscript.

Daniel J. Dekker
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td><strong>I. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Overview of Study</td>
<td>5</td>
</tr>
<tr>
<td><strong>II. REVIEW OF RELATED LITERATURE</strong></td>
<td>6</td>
</tr>
<tr>
<td>Development of the GAS</td>
<td>6</td>
</tr>
<tr>
<td>Reliability of the GAS</td>
<td>10</td>
</tr>
<tr>
<td>Validity of the GAS</td>
<td>18</td>
</tr>
<tr>
<td>Use of the GAS in State Mental Health Systems</td>
<td>45</td>
</tr>
<tr>
<td>Use of the GAS in Research Studies</td>
<td>52</td>
</tr>
<tr>
<td>MMPI Overall Severity Indexes</td>
<td>62</td>
</tr>
<tr>
<td><strong>III. METHOD</strong></td>
<td>65</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>65</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>66</td>
</tr>
<tr>
<td>Subjects</td>
<td>68</td>
</tr>
<tr>
<td>Procedure</td>
<td>72</td>
</tr>
<tr>
<td>Summary</td>
<td>74</td>
</tr>
<tr>
<td><strong>IV. RESULTS</strong></td>
<td>76</td>
</tr>
<tr>
<td>GAS Scores</td>
<td>76</td>
</tr>
<tr>
<td>MMPI Scores</td>
<td>79</td>
</tr>
<tr>
<td>Testing of Hypotheses</td>
<td>82</td>
</tr>
<tr>
<td>Additional Statistical Analyses</td>
<td>94</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. DISCUSSION</td>
<td>98</td>
</tr>
<tr>
<td>Summary of Study</td>
<td>98</td>
</tr>
<tr>
<td>Conclusions of Study</td>
<td>99</td>
</tr>
<tr>
<td>Issues in Interpretation of Results</td>
<td>104</td>
</tr>
<tr>
<td>Recommendations</td>
<td>109</td>
</tr>
<tr>
<td>Conclusion</td>
<td>110</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>112</td>
</tr>
<tr>
<td>A. GLOBAL ASSESSMENT SCALE</td>
<td>113</td>
</tr>
<tr>
<td>B. LETTER AND QUESTIONNAIRE TO STATE DEPARTMENTS OF MENTAL HEALTH</td>
<td>114</td>
</tr>
<tr>
<td>C. SUBJECT INFORMATION SHEET AND CONSENT FORM</td>
<td>117</td>
</tr>
<tr>
<td>D. MMPI PROFILE OF MEAN SCORES OF SAMPLE</td>
<td>119</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>120</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Inter-Rater Reliability Coefficients for GAS</td>
<td>12</td>
</tr>
<tr>
<td>2.</td>
<td>Concurrent Validity Coefficients for GAS</td>
<td>24</td>
</tr>
<tr>
<td>3.</td>
<td>Change in Validity Coefficients Due to Statistical Artifact as Compared to Actual Change</td>
<td>31</td>
</tr>
<tr>
<td>4.</td>
<td>GAS Means and Standard Deviations from the Literature</td>
<td>39</td>
</tr>
<tr>
<td>5.</td>
<td>GAS Scores for Ingham CMHC Outpatient Samples</td>
<td>78</td>
</tr>
<tr>
<td>6.</td>
<td>MMPI Means and Standard Deviations</td>
<td>81</td>
</tr>
<tr>
<td>7.</td>
<td>Correlations between GAS and MMPI Overall Severity Measures</td>
<td>83</td>
</tr>
<tr>
<td>8.</td>
<td>Correlations between GAS and MMPI Clinical Scales</td>
<td>85</td>
</tr>
<tr>
<td>9.</td>
<td>Correlations between GAS and Groups of MMPI Clinical Scales</td>
<td>91</td>
</tr>
<tr>
<td>10.</td>
<td>Correlations between GAS and Two MMPI Validity Scales</td>
<td>92</td>
</tr>
<tr>
<td>11.</td>
<td>Factor Analysis of GAS and 12 MMPI Scales</td>
<td>95</td>
</tr>
<tr>
<td>12.</td>
<td>Multiple Regression Analyses on GAS</td>
<td>97</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure

1. Experimental Design of Study ........................................ 67
2. Frequency Distribution of GAS Scores .............................. 76
3. Frequency Distribution of MMPI Total Raw Scores ................ 80
CHAPTER I

INTRODUCTION

"Measuring the outcome of health and human service programs is an idea whose time has come" (Green, Nguyen, & Attkisson, 1979, p. 137). Over the past 10 years, program evaluation has become a popular concept in the field of mental health. Mental health programs can no longer assume permanent funding and unquestioned acceptance of their effectiveness. Accountability has become a byword, as public administrators, legislators, mental health boards, and citizens have requested information about how public mental health funds are being spent.

As mental health program evaluation systems have been established, needs have arisen for measurement instruments which are quick and easy to administer, but which communicate significant information about clients. The most popular instruments have been rating scales that are completed by clinical staff. Two main types of scales are available—multidimensional scales, which consist of ratings in several symptom areas, and single dimensional scales, which provide a global rating of the overall disturbance of an individual. Global scales have been most widely used because of their ease of implementation and their ability to integrate several facets of a client's functioning into a single measure. Global scales have been found attractive because they communicate individuals' overall level of disturbance in a numerical index which can be used to describe the clients served by an agency or the amount of change associated with treatment. Because global scales rely
on subjective judgments of professionals, though, their use can present special problems in terms of reliability and validity.

The global scale which has been most frequently adopted in program evaluation systems in recent years is the Global Assessment Scale (GAS), a rating scale measuring overall severity of psychiatric disturbance. The GAS was published in 1976 by Endicott, Spitzer, Fleiss, and Cohen, and was introduced primarily as a tool for clinical research. It has received considerable use in psychiatric research studies, but its greatest use has been in public mental health evaluation systems.

The GAS provides a global rating of an individual's degree of disturbance along a broad continuum from positive mental health to severe mental illness. It is a 100-point scale that ranges from a high end of superior emotional functioning in all areas (91-100 range) to a low end of severe emotional disturbance characterized by danger to self or others and the need for constant supervision (1-10 range). The GAS consists of a one page description of the typical symptoms and behaviors of individuals who are rated in each of the ten 10-point intervals of the scale (see Appendix A). A person completes the GAS by assigning an individual a number from 1 to 100 which describes that individual's level of severity of disturbance.

Since its publication the GAS has received widespread usage. A questionnaire sent to several state departments of mental health as part of the present study indicated that as of Fall, 1983, five states are using the GAS in their statewide evaluation systems for approximately 700,000 public mental health clients per year. A review of the literature indicated that the GAS has been used in over 90 published research
studies and in six doctoral dissertations. With an estimated one and a half million GAS ratings made per year, the GAS is one of the most frequently used clinical instruments in the United States today.

Statement of the Problem

In spite of the extensive use of the GAS, few people have examined the scale's psychometric properties to determine whether the GAS has value as a rating instrument. Many researchers have used the GAS, but few have taken time to do research on the scale itself to determine its merits. To some degree, the value of the GAS has been accepted blindly. One of the most important questions to be asked when an instrument such as the GAS is introduced is whether it is valid; that is, does it in fact measure what it is supposed to be measuring? Does the GAS really measure severity of disturbance? This question of validity has not yet been adequately answered for the GAS. The small amount of validity data which are available provide questionable support for the scale. Further research is obviously needed. The present study will examine the question of whether the GAS is a valid measure of overall severity of psychiatric disturbance.

In the original publication of the GAS, Endicott et al. (1976) cite some validity evidence, but the validity coefficients reported are low and are derived from a sample of exclusively psychiatric inpatients. An unpublished study conducted by the Michigan Department of Mental Health (Herman, 1982a) is the only other study that has specifically examined the GAS's validity. This study also reported rather unimpressive validity coefficients and was conducted on a sample of psychiatric
inpatients and aftercare group home residents. No studies have examined the question of the GAS's validity in an outpatient, higher functioning population, which is the population for which the GAS is most commonly used. Neither have any previous validity studies used an objective psychological test as the criterion measure of severity of disturbance.

The purpose of the present study was to investigate the concurrent and construct validity of the GAS by comparing GAS ratings of a sample of psychiatric outpatients to a more established and more standardized measure of severity of psychiatric disturbance. It was believed that worthwhile information about the validity of the GAS could be obtained by determining whether GAS scores had a meaningful correlational relationship with an accepted, standard measure of psychopathology. As Anastasi (1976) notes, this is a commonly used method of evaluating validity. The instrument which was chosen as the criterion measure of psychiatric disturbance was the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1940), a comprehensive psychological test which is the most frequently used personality inventory. The MMPI is supported as a test instrument by 40 years of research and nearly 6000 research references (Buros, 1978).

The main research question which was asked in the study was whether the GAS, as a measure of overall severity of disturbance, was related to several measures of severity of disturbance taken from the MMPI. It was hypothesized that the GAS would be significantly related to the MMPI indexes of overall disturbance. A secondary research question was also dealt with in the study: namely, what separate components of psychopathology contribute most to the concept of severity of disturbance? This question
was investigated by examining relationships between GAS and the individual scales on the MMPI, as well as relationships between GAS and groups of MMPI scales. It was hypothesized that some groups of scales would show greater relationships to the GAS than others.

Overview of Study

This chapter has provided an introduction to the study by presenting some background on the GAS, stating the problem, and presenting the study's research questions. Chapter II provides additional background for the study by reviewing the published literature describing the development, reliability, validity, and usage of the GAS. Chapter III of the dissertation describes the design and methodology of the study, and Chapter IV presents the results of the statistical analysis of the data. The final chapter, Chapter V, is a discussion of the results, their implications, and some recommendations arising from the study.
CHAPTER II

REVIEW OF RELATED LITERATURE

Development of the GAS

The Global Assessment Scale was published by Endicott, Spitzer, Fleiss, and Cohen in 1976, but was written sometime earlier by Spitzer, Gibbon, and Endicott. Printed copies of the scale are available from the authors, along with an instruction sheet and material giving more detailed descriptions of the scale's ten intervals and examples of patients who are rated in each interval (Spitzer, Gibbon, and Endicott, 1978). Also available is a booklet of 23 case vignettes keyed to the authors' GAS rating on each case (Biometrics Research, 1978), which is intended to be used for training purposes and for evaluating inter-rater reliability.

The authors of the GAS are associated with the New York State Psychiatric Institute in New York City, which is the setting where the original reliability and validity studies on the GAS were conducted. Drs. Spitzer and Endicott are also on the faculty of the Department of Psychiatry, Columbia University College of Physicians and Surgeons. Dr. Robert Spitzer is best known for his role as Chairperson of the Task Force on Nomenclature and Statistics of the American Psychiatric Association, the committee that developed the Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition (DSM-III) (American Psychiatric Association, 1980). Dr. Jean Endicott also served on this APA task force, and Dr. Spitzer, Dr. Endicott, and Ms. Gibbon all served on...
advisory committees which assisted in the development of DSM-III. Drs. Spitzer and Endicott are also well known in psychiatric circles as the authors of the Research Diagnostic Criteria (RDC) and the Schedule for Affective Disorders and Schizophrenia (SADS), two widely used instruments in psychiatric research.

The Health-Sickness Rating Scale

The GAS evolved, according to the authors, from the Health-Sickness Rating Scale (HSRS), another 100-point rating scale for severity of psychiatric disturbance. The HSRS was developed at the Menninger Foundation and published by Luborsky in 1962. It differs from the GAS in that there are only eight anchor points and the anchor points are unequally distributed across the scale's range. Each anchor point on the HSRS is defined by a behavioral statement and also by case examples which include descriptions of appropriate diagnoses. The problem that Spitzer, Gibbon, and Endicott saw in the HSRS was that the examples of diagnoses sometimes conflicted with the behavioral definitions of a scale point, causing considerable difficulty in choosing an appropriate rating. For example, a schizophrenic patient in remission could be judged as functioning at a level above 50, yet the scale examples on the HSRS indicate that psychoses should be rated below 50. To solve this problem, Spitzer et al. wrote the GAS, which deliberately used only behavioral descriptions for the anchor points and was therefore intended to be free from any reference or ties to diagnostic categories. Other changes which were incorporated into the development of the GAS included increasing the scale anchors from eight to ten, creating equal
intervals between the anchors, expanding the anchor definitions and making them more specific, and using interval ranges instead of just single anchor points.

Because the HSRS was the forerunner of the GAS, it is worth looking briefly at the research which has evaluated it and the acceptance which the scale has received. Luborsky and Bachrach (1974) summarized 18 studies which used the HSRS. Of these studies, 10 examined the inter-rater reliability between independent judges. The reliability coefficients obtained ranged from .65 to .94 and according to Luborsky and Bachrach were "distributed toward the upper end of this range" (p. 295). Luborsky's data on the validity of the HSRS are not quite as impressive. Most of the validity coefficients compare HSRS ratings to ratings made by judges on other personality variables. Some of the reported coefficients are high: .81 with ego strength, .84 with severity of symptoms, and .72 with quality of interpersonal relationships. However, the two ratings in these correlations were made by the same judge using the same body of data, resulting in obvious criterion contamination and questionable conclusions. In cases where HSRS ratings were compared to an independent criterion, correlations were lower. When therapist's HSRS ratings were compared to independent ratings made by nurses or other therapists, some typical validity coefficients reported by Luborsky and Bachrach were .55 with the Jenkins Symptom Rating Scale and .66 with the Ann Arbor Psychotic Confusion Scale. When HSRS ratings were compared to independent inventories filled out by patients, the reported validity coefficients were quite low: .38 with the Social Assets Scale, and .32 with the Baron Ego-Strength Scale.
The HSRS's use in many research studies demonstrates the degree of acceptance which the scale has received. The HSRS was endorsed by the NIMH Psychotherapy Outcome Measures Project (Waskow and Parloff, 1975) as the best global scale for use in psychotherapy outcome research. Another older NIMH publication (Hargreaves and Attkisson, 1974) also recommended the HSRS for outcome studies in community mental health centers. Although the HSRS was used frequently in the 1970's, in the 1980's it appears to have been superceded by the newer and supposedly improved Global Assessment Scale.

The Global Assessment Scale for Children and the Global Assessment Scale for Developmentally Disabled

Since the GAS was introduced, two adaptations of it have been developed which are worth mentioning, the Global Assessment Scale for Children and the Global Assessment Scale for Developmentally Disabled. The Global Assessment Scale for Children (GAS-Children) was written by Sorrells, Rothman, and Heldman (1976) at the Alameda County (California) Child and Family Mental Health Center. They revised the GAS to take into account the developmental influences and stages peculiar to children under 12 years of age. The GAS-Children still has ten 10-point intervals, but the descriptive paragraphs for each range are different.

A committee of the Michigan Department of Mental Health (1981) has revised the GAS-Children into another children's GAS form for use in Michigan's public mental health programs. The Michigan Department of Mental Health (1981) has also developed a Global Assessment Scale for Developmentally Disabled (GAS-DD). The GAS-DD has a similar 100-point
structure and evaluates overall functioning with regard to two major factors: daily living and self-care skills, and presence or absence of maladaptive behavior.

Reliability of the GAS

An examination of the psychometric worth of a rating scale must begin with a look at the scale's reliability. Reliability indicates the consistency or stability with which the scale measures a variable. It reflects the degree to which scores are repeatable on different occasions or by different raters. A scale must be reliable in order to be valid, so the reliability of a scale must be examined before its validity can be evaluated.

Although there are different types of reliability, the kind which is most relevant to a rating scale such as the GAS is scorer reliability, or inter-rater reliability. This is simply the degree of correlation between the scores assigned to an individual by two or more raters. It provides a measure of the degree of agreement between different judges who rate a client at the same point in time. Because GAS ratings are to some degree subjective and hypothetically could vary according to the characteristics of each rater, it is important that inter-rater reliability be evaluated. Inter-rater reliability coefficients are usually expressed as product-moment correlations or intraclass correlation coefficients.
Inter-Rater Reliability Studies

In the review of literature on the GAS, 13 published studies and 4 unpublished studies were located which cited an inter-rater reliability coefficient. Several of the studies cite more than one coefficient, since inter-rater reliability was sometimes tested on different samples of raters. The samples cited vary according to the rater's professional level, the amount of training in use of the GAS which the raters had received, and the source of the information which was being rated. Some of the studies compare agreement between 2 raters, while others compare the degree of agreement between up to 38 different raters. The studies which cite inter-rater reliability coefficients for the GAS are listed in Table 1.

An analysis of Table 1 indicates that the 31 inter-rater reliability coefficients range from a low of .33 to a high of .98, with a median inter-rater reliability of .80. A coefficient of .70 is usually considered to be the minimum necessary for an acceptable degree of reliability. Therefore, the evidence cited in Table 1, derived from a variety of different patients and raters, makes it possible to conclude that the GAS does have adequate inter-rater reliability.

Factors Influencing Reliability

An examination of the data in Table 1 from a more qualitative point of view makes it possible to make some generalizations as to the factors that influence inter-rater reliability. The first factor is the level of professional training and experience of the raters. The
<table>
<thead>
<tr>
<th>Reference</th>
<th>Reliability Coefficient</th>
<th>Number of Raters Compared</th>
<th>Source of Information Rated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carpenter, Sadler, Light, Hanlon, &amp; Kurland (1983)</td>
<td>.94</td>
<td>2</td>
<td>Inpatients rated by psychiatrist &amp; psychologist</td>
</tr>
<tr>
<td>Curran, Miller, Monti, Zwick, &amp; Stout (1980)</td>
<td>.80 to .90</td>
<td>2</td>
<td>Inpatients &amp; day patients rated by therapists</td>
</tr>
<tr>
<td>Endicott, Cohen, Nee, Fleiss, &amp; Herz (1979)</td>
<td>.37</td>
<td>2</td>
<td>Inpatients rated by therapist &amp; researcher</td>
</tr>
<tr>
<td>Endicott, Spitzer, Fleiss, &amp; Cohen (1976)</td>
<td>.76</td>
<td>2</td>
<td>Inpatients rated by researchers</td>
</tr>
<tr>
<td></td>
<td>.69</td>
<td>2</td>
<td>Case records of inpatients rated by research assistants</td>
</tr>
<tr>
<td></td>
<td>.91</td>
<td>2</td>
<td>Aftercare patients rated by research assistants</td>
</tr>
<tr>
<td></td>
<td>.61</td>
<td>4</td>
<td>Transcripts of interviews with normals rated by therapists</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>15</td>
<td>Case vignettes rated by a variety of persons</td>
</tr>
<tr>
<td>Filstead, Shadish, Crandell, &amp; Altman (1982)</td>
<td>.71</td>
<td>(for one untrained rater)</td>
<td>Videotapes of interviews rated by CMH clinical staff</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Reliability Coefficient</th>
<th>Number of Raters Compared</th>
<th>Source of Information Rated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fink, Braden, &amp; Qualls (1982)</td>
<td>.76 (for one trained rater)</td>
<td>17</td>
<td>Videotapes of interviews rated by CMH clinical staff</td>
</tr>
<tr>
<td>Fink &amp; Heckerman (1981)</td>
<td>.83 (for two untrained raters)</td>
<td>38</td>
<td>Videotapes of interviews rated by CMH clinical staff</td>
</tr>
<tr>
<td>Gardos, Cole, &amp; LaBrie (1982)</td>
<td>.83</td>
<td>2</td>
<td>Inpatients rated by research assistants</td>
</tr>
<tr>
<td>Goldbart (1978)</td>
<td>.82</td>
<td>2</td>
<td>Inpatients rated by psychiatrist &amp; research assistant</td>
</tr>
<tr>
<td>Green, Nguyen, &amp; Attkisson (1979)</td>
<td>.87</td>
<td>3</td>
<td>Aftercare patients rated by nurse &amp; psychiatrist</td>
</tr>
<tr>
<td>Goldbart (1978)</td>
<td>.41 (for one rater)</td>
<td>48</td>
<td>Case vignettes rated by untrained CMH clinical staff</td>
</tr>
<tr>
<td>Gully &amp; Harris (1982)</td>
<td>.65 (for two raters)</td>
<td>48</td>
<td>Case vignettes rated by untrained CMH clinical staff</td>
</tr>
<tr>
<td>Hargreaves &amp; Bottino (1977)</td>
<td>.88</td>
<td>2</td>
<td>Outpatients rated by CMH clinical staff</td>
</tr>
<tr>
<td></td>
<td>.61</td>
<td>2</td>
<td>Inpatients &amp; outpatients rated by untrained clinical staff</td>
</tr>
<tr>
<td>Reference</td>
<td>Reliability Coefficient</td>
<td>Number of Raters Compared</td>
<td>Source of Information Rated</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------</td>
<td>---------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>Herman (1981)</td>
<td>.33</td>
<td>4</td>
<td>Inpatients rated by state psychiatric hospital ward staff</td>
</tr>
<tr>
<td></td>
<td>.83</td>
<td>2</td>
<td>Outpatients rated by CMH clinical staff</td>
</tr>
<tr>
<td></td>
<td>.88</td>
<td>2</td>
<td>Outpatients rated by CMH clinical staff</td>
</tr>
<tr>
<td>Hurt, Friedman, Clarkin, Corn, &amp; Aronoff (1982)</td>
<td>.90+</td>
<td>4</td>
<td>Inpatients rated by therapists</td>
</tr>
<tr>
<td>Owen (1983)</td>
<td>.63 to .84</td>
<td></td>
<td>Case vignettes rated by 27 groups of CMH clinical staff</td>
</tr>
<tr>
<td>Shenoy, Sadler, Goldberg, Hamer, &amp; Ross (1981)</td>
<td>.89 to .98</td>
<td>2</td>
<td>Aftercare patients rated by therapists</td>
</tr>
<tr>
<td>Sorenson, Hargreaves, &amp; Friedlander (1982)</td>
<td>.60</td>
<td>2</td>
<td>Inpatients rated by untrained clinical staff</td>
</tr>
<tr>
<td></td>
<td>.70</td>
<td>2</td>
<td>Outpatients rated by untrained clinical staff</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
lowest reliability coefficient in Table 1, .33 (Herman, 1981), was obtained by ward staff at a state psychiatric hospital. The second highest coefficient, .94 (Carpenter, Sadler, Light, Hanlon, & Kurland, 1983), was obtained by psychiatrists and PhD psychologists. Many of the coefficients in the middle range were taken from rater samples of community mental health clinical staff, which it is assumed are mostly master's level therapists. This relationship between amount of clinical experience of raters and the magnitude of reliability of global scales is also pointed out by Luborsky and Bachrach (1974) and Newman (1980).

Another factor that influences the degree of reliability of the GAS is the amount of training that raters have had in the use of the scale. In some of the studies in Table 1, untrained raters were deliberately used so as "to simulate routine use of the instrument on a broad scale where inservice training would be uneven or lacking" (Sorensen, Hargreaves, & Friedlander, 1982, p. 341). Five out of the six reliability coefficients from studies which used untrained raters are below the median coefficient of .80. Further support for the significance of training is seen in the two studies which made direct comparisons between the amount of training in use of the GAS given to the raters. Filstead, Shadish, Crandell, and Altman (1982) obtained an inter-rater reliability coefficient of .71 for an untrained rater, as compared to a coefficient of .76 for a trained rater. Owen (1983) compared the reliability coefficients obtained in rating a series of standard case vignettes between eight community mental health programs which received a full GAS training session of 2 1/2 hours and eight community mental health programs which received an abbreviated version....
of the same training. The full training included a discussion of strategies for minimizing error variance, a discussion of problem situations, and practice in rating written, video, and movie case examples. The programs receiving the full training had an average inter-rater reliability coefficient of .75, while the programs receiving the abbreviated training had an average reliability coefficient of .69.

A third factor influencing the size of the reliability coefficient, which is apparent from Table 1, is that reliability increases when more than one rater is used to obtain a client's GAS score. Green, Nguyen, and Attkisson (1979) used generalizability theory to statistically obtain estimated reliabilities in a community mental health center. When the estimate of reliability for one rater was unacceptably low (.41), they calculated an estimate of reliability for having two raters rate each client and using the average rating as the client's score. By reducing the residual variance by one-half in this manner, they raised the reliability to .65 for two raters. Likewise, Filstead et al. (1982) raised their interrater reliability from .71 to .83 by substituting GAS scores made by one rater with GAS scores based on the average score of two raters. This procedure of increasing reliability through using two raters may be applicable in inpatient settings where several professionals have contact with a client, but it does not seem to have much practical value in outpatient settings where clients routinely see only one therapist.

Another significant factor in the magnitude of reliability coefficients appears to be the setting in which the study was conducted. Nearly all of the studies in Table 1 which were conducted in research-oriented, university teaching hospitals obtained inter-rater reliabilities above
the median coefficient of .80. While some of the studies conducted in state hospitals and community mental health clinics also obtained reliabilities above the median, there were more from this group below the median than from the former group. This difference is likely due to differences in total milieu between the facilities, and it points out a difficulty in generalizing from research-oriented settings to some community ones.

Endicott et al. (1976) propose an additional factor influencing the GAS's reliability by stating that inter-rater reliability of the GAS increases as a function of increased heterogeneity in the severity level of the subject sample. This relationship with heterogeneity of the sample's GAS scores was not necessarily found to be true, however, in other reliability studies. The results obtained by Endicott et al. (1976) are more likely due to a statistical artifact of the correlation coefficient rather than to any real differences in inter-rater agreement between heterogeneous and homogeneous populations. As discussed by McNemar (1969), the magnitude of the correlation coefficient always increases as the degree of heterogeneity, or variance, of the variables being correlated increases. When an increase in correlation coefficients is seen between one sample which is homogeneous and another sample which is more heterogeneous, this does not necessarily mean that there is a stronger relationship between the two variables being correlated. It may only mean that a statistically larger coefficient is obtained due to an artifact of the correlation coefficient.

Other researchers have found additional factors which influence the magnitude of inter-rater reliability of global rating scales. These
factors have not been tested on the GAS, but their significance could well be generalized to the GAS. Luborsky and Bachrach (1974) found that when the amount and comprehensiveness of a rater's knowledge about a client was greater, the reliability of the HSRS was increased. Newman (1980) notes that inter-rater reliability for a global scale increased greatly when it was used routinely in case management and review processes in an agency. Newman also cites an example of reliability increasing significantly when a local program had an opportunity to adopt its own version of a global scale.

**Summary of Reliability**

In summary, it can be said that the GAS has been shown to have an adequate degree of reliability, and that its reliability is enhanced by greater experience, training, and knowledge of the raters. Reliability, however, is only a first step in demonstrating the value of a psychometric instrument. Reliability indicates that a scale measures consistently, but it does not say what the scale actually measures. A reliable scale could be measuring a construct different from that which it is supposed to be measuring. In order to evaluate what the GAS measures, its validity must be examined.

**Validity of the GAS**

The crucial issue in the determination of the worth of a rating scale is an evaluation of its validity. Validity involves determining what the scale actually measures. It involves finding out what characteristic or trait is being measured by scores on the scale. A scale
which is valid will measure what the scale authors or the title of the scale purports it to measure. A scale which is not valid will be measuring some extraneous variable or just error variance. By evaluating the validity of a scale, it is possible to determine whether scores on the scale actually mean what they are supposed to mean, and whether the scale is therefore useful for the purpose for which it is being used.

As was noted in Chapter I, the GAS has received quite extensive usage in recent years, but strong support for its validity has not yet been demonstrated. Researchers have been quick to use the scale, but slow to evaluate its worth. The fact that the GAS's use appears to have gotten ahead of a comprehensive evaluation of its validity is an interesting comment on the social psychology of research instruments. Apparently, some researchers see an instrument being used by others or cited in published literature and assume on the basis of this that it is a worthwhile and valid instrument. Or perhaps the widespread usage of the GAS without strong validity support is just an indication of the current state of the art of global rating scales: in spite of a lack of support for the GAS, there is not any better instrument available and so the GAS is the best instrument to use. Another factor which could account for this phenomenon regarding the GAS's use is that in many cases state departments of mental health, the largest users of the GAS, do not have the necessary financial and human resources to conduct basic research such as validity studies on the psychometric instruments they use.

The literature on the GAS contains various kinds of data which can be used to evaluate its validity. These data will be reviewed in categories,
according to the three different types of validity specified in the APA Standards for Educational and Psychological Tests (1974). Content validity will first be analyzed, then criterion-related validity, and construct validity. It should be remembered, though, that "construct validity is a comprehensive concept which includes the other types" (Anastasi, 1976, p. 159). All validity of a test or scale relates to demonstrating its construct validity, which is the extent that the instrument measures the theoretical construct or trait which it is supposed to measure. Since the GAS purports to measure overall severity of psychiatric disturbance, an evaluation of its validity involves determining in various ways to what extent it measures this construct.

**Content Validity**

Content validity is the simplest type, involving systematically examining the scale's content to determine whether it covers a representative sample of the behavior domain to be measured. In looking at the GAS, it can be seen that the scale includes descriptions of behaviors covering the broad continuum from mental health to mental illness. As Ciarlo, Edwards, Kiresuk, Newman, and Brown (1981) point out, though, the scale's descriptions are less well-elaborated in the upper portion of the scale than in the lower part. The upper half of the scale seems overly general and appears to leave out some common minimally impairing symptoms seen in outpatient clients. The lower half of the GAS cites 23 specific behavioral examples to illustrate the five categories from 1-50, while the top half of the scale cites only 7 behavioral examples in the 50-100 range. It would appear that the GAS's content validity...
could be improved by adding additional specific examples of behaviors and symptoms which are included in the 50-100 range. It seems that the content validity of the GAS is good for serious psychiatric impairments, but poor for moderate and mild impairments.

**Criterion-Related Validity**

Criterion-related validity involves comparing scores on a scale to an independent measure of the same construct which the scale being evaluated measures. This independent measure is called the criterion. When the criterion measure and the scale scores are obtained at the same point in time, it is concurrent validity which is being evaluated. In validity studies of rating scales such as the GAS, concurrent validity is the type that is most frequently examined and that is most important. By comparing the relationship of scores on an untested rating scale either to scores obtained at the same time on an instrument which is known to measure a certain variable, or to known contrasted groups, important information about the new scale can be obtained. As Cronbach and Meehl (1966) point out, "Studies relating a new test instrument to more established approaches serve to expand the nomological network which describes the construct validity of the new instrument" (p. 291).

Because some of the validity data on the GAS compare the relationship between GAS scores and criteria with questionable independence from GAS ratings, a few words should be said about choice of an independent criterion and avoiding criterion contamination. In order to accurately validate an instrument, the criterion used must be as independent as possible. If scores on the criterion are influenced by scores
on the original instrument, then contamination occurs and the correlation between the instrument and the criterion may be spuriously high. If, for example, a judge is making ratings which will be used to validate a test and he has prior knowledge of the test's scores, his ratings are contaminated and may be inaccurate. Examples of criterion contamination were seen in validity studies of the HSRS, where HSRS ratings were correlated with other ratings made at the same time by the same judge using the same body of information.

Concurrent Validity Studies

The literature on the GAS includes only two studies which specifically examine the concurrent validity of the scale. A few additional studies cite a concurrent correlation between the GAS and another psychometric instrument.

The Original Validation Study of the GAS. The major treatment of the GAS's concurrent validity is given in the original publication of the scale (Endicott et al., 1976). Validity coefficients are cited from a sample of inpatients \(n = 58-107\), in which GAS scores are compared to scores on three other instruments, the Family Evaluation Form (FEF), the Mental Status Examination Record (MSER), and the Psychiatric Status Schedule (PSS). Concurrent correlations between GAS and these other instruments are reported at the time of psychiatric admission and at the time of six months after admission, at which point most of the subjects were no longer patients.
The concurrent validity coefficients obtained by Endicott et al. (1976) are listed in Table 2. Since more severe psychopathology is indicated by low scores on the GAS and high scores on the other instruments, expected correlations are in the negative direction. The subjects in this study were all psychiatric inpatients, most of whom had a diagnosis of schizophrenia. The mean GAS score at admission was 31, with a standard deviation of approximately 10, indicating that the sample was low functioning. The criteria used were all independent, since the correlations are between ratings given by a therapist and a research assistant, by a staff person and a family member, or by two different research assistants.

The FEF criterion is a total score of patient psychopathology that was reported by a family member of each patient who was interviewed with the Family Evaluation Form (Spitzer, Gibbon, & Endicott, 1971). The MSER scores used as criteria are taken from the Mental Status Examination Record (Endicott, Spitzer, & Fleiss, 1976), an objectively scored record of a mental status examination. The MSER Severity scale is merely a 7-point rating scale for overall severity of illness, ranging from "not ill at all" at one end to "among the most extremely ill" at the other end. A possible methodological problem is seen in the use of this MSER Severity scale as a criterion measure. Since the MSER Severity scale and the GAS are both global rating scales ranging from health to illness, they are therefore very similar in nature, and a correlation between them hardly indicates a comparison between two different measures of the same construct. A case could be made that the correlation between GAS and the MSER Severity scale, the highest validity coefficient reported.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Validity Coefficient</th>
<th>Criterion Used</th>
<th>Subjects</th>
<th>Independent Criterion?</th>
<th>Time of Both Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endicott, Spitzer, Fleiss, &amp; Cohen (1976)</td>
<td>-.19</td>
<td>FEF</td>
<td>Inpatients</td>
<td>Yes</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>-.25</td>
<td>FEF</td>
<td>Inpatients</td>
<td>Yes</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>-.44</td>
<td>MSER Severity Scale</td>
<td>Inpatients</td>
<td>Yes</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>-.11 mean</td>
<td>MSER (other scales)</td>
<td>Inpatients</td>
<td>Yes</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>-.22 mean</td>
<td>PSS</td>
<td>Inpatients</td>
<td>Yes</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>-.45</td>
<td>FEF</td>
<td>Ex-patients</td>
<td>Yes</td>
<td>6 months after admission</td>
</tr>
<tr>
<td></td>
<td>-.52</td>
<td>FEF</td>
<td>Ex-patients</td>
<td>Yes</td>
<td>6 months after admission</td>
</tr>
<tr>
<td></td>
<td>-.62</td>
<td>MSER Severity Scale</td>
<td>Ex-patients</td>
<td>Yes</td>
<td>6 months after admission</td>
</tr>
<tr>
<td></td>
<td>-.27 mean</td>
<td>MSER (other scales)</td>
<td>Ex-patients</td>
<td>Yes</td>
<td>6 months after admission</td>
</tr>
<tr>
<td></td>
<td>-.37 mean</td>
<td>PSS</td>
<td>Ex-patients</td>
<td>Yes</td>
<td>6 months after admission</td>
</tr>
</tbody>
</table>
Table 2 (Continued)
Concurrent Validity Coefficients for GAS

<table>
<thead>
<tr>
<th>Reference</th>
<th>Validity Coefficient</th>
<th>Criterion Used</th>
<th>Subjects</th>
<th>Independent Criterion?</th>
<th>Time of Both Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herman (1982a)</td>
<td>-.31 mean</td>
<td>Self-Care Skills</td>
<td>Inpatients &amp; group home residents</td>
<td>No</td>
<td>During long-term treatment</td>
</tr>
<tr>
<td>Hurt, Friedman, Clarkin, &amp; Aronoff (1982)</td>
<td>-.63 to -.73 mean</td>
<td>Maladaptive Behavior Measures</td>
<td>Inpatients &amp; group home residents</td>
<td>No</td>
<td>During long-term treatment</td>
</tr>
<tr>
<td>Hurt, Friedman, Clarkin, &amp; Aronoff (1982)</td>
<td>-.49 mean</td>
<td>Community Living Skills</td>
<td>Inpatients &amp; group home residents</td>
<td>No</td>
<td>During long-term treatment</td>
</tr>
<tr>
<td>Hurt, Friedman, Clarkin, &amp; Aronoff (1982)</td>
<td>-.49 mean</td>
<td>Community Living Skills</td>
<td>Inpatients &amp; group home residents</td>
<td>No</td>
<td>During long-term treatment</td>
</tr>
<tr>
<td>Hurt, Friedman, Clarkin, &amp; Aronoff (1982)</td>
<td>-.63 to -.73 mean</td>
<td>Maladaptive Behavior Measures</td>
<td>Inpatients &amp; group home residents</td>
<td>No</td>
<td>During long-term treatment</td>
</tr>
<tr>
<td>Hurt, Friedman, Clarkin, &amp; Aronoff (1982)</td>
<td>-.63 mean</td>
<td>Hamilton Depression Rating Scale</td>
<td>Inpatients</td>
<td>Not clear</td>
<td>Within 1 month after admission</td>
</tr>
</tbody>
</table>

Note. More severe psychopathology is indicated by low scores on the GAS and high scores on the other instruments, so expected correlations are in the negative direction.

At each time period, is more similar to an inter-rater reliability coefficient than to a validity coefficient.

The other scales of the MSER cited in Table 2 were derived through factor analysis and each represent a separate psychiatric symptom cluster. The mean validity coefficient for the 20 other MSER scales at the admission rating was -.11. The highest correlation was -.37 with the...
Cognitive Disorganization and Unusual Thoughts-Delusions factors, and the lowest correlation was .00 with Sleep-Appetite Disturbance.

The third instrument which was used as a criterion in the original validation study was the Psychiatric Status Schedule (Spitzer, Endicott, Fleiss, & Cohen, 1970), a structured interview schedule. The correlations between the five scales on the PSS and the GAS at the time of admission averaged -.22. The highest coefficient was -.47 with the Reality Testing scale and the lowest coefficient was -.03 with the Impulse Disorder scale.

If an evaluation is made of the validity coefficients from the Endicott et al. study in Table 2, those reported for the admission rating must be judged to be low. The mean coefficient for the FEF and MSER overall severity measures is -.29; the mean of all of the coefficients from the three criterion instruments is -.15. These correlations indicate that GAS scores account for an average of only 8% of the variance in the other overall severity measures and an average of only 2% of the variance in the 28 total scales. These data are not very supportive of the GAS and do not give much endorsement to its use. They indicate that when the other instruments are used as criteria, decisions made on the basis of the GAS are barely better than chance.

The results reported by Endicott et al. (1976), however, indicate that the validity coefficients of the concurrent ratings done at the time of 6 months post-admission are higher. All of the single correlations and mean correlations reported in Table 2 increase from those rated at the time of admission to those rated at 6 months. The mean validity coefficient for the three overall severity measures (FEF and
MSER Severity) increases to -.53, and the mean correlation of all of the 28 scales increases to -.32. Endicott terms these coefficients "moderate" (Endicott et al., 1976, p. 771), and "quite high" (Endicott, 1982). A similar significant change in the validity coefficients of the HSRS between an admission rating and a 1-year post-admission rating is reported by Hargreaves (1983).

The problem which the user or potential user of the GAS is left with is deciding whether to use the admission or the 6-month set of correlations in evaluating the worth of the GAS. Which set of coefficients is more meaningful and which indicates the true validity of the GAS? Because this question is quite crucial in deciding what the degree of validity support is for the GAS in the literature, as well as for interpreting the results of the present study (which collected data at the time of admission), it will be dealt with in further detail.

Endicott et al. (1976) use the 6-month validity coefficients as evidence to support their contention that the GAS is a valid and worthwhile instrument. Many other authors who have used the GAS have also referred to these 6-month validity coefficients to justify the GAS's use and to establish that the GAS is a valid instrument which shows moderate correlations with independent measures of severity of disturbance. These individuals, however, erroneously ignore the lower admission set of validity coefficients.

A review of the uses of the GAS later in this chapter indicates that the GAS is usually used in either of two situations: (a) at the time of psychiatric admission for the purpose of describing or classifying patient groups, or (b) at both the time of admission and at the
time of discharge for the purpose of comparing pre- to post- scores and calculating a measure of client change. In either of these cases, the admission score is utilized and is a crucial element. The GAS admission validity coefficients are therefore directly relevant to the actual conditions under which the GAS is usually used. The 6-month coefficients reported by Endicott et al., on the other hand, were obtained 6 months after admission, when most of the subjects were living in the community and were no longer patients. The GAS is infrequently used under those conditions. It does not seem justifiable to generalize from the 6-month ratings of nonpatients or ex-patients to usual usage of the GAS with current patients.

Endicott et al. explain away the low admission validity coefficients by saying they were due to "conditions of restricted variability" (p. 769) at the time of admission. The explanation they give for the coefficients increasing at six months is:

This is primarily accounted for by the greater heterogeneity of the patients' severity scores at six months. On admission, almost all patients have scores below 50. At six months, while some patients remain well below 50, others have improved sufficiently to have scores well above 50. The standard deviation of the GAS ratings thus increases from approximately 10 on admission to approximately 15 at six months. (p. 769)

Endicott et al., then, take the position that the GAS admission standard deviation of 10 is "restricted," which causes the admission validity coefficients to be artificially attenuated, while the 6-month standard deviation of 15 is typical and consequently the 6-month validity coefficients are more "true".
It seems arbitrary to judge the 6-month standard deviation as being typical and the admission standard deviation as being restricted. It could just as well be stated that the admission standard deviation of 10 is typical for psychiatric patients and the 6-month standard deviation of 15 is inflated. The answer can be determined only by examining the reported GAS standard deviations of other psychiatric patient groups. Table 4, which reports 20 GAS standard deviations from the literature from ratings at various time periods, indicates that the mean standard deviation is 10.7 and the median standard deviation is 10.8. Therefore, Endicott et al.'s admission standard deviation of 10 is not actually restricted, but is typical. The admission validity coefficients cannot be dismissed because of "restricted variability".

The 6-month standard deviation of 15, on the other hand, appears to be unusually high, probably because it was obtained on an expatient sample. The 6-month validity coefficients are consequently likely to be artificially inflated. It must be concluded that the admission validity coefficients are the best estimate of the GAS's concurrent validity.

As McNemar (1969) explains, the increase seen in a correlation coefficient when the variance of the sample studied increases is actually a statistical artifact of the correlation coefficient. The value of $r$ will always increase when the standard deviation of the sample increases, irrespective of whether there is any actual change in the strength of the relationship. If we take Endicott et al.'s admission validity coefficients and observe how the coefficients increase as the sample's standard deviation increases, some interesting insights can be seen. By putting the coefficients into a formula given by McNemar (1969, p. 162), it
becomes possible to see how much the coefficients would increase only from the statistical artifact when the standard deviation of the sample is increased from 10 to 15 (the standard deviations of the admission sample and the 6-month sample).

The results of this calculation are shown in Table 3. As expected, the validity coefficients all increase when the standard deviation is increased from 10 to 15, but none of the coefficients increase to the level of Endicott et al.'s actual 6-month coefficients. The change caused by the statistical artifact represents only a part of the actual change between the two sets of coefficients. For three of the coefficients, the artificial change is less than one-half of the actual change. It must be concluded that there is also another significant factor which is responsible for the increase between the admission coefficients and the 6-month coefficients. The conclusion can also be made that Endicott et al. are incorrect in their statement that the 6-month validity coefficients are higher primarily because of the increase in the standard deviation.

There is no direct evidence for what this additional factor is that causes an increase in validity coefficients over time. However, a hypothesis can be made that the additional factor is the increased familiarity and knowledge about a client that a rater acquires over time and through repeated contacts. It seems plausible that familiarity would have a significant effect on validity. In usual use of the GAS, an admission rating is made on the basis of a one hour intake interview. At this time, clients are often defensive and on their guard because treatment is a new experience. It is difficult for a therapist or
TABLE 3
Change in Validity Coefficients Due to Statistical Artifact as Compared to Actual Change

<table>
<thead>
<tr>
<th>Criterion Variable</th>
<th>Admission Coefficient&lt;sup&gt;a&lt;/sup&gt; (SD=10)</th>
<th>Statistically Adjusted Coefficient for SD=15</th>
<th>6-Month Coefficient Obtained&lt;sup&gt;a&lt;/sup&gt; (SD=15)</th>
<th>Difference between 6-month Coefficient and Adjusted Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEF</td>
<td>-.19</td>
<td>-.28</td>
<td>-.45</td>
<td>.17</td>
</tr>
<tr>
<td>FEF</td>
<td>-.25</td>
<td>-.36</td>
<td>-.52</td>
<td>.16</td>
</tr>
<tr>
<td>MSER Severity</td>
<td>-.44</td>
<td>-.59</td>
<td>-.62</td>
<td>.03</td>
</tr>
<tr>
<td>MSER Mean</td>
<td>-.11</td>
<td>-.16</td>
<td>-.27</td>
<td>.11</td>
</tr>
<tr>
<td>PSS Mean</td>
<td>-.22</td>
<td>-.32</td>
<td>-.37</td>
<td>.05</td>
</tr>
</tbody>
</table>

<sup>a</sup>Endicott, Spitzer, Fleiss, & Cohen, 1976

researcher to gather sufficient information in this short time to make a valid rating. As has been seen, the validity of admission GAS ratings is low. However, after two or three sessions, clients usually become more accustomed to the therapy experience and are less defensive. At this point, the rater has also had more time to obtain information. It is likely that GAS ratings made later in treatment will be more valid. Research has shown that reliability is affected by the rater's amount of knowledge about a client (Luborsky & Bachrach, 1974), but its effect on validity has not yet been studied.
One final worthwhile item to note in the Endicott et al. (1976) study is the pattern of concurrent correlations within the different scales on the MSER and the PSS. In general, the highest correlations were obtained on scales that measured psychotic symptoms, while lower correlations were obtained on scales that measured neurotic symptoms. As has been noted, the highest correlation coefficients on the MSER were with the Cognitive Disorganization scale ($r = -.37$) and the Unusual Thoughts-Delusions scale ($r = -.37$). Low correlations were obtained between GAS scores and the Anxiety scale ($r = .03$) and the Depressive Ideation-Mood scale ($r = .17$). On the PSS highest correlations were obtained with Reality Testing ($r = -.47$) and Behavioral Disturbance ($r = -.42$). The correlation with the Subjective Distress scale on the PSS was only $-.07$.

A similar pattern of correlations was obtained in a study by Battista (1982) which will be discussed later. He correlated GAS scores with concurrent ratings of use of ego functions. Three psychotic level ego functions had a mean $r$ of $-.57$ with the GAS, while four neurotic level ego functions had a mean coefficient of only $-.06$.

These patterns of individual correlations of other instruments with the GAS suggest that the construct which the GAS measures is more related to psychotic symptoms than to neurotic symptoms. Ideally, a measure of severity of disturbance should relate equally to psychotic and neurotic characteristics, since both are included in the continuum between mental health and mental illness and both types of disorders can range from mild to severe states. Perhaps the above results are a function of the mostly schizophrenic sample which Endicott et al.
studied, or perhaps they are due to the content of the GAS being better elaborated in the psychotic range. Only further research will be able to answer these questions and provide a better understanding of this characteristic of the GAS.

Herman's Michigan Department of Mental Health Study. The second study which addresses the issue of the GAS's concurrent validity is an unpublished study by Herman (1982a) conducted for the Michigan Department of Mental Health. This study was also done on a low-functioning population, consisting of 824 state psychiatric hospital inpatients and aftercare community group home residents. The mean GAS score of the sample was 34. Herman cites several concurrent validity coefficients for the GAS, which were obtained during long-term treatment. GAS scores were correlated with ratings made by staff members on five self-care behaviors (e.g. dressing, bathing), five community living skills (e.g. food preparation, shopping), and records of occurrence and severity of maladaptive behaviors. Expected correlations are again in the negative direction.

Results of Herman's study, which are shown in Table 2, indicate that GAS scores have a moderate correlation with the maladaptive behavior measures and community living skills. Coefficients with the maladaptive behavior measures ranged from -.45 to -.52, and averaged -.49. Coefficients with the community living skills ranged from -.46 to -.54, and also averaged -.49. The correlations between GAS scores and ratings on self-care skills, however, were lower, averaging -.31. When a multiple regression analysis was done combining the ratings for all three areas,
a multiple \( R \) of .62 was obtained, indicating that the three areas together accounted for 38% of the variance in GAS scores.

While the coefficients obtained by Herman are generally higher than those obtained by Endicott et al. (1976), a methodological problem is seen in the study which could question its results. The GAS ratings and the ratings on the various behavior dimensions were all made by the same rater at the same time and were based on the same information about a client. Since the rater was aware of his own GAS score when he made the criterion ratings, the criteria were not independent and there was contamination. The raters undoubtedly attempted to be objective in their ratings, but the limits of this objectivity can be questioned. The Herman (1982a) study adds worthwhile information to the question of the GAS's validity, but its results could be spuriously high.

**Other Concurrent Validity Coefficients.** A few additional studies in the literature cite a concurrent correlation between GAS and another psychometric measure. While the purpose of none of these studies was to examine the GAS's validity, the reported correlations can add additional information to an understanding of the validity of the GAS.

A study by Hurt, Friedman, Clarkin, Corn, and Aronoff (1982) correlated both actual and extracted scores on the Hamilton Depression Rating Scale (HDRS) with GAS scores. The HDRS is a commonly used rating scale for the severity of depression. Correlations with the GAS ranged from -0.63 to -0.73. It is not clear from the information given whether or not the same judge rated subjects on both scales.

Six recent studies report the development of a new psychometric instrument and then use the GAS as a standard to compare the newly developed
Concurrent correlations between GAS and the other instrument are reported to develop construct validity for the instrument. This use of the GAS as a validity standard for other instruments is an indication of the degree to which the GAS has achieved acceptance among researchers. It is ironic that the GAS is used to establish validity for other scales before its own validity has been well established.

Because these newly introduced scales are relatively unknown and untested, it is not advisable to attach much weight to their correlations with the GAS. The studies are worth looking at briefly, however. In a study which has already been mentioned, Battista (1982) correlated a new scale he developed, the Ego Function Inventory, with the GAS. Correlations between GAS and the twelve pathological ego functions which were rated averaged -.25, and ranged from -.66 with Delusional Projection to +.29 with Intellectualization. The study's design used the same rater, the client's therapist, for both scales.

In another study, Lefkovit, Morrison, and Davis (1982) developed an instrument called the Assessment of Current Functioning Scale (ACFS). The ACFS was intended to be a global measure of a client's degree of "role performance," which is the extent to which the client's behavior meets societal and personal needs. The ACFS correlated .87 with the GAS, which is used by the authors to claim concurrent validity for the ACFS. However, since the client's therapist rated subjects on both scales at the same time, criterion contamination is likely and the results may be spuriously high.

Sorenson, Hargreaves, and Friedlander (1982) evaluated the Children's Impairment Scale (CIS), a rating scale for severity of children's
psychiatric disorder which was developed for possible statewide use in California. A group of adolescents who were appropriate for both the CIS and the GAS were rated on the two scales. The correlation of the composite CIS with the GAS was -.72; correlations of the individual CIS subscales ranged from -.42 to -.79. Since the same clinician rated subjects on both scales, criterion contamination and spurious results are again possible in this study.

A clinical rating scale developed by Filstead et al. (1982) was evaluated by having therapists rate videotapes of intake interviews on both the GAS and the new scale. The Overall Impairment Index on Filstead's scale correlated .58 with the GAS, and the correlations of the individual subscales were between .42 and .69. Again, criterion contamination appears likely in this study.

Stone (1979) constructed a Temperament Index, which reflects the number of temperament characteristics possessed by clients. He rated his clients on both the GAS and the Temperament Index and obtained a correlation of -.61 between the two measures. In the last study, Fawcett, Clark, Schneftner, and Gibbons (1983) developed a Pleasure Scale, which measures a client's capacity to experience pleasure. Scores on the Pleasure Scale were correlated with several measures, including the GAS, with which a coefficient of -.12 was obtained.

Summary of Concurrent Validity. A review of the concurrent validity coefficients which have been cited on the GAS indicates that the methodologically "pure" correlations are low and account for a small portion of the variance in GAS scores. Other coefficients have been
cited which are in the moderate or even high ranges, but these correlations are tentative because of problems such as criterion contamination in the studies. Most of the data come from studies on inpatients and aftercare group home residents. The pattern of concurrent correlations does suggest that GAS scores are more highly related to psychotic symptoms than to neurotic or adjustment problem ones. Data also suggest that size of validity coefficients may be related to the rater's degree of familiarity with the subject. In sum, support given for the GAS by concurrent validity coefficients is rather weak, but it is also inconclusive.

Contrasted Groups Validity

Other methods can also be used to demonstrate criterion-related validity, as Anastasi (1976) points out. One method which has been more successful for the GAS than concurrent correlations is the method of contrasted groups. This involves comparing scores between known groups of subjects who would be expected to differ on the construct measured by a scale. The contrasted groups, then, become the criterion for scores on the scale. Contrasted groups validity for the GAS is seen in the significantly different mean scores reported in the literature for different groups of psychiatric patients and normals.

Some typical mean GAS scores for different client status groups are listed in Table 4. As can be seen, inpatient means are usually between 29 and 39, means for acute outpatients are usually in the 53 to 60 range, and normal subjects have a mean above 73. Mean scores for these three groups are generally clustered together in the range which
would be expected for each group. Only two mean scores overlap with those of another client status group, and only two scores are outside the clearly separated ranges noted. This suggests adequate validity of the GAS for distinguishing between these patient groups.

One problem in using this data to support the GAS is that contrasted groups validity only distinguishes between groups with broad differences. It does not demonstrate validity of the GAS for measuring severity of disturbance when the differences between subjects are smaller, as is frequently the case. The contrasted groups data is also valid for making conclusions only about mean scores of groups of clients; it does not demonstrate validity for distinguishing between scores of individual clients.

Construct Validity

While the previously cited material on validity all contributes to an understanding of the GAS's construct validity, some additional construct validity evidence is provided by methods which are neither content nor criterion-related.

Reflection of Change During Treatment

One additional method which can be used to examine validity is to evaluate the effect that a variable which is expected to cause change has on a rating scale's scores (Anastasi, 1976, p. 155). Psychiatric treatment is expected to cause change in clients' level of severity of disturbance; therefore, treatment would be expected to cause change on the GAS. If an increase in GAS scores can be demonstrated following treatment, additional evidence is given for the GAS's validity.
### TABLE 4
GAS Means and Standard Deviations from the Literature

<table>
<thead>
<tr>
<th>Reference</th>
<th>Mean</th>
<th>SD</th>
<th>Patient Group</th>
<th>Time of Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andreasen &amp; Hoenk (1982)</td>
<td>86.8</td>
<td>~12.6</td>
<td>Normals</td>
<td></td>
</tr>
<tr>
<td>Bassuk, Winter, &amp; Apsler (1983)</td>
<td>47.0</td>
<td>~16.6</td>
<td>Emergency service out-patients</td>
<td>Admission</td>
</tr>
<tr>
<td>Braff (1981)</td>
<td>34.7</td>
<td>10.6</td>
<td>Inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Braff, Glick, &amp; Griffin (1983)</td>
<td>39.0</td>
<td>~13.0</td>
<td>Inpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td>Carpenter, Sadler, Light, Hanlon, &amp; Kurland (1983)</td>
<td>38.1</td>
<td>~11.0</td>
<td>Chronic schizophrenic outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td>Curran, Miller, Monti, Zwick, &amp; Stout (1980)</td>
<td>35.7</td>
<td>-</td>
<td>Inpatients &amp; day patients</td>
<td>Admission</td>
</tr>
<tr>
<td>Fink, Braden, &amp; Qualls (1982)</td>
<td>29.0</td>
<td>~6.0</td>
<td>Inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>~11.0</td>
<td>Inpatients</td>
<td>Discharge</td>
</tr>
<tr>
<td>Goldstein (1980)</td>
<td>~22.0</td>
<td>-</td>
<td>Acute schizophrenic inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Gully &amp; Harris (1982)</td>
<td>56.9</td>
<td>13.9</td>
<td>CMH outpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Harley (1983)</td>
<td>55.1</td>
<td>10.2</td>
<td>CMH outpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Herz, Szymanski, &amp; Simon (1982)</td>
<td>62.7</td>
<td>-</td>
<td>Chronic schizophrenic outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td>Horowitz, Krupnick, Kaltreider, Wilner, Leong, &amp; Marmar (1981)</td>
<td>59.4</td>
<td>8.5</td>
<td>Outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td></td>
<td>72.9</td>
<td>10.9</td>
<td>Normals</td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Mean</th>
<th>SD</th>
<th>Patient Group</th>
<th>Time of Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iacono, Peloquin, Lumry, Valentine, &amp; Tuason (1982)</td>
<td>46.2</td>
<td>~ 9.2</td>
<td>Inpatients &amp; outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td></td>
<td>78.7</td>
<td>~ 8.0</td>
<td>Ex-patients</td>
<td>Not in treatment</td>
</tr>
<tr>
<td>Kaplan, Pelcovitz, Salzinger, &amp; Ganeles (1983)</td>
<td>58.3</td>
<td>-</td>
<td>Outpatients</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>79.6</td>
<td>-</td>
<td>Normals</td>
<td></td>
</tr>
<tr>
<td>Kass, Charles, Walsh, &amp; Barsa (1983)</td>
<td>59.6</td>
<td>-</td>
<td>Outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td>Larsen (1978)</td>
<td>55.9</td>
<td>-</td>
<td>CMH outpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Meltzer, Busch, Tricou, &amp; Robertson (1982)</td>
<td>29.4</td>
<td>9.5</td>
<td>Schizophrenic inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Shenoy, Sadler, Goldberg, Hamer, &amp; Ross (1981)</td>
<td>61.5</td>
<td>12.1</td>
<td>Chronic schizophrenic outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td>Sorensen, Hargreaves, &amp; Friedlander (1982)</td>
<td>~ 39.0</td>
<td>~ 12.2</td>
<td>Inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td></td>
<td>~ 59.0</td>
<td>~ 11.5</td>
<td>Outpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Stone (1979)</td>
<td>47.1</td>
<td>-</td>
<td>Outpatients</td>
<td>In treatment</td>
</tr>
<tr>
<td>VanPutten, May, Marder, &amp; Wittman (1981)</td>
<td>31.0</td>
<td>-</td>
<td>Schizophrenic inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Weckowicz, Tam, Mason, &amp; Bay (1978)</td>
<td>51.1</td>
<td>7.1</td>
<td>Inpatients</td>
<td>Admission</td>
</tr>
<tr>
<td>Weiner, Marten, Wochnick, Davis, Fishman, &amp; Clayton (1979)</td>
<td>53.6</td>
<td>~10.7</td>
<td>Severest episode of doctors</td>
<td>Retrospective</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
However, the possible sources of error in studies of psychiatric treatment diminish the strength of the validity evidence which they provide. One potential source of error in these studies is that the subjects' therapist was usually the rater. Since a therapist's sense of self-worth is often threatened if he does not see positive change in his clients during treatment, therapist-raters may have tended to inflate post-treatment scores. By using the same individual to rate clients before and after treatment, the ratings are also left open to contamination.

Larkin (1979) reports that GAS scores of a group of hospitalized schizophrenic patients were significantly greater 4-6 weeks after discharge than they were at the time of admission. Fink, Braden, and Qualls (1982) found that the GAS scores of a group of acute inpatients increased from a mean of 29 at admission to a mean of 50 after an average average hospitalization of 3 weeks. Herz, Endicott, and Spitzer (1975) report that the mean GAS scores of the inpatients used in the original validation sample of the GAS were 31 at admission, 45 at 3 weeks after admission, and 51 at 12 weeks after admission.

Goldstein (1980) followed a group of acute schizophrenics from admission to a follow-up 6 months later. The mean GAS at admission was approximately 22. After 2 months of inpatient and intensive outpatient aftercare treatment, GAS scores increased to an average of 51, and at the follow-up point the mean was 60. Another group of inpatients and day hospital patients was followed by Curran, Miller, Monti, Zwick, and Stout (1980), who report an admission GAS mean of 36 and a 2-month post-discharge mean of 62. Gudeman, Dickey, Rood, Hellman, and Grinspoon
(1981) report that GAS scores of a group of chronic schizophrenics increased from 25 before placement in an experimental quarterway house program to 38 after 1 year in the program.

Some additional studies report changes in GAS scores during outpatient treatment. Stone (1983) reports a 16 point increase in GAS scores for a group of affectively ill borderline patients he treated. Rehm (1981) found that depressed outpatients who were treated with a self-control therapy program showed an increase in GAS scores. Finally, unpublished data from Ingham Community Mental Health Center, the site of the present study, indicate an average increase in GAS scores of 1.2 points per outpatient session (Kriauciunas, 1981).

**Similar Pattern of Scores as Other Measures**

A final method which can be used to gain information about the GAS's construct validity is to examine the relationship of GAS scores to other variables which measure severity of disturbance in studies where concurrent correlations are not cited. This involves determining whether GAS scores and other measures of psychopathology show similar patterns of change and give identical indications of statistical significance. Some of the studies in this group give further support for the GAS, while others provide additional evidence that questions the GAS's validity.

Rehm (1981) found that his clients displayed significant change on the GAS as well as on another rating scale, the Hamilton Depression Rating Scale, and on two self-report measures, the Beck Depression Inventory and the Pleasant Events Schedule. Fink et al. (1982) cite a
similar pattern of scores on both the GAS and the Target Symptom Rating Scale. Shenoy, Sadler, Goldberg, Hamer, and Ross (1981) found no difference between two groups of patients on the GAS in several ratings, and had this confirmed by similar results on the SADS.

The scale that has been used most frequently along with the GAS is the Brief Psychiatric Rating Scale (BPRS), an instrument for rating psychiatric symptoms in 16 areas. A problem in comparing GAS and BPRS scores is that the two scales were usually rated by the same rater, producing non-independent measures. The studies which used both instruments, however, are still worth examining. King and Goldstein (1979) found a significant difference in GAS scores between two groups of former inpatients in a follow-up rating. A significant difference was also seen at this time in the BPRS Thought Disorder factor, but no significant differences were seen in the three other BPRS factors, Anxiety-Depression, Hostility-Suspiciousness, and Withdrawal. Bassuk and Gerson (1980) obtained similar results in using the GAS and BPRS with two groups of emergency service outpatients. GAS, total BPRS score, and the BPRS Unusual Thought Content scale all showed significant differences, but the BPRS Anxiety and Depressive Mood scales were non-significant. These results again suggest that the construct measured by the GAS is more related to psychotic symptoms than to neurotic symptoms.

However, the lack of a relationship between GAS and neurotic symptoms is questioned by a study by Horowitz, Krupnick, Kaltreider, Wilner, Leong, and Marmar (1981). In their use of the GAS and BPRS to compare groups of out-patients and normals who had both experienced recent stress, they found that the outpatients had greater symptoms on the GAS, total
BPRS, BPRS Anxiety scale, and BPRS Depressive Mood scale. Since they did not use the eight psychotic scales on the BPRS, relationships with those scales are not known.

Other studies show dissimilar patterns of scores on the GAS and concurrent measures of severity of disturbance. Larsen (1978) found that a pre-therapy preparation interview for outpatients made a significant difference in the outcome of therapy as measured by the Symptom Checklist and the BPRS Depression scale. The GAS, however, did not show a significant positive effect. Baron, Gruen, and Asnis (1982) compared two groups of chronic schizophrenics at admission. The groups did not differ in a symptom profile or in the presence of Schneiderian first rank symptoms, but they did differ on the GAS. Kanas, Rogers, Kreth, Patterson, and Campbell (1980) found a significant difference between group therapy recipients and non-recipients on the Overall Severity of Illness scale of the Psychiatric Evaluation Form, but they did not find a difference on the GAS. Similar results are seen in a study by Rounsaville, Weissman, Wilber, and Kleber (1982), who compared three groups of drug abusers on several measures. No significant difference was seen in GAS scores, but significant differences were displayed by all of the other measures used, including the Beck Depression Inventory, the Maudsley Personality Inventory Neuroticism scale and Extroversion scale, and the Addiction Severity Index.

Summary of Validity

At the present time, validity evidence for the GAS is inconclusive. Some research supports the validity of the GAS, while other research
does not. Concurrent correlations of the GAS with other measures are low to moderately high. Several studies have lacked an independent criterion of comparison, causing their results to be questionable. Most research has been done on inpatient samples, which cannot always be used to generalize to outpatient clients. It appears that the GAS is more effective in making broad discriminations between clients than in making finer ones. Some research indicates that the GAS is more related to psychotic symptoms than to neurotic ones, but it is uncertain whether this is a function of the samples studied or an innate characteristic of the scale. There are still many unanswered questions concerning the validity of the GAS. It is clear that further research is needed.

Use of the GAS in State Mental Health Systems

As was noted in Chapter I, the most widespread use of the GAS is as an evaluation instrument in state mental health systems. Although the GAS was originally proposed by Endicott et al. (1976) as a research and clinical tool, that has not in the strict sense been its major use. The GAS has been taken up by the new and growing field of mental health program evaluation and put to use in statewide evaluation systems in several states in numbers totalling hundreds of thousands of GAS ratings per year. As stated by Lefkovitz et al. (1982), "program evaluation in community mental health centers has proliferated in recent years" (p. 297).

The major recent impetus for mental health program evaluation was U.S. Public Law 94-63, the Community Mental Health Centers Amendments of 1975. This law mandated evaluation studies of the entire range of operation of community mental health centers receiving federal funds.
(Hargreaves & DeLay, 1979). It stated that in all these CMHC's, an amount of no less than 2% of the previous year's total operating budget must be spent on program evaluation and utilization review activities. Passage of this law resulted in a major new emphasis on program evaluation. A detailed discussion of the provisions of PL 94-63 affecting program evaluation is provided by Windle and Ochberg (1975).

In the latter part of the 1970's, requirements for mandatory program evaluation from the federal level combined with tightened funding and movements for increased accountability of mental health programs on the state and local levels to create a new mood in public mental health administration. Wilson's (1979) succinct description of the situation in Colorado could well be generalized to most states:

Over the last several years, accountability has come of age in Colorado. Gone are the days of "expert testimony", automatic budget increases, unquestioned acceptance of process figures, and across the board per capita funding. Enter the era of performance contracts, unit costs, cost effectiveness, and treatment outcome measures. (p. 1)

As program evaluation activities increased rapidly in the 1970's, needs were created for more sophisticated evaluation methods and instruments. State legislators and budget analysts became increasingly knowledgeable about outcome evaluation issues and asked for more objective data which could be used to base policy and funding decisions on. At the same time, individuals well-trained in the traditional disciplines of psychology and sociology, as well as individuals trained in the specialty area of mental health program evaluation, began to occupy program evaluation positions in both state level offices and local CMHC's. An interest evolved in developing new methods for measuring clients' level
of severity of disturbance and for measuring the amount of change in client severity level which occurred during treatment.

Many state and local programs turned to global rating scales for an evaluation measure of level of disturbance, because of global scales' simplicity and ease of implementation. The scale which was most frequently adopted by state evaluation systems was the GAS. Other global scales have been available, such as the Carter-Newman Level of Functioning Scale, the Index of Well-Being, and the Global Improvement Scale, but use of none of them has been as extensive as that of the GAS. The GAS was apparently thought to provide the best balance between an instrument which could be administered quickly, but which was comprehensive enough to integrate several facets of a client's functioning into a single measure. Some programs have opted for longer, multidimensional scales, such as the Denver Community Mental Health Questionnaire, the Colorado Client Assessment Record, or the Oregon Quality of Life Questionnaire, but their use has also not been as extensive as the GAS's.

Extent of GAS Usage

It is difficult to obtain information about the extent of use of statewide evaluation instruments, since studies from state departments of mental health are rarely published in academic journals. The investigator wished to determine how many states were using the GAS on a statewide basis, so a questionnaire was sent to several departments of mental health. A list of potential states was obtained from the Michigan Department of Mental Health, and questionnaires were sent to them. Each department was asked to list any other states they were aware of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
that used the GAS. Names of states using the GAS were also requested from the Division of Biometry of the National Institute of Mental Health. Altogether, 15 states were surveyed, which should have included all of those using the GAS. Replies indicated that the GAS is being used state-wide in five states. These states and the extent of their GAS use are:

- Arizona - used for all outpatients and inpatients
- California - used for all outpatients at admission and discharge
- Michigan - used for all outpatients and inpatients at admission, at updates every three months, and at discharge
- Utah - used for all outpatients at admission and discharge
- Washington - used for all outpatients

Numbers of clients to whom GAS ratings are given annually were available for the two largest states, California and Michigan. These two states comprise 80% of the population of the five states. For California and Michigan, the percentage of the total state population (1980 Census) represented by the number of clients given GAS ratings annually was calculated. The individual state percentages (1.69% and 1.77%) averaged 1.73%. This proportion was applied to the total population of the five states, resulting in an approximate total of 713,000 clients who receive GAS ratings every year. Considering that most clients are rated from two to four times per year on the GAS, a conservative estimate would place the total number of GAS ratings given annually at 1.5 million. With this extensive use, the GAS is one of the most frequently used clinical instruments in the United States today.
According to Green and Nguyen (1977), the GAS was first mandated for all county community mental health programs in California in 1977. Michigan introduced the GAS in 1979 in six counties, including Ingham county, the site of the present study. The investigator's use of the GAS during his pre-doctoral internship year at Ingham CMHC provided the impetus for the present study. Michigan adopted statewide mandatory use of the GAS in 1982. In Utah, statewide use of the GAS was adopted in 1983.

Specific Uses of GAS in State Programs

Because of the lack of published material about the GAS from state programs, the exact uses which are given to GAS scores are in some cases ambiguous. This is unfortunate for the purposes of this study because a discussion of validity should ideally be tied to the usage of an instrument; i.e., validity should be discussed in terms of validity for particular types of decisions (American Psychological Association, 1974). Newman (1980) states that global scales can be used in mental health programs for treatment evaluation, program planning, budgeting, policy development, and cost effectiveness studies. Unpublished materials from some state departments of mental health give examples of these uses.

In Michigan, the GAS is labeled as a tool for "improving management and decision-making capacity" (Michigan Department of Mental Health, 1981, p. 1). Richard Spates (1982), a Department of Mental Health administrator, stated that one of the major uses of the GAS is to provide an outcome orientation for budgeting decisions. Prior to implementation
of the GAS, he stated, funding decisions in the one-half billion dollar
department budget needed to be made mostly on "hearsay." The GAS has
provided the department with a client-related measure which can be used
in making budget and program decisions.

Michigan Department of Mental Health policy requires all local CMH
boards to classify clients served into a GAS/PPB matrix (Michigan Depart­
ment of Mental Health, 1981). This matrix is a table indicating numbers
of clients served in each of five 20-point ranges of the GAS. Each of
the GAS ranges is further subdivided into five Program Planning and
Budgeting (PPB) objectives, which indicate the major reason for providing
service to a given client (Prevention, Psychosocial Adjustment, Crisis
Resolution, Rehabilitation/Habilitation, and Maintenance). The GAS/PPB
matrix, then, provides a descriptive overview of the level of disturbance
of all clients served by an agency, a CMH board, or the entire state.

GAS/PPB matrices are used to determine whether the Department of
Mental Health is serving the prioritized groups of clients, as they are
defined in the Michigan Mental Health Code, Public Act 258, and in the
Program Policy Guidelines of the department. These documents state that
within the area of mental illness, priority is to be given to the more
severely disabled. GAS data, then, indicate whether agencies, boards,
and the department are providing services to clients of lower severity
levels in a manner consistent with stated policy. GAS/PPB data are
also combined with cost data, such as cost per client and total cost to
determine for which clients public mental health dollars are being
spent. These data can also be used to compare the cost effectiveness
of treatment between clients of different functioning levels.
The GAS/PPB Workbook of the Michigan Department of Mental Health (1981) also indicates other uses which are given to the GAS. It is used to identify program areas of greatest need in planning. It can be used to evaluate change in client functioning level over periods of time. The GAS is used to compare the cost of treating a given level of clients in different settings, such as state institution versus community care. It is suggested that the GAS could be used to evaluate ward staffing levels in state institutions in relation to differing levels of client disturbance. Some local CMH boards have used GAS scores to determine eligibility for entry into programs, and others have proposed using the GAS in determining priority for treatment if reduced funding necessitates limiting the numbers of clients served (DeRath, 1981). The Michigan Department of Mental Health has also used the GAS in a residential services continuum, which specifies appropriate GAS ranges of typical clients in different residential settings (Herman, 1982b). This continuum is intended to provide guidelines for placement of clients in the least restrictive environment, and when combined with cost data, to provide a means for selecting the lowest cost option for providing the necessary level of care.

Program evaluation materials received from the State of Utah indicate that the GAS is used to evaluate treatment outcome through indices such as percentage of patients displaying improvement and diagnostic/treatment factors associated with improvement. Policy materials received from the State of Washington indicate that the GAS was until recently used there to define a "seriously disturbed person" for legislative purposes (GAS score of 50 or less). A new state law, however, changed the definition to include other broader criteria.
Review of GAS in National Institute of Mental Health Publications

Some NIMH publications have commented on the use of the GAS as an evaluation instrument in public mental health programs. Hagedorn, Beck, Neubert, and Werlin (1976) endorsed the GAS as having the best "methodological reliability and validity, capacity to detect changes in client status, and cost of administration" (p. 201) in their review of program evaluation techniques for community mental health centers. Hargreaves, McIntyre, Attkisson, and Siegel (1977) included the GAS in their recommended "core battery" of outcome measurement instruments for use in community mental health program evaluation. A recent, still unpublished review of mental health outcome measurement techniques by Ciarlo et al. (1981) gave a generally favorable review to the GAS. This report is the first to cite empirical studies using the GAS; however, the references are current only through mid-1980.

Use of the GAS in Research Studies

A review of the literature between 1975 and June, 1983 indicates that the GAS has been used in 91 published research studies, in six doctoral dissertations, and in a few unpublished studies. Many of these studies have been cited in the review of the GAS's reliability and validity. The remaining studies have little to add about the psychometric properties of the GAS, which is the focus of the present investigation. These other studies can, however, add to an understanding of the GAS by giving additional examples of how the GAS is being used. They also demonstrate the degree of importance of the decisions and conclusions which are being based on GAS scores.
Unfortunately, most of the research using the GAS comes out of the psychiatric literature and consists of experimentally oriented studies. As has been seen, the greatest use of the GAS numerically is in outpatient community mental health programs. Few published studies originate from these settings, and relatively few studies are found in the community mental health and program evaluation literatures. This is unfortunate in that it necessitates often generalizing results from inpatient and research oriented settings to other settings where conditions may be different.

The studies which have used the GAS will be categorized into major groups and briefly cited.

**Measure of Change During Treatment**

The most common use of the GAS in research is as an outcome measure to evaluate the degree of clinical change of psychiatric patients who have been given some type of treatment. A significant change in GAS scores is used to conclude that the treatment was effective; the lack of a change on the GAS suggests that the treatment was not effective. Frequently scores are compared between an experimental group and a control group. Many of the treatments which are evaluated with the GAS are newly developed or innovative ones. Since some of them have far-ranging implications, the import of the decisions based on GAS scores is frequently great.
Change Due to Psychotherapy

Many of the studies using the GAS have evaluated the effectiveness of some type of psychotherapy. Some of these studies have been cited in the section on the GAS's validity. King and Goldstein (1979) used GAS scores to support the effectiveness of crisis-oriented family therapy. An innovative quarterway house program was found to be more effective than inpatient treatment through a comparison of changes in GAS scores by Gudeman et al. (1981). GAS scores were used by Rehm (1981) to support the efficacy of a self-control therapy program for depression. A series of articles by Herz and his colleagues (Endicott, Cohen, Nee, Fleiss, & Herz, 1979; Herz, Endicott, & Gibbon, 1979; Herz et al., 1975; Herz, Endicott, & Spitzer, 1977) used changes in GAS scores over a 2-year period to conclude that brief hospitalization is preferable to longer-term hospitalization for most patients. Caton (1982) also supported short-term hospitalization through use of the GAS.

Several other studies have used the GAS to determine the value of a particular treatment. Masnik, Olarte, and Rosen (1980) used GAS ratings to evaluate a coffee group program for chronic schizophrenics. Kanas, Kreth, Rogers, Patterson, and Campbell (1978) evaluated two different types of group therapy with the GAS. The efficacy of a cognitive/behavioral treatment program for depression was evaluated through the use of GAS scores by McDonald (1978). Dagadakis, Hunt, and Carr (1981) examined changes in GAS scores to evaluate the results of their psychotherapy outcome study.
Gully and Harris (1982) evaluated the effects of four different fee-setting procedures in a community mental health center through comparing clients' changes in GAS scores. McRae (1979) also used GAS scores to evaluate the impact of fee assessment on psychotherapeutic outcome.

Two studies have utilized the GAS to evaluate innovative treatments. Williams, Gold, Shrout, Shaffer, and Adams (1979) used increases in GAS scores to demonstrate the effectiveness of psychotherapy in improving seizure control for patients with uncontrolled seizures. Carpenter, Sadler, Light, Hanlon, and Kurland (1983) used GAS scores as one measure to evaluate the controversial use of hemodialysis to treat schizophrenia. Nonsignificant changes in GAS scores were used to conclude that hemodialysis was not effective. Finally, the use of the GAS is reported by Shapiro and Keller (1981) in the National Institute of Mental Health's study of the psychobiology of depression, an ongoing, 2-year longitudinal study involving 1000 patients nationwide.

**Change due to Medication**

Another category of studies which has used the GAS to evaluate clinical change during treatment is drug studies. GAS ratings before and after treatment with a medication, sometimes an experimental medication or an experimental dosage, are used to help determine the effectiveness of the drug.

Two studies used GAS scores to evaluate the use of Lecithin in the treatment of mania (Cohen, Lipinski, & Altesman, 1982; Cohen, Miller, Lipinski, & Pope, 1980). Meltzer, Tueting, and Jackman (1982) and
VanPutten and Sanders (1975) evaluated the effect of Lithium by examining changes in GAS scores, and Braden, Fink, Qualls, Ho, and Samuels (1982) used the GAS to compare the relative effectiveness of Lithium and chlorpromazine in the treatment of schizophrenia and mania. Another medication that was evaluated with the GAS for potential use was Des-tyrosine-gamma-endorphin (Meltzer, Busch, Tricou, & Robertson, 1982).

Brown and Qualls (1981) used GAS results to compare the responses of Dexamethasone Suppression Test suppressors and nonsuppressors to different antidepressant medications. Results suggested that suppressors and nonsuppressors each respond better to different antidepressants.

Two studies used GAS scores to test the interesting hypothesis that a patient's subjective response to a test dose of medication at the beginning of treatment can be used to predict his or her clinical outcome at discharge. VanPutten and May (1978) determined that the subjective response of patients to an initial test dose of Thorazine was correlated with outcome as measured by GAS score change. Fink et al. (1982) found the same relationship to be true for a mixed group of medications.

Two final studies used clinical change as measured by the GAS to evaluate the effectiveness of partial medication programs. Herz, Szymanski, and Simon (1982) found that GAS levels did not change in some stable chronic schizophrenics when they were given medication intermittently. It was determined by Shenoy et al. (1981) that a group of chronic schizophrenics had no relapses in clinical condition, as measured by GAS scores, when they were given a 6-week drug holiday.
Comparison of Severity of Disturbance Between Experimental Groups

A large category of studies has used the GAS to compare the level of severity of disturbance between two or more experimental groups, or between an experimental group and a control group. These comparisons are sometimes used to draw significant conclusions about the subjects.

Bassuk used the GAS to compare the severity of disturbance of different patient groups at Beth Israel Hospital in Boston. She compared geriatric patients and younger patients in the psychiatric emergency service and found no significant difference in GAS scores (Bassuk, Minden, & Apsler, 1983). She did find, however, that repeat patients in the psychiatric emergency service had significantly higher GAS scores than first-time patients (Bassuk & Gerson, 1980). When GAS scores of emergency service clients at the Boston hospital were compared to those at a similar hospital in England, no difference was found (Bassuk, Winter, & Apsler, 1983).

Two studies used the GAS to compare characteristics of groups of drug addicts. Murphy, Rounsaville, Eyre, and Kleber (1983) found that addicts who had made suicide attempts had significantly lower GAS scores than those who had not. However, Rounsaville, Weissman, and Kleber (1982) did not find a difference between opiate addicts who were also alcoholics and those who were not. In another study, McRae (1983) found a significantly lower mean GAS score among clients who were admitted to a state hospital than among clients who applied for admission but were not accepted. VanPutten, Crumpton, and Yale (1976) discovered that drug refusers had lower GAS scores at discharge than drug compliers.
A difference in pretreatment GAS scores was found between syntonic and dysphoric responders to a medication test dose by VanPutten, May, Mar­der, and Wittman (1981).

In probably the only study which has used the GAS with only normal level subjects, Welner, Marten, Wochnick, Davis, Fishman, and Clayton (1979) compared the severity of worst depressive episodes between female MD's and female PhD's. Horowitz et al. (1981) found that a group of psychiatric outpatients who had recently experienced the death of a parent had a significantly lower mean GAS score than a group of non-patient normals who had also recently experienced parental death. Phillips (1979) compared GAS scores between clients who dropped out of outpatient therapy and those who continued, and found a difference only after the second session. Adherers to outpatient treatment were found to have higher GAS scores than nonadherers by Fink and Heckerman (1981). Curran et al. (1980) found that socially inadequate patients and a general inpatient population differed on the GAS at one out of three rating times.

Higher mean GAS scores were found among the parents of normal children than among parents of abused and neglected children by Kaplan, Pelcovitz, Salzinger, and Ganeles (1983). Doane, Jones, Fisher, Ritzer, Singer, and Wynne (1982) compared the GAS scores of parents of low functioning at-risk children to the parents of higher functioning at-risk children, but did not find a difference.

Two final studies in this category also compared GAS scores of two experimental groups, but for the purpose of examining differences or biases in the raters instead of differences in the client groups. Bassuk
and Apsler (1983) compared the GAS scores given to rape victims by male and female therapists. The fact that male therapists gave higher GAS scores led them to conclude that there may be a sex bias in rape counseling. In a similar study, Weiner (1978) found no sex bias in the GAS ratings given by male and female therapists to standard case descriptions. However, he did find a difference in one set of GAS ratings given by androgynous therapists as compared to other therapists.

Correlation of Severity of Disturbance with Another Variable

The GAS has been used as a measure of severity of disturbance in a few studies where the relationship of another variable to level of psychopathology was examined.

Some studies have examined the relationship between the severity level of parents who had experienced at least one psychiatric hospitalization and characteristics of their children, who were considered to be at-risk. Two studies used the GAS in investigating the relationship between level of psychopathology of parents and the behavioral competence and functioning level of their children (Harder, Kokes, Fisher, & Strauss, 1980; Kokes, Harder, Fisher, & Strauss, 1980). Erlenmeyer-Kimling, Cornblatt, and Fleiss (1979) compared GAS scores of former inpatient parents to the speed of recovery of the electrodermal response in their children. GAS scores were used by Romano and Geertsma (1978) as weighted factors in two scales they developed to be used on previously hospitalized parents in predicting the mental health outcome of their at-risk child.

In a different area of research, Braff and Sacuzzo (1982) used GAS scores to correlate level of severity of disturbance with speed of information processing in schizophrenics.
Screening Instrument to Select Subjects

Another group of studies has used the GAS as a screening instrument to select subjects of a certain level of severity of disturbance for experimental groups. This selection of subjects was done so that results of the studies could be generalized to particular individuals, or so that the experimental groups could be equated for severity and severity level ruled out as a confounding factor.

Several studies have limited their subjects to remitted psychiatric patients so that results could be generalized to this particular group. These studies used GAS scores to define the status of being "in remission." Iacono (1982) and Iacono, Tuason, and Johnson (1981) used a GAS score of 60 or greater as a cut-off for choosing schizophrenics who were in remission. In studies which Iacono did on patients with major affective disorder in remission, the subjects were required to have a GAS score of 70 or greater (Iacono, Lykken, Peloquin, Lumry, Valentine, & Tuason, 1983; Iacono, Peloquin, Lumry, Valentine, & Tuason, 1982; Iacono & Tuason, 1983). Asarnow and MacCrimmon (1981) also used GAS scores as a criterion for being in remission, but they do not state what cut-off score was used.

Cohen, Lipinski, Harris, Pope, and Friedman (1980) selected their subjects on the basis of GAS scores. Good responders to medication were defined as patients with 20-50 points GAS increase while on medication, and poor responders were defined as patients who showed less than 5 points GAS increase. The GAS was also used as a selection instrument by Friedman, Vaughan, and Erlenmeyer-Kimling (1982), who
used only schizophrenic subjects whose GAS scores from different raters were all in the mentally ill range.

Other researchers have used GAS scores for selecting subjects in order to equate the experimental groups for level of severity of disturbance at the beginning of the study. By choosing subject groups with an equivalent level of disturbance, psychopathology level was able to be ruled out as a confounding factor in the experimental design. Examples of this use of the GAS are seen in studies by Geyer and Braff (1982); Latimer, Sarna, Campbell, Latimer, Waterfall, and Daniel (1981); Margo and McMahon (1982); and Weckowicz, Tam, Mason, and Bay (1978).

Three additional studies used the GAS to eliminate the possible confounding effect of the severity of disturbance variable in their experimental design, but did this through statistical use of covariance rather than through screening of the groups. The covariance procedure allowed researchers to keep the natural groups intact, by statistically controlling for the extraneous effects of differences in level of psychopathology between the groups. Use of GAS scores as a covariate is seen in studies by Abramowitz, Davidson, Greene, and Edwards (1980); Braff and Saccuzzo (1981); and Edwards, Greene, Abramowitz, and Davidson (1979).

Measure to Describe Subjects

A last group of studies has used GAS scores as a means to describe the subjects in the study. Sometimes GAS scores are given to validate subjective descriptions of the subjects, such as "moderately disturbed patients." Other times scores are given just to communicate the level of psychopathology of the subjects to the reader.

Four final studies mention that the GAS was used as a measure of severity of disturbance in the study, but no scores or results are given (Caton, 1981; Harding & Brooks, 1980; Leader, 1978; Szymansk, Simon, & Gutterman, 1983).

MMPF Overall Severity Indexes

Four indexes of overall severity of disturbance taken from the MMPI were used as criterion variables in the study. Literature supporting use of these indexes as measures of severity of psychiatric disturbance will be briefly reviewed.

One measure of overall disturbance used was the total raw score on the clinical scales of the MMPI, excluding scale Mf. This was simply the total number of clinical scale items which were answered in the deviant, or scoreable direction. Gallagher (1953) used the sum of deviant items on all the MMPI scales as a Maladjustment Score, and recommended using this as an index of severity of maladjustment. The same total score index was used as a measure of personality deviancy by Berg (1955, 1957) in his formulation of the deviation hypothesis. Meehl and Hathaway (1946) analyzed the sum of deviant responses in extreme groups to derive a plus scale, which they used as a measure of response deviation.
The second MMPI index of severity of disturbance used was the total T-score of the nine clinical scales. This measure is the same as the total raw score, except that K-corrections have been added and the scores have been standardized into a normal distribution.

The Scatter Index was used as the third measure of overall severity of disturbance. This index, which was introduced by L'Abate in 1962, consists of the sum of the absolute values of the T-score deviations from the mean T-score of 50 on the clinical scales. It is, therefore, a measure of scatter from the mean, and as such considers deviations below the mean equal to deviations above the mean. L'Abate found that the Scatter Index differentiated well between groups of different psychiatric patients and normals. The Scatter Index was also used as a measure of severity of disturbance by Craddick (1963) and Stone (1964).

A similar measure of scatter from the mean, the Clerical Index, was used by Levitan, Goldfarb, and Jacobs (1959), and was discovered to be highly correlated with ratings of the degree of maladjustment of MMPI profiles.

The final index used as a criterion measure of overall severity of disturbance was the MMPI F scale. This scale was originally formulated on the basis of its ability to discriminate between the Minnesota normal group and patient groups (Dahlstrom, Welsh, & Dahlstrom, 1972). It has been used frequently in research studies as a measure of severity of disturbance. Blumberg (1967) found that F-scale scores were related to three other measures of psychopathology. The F scale was found to be correlated with clinicians' judgments of severity of psychopathology by Sines and Silver (1963). Winter and Stortroen (1963) compared the ability
of 10 MMPI indexes to discriminate between patients of varying degrees of disturbance, and found that the F scale had the highest discriminative power. Shaffer, Ota, and Hanlon (1964) determined that the F scale was one of the three best. Others who have found that the F scale is a valid index of psychopathology are Gross (1959) and Lutzker (1961).
CHAPTER III

METHOD

The problem which is the background for the study and a statement
of the study's research questions have been presented in Chapter I. In
brief, the problem is that past evidence for the GAS's validity is mar-
ginal and inconclusive. Most of the concurrent validity data on the
GAS come from studies of inpatient and aftercare clients, in spite of
the fact that greatest use of the GAS is with acute outpatients. Many
concurrent validity studies have failed to use an independent criterion.
No previous studies have directly examined the concurrent validity of
the GAS in an outpatient sample, and no studies have used a self-report
psychological test for the criterion measure. In addition, evidence
regarding which facets of psychopathology the GAS measures is inconclu-
sive. It is these questions of whether the GAS is, in fact, a valid
measure of severity of disturbance and which facets of psychopathology
it measures that were investigated in the present study. This investi-
gation was done by comparing GAS scores to scores on an established
psychological test, the MMPI.

Hypotheses

The hypotheses examined in the study were divided into a primary
hypothesis and five secondary hypotheses.

The primary hypothesis was: There is a significant correlation be-
tween the GAS and each of four measures of overall severity of disturbance
taken from the MMPI. These four measures of overall severity from the
MMPI are Total T-score, Total Raw Score, Scatter Index, and F-scale Score.

The secondary hypotheses were:

a) There is a significant correlation between the GAS and each of
the nine standard clinical scales of the MMPI (scale Mf
excluded). These nine clinical scales are Hypochondriasis
(Hs), Depression (D), Hysteria (Hy), Psychopathic Deviancy
(Pd), Paranoia (Pa), Psychathenia (Pt), Schizophrenia (Sc),
Mania (Ma), and Social Introversion (Si).

b) There is a significant multiple correlation between the GAS
and the MMPI neurotic triad scales, scales Hs, D, and Hy.

c) There is a significant multiple correlation between the GAS
and the MMPI psychotic triad scales, scales Pa, Sc, and Ma.
There is also a significant multiple correlation between the
GAS and the psychotic tetrad scales, scales Pa, Pt, Sc, and
Ma.

d) There is a significantly greater multiple correlation between
both the psychotic triad and psychotic tetrad scales and the
GAS than there is between the neurotic triad scales and the
GAS.

e) There is a significant correlation between the GAS and each of
the two remaining validity scales of the MMPI, scales L and K.

Experimental Design

The design of the study consists of correlations between GAS and
several other variables. The statistical analyses are Pearson product
moment correlations ($r$) and multiple correlations ($R$). The design of the study is diagrammed in Figure 1.

### Severity Indices

<table>
<thead>
<tr>
<th>GAS Score</th>
<th>N = 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total T-score</td>
<td>$r$</td>
</tr>
<tr>
<td>Total Raw Score</td>
<td>$r$</td>
</tr>
<tr>
<td>Scatter Index</td>
<td>$r$</td>
</tr>
<tr>
<td>F-scale Score</td>
<td>$r$</td>
</tr>
</tbody>
</table>

### Clinical & Validity Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
</tr>
<tr>
<td>Hs</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>Hy</td>
<td></td>
</tr>
<tr>
<td>Pd</td>
<td></td>
</tr>
<tr>
<td>Pa</td>
<td></td>
</tr>
<tr>
<td>Pe</td>
<td></td>
</tr>
<tr>
<td>Sc</td>
<td></td>
</tr>
<tr>
<td>Ma</td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td></td>
</tr>
<tr>
<td>Neurotic Triad (Hs, D, &amp; Hy)</td>
<td>$R$</td>
</tr>
<tr>
<td>Psychotic Triad (Pa, Sc, &amp; Ma)</td>
<td>$R$</td>
</tr>
<tr>
<td>Psychotic Tetrad (Pa, Pt, Sc, &amp; Ma)</td>
<td>$R$</td>
</tr>
<tr>
<td>Nine Clinical Scales &amp; F-scale</td>
<td>$R$</td>
</tr>
<tr>
<td>All 12 MMPI Scales</td>
<td>$R$</td>
</tr>
</tbody>
</table>

$r$ = Pearson product moment correlation  
$R$ = Multiple correlation

Figure 1. Experimental design of study

The MMPI individual scale scores in the design are all T-scores. The Total T-score variable consists of the sum of a subject's T-Scores on all of the MMPI standard clinical scales except for Scale Mf (nine scales).
scales). Scale Mf was not used in any of the analyses because it is more a measure of interests and passivity rather than true psychopathology. The Total Raw Score variable is the total of a subject's raw scores on the nine clinical scales. The Scatter Index is the sum of the absolute values of the T-score deviations from \( T = 50 \) for the nine clinical scales.

For testing the hypotheses, a \( p < .05 \) level of probability was determined as the level required for rejection of the null hypotheses, and, thus, acceptance of the research hypotheses.

Subjects

The outpatient clients used as subjects in the study were 60 adult outpatients at the Ingham Community Mental Health Center in Lansing, Michigan. This comprehensive CMHC serves all of Ingham County, which includes the cities of Lansing and East Lansing, as well as other suburban and rural areas. The entire population of Ingham County is approximately 200,000. The Ingham CMHC serves a total of nearly 6,000 clients annually, who come from a mixture of different socio-economic backgrounds. The representativeness of the population served by the Tri-County Community Mental Health system, of which the Ingham CMHC is the largest agency, is demonstrated by the fact that this CMH board was chosen in 1977 by The Urban Institute, a Washington-based private research corporation, as one of two nationwide representative locations to pilot test a client outcome monitoring approach (Schainblatt & Hatry, 1979).

The clients in the study were representative of mild to moderately disturbed, acute, non-aftercare outpatients typically seen in a public
community mental health center in a medium-sized city. They were of mixed diagnoses and ages. The clients came from a variety of socio-economic backgrounds, ranging from upper-middle to lower classes.

All consecutive new admissions to the Adult Outpatient Unit for a period of 7 weeks were contacted and asked to participate in the study. This ensured a sample representative of the population of clients seen at the center. A new admission was defined as a client for whom a case was to be opened or re-opened. Clients were telephoned by the investigator from 1 to 3 days before their scheduled intake appointment and asked to participate. Ten to 15 percent of the clients could not be reached. Of those contacted, most agreed to participate in the study, but some did not show up for their intake appointments. Approximately 70% of those contacted eventually participated and were used as subjects. Subjects were required to be at least 18 years old and to have a minimum sixth-grade education, in order to insure sufficient reading ability to complete the MMPI.

It must be emphasized that the clients were in nearly all cases not aftercare or chronic clients. Aftercare clients who had formerly been in state hospitals were seen in a different location by a separate outpatient agency of the Tri-County Community Mental Health system. The only previously hospitalized outpatient clients seen at the center where the study was conducted were those who were referred from a small, short-term (2 week average stay), inpatient psychiatric unit located in a general hospital adjacent to the community mental health center. The sample included only two known individuals referred from this acute inpatient unit. Nearly all of the subjects in the study were first-time
admissions to the community mental health center. A count of the num-
ber of re-opened cases in the subject group was not kept, but a retro-
spective review indicated that only 5-10 of the clients had been pre-
viously treated at the center - most of them several years before their
current admission.

The diagnoses of the subjects were similar to those of a random sam-
ple of 1,206 adult outpatient clients admitted to the Ingham CMHC during
1981 (Coon, 1981). This study indicated the following breakdown of diag-
nostic categories: Adjustment Reactions (36%), Neurotic Disorders (21%),
Personality Disorders (19%), Acute Reaction to Stress (4%), Affective
Psychoses, Schizophrenic Disorders, and Paranoid States (4%), Distur-
bances of Conduct (3%), Sexual Disorders and Deviations (3%), other mis-
cellaneous diagnoses (10%).

The present sample was considered to be representative of a broad
range of different types of outpatient clients. Approximately one-half
of the clients were employed or attending school full-time, and one-half
were not employed. Several of the clients had some college education
or were college graduates, and several clients were living on public
assistance. The sample included at least one individual from each of
the following groups: factory worker, store clerk, psychiatric aide,
paramedic, repairman, secretary, homemaker, single-parent mother living
on public assistance, recently laid off and long-term unemployed fathers,
Social Security and SSI recepients, law student, undergraduate college
student, lower and mid-level office employees (both government and pri-
ivate business), and small business owner.
The mean age of the sample of 60 clients was 33.7 (SD = 11.7). The range of ages was 19 to 81 years old. Age was the only variable in the study which was significantly skewed, because ages were clustered in the 19-38 year old group. Because age was found to have a correlation with GAS of only -.144 (n.s.) in the study, it was not regarded as a variable that would cause any bias.

The outpatient sample included 19 males (32%) and 41 females (68%). This distribution is more heavily weighted for females than the 60% female-40% male distribution most commonly reported for outpatient populations. However, t-tests between males and females on all of the 18 variables analyzed in the study indicated no significant differences. Consequently, the higher proportion of females in the study was not regarded to be a biasing factor.

Raters

Raters used in the study were regularly employed therapists in the Adult Outpatient Unit at Ingham CMHC. Twelve raters were used, and each rated her or his own clients as part of the normal intake procedure of the agency. The range of cases rated by each therapist was 1-12, with a mean number of 5 cases rated by each therapist.

Nine of the raters possessed a master's degree in a mental health field (Clinical Psychology, Counseling Psychology, or Social Work) and three raters had a doctoral degree in either Clinical or Counseling Psychology. Three of the master's level therapists had completed all requirements for a doctoral degree in psychology except for the dissertation. The mean number of years of post-Master's degree work experience

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
in mental health for the raters was 10.8 years, with a lower limit of 6 years. The mean number of years which the therapist-raters had been employed by the Tri-County Community Mental Health Board was 8.5 years, with a lower limit of 3 years. These statistics support the assumption that the raters were all experienced mental health professionals who were familiar with the client population seen at Ingham CMHC.

All raters were required to have attended a training session in the use of the GAS. This requirement was made in order to ensure that an adequate standard of reliability was being followed in the assignment of ratings. As has been noted in the review of the GAS's reliability, training has a positive effect on reliability levels. All but one of the 12 raters had attended a training session sponsored by the Michigan Department of Mental Health when the GAS was first introduced at Ingham CMHC in 1979. These 11 raters had been using the GAS regularly with their client loads over the previous 3 1/2 years, so they were experienced in use of the scale. One rater began employment after the GAS was introduced, so he was trained individually in the use of the GAS during orientation. He had 3 years experience using the GAS.

Procedure

When clients were initially contacted on the telephone by the investigator to be asked to participate in the study, they were presented with the information contained on the sheet, "Information on Client Survey" (see Appendix C). The points regarding voluntary participation and confidentiality were carefully explained. Clients were asked to come in one hour early or to stay one hour late after their first appointment to take the MMPI.
When a client came in for his or her intake (first) appointment, he/she was administered the MMPI by the investigator. In a few cases, clients were unable to take the MMPI on the same day as the intake appointment, but were able to return within the next few days to complete it. In no cases was the MMPI administered more than five days after the intake appointment.

The rights of human subjects were safeguarded throughout the entire experimental procedure. Before taking the MMPI, clients were asked to read the "Information on Client Survey" sheet and were given an opportunity to ask questions. Immediately following this, they were asked to read and sign the "Consent Form for Client Survey" (see Appendix C). The information sheet and the consent form contained the six basic elements of informed consent as outlined in the research standards of the National Institutes of Health, U.S. Department of Health, Education, and Welfare. The consent forms, as well as other human subject considerations of the research, were reviewed and approved by the Research Committee of Ingham CMHC and the Human Subjects Review Board of Western Michigan University.

Clients were given a GAS rating by their therapist after their one-hour intake appointment. This procedure is followed as part of the normal process of opening a case in the agency. GAS ratings were retrieved by the investigator, and only after a client's GAS rating was recorded was the client's MMPI profile given to the therapist for his/her information. Therapists were not allowed to see a client's scored MMPI profile before turning in the GAS score. Therefore, the raters were blind in making the GAS ratings, and the criterion in the study, the MMPI, was independent of the assignment of GAS scores.
The investigator had no further contact with the clients after the MMPI administration. Clients were told that if they wished to receive feedback on their MMPI results, they should discuss this question with their therapist. Extent and timing of relating MMPI results to clients was left up to the judgment of each therapist.

The MMPI's were hand-scored by either the investigator or an agency secretary who was an experienced scorer. MMPI raw scores were converted to T-scores on a personal computer, using the T-score conversion tables from the MMPI Manual (Hathaway & McKinley, 1967). The MMPI T-scores and the GAS scores were then coded onto computer sense-scan sheets, from which the data were read and transferred to the Western Michigan University DEC-10 computer. All data were analyzed on the Western Michigan University computer, using programs from Statpack Statistical Package (Houchard, 1974).

Statistics used in analyzing the data were Pearson product moment correlations, multiple correlations, multiple regression analyses, factor analysis, t-tests, and evaluation of significance of differences between correlations.

Summary

It was hypothesized that the GAS is significantly correlated with four measures of overall severity of disturbance from the MMPI, as well as with the individual MMPI scales, and the MMPI neurotic and psychotic groups of scales. The MMPI was administered to 60 representative community mental health outpatients during their intake visit. Twelve therapists who were experienced in using the GAS rated their clients after a
1-hour intake session. Various correlations were calculated between the MMPI and GAS data.
CHAPTER IV

RESULTS

GAS Scores

The mean GAS score for the 60 outpatients in the study was 55.7 (SD = 10.6). Scores ranged from a low of 32 to a high of 81. The distribution of scores closely approximated a normal curve, as is illustrated by the frequency distribution of scores in Figure 2.

**** BAR GRAPH FOR VARIABLE: GAS  ****

<table>
<thead>
<tr>
<th>RANGE OF VALUES</th>
<th>FREQ</th>
<th>PCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.00 - 32.00</td>
<td>1</td>
<td>1.7</td>
</tr>
<tr>
<td>33.00 - 36.00</td>
<td>2</td>
<td>3.3</td>
</tr>
<tr>
<td>37.00 - 40.00</td>
<td>1</td>
<td>1.7</td>
</tr>
<tr>
<td>41.00 - 44.00</td>
<td>4</td>
<td>6.7</td>
</tr>
<tr>
<td>45.00 - 48.00</td>
<td>4</td>
<td>6.7</td>
</tr>
<tr>
<td>49.00 - 52.00</td>
<td>13</td>
<td>21.7</td>
</tr>
<tr>
<td>53.00 - 56.00</td>
<td>8</td>
<td>13.3</td>
</tr>
<tr>
<td>57.00 - 60.00</td>
<td>11</td>
<td>18.3</td>
</tr>
<tr>
<td>61.00 - 64.00</td>
<td>5</td>
<td>8.3</td>
</tr>
<tr>
<td>65.00 - 68.00</td>
<td>5</td>
<td>8.3</td>
</tr>
<tr>
<td>69.00 - 72.00</td>
<td>2</td>
<td>3.3</td>
</tr>
<tr>
<td>73.00 - 76.00</td>
<td>1</td>
<td>1.7</td>
</tr>
<tr>
<td>77.00 - 80.00</td>
<td>2</td>
<td>3.3</td>
</tr>
<tr>
<td>81.00 - 84.00</td>
<td>1</td>
<td>1.7</td>
</tr>
</tbody>
</table>

---

60 10.0 20.0 30.0 40.0

PERCENTAGE

Figure 2. Frequency distribution of GAS scores.

76
To determine whether the sample was representative of the overall population of outpatients seen at Ingham Community Mental Health Center, the study's sample was compared to three previous samples of clients from the Adult Outpatient Unit. These groups are listed in Table 5. T-Tests between means of the dissertation group and each of the three other groups were not significant. Results indicate that the present sample is similar to other outpatient groups from the same agency, both in average GAS score and in variability.

The GAS mean and standard deviation of the study's sample were also compared to the outpatient samples from the literature in Table 4. Results indicate that data of the present sample are also typical of samples from other similar settings. For eight acute outpatient groups in the literature, including ratings made both at admission and during treatment, the mean GAS score is 56.2 and the average standard deviation is 11.2. These data indicate that scores of the present sample are not restricted in variability as compared to other samples reported in the literature.

GAS means of males and females in the present study showed little difference. The mean for the 19 males was 53.47; mean for the 41 females was 56.76. This difference was not significant, $t(58) = 1.12$, n.s.

Reliability of GAS Ratings

An adequate level of inter-rater reliability was assumed for the 12 raters in the study. As was noted in the review of the GAS's reliability in Chapter I, the GAS has demonstrated good inter-rater
TABLE 5

GAS Scores for Ingham CMHC Outpatient Samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissertation</td>
<td>60</td>
<td>55.72</td>
<td>10.62</td>
<td></td>
</tr>
<tr>
<td>Kriauciunas (1981)</td>
<td>100</td>
<td>55.99</td>
<td>a</td>
<td>.197, n.s.</td>
</tr>
<tr>
<td>Kriauciunas (1981)</td>
<td>100</td>
<td>55.00</td>
<td>a</td>
<td>.525, n.s.</td>
</tr>
</tbody>
</table>

*a Not available

Reliability with different types of raters and with different patient groups. Reliability has been shown to be best with raters who are experienced professionals and who are trained in the use of the GAS. Both of these characteristics were descriptive of the raters in the present study, as was explained in Chapter III. It was therefore not believed necessary to test reliability levels prior to the study.

The data reported by Harley (1983) from a large group of outpatient clients at Ingham CMHC include 668 cases who were rated by eight of the raters used in the present study. All of these cases were admitted to Ingham CMHC before the present study was conducted. An analysis of variance between mean GAS scores of the eight raters is significant, \( F (7, 660) = 10.85, p < .001 \). However, this cannot be used to conclude

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
that there is bias or unreliability in the ratings, since assignment of clients to therapists by the intake office is not random. More severely disturbed clients are often deliberately assigned to particular therapists. The significant analysis of variance, therefore, probably indicates a difference between the client loads of individual therapists rather than a lack of reliability. Interestingly, the mean of this group of 668 cases was 55.75 (SD = 10.93), almost identical to the mean and standard deviation of the present study.

**MMPI Scores**

The 13 standard scales of the MMPI were scored, as well as three MMPI composite indexes, Total T-Score (TotT), Total Raw Score (RS), and Scatter Index (Scin). These composite indexes, which have been described in Chapter III, were used as the MMPI measures of overall severity along with the F-scale T-score. The individual scales and composite indexes each had an approximate normal distribution, as indicated by a coefficient of skewness of less than .90 for all variables. A frequency distribution of the MMPI total raw scores is shown in Figure 3.

The means and standard deviations of the MMPI scales are reported in Table 6. The average T-score for the nine clinical scales used was 66.0, and the average standard deviation was a T-score of 12.4. Scale Mf is shown in Table 6, but it was not used in any of the analyses, as has been previously explained. Mf was found to have little relationship to GAS (r = .096).

The standard deviations of the MMPI scores indicate additional information about the variability of the sample. MMPI T-scores are standard
**** BAR GRAPH FOR VARIABLE: RS  ****

<table>
<thead>
<tr>
<th>RANGE OF VALUES</th>
<th>FREQ</th>
<th>PCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.00 - 110.2</td>
<td>2</td>
<td>3.3 IXX</td>
</tr>
<tr>
<td>110.2 - 124.4</td>
<td>1</td>
<td>1.7 IXX</td>
</tr>
<tr>
<td>124.4 - 138.6</td>
<td>4</td>
<td>6.7 IXXXXX</td>
</tr>
<tr>
<td>138.6 - 152.8</td>
<td>3</td>
<td>5.0 IXXXX</td>
</tr>
<tr>
<td>152.8 - 167.0</td>
<td>4</td>
<td>6.7 IXXXXX</td>
</tr>
<tr>
<td>167.0 - 181.2</td>
<td>5</td>
<td>8.3 IXXXXX</td>
</tr>
<tr>
<td>181.2 - 195.4</td>
<td>9</td>
<td>15.0 IXXXXXXXXXX</td>
</tr>
<tr>
<td>195.4 - 209.6</td>
<td>6</td>
<td>10.0 IXXXXX</td>
</tr>
<tr>
<td>209.6 - 223.8</td>
<td>5</td>
<td>8.3 IXXXXX</td>
</tr>
<tr>
<td>223.8 - 238.0</td>
<td>3</td>
<td>5.0 IXXXX</td>
</tr>
<tr>
<td>238.0 - 252.2</td>
<td>8</td>
<td>13.3 IXXXXXXXXXX</td>
</tr>
<tr>
<td>252.2 - 266.4</td>
<td>5</td>
<td>8.3 IXXXXX</td>
</tr>
<tr>
<td>266.4 - 280.6</td>
<td>3</td>
<td>5.0 IXXXX</td>
</tr>
<tr>
<td>280.6 - 294.8</td>
<td>1</td>
<td>1.7 IXX</td>
</tr>
<tr>
<td>294.8 - 309.0</td>
<td>1</td>
<td>1.7 IXX</td>
</tr>
</tbody>
</table>

---

60

5.0 10.0 15.0 20.0
PERCENTAGE

Figure 3. Frequency distribution of MMPI total raw scores.

scores which were normed on a population of normal adults when the MMPI was introduced. By definition, a T-score has a mean of 50 and a standard deviation of 10. Since a standard deviation of 10 was the degree of variability of the MMPI norm group, the dissertation sample, with an average standard deviation of 12.4, has greater variability than found in a normal population. Eight of the nine MMPI clinical scales have standard deviations of greater than 10. These data confirm the earlier impression, based on GAS scores, that the sample was in fact heterogeneous in level of severity of disturbance and was not restricted in variability.
Mean scores of males and females were compared on all of the MMPI scales, and no significant differences were found. The total T-scores for males and females were nearly identical, 593.6 for females and 594.1 for males.
Testing of Hypotheses

The primary hypothesis of the study was that there is a significant correlation between GAS and each of the four MMPI measures of overall severity of disturbance. These four measures taken from the MMPI were Total T-Score (TotT), Total Raw Score (RS), Scatter Index (Scin), and F-scale Score. Since GAS and the severity measures each had an approximate normal distribution, the assumption of normality of distributions was satisfied. The correlation coefficients between GAS and the MMPI overall severity measures are listed in Table 7.

Results indicate that correlations of GAS with the four severity measures are all significant. Therefore, the null hypothesis is rejected and the study's primary hypothesis is confirmed. The correlations between GAS and Scin, TotT, and F-scale are similar, and are significant at the .05 level of probability. The correlation of GAS with RS is larger and is significant at the .01 level of probability. Correlations between GAS and the overall measures are generally higher than the correlations of GAS with the individual MMPI scales, which are listed later in Table 8. The four severity measure correlations are all among the six highest single variable coefficients in the study.

Although the correlations between GAS and the four overall severity indexes are significant, the question can be raised whether they are meaningful correlations. A statistically significant correlation is not necessarily a meaningful correlation. The meaningfulness of the coefficients can be evaluated by examining the proportion of the variance of the severity indexes which GAS accounts for. Variance
TABLE 7

Correlations between GAS and MMPI Overall Severity Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation Coefficient (r)</th>
<th>Coefficient of Determination (r^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS</td>
<td>-.364**</td>
<td>.132</td>
</tr>
<tr>
<td>Scin</td>
<td>-.302*</td>
<td>.091</td>
</tr>
<tr>
<td>TotT</td>
<td>-.294*</td>
<td>.086</td>
</tr>
<tr>
<td>F Scale</td>
<td>-.294*</td>
<td>.086</td>
</tr>
</tbody>
</table>

df = 58

**p < .01
*p < .05

accounted for is equal to $r^2$, which is the coefficient of determination. The coefficients of determination in Table 7 indicate that the GAS accounts for between 9% and 13% of the variance in the MMPI overall severity indexes. This percentage is not impressive, and it questions whether the significance of the information communicated in GAS scores is worth the effort and expense required to produce them.

It can be noted that the MMPI validity coefficients are slightly higher than the validity coefficients obtained at the time of admission by Endicott et al. (1976) in the original validation study of the GAS (see Table 2). However, the MMPI coefficients are not markedly higher, and do not provide significantly more support for the concurrent validity.
of the GAS than was provided by Endicott et al. The MMPI results essentially confirm results of the original validation study indicating that concurrent validity of the GAS at the time of admission is weak. The present results indicate that concurrent validity of the GAS at admission is not significantly better for outpatients than it is for inpatients.

It has been demonstrated that the variability of the sample was not restricted, so the MMPI validity coefficients cannot be disregarded because of an artifact of the correlation coefficient due to restricted variability. The MMPI sample was similar in variability to other GAS samples reported in the literature, and it had a greater degree of variability in scores than the MMPI norm group.

Secondary Hypotheses

The secondary hypotheses concerned correlations between GAS and the individual MMPI scales, as well as correlations between GAS and groups of MMPI scales. Since each of the individual MMPI scales had an approximate normal distribution, correlations could be calculated.

Hypothesis a

Hypothesis a was that GAS is significantly correlated with each of the nine MMPI clinical scales. This hypothesis was partially confirmed and partially denied, since three clinical scales had significant correlations with GAS and six scales did not. The significant scales were Pd, Sc, and Si. Scales Hs, Pa, D, Pt, Hy, and Ma each had non-significant correlations with GAS. These coefficients are shown in Table 8.
TABLE 8

Correlations between GAS and MMPI Clinical Scales

<table>
<thead>
<tr>
<th>MMPI Scale</th>
<th>Correlation Coefficient (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pd</td>
<td>-.352**</td>
</tr>
<tr>
<td>Sc</td>
<td>-.264*</td>
</tr>
<tr>
<td>Si</td>
<td>-.254*</td>
</tr>
<tr>
<td>Hs</td>
<td>-.214</td>
</tr>
<tr>
<td>Pa</td>
<td>-.182</td>
</tr>
<tr>
<td>D</td>
<td>-.163</td>
</tr>
<tr>
<td>Pt</td>
<td>-.162</td>
</tr>
<tr>
<td>Hy</td>
<td>-.148</td>
</tr>
<tr>
<td>Ma</td>
<td>-.033</td>
</tr>
</tbody>
</table>

df = 58
**p < .01
*p < .05

It is interesting that the Pd scale (Psychopathic Deviancy) had the highest correlation with GAS. The Pd scale also had the highest mean of all the MMPI scales (see Table 6). This scale is usually considered to indicate anger, rebelliousness, dislike of authority and regulations, lack of social conformity, and a tendency to get into scrapes (Dahlstrom, Welsh, & Dahlstrom, 1972; Webb, McNamara, & Rodgers, 1981). The feelings
of high Pd clients are frequently expressed in acting out behavior, such as unconventional dress/appearance and speech, heavy drinking or drug use, and poor work and marital adjustment. One thing that distinguishes the Pd scale from the other scales is that the acting out characteristics are frequently visible, and are often expressed in specific behaviors which a therapist would see or question a client about in an initial interview. In contrast to this, personality characteristics indicated by some of the other MMPI scales are underlying traits which would not be as easily discerned in an initial interview. Perhaps this visibility of the Pd symptoms accounts for why the Pd scale displayed the highest relationship to GAS ratings at admission.

In relation to the other MMPI scales, the Pd scale stands by itself. It is not included in the neurotic or psychotic groups of scales later analyzed, and it did not show up in the major factors of the factor analysis performed. The Pd scale does, though, have a higher correlation with GAS than three of the four overall severity indexes and has almost as much common variance with GAS as the highest composite index, Total Raw Score.

The fact that the Sc scale (Schizophrenia) was significantly correlated with GAS was not surprising. This result confirms previous research indicating that the GAS is related to symptoms of unusual thought content, disorganized thinking, and poor reality testing - delusions. All of these characteristics can be included within the scope of the Sc scale. The Sc scale can also indicate characteristics of social alienation, isolation, and general dissatisfaction (Webb et al., 1981), which are related symptoms that may be more commonly seen among outpatients.
What is notable about the significant correlation between GAS and Sc is it indicates that the relationship of GAS to psychotic symptoms is true not only among inpatients that are mostly schizophrenic, but it is also true among acute outpatients. This finding suggests that previous results identifying a GAS-psychotic relationship were not caused solely by the schizophrenic samples studied, but rather that the relationship is a characteristic of the GAS. The GAS has this characteristic probably because it contains a number of specific and well-elaborated examples in the psychotic ranges of the scale, as was discussed in the review of the GAS's content validity.

The third individual MMPI scale that displayed a significant correlation with GAS was the Si scale, which indicates a tendency to be uncomfortable in social situations and introverted. High scorers on Si may have schizoid tendencies and be withdrawn (Webb et al., 1981). The significant relationship between the Si scale and GAS is probably more due to the fact that low Si clients, who are generally sociable and outgoing, tended to be rated healthy, rather than the fact that high Si clients were seen as being very disturbed. The GAS's content places considerable importance on interpersonal relationships ("has meaningful interpersonal relationships," "few friends," "impairment in family relations," "avoids friends"), and this probably also accounts for the significant correlation between GAS and Si.

The lack of significant correlations between GAS and the other individual MMPI scales confirms previous findings that GAS scores have little relationship to neurotic symptoms. Correlations with two MMPI "neurotic" scales, Hy (Hysteria) and D (Depression) were both low, as
was the correlation with Pt (Psychasthenia), which indicates ruminative thinking and worrying. The coefficient between GAS and Hs (Hypochondriasis) was slightly higher, but also non-significant. These low correlations may be due to the fact that the GAS has fewer behavioral examples and is less well elaborated in the upper half of the scale, which is where most neurotic disorders fall. The lack of specificity and the lack of benchmarks on the scale for symptoms characteristic of neurotic syndromes may be the reasons why the variances of neurotic scales and the GAS have so little overlap.

Two remaining scales, Pa (Paranoia) and Ma (Mania), also had non-significant correlations with GAS. The GAS includes statements directly relevant to paranoid ideation and behavior only in the severely disturbed ranges below 30. Perhaps this is why paranoid symptoms in an outpatient sample showed little relationship to GAS scores. Another reason for the low correlation with Pa may be that mild paranoid symptoms in a client are not always perceivable in a first session. By the fact of being paranoid, a client is guarded and defensive, and it usually takes a therapist a few sessions to see through this.

There is no obvious reason why the correlation of GAS with the Ma scale was so low. A possibility is that mildly manic individuals do not usually come in voluntarily for treatment, and when they do, their high energy level may initially be seen as adaptive, rather than disabling. The question remains why clients with an abnormally low energy level in the outpatient sample did not tend to be seen as disturbed. This is again probably due to the fact that the GAS's content does not include statements relevant to this syndrome in the mildly disturbed ranges.
It must be emphasized that even though some MMPI scales displayed statistically significant correlations and larger correlations than other scales, the magnitude of none of the individual correlations was very high. The distinction between a statistically significant correlation and a meaningful correlation which was made for the overall severity indexes applies to the individual correlations as well. GAS scores accounted for no more than 13% of the variance in any individual MMPI scales, which is not a sufficient basis on which to make a decision of any importance.

**Hypotheses b, c, & d**

Secondary hypotheses b and c predicted significant multiple correlations between GAS and each of three groups of MMPI clinical scales, the neurotic triad, the psychotic triad, and the psychotic tetrad. Hypothesis d predicted that correlations with the two psychotic groups would be greater than the correlation with the neurotic group. The multiple correlations of GAS with these scales are listed in Table 9.

The multiple correlations are interesting, but they add little additional information than has already been given by correlations of GAS with the individual MMPI scales. The correlation of the neurotic triad is significant at the .05 level of probability, thereby confirming hypothesis b. This multiple coefficient of scales Hs, D, and Hy, however, is barely more than the coefficient for scale Hs alone (an increase from -.214 to -.226). This small difference is because of the considerable intercorrelation and common variance between the three scales, particularly between Hs and Hy. Although the correlation of GAS with
the neurotic triad is statistically significant, it reinforces previous impressions that GAS scores are not highly related to neurotic symptoms.

The multiple correlations of GAS with the psychotic triad and psychotic tetrad, as shown in Table 9, also add little new information. The multiple coefficient of Scales Pa, Sc, and Ma is again barely more than for the highest individual correlation alone, Scale Sc (an increase from -.264 to -.269). This is because of the high intercorrelation between Sc and Pa, and because Ma had virtually no relationship to GAS. Adding Scale Pt to this group does raise the multiple correlation somewhat, to .304.

The correlations of both psychotic groups of scales with GAS were statistically significant, confirming hypothesis c. This result gives additional support to previous conclusions that GAS scores are related to psychotic symptoms. The correlations of GAS with the two psychotic groups of scales, though, were not significantly greater than the correlation of GAS with the neurotic triad, thereby rejecting hypothesis d ($z = .256, n.s.; z = .459, n.s.$). Although the correlations of the psychotic groups were numerically greater, as was predicted, the difference was not large enough to be significant.

**Hypothesis e**

Hypothesis e stated that GAS is significantly related to the two remaining validity scales, scales L and K. This hypothesis was inserted for exploratory purposes. The results, however, turned out to be a surprise and are among the most significant findings of the study. It was discovered that the K scale had the highest correlation with GAS of any
TABLE 9

Correlations between GAS and Groups of MMPI Clinical Scales

<table>
<thead>
<tr>
<th>Scale Group</th>
<th>Multiple Correlation Coefficient ((R))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurotic Triad</td>
<td>-.226*</td>
</tr>
<tr>
<td>(Hs, D, Hy)</td>
<td></td>
</tr>
<tr>
<td>Psychotic Triad</td>
<td>-.269*</td>
</tr>
<tr>
<td>(Pa, Sc, Ma)</td>
<td></td>
</tr>
<tr>
<td>Psychotic Tetrad</td>
<td>-.304**</td>
</tr>
<tr>
<td>(Pa, Pt, Sc, &amp; Ma)</td>
<td></td>
</tr>
</tbody>
</table>

\(df = 58\)

**\(p < .01\)

*\(p < .05\)

of the variables in the study. The \(L\) scale was found to have the fourth highest coefficient of the 12 MMPI scales. These results are shown in Table 10.

The MMPI \(K\) and \(L\) scales both give an indication of the test subject's desire to present him or herself in a good light. In the extreme high ranges, they indicate a tendency to "fake good", or to slant test responses in such a way as to make the subject appear to look better than is actually the case. In the lower, more common ranges, they are an indication of the degree of defensiveness and guardedness of the test subject.
TABLE 10

Correlations between GAS and Two MMPI Validity Scales

<table>
<thead>
<tr>
<th>MMPI Scale</th>
<th>Correlation Coefficient (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>+.417**</td>
</tr>
<tr>
<td>L</td>
<td>+.280*</td>
</tr>
</tbody>
</table>

\[ df = 58 \]
\[ **p < .01 \]
\[ *p < .05 \]

The difference in the interpretations given to scales K and L is in the degree of subtlety of the client's desire to make him or herself look good. The L scale indicates a very naive, simple attempt to appear virtuous, while the K scale indicates a much more subtle and sophisticated attempt to cover up one's weaknesses (Dahlstrom et al., 1972). In clinical practice, the L scale reflects defensiveness usually only among individuals of limited education, intelligence, or cultural background. The K scale, on the other hand, is the scale which is used as a measure of defensiveness for most clients who are of adequate sociocultural backgrounds. In an outpatient population such as the one in the present study, the K scale is the best indicator of the defensiveness and guardedness of a test subject.

For outpatient clients, a low K score indicates an individual with some lack of normal defensiveness and low ego strength. A score in the
middle range indicates appropriate defensiveness for middle and upper class individuals, as expressed by a balance between self-protection and self-disclosure. A high $K$ score is indicative of a moderate to highly defensive individual who is rigid and makes a consistent effort to maintain an appearance of complete adequacy, self-control, and effectiveness.

What is interesting about the results in Table 10 is that GAS scores displayed more common variance with the $K$ scale than with any other variables in the study, including the four overall severity indexes. Of all the MMPI measures, GAS ratings were most related to the degree of defensiveness and guardedness of clients. When the raters judged the overall severity of disturbance of clients, they responded more to the degree of the client's defensiveness than they did to the client's psychiatric symptoms. Clients who were guarded and defensive tended to be rated as being more healthy, and clients who were less defensive tended to be rated as being more severely disturbed, often irrespective of the actual amount of psychopathology which was present. This suggests that GAS ratings at the time of admission in an outpatient population reflect more the degree of defensiveness of clients than they do the clients' true level of psychopathology.

The preceding conclusion has implications for the question of whether an admission GAS rating or a rating done later in treatment can be expected to be more valid, since defensiveness is usually high at the time of admission and decreases as familiarity develops. That question will be discussed in Chapter V.
Additional statistical analyses were done to further evaluate the relationship of the K and L scales to GAS. Both the factor analysis and the multiple regression analyses, which are reported in the next section, further emphasize the fact that these two scales display a stronger relationship to GAS than the MMPI clinical scales.

Additional Statistical Analyses

A factor analysis was calculated to obtain additional information concerning which variables GAS shared common variance with. Results of the factor analysis are shown in Table II.

The MMPI "performed" as expected in the factor analysis, as seen in the four factors which were produced. The "psychotic" scales clustered together, as did two different types of "neurotic" scales and the defensiveness variable. None of the three symptom factors loaded with GAS to a significant degree (highest loading was -.234 with Factor 1), but the Guardedness-Defensiveness factor loaded .773 with GAS. The results indicate that in terms of common variance the GAS was not closely related to clusters of scales measuring psychiatric symptoms, but the GAS was highly related to a cluster of variables measuring guardedness and defensiveness. These data provide additional support for the conclusions made in the discussion of hypothesis e.

Multiple Regression Analyses

Since scales K and L were significantly related to GAS, it was thought worthwhile to compare the multiple correlation of these two scales with GAS to the correlation of the two best predictor remaining
TABLE 11

Factor Analysis of GAS and 12 MMPI Scales

<table>
<thead>
<tr>
<th>Factor</th>
<th>Loading with GAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1 - Hy, Hs (Neurosis: repression, somatization)</td>
<td>-.234</td>
</tr>
<tr>
<td>Factor 2 - F, Pa Sc, Ma (Psychosis, deviancy)</td>
<td>.082</td>
</tr>
<tr>
<td>Factor 3 - Si, D, Pt (Neurosis: withdrawal, depression, anxiety)</td>
<td>.136</td>
</tr>
<tr>
<td>Factor 4 - L, K (Guardedness, defensiveness)</td>
<td>.773</td>
</tr>
</tbody>
</table>

scales with GAS. This allowed a comparison between the strength of relationship of scales $K$ and $L$ to that of the two best clinical scales. In order to find the two best clinical scales, a stepwise multiple regression analysis was done on the remaining 10 MMPI scales with GAS as the dependent variable. The pair of best predictor scales was $Pd$ and $Si$, which had a multiple correlation with GAS of .407. In comparison to this, scales $K$ and $L$ had a higher multiple correlation, .448. These data agree with results of the factor analysis, which show GAS to be more highly related to scales $K$ and $L$ than to the clinical scales.

The stepwise multiple regression analysis identified that the nine clinical scales and F-scale together accounted for 29.3% of the variance.
in GAS scores. Two additional multiple regression analyses were performed to determine the amount of variance in GAS scores which could be accounted for by the nine clinical scales as a group and by all 12 of the MMPI scales together. The results of these analyses are shown in Table 12. The data indicate that all 12 MMPI scales together account for 39.8% of the variance in GAS scores.

**Dissimulation Test**

A criticism that is frequently made of the use of self-report measures in research is that some subjects do not respond honestly. The argument is that some subjects overexaggerate their ills in an attempt to gain attention or self-pity, while other subjects overly deny problems because of being too guarded. As a consequence, some researchers argue, self-report measures lack accuracy.

In order to defend this study's use of the MMPI, a self-report measure, against this criticism, an additional set of analyses was performed. The MMPI is one of the few tests which contains built-in validity scales, which allow a researcher to screen out test profiles which could be categorized as "fake good" or "fake bad." This is usually accomplished by setting cutoff scores on the F-K (Dissimulation) Index (Gough, 1950).

Using cutoff scores of ± 14 or greater on the MMPI Dissimulation Index, which are fairly strict limits, the 60 original subjects were screened. Eight subjects who did not meet the cutoffs were removed, leaving 52 subjects. The correlations between GAS and the MMPI scales were then re-run on this smaller group of subjects. Results indicated
TABLE 12

Multiple Regression Analyses on GAS

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Multiple Correlation Coefficient ($R$)</th>
<th>Coefficient of Determination ($R^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nine clinical scales</td>
<td>.512**</td>
<td>.262</td>
</tr>
<tr>
<td>Nine clinical scales</td>
<td>.541**</td>
<td>.293</td>
</tr>
<tr>
<td>&amp; F scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All 12 MMPI scales</td>
<td>.631**</td>
<td>.398</td>
</tr>
</tbody>
</table>

$df = 58$

* $P < .05$
** $P < .01$

that all of the correlations with GAS went down, except for one scale, $Hy$, which increased slightly. This demonstrates that when cutoffs were used to include only subjects who could definitely be judged as responding honestly, the GAS displayed even weaker relationships with the MMPI variables.
CHAPTER V

DISCUSSION

Summary of Study

The study began by asking the question whether the Global Assessment Scale (GAS) is a valid measure of overall severity of psychiatric disturbance. It was pointed out that in spite of widespread usage of the GAS, little has been known about the scale's validity. No published reviews of the GAS were found in the literature. Therefore, it was necessary to do a comprehensive review of studies citing reliability and validity data for the GAS, and material describing use of the GAS in mental health program evaluation and in psychiatric research. The literature review itself provided considerable information related to the question of the GAS's validity.

The empirical portion of the study involved testing the concurrent validity of the GAS by comparing clients' GAS scores to several MMPI variables. Both GAS scores and MMPI profiles were obtained from 60 community mental health outpatients at the time of their admission to treatment. Correlations were calculated between GAS scores and four MMPI overall severity indexes, as well as between GAS and the individual MMPI scales. Results of correlations between GAS and the severity indexes ranged from -.29 to -.36, indicating only 9% to 13% common variance between the measures. Correlations between GAS and the individual MMPI clinical scales were also under .36, and reinforced impressions from the literature that GAS scores are more related to psychotic symptoms.

98

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
than to neurotic ones. Correlations between GAS and the MMPI validity scales surprisingly indicated that GAS scores were more related to measures of guardedness and defensiveness than they were to the overall severity indexes or the individual clinical scales. This finding was supported by a factor analysis, which showed GAS clustering with the defensiveness variables. A multiple regression analysis indicated that all 12 MMPI scales correlated .63 with the GAS.

Conclusions of Study

The question of whether the GAS is a valid measure of severity of disturbance, which was posed in the Introduction, will now be returned to in light of the study's findings. What can be concluded about the validity of the GAS as a result of this study? As noted in the APA Standards for Educational and Psychological Tests (1974), conclusions made about the validity of a test instrument should ideally refer to its validity for certain uses or for particular types of decisions:

Any study of test validity is pertinent to only a few of the possible uses of or inferences from the test scores. It is incorrect to use the unqualified phrase "the validity of the test." No test is valid for all purposes or in all situations or for all groups of individuals. (p. 31)

Therefore, statements made about the GAS's validity from the results of the present study will be directed toward the different types of usage which can be given to GAS scores.

A problem with tying a discussion of the GAS's validity to the usage given the scale is that its use in some program evaluation settings is not explicit. As was noted in Chapter II, the lack of published
studies from state mental health departments creates some ambiguity about the uses of GAS scores in state evaluation systems. Some examples of decisions made on the basis of the GAS by the Michigan Department of Mental Health, however, were discussed, and several other examples of uses given to the GAS in clinical research settings were seen in the literature review.

As was stated, results of the study indicated that correlations between GAS and the overall severity indexes from the MMPI were low. The highest coefficient was .36, which indicates that the two measures of overall severity of disturbance have only 13% of their variance in common. If the MMPI is taken as an accepted standard of psychopathology, as was assumed at the beginning of the study, the results mean that the GAS has only a 13% overlap with the standard measure. The remaining 87% of the GAS's variance is due to its measurement of some extraneous variable, or it is error variance. By psychometric standards, these statistics are unacceptably low. They suggest that the GAS is not a valid measure of the construct of severity of disturbance. Results of the present study are confirmed by other concurrent validity coefficients reported in the literature, which are even lower.

However, other data which have been reviewed in the literature indicate that GAS mean scores of contrasted groups of clients and normals may have some accuracy. In Chapter II it was seen that the mean GAS scores of groups of different client status (inpatient, outpatient, normal) were clustered together in a defined range, with good distance from the other client status groups (see Table 4). Nearly all of the means were in the range which would be expected for that client status.
These data demonstrate the GAS's ability to distinguish between contrasted groups, which is also a way of indicating validity. The data suggest that the GAS has validity for distinguishing between mean scores of client groups which are known to be different.

Mean score results from the present study reinforce conclusions made from the literature about the apparent validity of GAS mean scores. The mean of the 60 outpatients in the sample was 55.7, very close to the overall mean of 56.1 reported for nine other outpatient groups in the literature. More significant is the fact that the level of disturbance indicated on the GAS by this mean score of 55.7 corresponds to the overall level of disturbance indicated by the hypothetical MMPI profile made from the means of the MMPI scores of the sample (see Appendix D). A GAS score assigned to a hypothetical client having the MMPI profile of the mean MMPI scores of the sample would probably be close to a score of 56.

The conclusion that can be made from this discussion is that the GAS displays poor validity in a concurrent correlation with another psychometric instrument, but yet it displays adequate validity when mean scores of groups are compared. What seems to be happening statistically is that individual scores on the GAS and another instrument do not vary similarly, or along with each other, but yet the discrepancies in common varying cancel each other out and become insignificant when the mean scores of groups are compared. The occurrence of this phenomenon can be used to make different conclusions about the validity of the GAS for different types of usage.
The data from both the literature and the present study suggest that the GAS has acceptable validity for uses where mean scores of groups of clients are compared to each other, or for uses where mean scores of a group are compared across different points in time. However, low concurrent correlations found between GAS and the MMPI in the present study indicate that the GAS has unacceptably low validity for making judgments about individuals on the basis of GAS scores. These low correlations indicate that individual scores can have a large discrepancy between levels of severity of disturbance reported by the GAS and the MMPI, and that there is, then, considerable potential for error if GAS scores are used to make decisions about individuals. It can therefore be concluded that the GAS probably has acceptable validity for those uses where group mean scores are used to make decisions, but it has unacceptable validity for those uses where individual scores are used as a basis for making decisions.

Specific examples of usage of the GAS for both groups and individuals were noted in Chapter II. In mental health programs, GAS scores of groups are used by the Michigan Department of Mental Health in a matrix to picture the level of disturbance of clients served by an agency, a CMH board, or the entire state. For these group categories, the GAS is probably accurate in communicating the desired information and it appears to have adequate validity for such purposes. Other similar legitimate uses of the GAS would include comparing costs of service between different client groups of comparable GAS levels, or determining the proportion of clients served in a state that display significant change, as is done in Utah. In research settings, the validity of the
GAS would seem acceptable for evaluating the change in mean scores of
groups of clients who have received some treatment, or for comparing
mean GAS scores between two experimental groups.

However, for uses of the GAS where decisions are made on the basis
of an individual's GAS score, definite problems are evident because of
the low concurrent correlations between GAS and other measures of sever­
ity of disturbance. For situations where the GAS score of an individual
client is used to make a decision about the client, the GAS's validity
does not appear adequate enough to justify its use. The discrepancies
seen in the study between severity of disturbance levels indicated by
the GAS and those indicated by the MMPI are too large to base significant
decisions about an individual on GAS scores. Data indicate that it is
too likely an individual with a high GAS score could, in fact, be severe­
ly disturbed, and vice versa.

Some examples of questionable use of the GAS to make decisions about
individuals were seen in Chapter II. The Michigan Department of Mental
Health (1981) states that some local, CMH boards use cutoff GAS scores
of individuals to determine eligibility for service or eligibility for
certain programs. It has been proposed in Michigan to use an outpatient's
GAS score to assign him or her a priority level for treatment, and to
use GAS cutoffs as entry and exit criteria for community residential
settings. In Washington, an individual's GAS score was until recently
used to legally define whether he/she was a "seriously disturbed person."
In research studies, individual GAS scores have been used to screen sub­
jects for experimental groups. Results of the present study question
all of these uses of GAS scores. Their further use is discouraged, and
and it is recommended that they be re-evaluated in view of data that indicate the low validity of the GAS for these purposes.

Issues in Interpretation of Results

Four major issues which could affect the interpretation of the study's results will be discussed.

Time Period of Concurrent Administration

The significance of the time period at which the GAS's concurrent validity is measured was discussed at some length in Chapter II. Endicott et al. (1976) reported a significant increase in concurrent validity coefficients of the GAS between a measurement done at the time of admission and a measurement done 6 months after admission. Hargreaves (1983) reported a similar increase in validity coefficients of the HSRS from an admission to a 1-year rating. As was discussed in Chapter II, this increase appears to be due to two factors: the increased variability in GAS scores at a 6-month or 1-year follow-up, and the increased familiarity between rater and client which occurs over time.

The concurrent correlations of the present study were obtained at the time of admission to psychiatric treatment. In a strict sense, their generalization to other situations should be limited to only other assessments done at the time of admission. As has been demonstrated in the literature review, though, most of the GAS's use involves either an admission rating or an admission-discharge difference score, and so conclusions which have been drawn about the GAS's validity can legitimately be generalized to most uses of the GAS. Statistical data describing the
sample of the present study indicate that the sample was not restricted in variability in comparison to other psychiatric patient groups or to a normal population. The low coefficients obtained between the GAS and the MMPI cannot, therefore, be attributed to restricted variability or excessive homogeneity of the sample.

Significant questions still remain concerning what changes occur in the GAS's concurrent validity coefficients in measurements done after the time of admission, and why these changes occur. The present study had originally intended to investigate these questions further by obtaining two samples - one of concurrent ratings done at the time of admission and one of ratings obtained 3 months after admission. However, anticipated difficulties in locating clients at the 3-month point and in obtaining cooperation of both raters and clients caused this plan to be dropped. It is recommended that a future study make this comparison between concurrent correlations of the MMPI and the GAS at admission and at a later time in treatment. In Chapter II the hypothesis was made, based on data of the original validation study of the GAS, that concurrent validity coefficients may increase over time because of the increased familiarity that occurs between therapist and client. It would be helpful for a future study to test this hypothesis.

**Relationship of GAS Scores to Defensiveness**

An issue which is related to the increases observed in validity coefficients over time is the finding of the present study that GAS scores at admission had the greatest relationship to MMPI variables measuring guardedness and defensiveness. GAS ratings in the study...
were more closely related to the degree of defensiveness of clients than they were to the clients' actual psychopathology.

This relationship between GAS and defensiveness provides a possible explanation why GAS admission validity coefficients are low. When clients come to an intake session, they frequently feel apprehensive because of being in a new situation. Their initial posture of defensiveness makes it more difficult for a therapist to do an accurate assessment. The therapist usually has only one hour to do an intake interview, which is not sufficient time to gather a great deal of information about the client. In view of these factors, it is not surprising that admission ratings have low validity. As clients become accustomed to therapy and defensiveness decreases throughout treatment, a therapist will become able to more accurately perceive the client's severity of disturbance and will make a more accurate rating. This may partially account for why GAS ratings made later in treatment have been found to be more valid.

The close relationship between GAS and defensiveness in the present study, and the probable error variance which this factor introduced into the ratings, make it appropriate to recommend that future GAS ratings be made in the second or subsequent client contact whenever possible. Results suggest that if GAS ratings are made later in treatment, they will likely have a higher degree of validity.

**Question of Generalization**

Another issue relevant to the interpretation of the study's results is the question of generalization. To what extent can results from the study be generalized to other settings?
In Chapter III it was stated that the community mental health center where the study was conducted was representative of other mental health centers in medium-sized cities. Reasons for this were that the demographic characteristics of the clients served by Ingham CMHC are diverse, and the types of services offered by the center are typical of most comprehensive CMHCs. Traditional test theory states that results of the present study could be generalized to other similar settings, and that equivalent results would be obtained in another setting to the degree that all significant variables were similar. Results of the study would not be considered likely to vary significantly in most other community mental health settings.

One possible unusual characteristic of the setting of the present study was the above average educational level of the staff used as raters (one-half were doctoral or all but dissertation level therapists). In other centers where staff are not as highly trained, validity coefficients could be lower. Another relevant factor is the observation made in Chapter II that GAS reliability coefficients are sometimes lower in community settings than they are in research-oriented university settings. If that the same difference occurs in validity coefficients, and it is possible that slightly higher validity coefficients could be obtained in a more research-oriented milieu.

**Soundness of the MMPI as the Validation Criterion**

At the beginning of the study, it was stated that the MMPI was chosen as the criterion measure because of its broad acceptance, wide usage, and extensive research support. The appropriateness of the MMPI
for the criterion measure in the study was assumed. This assumption, however, could be questioned on some points, since the MMPI is not error-free. For the purposes of this study, though, the MMPI was considered to be the best and most practical instrument available to use as the criterion. All of the MMPI data in the study were congruent with expected results based on the theory of the test.

Probably the main criticism which could be made of the MMPI is the reliance on patient self-reports as measures of psychopathology. It cannot be denied that response bias of test subjects can cause some clients to overexaggerate their ills and cause other clients to excessively deny problems. The effect of excessive response bias, however, was eliminated in the dissimulation test, as was described in Chapter IV. Results indicated that when highly exaggerating and highly denying clients were removed from the subject group, most of the correlations between GAS and the MMPI scales decreased, indicating an even weaker relationship.

Another point of disagreement could be made in the interpretation given in the study to the K-scale as a measure of defensiveness. The K-scale has the peculiar characteristic that excessively low scores, indicating a serious deterioration of defenses, are as undesirable as excessively high scores, indicating rigidity. Clinicians sometimes view K as a measure of health and ego-strength. This can be legitimate when comparing a very low to a moderate score. However, in the present study, the K scores were spread out across the entire range of the scale, making such an interpretation for K not plausible. In addition, the fact that the K scale was closely related to the L scale, which indicates more
blatant evasiveness, confirms the defensiveness interpretation which was given to the K scale.

Recommendations

Two recommendations regarding usage of the GAS have already been made in this chapter; they will be repeated. A third recommendation grows out of the analysis of the GAS's content validity in Chapter II.

1. It is recommended that if a revision of the GAS is undertaken, the descriptive paragraphs in the upper half of the scale be expanded to include more specific behavioral examples. The content analysis of the GAS indicated that the range of the scale from 50-100 contained only 7 specific behavioral examples, as compared to 23 behavioral examples in the lower half of the scale. This difference is one possible reason why the GAS correlated so weakly with the MMPI scales measuring neurotic syndromes. The upper half of the GAS seems too general and could be improved by including more symptoms typically seen in outpatient clients. The GAS was developed in an inpatient setting and was originally validated on an inpatient population, which probably accounts for its emphasis on serious psychopathology. In recent years, though, the GAS's major use has been in outpatient community mental health settings. It should be revised to be more relevant to an outpatient population.

2. It is recommended that distinction be made between decisions made on the basis of group GAS scores and decisions made on the basis of individual GAS scores. Because of the low agreement between individual GAS scores and independent measures of severity of disturbance, it is recommended that decisions made for individuals on the basis of GAS
scores be discontinued. This would include screening individual clients for eligibility for service and placing clients in treatment programs on the basis of their GAS scores.

3. It is recommended that whenever possible, GAS scores be administered at the second or subsequent client contact session. Reduced defensiveness of clients in the later sessions and additional time for therapists-raters to gather information should lead to more valid ratings.

Conclusion

In conclusion, the point must be made that an evaluation of the GAS's validity, or its overall worth, cannot be completely good or completely bad. This study has attempted to point out both positive and negative aspects of the GAS and its use.

On the positive side, it appears that some worthwhile and creative uses have been given to the GAS. One of its major uses in research settings, the evaluation of psychotherapeutic treatments, is a positive one that undoubtedly has had and will continue to have humane benefits. In the field of program evaluation, the GAS has been put to use in numerous ways by the State of Michigan in its attempt to deal with the financial crisis it has faced in recent years, and as a means to bring services to clients who are most needy. The GAS appears to have demonstrated adequate validity for most of these uses.

On the negative side is the fact that the GAS has not demonstrated good concurrent validity with another psychometric measure of severity of disturbance in any study with psychiatric patients. Obvious discrepancies
between GAS scores and other accepted measures of psychopathology suggest that the GAS is being used for some purposes for which it has not demonstrated adequate validity. The use of cutoff GAS scores to determine eligibility for service in mental health settings may have inadvertently resulted in denial of services to those who need them. Legally defining a seriously disturbed person on the basis of his or her GAS score may have resulted in the mislabeling of some individuals. Hopefully, usage of the GAS in situations for which it has not demonstrated validity will be discontinued.

This has been a study of the validity of a new rating scale. A false belief some individuals have is that validity is a black or white concept. A determination of validity, however, actually involves subjective personal judgments. As stated in the APA Test Standards,

> Validity is itself inferred, not measured....validity for a particular aspect of test use is inferred from (a) collection of (data). It is, therefore, something that is judged as adequate, marginal, or unsatisfactory. (p. 25)

Hopefully the reader will make his or her own inferences from the data of this study and be better able to make an informed judgment about the validity of the GAS.
APPENDIX A

Global Assessment Scale (GAS)

Robert L. Spitzer, M.D., Miriam Gibbon, M.S.W., Jean Endicott, Ph.D.

Rate the subject's lowest level of functioning in the last week by selecting the lowest range which describes his functioning on a hypothetical continuum of mental health-illness. For example, a subject whose "behavior is considerably influenced by delusions" (range 21-30), should be given a rating in that range even though he has "major impairment in several areas" (range 31-40). Use intermediary levels when appropriate (e.g., 35, 58, 62). Rate actual functioning independent of whether or not subject is receiving and may be helped by medication or some other form of treatment.

Name of Patient ______________ ID No. ______________ Group Code ______________

Admission Date __________ Date of Rating __________ Rater ______________

GAS Rating: ______________

100 Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his warmth and integrity. No Symptoms.

90 Good functioning in all areas, many interests, socially effective, generally satisfied with life. There may or may not be transient symptoms and "everyday" worries that only occasionally get out of hand.

81 No more than slight impairment in functioning, varying degrees of "everyday" worries and problems that sometimes get out of hand. Minimal symptoms may or may not be present.

71 Some mild symptoms (e.g., depressive mood and mild insomnia) OR some difficulty in several areas of functioning, but generally functioning pretty well, has some meaningful interpersonal relationships and most untrained people would not consider him "sick."

60 Moderate symptoms OR generally functioning with some difficulty (e.g., few friends and flat affect, depressed mood and pathological self-doubt, euphoric mood and pressure of speech, moderately severe antisocial behavior).

50 Any serious symptomatology or impairment in functioning that most clinicians would think obviously requires treatment or attention (e.g., suicidal preoccupation or gesture, severe obsessional rituals, frequent anxiety attacks, serious antisocial behavior, compulsive drinking, mild but definite manic syndrome).

40 Major impairment in several areas, such as work, family relations, judgment, thinking or mood (e.g., depressed woman avoids friends, neglects family, unable to do housework), OR some impairment in reality testing or communication (e.g., speech is at times obscure, illogical or irrelevant), OR single suicide attempt.

30 Unable to function in almost all areas (e.g., stays in bed all day) OR behavior is considerably influenced by either delusions or hallucinations OR serious impairment in communication (e.g., sometimes incoherent or unresponsive) or judgment (e.g., acts grossly inappropriately).

20 Needs some supervision to prevent hurting self or others, or to maintain minimal personal hygiene (e.g., repeated suicide attempts, frequently violent, manic excitement, smears feces), OR gross impairment in communication (e.g., largely incoherent or mute).

10 Needs constant supervision for several days to prevent hurting self or others (e.g., requires an intensive care unit with special observation by staff), makes no attempt to maintain minimal personal hygiene, or serious suicide act with clear intent and expectation of death.
Dear Director:

I am a doctoral student in Counseling Psychology at Western Michigan University. For my doctoral dissertation I am doing a validity study of the Global Assessment Scale (GAS), a rating scale for severity of psychiatric disturbance. The GAS, which was introduced by Endicott, Spitzer, Fleiss, and Cohen in 1976, is used by some state Departments of Mental Health as an evaluation instrument in statewide recording systems. In our Michigan Department of Mental Health GAS ratings are given to all state facility inpatients and community mental health outpatients at admission, at each 3-month update, and at discharge.

I am attempting to survey the extensiveness of nationwide usage of the GAS for my dissertation. Reports from either the psychiatric literature or hearsay from program evaluators in the Michigan DMH office indicate or suggest that your state uses the Global Assessment Scale to some extent. I would very much appreciate your answering the two short questions about the GAS on the enclosed survey sheet and returning the form to me as quickly as possible.

My study consists of evaluating the concurrent validity of the GAS by comparing GAS ratings of a group of adult outpatients seen in a large community mental health center (Ingham CMHC, Lansing, Michigan) to scores which the same clients obtained on the MMPI (Minnesota Multiphasic Personality Inventory). Our early results indicate rather disappointingly low correlations (under .40) between the GAS and several severity of disturbance measures taken from the MMPI, as well as low correlations between the GAS and individual scales on the MMPI. One interesting finding is that GAS ratings are more highly related to the validity scales on the MMPI measuring guardedness and defensiveness than they are to actual psychopathology. That is, clients who were more guarded and defensive tended to be rated as more psychologically healthy, irrespective of their actual psychopathology.
My main reason for undertaking this research is that no validity studies of the GAS have been reported in published journals since the GAS was introduced in 1976. If perhaps your office has conducted any unpublished studies related to the validity (or reliability) of the Global Assessment Scale, would you please send a copy to me? I would be glad to also send to you a written report of my results when they are complete. Please indicate this on the enclosed sheet and fill in your name and address.

Thank you for your cooperation.

Sincerely,

Daniel Dekker, M.S.

DD: sd

Enclosure
SURVEY SHEET

1.) Is the Global Assessment Scale (GAS) used in your state Department of Mental Health?

☐ yes  ☐ no - do not fill out questions 2. and 3.; please return questionnaire

2.) To what extent is the Global Assessment Scale used in your state mental health system (including inpatient facilities and outpatient community mental health programs)?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

3.) To what extent is the use of the GAS or any other global psychiatric rating scale mandatory in your state?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ I would like a copy of your research results on the GAS when they become available. Please send to:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ I do not wish to have a copy of your research results.

Person completing this form: ____________________________

Title: ____________________________
APPENDIX C

Subject Information Sheet and Consent Form

INFORMATION ON CLIENT SURVEY

The Ingham Community Mental Health Center occasionally conducts surveys of its clients to obtain information which will help to improve services. You are being asked to help today by taking a personality inventory which is part of a survey.

The inventory you will be given, which is called the MMPI (Minnesota Multiphasic Personality Inventory), consists of several questions about yourself which need to be answered true or false. It will take you about one hour to complete it. After you complete it, you will then go to see your counselor.

Participation in this survey is voluntary. If you choose not to participate, this will not affect the way you are treated by your counselor or the mental health center in any way. If you do take the inventory, the results will be given to your counselor, and they may help your counselor to better understand you and the problems you are having.

If you wish to find out your results, you should ask your counselor and he or she will discuss this question with you. Your results will be kept confidential and will not be seen by anyone except the professional staff of this center.

If you decide to withdraw from the survey after beginning, you are free to do so. However, the results will not be useful to the center unless you complete the entire inventory.

This survey is being conducted by Dan Dekker, who is the person who gave you this information sheet. If you have any questions, please ask him.

You may keep this information sheet for your own use. Thank you for your participation.
CONSENT FORM FOR CLIENT SURVEY

I, __________________________________, have read the sheet entitled "Information on Client Survey" and understand its contents. I voluntarily agree to participate in this survey by taking the MMPI (Minnesota Multiphasic Personality Inventory). I understand that my results will be confidential and will be available only to my counselor or other professional staff of this center. In any other use which is made of the results, my name and other identifying information will be removed. I have been given an opportunity to ask questions and they were answered to my satisfaction. I understand that I am free to withdraw from this survey at any time.

Signed __________________________________ Date ____________

Witness __________________________________ Date ____________

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX D

MMPI Profile of Mean Scores of Sample

The Minnesota Multiphasic Personality Inventory
Starke R. Hathaway and J. Chamley McKinley

Score's Initials

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
BIBLIOGRAPHY


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Craddick, R. A. MMPI scatter of psychopathic and non-psychopathic prisoners. Psychological Reports, 1963, 12, 238.


Endicott, J. Personal communication, December 7, 1982.


Green, R., Nguyen, T., & Attakissoon, C. Harnessing the reliability of outcome measures. Evaluation and Program Planning, 1979, 2, 137-142.


Hargreaves, W. A. Personal communication, October 4, 1983.


Herman, S. E. Global assessment scales: Validity studies on the GAS-MI and GAS-DD scales. Unpublished manuscript, 1982. (Available from Research and Evaluation Division, Michigan Department of Mental Health, Lewis Cass Building, Lansing, MI 48926). (a)

Herman, S. E. Residential service continua: Eastern region field test. Unpublished manuscript, 1982. (Available from Research and Evaluation Division, Michigan Department of Mental Health, Lewis Cass Building, Lansing, MI 48926). (b)


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Kanas, N., Rogers, M., Kreth, E., Patterson, L., & Campbell, R. The effectiveness of group psychotherapy during the first three weeks of hospitalization. Journal of Nervous and Mental Disease, 1980, 168, 487-492.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Spates, R. Personal communication, November, 1982.


Van Putten, T., Crumpton, E., & Yale, C. Drug refusal in schizophrenia and the wish to be crazy. Archives of General Psychiatry, 1976, 33, 1443-1446.


Windle, C., & Ochberg, F. M. Enhancing program evaluation in the community mental health centers program. Evaluation, 1975, 2, 31-36.