77-28,645

JORDAN, Lawrence Arthur, 1938-
THE USE OF COVARIANCE STRUCTURE ANALYSIS FOR THE EXPLORATION OF ACQUIESCENCE AND OTHER RESPONSE STYLE ISSUES IN FACETED TEST DATA.

City University of New York, Ph.D., 1977
Psychology, general

Xerox University Microfilms, Ann Arbor, Michigan 48106

THE USE OF COVARIANCE STRUCTURE ANALYSIS FOR THE EXPLORATION OF ACQUIESCENCE AND OTHER RESPONSE STYLE ISSUES IN FACETED TEST DATA

by

LAWRENCE A. JORDAN

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1977

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

8/10/77 — Samuel Messick

date — Chairman of Examining Committee

8/17/77 — Florence L. Denmark

date — Executive Officer

John S. Antrobus

Samuel Messick

Donald A. Rock

Supervisory Committee

The City University of New York

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to many people:

*To Samuel Messick, Douglas Jackson, Darrell Bock, and Karl Jöreskog, whose teaching and research set high standards for excellence, and whose writings helped me to address the issues studied with this dissertation.

*To Leonard S. Kogan, whose counsel I have missed since his untimely death.

*To my dissertation committee, Samuel Messick, Donald A. Rock, and John A. Antrobus, for their help and encouragement throughout this project.

*To Charles P. Smith and Alan L. Gross, for serving as consultants at the dissertation defense.

*To Martin Morf and Douglas Jackson, for graciously allowing me to reanalyze their data as part of a pilot project.

*To Annette Benedict, Rochelle Brief, Valerie Jordan, Diane Krooth, Wanda Rapaczynski, Gail Wasserman and Sue Zalk, for allowing me to test in their classrooms, and to their students, for agreeing to serve as subjects.

*To my loving wife, Valerie, for her advice and help at every stage of the dissertation.

The dissertation is affectionately dedicated to my parents, Arthur and Katherine Jordan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

THE USE OF COVARIANCE STRUCTURE ANALYSIS FOR THE EXPLORATION OF ACQUIESCENCE AND OTHER RESPONSE STYLE ISSUES IN FACETED TEST DATA

by

Lawrence A. Jordan

Adviser: Professor Samuel Messick

This dissertation was designed to test the two-factor theory of acquiescence, using faceted test data. More generally, it was intended as an exposition of the statistical and methodological problems posed by faceted test data.

The two-factor theory of acquiescence states that there are two kinds of acquiescent response styles, which may be distinguished using two kinds of item reversals. Polar opposite reversals (_I am X_ and _I am Y_, where _X_ and _Y_ are antonyms, for example) are suitable for detecting "agreement" acquiescence, and negation reversals (_I am X_ and _I am not X_, for example) are suitable for detecting "endorsement" acquiescence. Agreement is thought to be inversely related to verbal ability, while endorsement is thought to be positively related to impulsivity and rapid responding. The two-factor theory may be studied by designing faceted test data, in which the same item content is expressed with both polar opposite (true or false keyed) and negation (positively or negatively phrased) reversals.

Chapter 2 reviews the component analysis problems which are posed by faceted test data. In addition to individual differences in test content, faceted test data are presumably also influenced by individual differences in response style tendencies associated with the design facets, and it is useful to estimate the size of the variance components for these individual differences. These component estimates have immediate implications for the reliability of measures, for example. Under stated assumptions, determinate component analysis solutions exist for certain low-dimension designs. When the assumptions fail or high-dimension designs are used, however, determinate solutions may be difficult to obtain using classical least-squares methods. Fortunately, recent work by Bock and Jöreskog and their colleagues has led to very general component analysis methods, as a special case of maximum-likelihood covariance structure analysis. These new methods yield component estimates under a wide range of assumptions, and also provide ways of comparing the merits of differing models for the same data.

As a pilot study, covariance structure analysis was applied to data which were originally collected by Morf and Jackson as a test of the two-factor theory of acquiescence. The results indicated that their design could be improved by counterbalancing item reversals so that the response style facets are not systematically confounded with differences in item content.

New data for testing the two-factor theory were collected from 199 undergraduates at several colleges in the New York City area, using scales based on the Personality Research Form Play and Understanding scales. Strong agreement components were found, and agreement had the predicted negative correlation with verbal ability. The endorsement components were small, however, and unrelated to measures of impulsivity and rapid responding. Several minor response styles were also predicted, for item format (self-descriptive vs. attitude format) and for overgeneralization (absolute vs. relativistic phrasing, using a faceted version of the F-scale). Of the four response styles--agreement, endorsement, item format and overgeneralization--only agreement is present in the data in any strength. For the F-scale, however, it was possible to fit models in which the agreement components for absolutely-worded scales were greater than the ones found with relativistically-worded scales.

An unexpected feature of the data was that subjects from one college were markedly lower in verbal ability than the other subjects--so much lower, that it was necessary to analyze the data separately within the low and high ability groups. For all the measures, variance components for content and error were similar in size for each of the groups, but the agreement component was about twice as large in the low-ability group than in the high-ability group. This indicates that agreement is more of a measurement problem in low-ability groups.

INTRODUCTION

This dissertation aims at a merger of two separate lines of inquiry. The first line is primarily substantive, and concerned with the interpretation of the response style of acquiescence in test data. The second line is primarily methodological, and concerned with the use of component and covariance structure analysis for addressing response style issues. Response styles may be defined as individual differences in response to variations in test item format, where the variations have no conceptual relationship to the nominal content of a test. It has long been known that such variations can systematically affect the rank order of scores and the apparent reliabilities of tests, and that they introduce extraneous variance in test scores which needs to be detected and controlled (Cronbach, 1942). A more recent view is that the individual differences which give rise to response styles can also be informative about personality in their own right (Jackson & Messick, 1958).

The first chapter reviews the literature on response styles. Acquiescence was first detected in tests of ability and achievement, but has since been studied more intensively in personality tests such as the California F-scale, MMPI, and PRF. Initially, acquiescence was defined as the tendency to respond _true_, _yes_, _agree_ or _like_ on tests with bipolar response options. Later work has suggested the presence of at least two kinds of acquiescence, which I will call _agreement_ and _endorsement_. Agreement acquiescence has been called "interpretive acquiescence" (Messick, 1967), "true responding" (Morf, 1968), or "agreement acquiescence" (Bentler, Jackson, & Messick, 1971), and is operationally defined by differences in scores on _true_ and _false_ keyed scales for the same content. Endorsement acquiescence has been called "trait endorsement" or "trait acceptance" (Jackson & Messick, 1965), "impulsive acquiescence" (Messick, 1967), "item endorsement" (Morf, 1968), or "acceptance acquiescence" (Bentler, et al., 1971), and is operationally defined by differences in scores on _positively_ and _negatively_ worded scales for the same content. Agreement and endorsement were often confounded with one another, and with the nominal content of tests, in early studies of acquiescence. The first chapter ends with a discussion of the Morf (1968) study, which provided evidence for the separation of agreement, endorsement, and scale content, but evoked a great deal of critical discussion (Bentler, 1973; Bentler, et al., 1971, 1972; Block, 1971, 1972; Jackson & Morf, 1973, 1974; Morf & Jackson, 1972; Samelson, 1972). Review of this material indicates that Morf's basic design can be improved to provide better tests of response style hypotheses.

In the second chapter, recent developments in the literature on component and covariance structure analysis are reviewed. Morf had analyzed his data using factor analysis, with limited use of formal significance tests, but the newer methods of component and covariance structure analysis--which were not highly developed when Morf analyzed his data--provide more rigorous, statistically-based tests of response style hypotheses. What is here called <u>component structure</u> analysis is less a systematic method of analysis than a heuristic approach to the analysis of faceted test data. It is based on the work of Bock and his colleagues (Bock, 1960, 1964; Bock, Dicken & van Pelt, 1969; Chapman & Bock, 1958), and provides significance tests for structural hypotheses and ways of investigating response style issues. Component structure analysis has been largely superseded by recent developments in <u>covariance structure</u> analysis (Bock & Bargmann, 1966; Jöreskog, 1967, 1969, 1970), which provide comprehensive maximum-likelihood methods for testing detailed hypotheses about the covariance structure of data. In a reanalysis of the Bock et al. data, Bramble and Wiley (1974) have shown that covariance structure analysis may be applied to response style hypotheses. Chapter 2 provides a thorough review of covariance structure analysis as applied to analyses of faceted test data.

The third chapter reports the results from a reanalysis of Morf's (1968) dissertation data, and this reanalysis served as a pilot study of the methods used in the present dissertation. Part of Morf's database involved scales in which the same PRF content was expressed in either <u>true</u> or <u>false</u> (agreement facet), <u>positive</u> or <u>negative</u> (endorsement facet), and <u>self-descriptive</u> or <u>attitude</u> (form facet) format. On examination, his scales were found to be counterbalanced for the form facet (the same item stem was used to provide both a <u>self-descriptive</u> and an <u>attitude</u> item), but not counterbalanced for the agreement and endorsement facets (items defining these facets were based on entirely different stems). The lack of counterbalancing in the Morf data is a major design flaw. As shown in Chapter 3, the covariance structures for most of Morf's sets of scales are reproduced better by fitting models to the item stem groups than to the response style facets. The chapter ends with a discussion of a balanced incomplete block design intended to assure that the design facets are not represented by items which are based on systematically different stems.

Chapter 4 presents the procedure and hypotheses for the present dissertation. With items counterbalanced across design facets, it is possible to make a defensible test of the two-factor (agreement and endorsement) theory of acquiescence. The present dissertation also tests the hypotheses that agreement is negatively related to verbal ability, and endorsement is positively related to tempo, speed of response, and impulsivity.

Chapter 5 presents the results. A complication for the analysis was that there were two distinct subsamples in the data collected for the dissertation, a low-ability subsample from a local "technical" college, and a high-ability subsample from several "academic" colleges. The analysis indicates that there is very little endorsement acquiescence in the data. There are strong agreement acquiescence effects in the data, especially in the low-ability group. The analysis confirms the hypothesis of a negative linear relationship between agreement and verbal ability, and also reveals a strong nonlinear relationship such that the low-ability group is much more variable in agreement tendency than the high-ability group. A summary of the results appears in Chapter 6, which concludes the dissertation.

Chapter 1

The Response Style of Acquiescence

Response styles arise from individual differences in responding to features of test item format which have no conceptual relationship to the nominal content of a test. True-false, Likert, multiple-choice, and other item formats elicit consistent kinds of response habits, and individual differences in these habits can lead to spurious correlations among tests having the same format or among subscales of a test battery. It has long been known that variations in test item format can systematically affect the rank order of scores and apparent reliability of a test, and that they introduce extraneous variance which needs to be detected and controlled. Cronbach's (1942, 1946, 1950) reviews summarized the early research on response styles, which had revealed noncontent response tendencies of acquiescence, evasiveness, working for speed rather than accuracy, guessing when uncertain, and checking many items on lists. He recognized that response styles could be either "mere incidental sources of error" or reflections of "deeper personality traits," but thought that they should be regarded primarily as an "enemy to validity" Cronbach, 1950). A more recent view is that the individual differences which give rise to response styles can be informative about personality in their own right. (Damarin & Messick, 1965; Jackson & Messick, 1958, 1961; Messick, 1967).

Historically, noncontent responding was first attributed to response "set," defined as "any tendency causing a person to consistently give different responses to test items than he would when the same content is presented in a different form" (Cronbach, 1946). This definition implies a distinction between "content" and "set," and the first prerequisite for research on noncontent responding is a design which enables us to distinguish between these two sources of variance. Jackson and Messick made a similar distinction between "content" and "style"--

> between the interpretation of behavior in terms of (a) the content of "needs" and of cognitive structures generally and in terms of (b) characteristic styles or response and action. . . . One may legitimately ask not only what a person says and does (the particular content of his statements and actions) but how he acts (his characteristic mode or style of expression).
>
> What is conceptually a relatively sharp distinction is typically blurred and confounded in a particular concrete act; the what and how are fused in a given goal-directed response (Jackson & Messick, 1958, p. 243).

Fixed-alternative tests offer rather little scope for stylistic responding, and are intended to reduce the opportunities for free expression which are present in essay or sentence completion tests. A researcher wants a tractable measure of the content he is interested in, with as little noise as possible, but unfortunately, even fixed-alternative tests are susceptible to various kinds of distortion. From a response-style point of view, these distortions represent stylistic responding, which may not be related to the kind of content a researcher wants to measure, but can nevertheless be informative about traits of the subjects. Jackson and Messick (1961) made a further distinction between response "style," representing variance associated with variations in item format and having some generality beyond a single test or single occasion of measurement, and response "set," representing form-specific or occasion-specific variance (which would normally be treated simply as error variance). This paper adopts the term "response style" as a generic term for reliable noncontent responding, and focussed primarily on the response style of acquiescence. Acquiescence was first detected in tests of ability and achievement, but has since been studied more intensively in personality tests such as the California F-scale, Minnesota Multiphasic Personality Inventory (MMPI), and Personality Research Form (PRF).

Acquiescence and Ability Assessment

An early finding was Fritz's (1927) discovery that on a very difficult, balanced, true-false information test, 64% of the wrong answers had been marked true, while only 36% had been marked false. This was interesting; as Damarin and Messick remarked: "Nothing in classical test theory would lead us to suppose that the incorrectly answered items in true-false tests would display other than half 'true' and half 'false' responses" (Damarin & Messick, 1965, p. 4). Fritz's finding was replicated and extended to other kinds of bipolar response options, and Lentz (1938) suggested the name "acquiescence" for the tendency to favor the positive responses true, yes, like, agree, etc., in such tests. A typical finding is that in true-false tests of ability, the false-keyed items are more reliable and correlate more highly with the total score and with other ability measures than the true-keyed items (Wesman, 1947).

Cronbach introduced his discsussion of a study as follows: "The procedure was the usual one: to obtain a 'bias' score for each individual and determine its reliability. If the score is reliable, the response set is proved to exist" (Cronbach, 1950, p. 8). Notice that there must be a reliable bias score (e.g., true minus false responses, if the suspected source of bias is acquiescence) before we can say that the data are affected by a response style. A type of item variation which merely shifts a scale mean upward or downward, by increasing the difficulty of items, for example, will not yield a reliable bias score. Fritz had interpreted his finding in terms of a general tendency to respond true rather than false--perhaps because we habitually say "True or false?" rather than "False or true?" Such a general tendency could account for the mean difference between the true and false scales in Fritz's data; but to conclude that a response style is present, we must find systematic individual differences in the tendency, which would show up as a reliable bias score, a nonzero variance component for bias, or a person-by-measures interaction, depending on how the data were analyzed. In other words, a response style interpretation of the data is that there are some persons who favor true over false, while others favor false over true. Couch and Keniston (1960) dubbed the two kinds of people "yeasayers" and "naysayers," and acquiescence refers to the individual differences dimension running from yeasaying to naysaying. Some researchers (e.g., Peabody, 1964, 1966) have attempted to provide separate scores for yeasaying and naysaying ("acquiescence" and "negativism"), but the best measure of the naysaying tendency is simply a measure of yeasaying with reflected scoring (Bock, Dicken, & VanPelt, 1969). The traditional way of measuring acquiescence uses true minus false bias scores, which means that persons with high scores will be yeasayers.

Cronbach's usual method for establishing that a response style is present was to compute split-half or coefficient-α reliability for a bias measure. A roughly equivalent procedure is that of showing that two variations of a scale correlate less highly than their respective reliabilities will allow, which implies that another source of variance besides common content must be present in the data. Neither of these methods yields significance tests for the presence of the suspected response style. Variance component and covariance structure analyses provide more flexible and sophisticated ways of addressing response style issues, as discussed in detail in Chapter 2.

Another point which may be mentioned in the context of ability assessment is that acquiescence can only operate when subjects are uncertain about the "correct" answer to an item. On a balanced true-false vocabulary test, for example, a subject who knows all the answers will not have an opportunity to display his acquiescence; less obviously, a subject who gets every answer wrong will also be treated as neutral in acquiescence. From the standpoint of content, of course, the same thing is true: A subject who is maximally acquiescent and who knows none of the answers--and would presumably respond true to all the questions--would get half the items right on a balanced test, and thus would be scored as somewhat neutral in verbal ability (depending on the difficulty of the test). Thus, scores on a test and acquiescence measured by true and false variants of the test will tend to be nonlinearly related (Cronbach, 1950; Damarin & Messick, 1965; Messick, 1967). Formulas for estimating acquiescence as a weighted function of content have been proposed (Helmstadter, 1957; Messick, 1961). These formulas have the drawback that the functional form of the relationship between acquiescence and a particular kind of content must be presumed known, and if the functional form is incorrectly specified, the formulas will introduce a nonlinear relationship between content and acquiescence, where none existed before. This problem has never been solved satisfactorily, and most researchers have handled the problem by ignoring it and using standard linear models.

The trade-off between content and acquiescence does have implications for test and research design. Cronbach was primarily concerned with content validity, and since the influence of response styles is greatest with difficult or ambiguous items or ambiguous instructions, he recommended that items and instructions be as unambiguous as possible, and that items should not be too difficult for the students tested. He recommended the use of "do guess" instructions, and the use of forced-choice and multiple-choice formats whenever possible. Multiple-choice items, in particular, are relatively free from response sets except with very difficult items or very poor students (Cronbach, 1950). Multiple-choice tests are the most popular kind of fixed-alternative ability tests, perhaps partly as a result of the findings on response style. The multiple-choice format is not well-suited to personality measurement, however, and I know of no standardized multiple-choice personality test. Accordingly, problems of acquiescence are more acute for personality assessment than for ability assessment.

Acquiescence and Personality Assessment

The F-scale. Following the publication of The authoritarian personality (Adorno, et al., 1950), the F-scale and other measures developed for this work were correlated with almost every variable extant. Reviewers of this material (e.g., Titus & Hollander, 1957; Byrne, 1966, pp. 237-283) have found it difficult to make a substantive interpretation of the results. Titus and Hollander concluded:

> Without entertaining a broad critique, one may nonetheless be awed by the massive and amorphous area which has been touched upon by F-scale researchers. To comprehend it requires a theoretical substructure which is not yet available, except in kaleidoscopic form (Titus & Hollander, 1957, p. 472).

Careful work had gone into the construction of the F-scale. Adorno et al. had postulated on the basis of psychoanalytic theory that authoritarian or "fascist" personality involved a number of interrelated features: conventionalism, authoritarian submission, authoritarian aggression, destructiveness and cynicism, preoccupation with power and "toughness," superstition and stereotypy, anti-intraception, projectivity, and an exaggerated concern about the sexual behavior of others. Accordingly, items were written for each of these traits, and combined into the F-scale. For purposes of detecting "fascists" in the general population, it is reasonable to construct a test which is factorially complex. For purposes of integrating a theory of authoritarian personality with the rest of personality theory, however, it is a terrible strategy, since it is difficult to tell which components of a factorially complex test are responsible for its correlations with other measures. Factor analyses of the F-scale have confirmed the multidimensional nature of the scale (Bendig, 1959, 1960; Camilleri, 1959; Krug, 1961). While a number of critics have suggested the need for separate measures of the F-scale components (e.g., Jackson & Messick, 1958), no one has attempted to construct these. Krug and Moyer (1961) correlated factor subscores from the F-scale with subscales of the Edwards Personal Preference Schedule and Guilford-Zimmerman Temperament Survey and found, as predicted, that the factor subscores were not all related to the other personality measures in the same way.

Shortly before Titus and Hollander's review, acquiescence was also found to be a component of the F-scale (Bass, 1955, 1957; Chapman & Campbell, 1957; Christie, Havel, & Seidenberg, 1958; Jackson & Messick, 1957; Jackson, Messick, & Solley, 1957; Leavitt, Hax, & Roche, 1955; Messick & Jackson, 1957). All of the items in the F-scale were <u>true</u> (or <u>agree</u>) keyed, so that acquiescence tendencies would be confounded with authoritarianism. The usual method for showing the presence of acquiescence in the F-scale involved administering the original F-scale and <u>false</u> keyed reversals to the same subjects. In a review and reanalysis of nine studies using F-scale reversals, Chapman and Bock (1958) found that acquiescence consistently accounted for 18-39% of the reliable variance of the original (<u>true</u> keyed)

F-scale, while content accounted for 36-52%; the remainder was accounted for by the content-acquiescence covariance. (See Chapter 2, where the Chapman-Bock analyses are discussed in detail.)

The components found for the F-scale are among the largest acquiescence components found for any published scale.

Adorno et al. were concerned about the dangers of writing an all true-keyed scale, but what happened is that the false-keyed items, originally provided, dropped out during the item analysis and revision process, owing to their relatively poor correlations with the total scale score. Because of the substantial acquiescence components, acquiescence made a noticeable contribution to the reliability of the all-true-keyed scale. The F-scale was intended to measure a factorially complex syndrome involving about nine separate traits, and Adorno et al. should probably have been satisfied with a test having a low internal consistency. It is well known that a factorially complex test can have a high correlation with a factorially complex criterion (such as "authoritarian personality"), but will necessarily have a low internal consistency. From this point of view, the quest for increased internal consistency of the F-scale was misguided, and resulted in substantial contamination of the measure by acquiescence. From a research point of view, it is equally misguided to suppose that the proper use of the F-scale is the "detection" of fascists in the general population, and its more likely uses would include validation of the theory of authoritarianism underlying the scale, and study of the relationships between the traits underlying the authoritarian syndrome and other personality variables. From this second point of view, it would be preferable to build homogeneous subscales for the constituent traits:

> The notion to be emphasized here is that the use of a single total score is an unfortunately simplified reflection of a dynamic theory that postulates a complex of traits (Messick & Jackson, 1958, p. 747).

Contamination of the F-scale by acquiescence tends to add to the difficulty of interpreting scores on the F-scale, since authoritarianism and acquiescence have different theoretical and practical implications.

Some researchers have argued that the confounding of content and style in the F-scale is advantageous, since: (a) acquiescence contributes to the empirical validity of the F-scale, and (b) acquiescence may have a theoretical relationship to the the trait of "authoritarian submission," which Adorno et al. had conceptualized as "a submissive uncritical attitude toward idealized moral authorities of the ingroup" (Gage & Chatterjee, 1960; Gage, Leavitt, & Stone, 1957; Leavitt, Hax, & Roche, 1955). Point (a) is true in the sense that the correlation between F-scale content and acquiescence leads to improved prediction of a criterion measure which is not also contaminated by acquiescence (Chapman & Bock, 1958). As for point (b), reversal studies show that there are some ideologically consistent subjects who will endorse authoritarian content regardless of the direction of keying, while others are acquiescent, endorse both true-keyed items and their reversals, and are ideologically confused (Christie et al., 1958). Both groups would tend to score high on the original true-keyed scale, however, and as Peabody (1966) has noted, "it makes both a theoretical and a practical difference whether they [the high scorers] are confused and apathetic or fanatical true believers."

Christie et al. (1958) reported difficulty finding undergraduate subjects with high scores on the original F-scale, who would "firmly reject" the reversals.

> These individuals who whole-heartedly and consistently endorsed the most discriminating F-scale items displayed an enthusiastically random pattern of response to the reversed items. Our interpretation is that they have a tendency to make extreme responses and are ideologically somewhat confused (Christie et al., 1958, p. 150).

Christie et al. had used Likert scaling, and so their data were subject to extremeness response style as well as acquiescence response style (e.g., Peabody, 1962). (This paper avoids the complications posed by extreme responding, by emphasizing test data in true-false rather than Likert format, and by interpreting results using Likert data along the acquiescence but not the extremity dimension.) Christie et al. had also avoided what Jackson and Messick (1958) called the "extremely-worded, cliche-ridden style" of the original F-scale, when they wrote their reversals, and had expressed the reversals in a false-keyed and also more cautious or "relativistic" style. What Christie et al. found, then, was a response pattern in which subjects used extreme categories to endorse extremely-worded items, but responded more neutrally and somewhat inconsistently to more cautiously-worded reversals. Jackson and Messick (1958) suggested that another potential style component of the F-scale might be based on extremity of wording. They suggested that this response style, which they called "overgeneralization," might be studied by means of a facet design in which F-scale items were reversed along the overgeneralization dimension (with absolutely-worded, extreme, dogmatic statements at one pole, and relatively-worded, cautious, probabilistic statements at the other), as well as along the acquiescence dimension. The overgeneralization component could be found in true-false data, since it is based on extremity of wording of the items, rather than extremity in the choice of Likert-scale response options.

Clayton and Jackson (1961) wrote absolute and relative reversals for F-scale items, which were later used by Morf (1968). Morf administered both the absolute and relative subscales to the same subjects, making it possible to estimate an overgeneralization component, and his F-scale data are discussed in some detail in Chapter 2 and again in Chapter 3, from a methodological point of view. A reanalysis of Morf's data provides no evidence for an overgeneralization component. Some suggestions are made in Chapter 3 which may improve the test of the overgeneralization hypothesis.

The MMPI. After the wave of research concerned with acquiescence in the F-scale, interest shifted to the role of acquiescence in the MMPI. The issues are more complicated for research using the MMPI, since at least two other response styles--social desirability (SD) and impression management--have been held to be determinants of MMPI responses. SD has been claimed by Edwards (1957) to be the major determinant of MMPI responses and, indeed, of responses on most personality inventories. It is marked by scales consisting of items for which the keyed response is both desirable, based on ratings of the desirability of the keyed response, and probable, based on actual response frequencies. In the MMPI, many of the clinical scales behave like SD scales with reflected scoring, owing to the highly undesirable content of the items. Impression management, on the other hand, is marked by scales such as the L ("lie") scale and Cofer, Chance and Judson's (1949) Mp ("positive malingering") scale. Persons scoring high on impression management scales are making desirable but improbable responses (e.g., responding false to "I do not always tell the truth," from the L scale, or true to "I am entirely self-confident," from the Mp scale). Few of the MMPI scales are designed to detect impression management, however, and it usually has a minor role in determining the MMPI factor structure.

The wave of research concerned with acquiescence in the MMPI began with Messick and Jackson's (1961) review and Jackson and Messick's (1961, 1962a, 1962b) factor analyses of MMPI responses for three large samples--one of prisoners, one of hospitalized mental patients, and one of college students. Jackson and Messick used several novel procedures for scoring the MMPI. The first involved separate scoring of the true and false items of the standard MMPI scales. Another involved the provision of five graded SD scales, all true-keyed and containing items which were relatively heterogeneous for content and homogeneous (at five graded levels) for rated SD. In each of the three samples, two large factors dominated the factor structure, accounting for about 75% of the common variance or 55% of the total variance. In the hospital and college samples, the larger of these factors ranked the scales fairly well in order of their average rated SD values, while the smaller factor completely separated the true and false subscales. In the prison sample, the same two factors appeared, but the size of the factors was reversed. Jackson and Messick interpreted the two factors as SD and acquiescence, because of the clear association between scale properties and factor loadings, and attributed them to response styles.

It had been known for some time that the MMPI was dominated by two large factors, and they had been labeled "Anxiety" and "Repression" by Welsh (1956), on the basis of their apparent content. Welsh had also obtained "pure factor" scales A and R, the first a nearly all-true scale with markedly undesirable content, and the second an all-false scale with items predominately neutral in rated SD. It had also been known that the MMPI lacks "simple structure" in Thurstone's (1947) sense, and that the test vectors for the first two factors fall into a distinctive "circumplex structure" (Schaefer, 1959; Kassebaum, Couch, & Slater, 1959). The Jackson-Messick analyses also revealed the circumplex ordering, which was even more accentuated by the separate keying of the true and false parts of the scales. Locating factors in a circumplex array is somewhat arbitrary, and Jackson and Messick used a patterned quartimax rotation to fit the two factors to an SD-acquiescence theory of the data. The acquiescence factor corresponds closely to Welsh's R scale (except that the R scale is false-keyed, and acquiescence is always reported in the true or yea-saying direction), and the SD factor corresponds closely to Welsh's A scale (except that the mostly-true A scale also has a modest loading on the acquiescence factor). Thus, the nominally content-based and the nominally style-based factors tend to coincide.

The Jackson-Messick findings concerning the true and false parts of the scales were unprecedented in the MMPI literature, and raised serious questions about a content interpretation of the data: When the true and false parts of the classical clinical scales were scored in the content direction, their correlations were low and often negative, instead of being strongly positive as we would expect for alternate measures of the same content. For example, the D, Hy, Ma, Mf, Pa, Pd and Si scales all had negative correlations between their true and false parts in the prison sample (Jackson & Messick, 1961).

Another kind of support for the acquiescence theory of the MMPI is provided by the correlations between the MMPI factors and independent measures of acquiescence. Couch and Keniston (1960) obtained a 360-item scale over extremely heterogeneous content called the "Overall Agreement Score" (OAS), and Bock et al. cite their finding of a correlation between the OAS and first-factor scale A ($r_{OAS,A} = .50$) as evidence for the acquiescence theory. Actually, since A is thought to be primarily a measure of SD responding, the correlation between the OAS and second-factor scale R ($r_{OAS,R} = -.34$) should provide a better estimate of the correlation between the supposedly content-free OAS and acquiescence on the MMPI.

The style interpretation of the MMPI evoked considerable controversy (Block, 1965, 1967; Bock et al., 1969; Dicken, 1967; Dicken & Van Pelt, 1967; Jackson, 1967a, 1967b, 1967c; Jackson & Messick, 1961, 1962a, 1962b, 1965; Lichtenstein & Bryan, 1965; Messick, 1967; Messick & Jackson, 1961; Rorer, 1965; Rorer & Goldberg, 1965; Tellegen, 1965). Jackson and Messick had encouraged readers to think of the main MMPI factors as all response style, as in the passage:

> In all three samples two very large factors appeared, identifiable as acquiescence and desirability. On the average, approximately three-quarters of the common variance and over half of the total variance was attributable to the two stylistic dimensions . . .
>
> In addition, several quite small factors were obtained in each of the analyses, some attributed to item overlap and some to consistent content responses (Jackson & Messick, 1962a, pp. 296-297).

Jackson and Messick always insisted that response styles have personological significance and even a certain amount of diagnostic import, since the "massive response sets apparently contribute to the differentiation of normals from mental patients and thus to convergent validity" (ibid.) But it was clearly not desirable to have only a few small factors attributable to distinctive psychopathological content in the MMPI. The variance which they attributed to the acquiescence factor in the three factor analytic studies ranged from 25% to 45% of the common variance.

In the mid-60's, several item reversal studies raised doubts about the acquiescence part of the response style theory of the MMPI, since very high original-reversed correlations were found for the MMPI scales (Dicken & Van Pelt, 1967; Lichtenstein & Bryan, 1965; Rorer, 1965). Only Rorer estimated variance components for his data, and he found acquiescence components of only 4% and 5% for the A and R scales, as against 85% and 69% for the content components (the remainder being attributable to error of measurement). The largest acquiescence component was found for Jackson-Messick scale Dy-3, which was estimated to have 10% of its variance attributable to acquiescence as against 75% attributable to content. The sizes of the original-reversed correlations in the other reversal studies suggest that the estimates of variance components for these other studies would comparable to those found by Rorer. Rorer chose to overlook the small acquiescence components which he had found, and concluded:

> If due allowance were made for features of the design which tended to inflate these acquiescence estimates, it seems unlikely that any scale variance would be attributed to acquiescence (Rorer, 1965, p. 145).

For a variety of reasons, Rorer (1965) and Block (1965) rejected the response style interpretation of the MMPI, and strongly defended a content interpretation of the first two factors. With respect to acquiescence, the evidence included: (a) the results of the reversal studies showing small acquiescence components; (b) Block's showing that essentially the same factor structure found by Jackson and Messick could be found when scales having equal numbers of true and false items, which were presumably balanced for acquiescence, were used; and (c) Block's showing that the R, Dy-3 and other scales marking the acquiescence factor could be interpreted as content-based measures of impulsivity or, with reflected scoring, as measures of what Block called "Ego-control."

The dispute whether the first two factors of the MMPI may be claimed for "content" or for "style" may be simply resolved if they are interpreted as content-style amalgams of some sort, which is essentially the position taken here. While a completely style-based interpretation of the MMPI seems too strong, it is clear that the association between the style and content of the scales is too regular to be a random occurrence. What is remarkable in the data is the clear association between the stylistic features of the scales (the direction of keying and rated SD of the items) and the content of the scales: That is a phenomenon which needs to be explained. Block (1965, p. 61) interprets the association as an "accident," to be explained by "the characteristics of the MMPI-item pool, in conjunction with some other understandings," and Rorer (1965, p. 141) interprets it as a result of "certain systematic characteristics of our language." Their semantic explanation of the confounding of keying and content does not go far beyond noting that such confounding is present, and seems ad hoc. And the puzzle remains: Why does the confounding occur in the first place? Why is it that:

> it is apparent from inspection of the item content of the A and R scales that more than direction of keying distinguishes the two scales. Without exception, the A scale items ask for a highly subjective appraisal of emotional well-being, internal states, and personal adequacy. The subject marking True on these items describes himself as mentally flaccid, anxiously indecisive and full of self-doubt, lacking in the emotional resources and energy to act, conforming, concerned with the possibility of criticism, overly sensitive to signs of disapproval from others, and troubled by his own mental contents and processes. In contrast, the R items deal with comparatively objective physical concerns, with emotionality and violence, and with self-affirmation, assertiveness, and social presence. The subject marking these items in the keyed-False direction characterizes himself as unexceptional, unaggressive, and equable in mood, and as unconcerned with competing or expressing himself in the social world. He is not required by these items to report on his self-esteem or on such patently dysphoric concerns as covered by the A scale. This degree of similarity of content within the two scales, and the clear contrast between the scales, would not be expected if the first factor represented exclusively acquiescence tendency (Bock et al., 1969, p. 128).

As Bock et al. show, however, small acquiescence components can have "appreciable effects" on scale characteristics. These effects help explain both the imbalance of keying on the MMPI and the correlations found between MMPI scales and the heterogeneous OAS measure. A detailed discussion of Bock's component analysis models will be presented in Chapter 2, but it will be helpful to draw some contrasts between the factor analytic and component analysis methods for studying response styles, in the next section.

Factor analysis vs. component analysis. The Jackson-Messick analyses suggested that acquiescence accounted for 25-45% of common-factor variance, while component analyses have suggested that it accounts for 4-5% of scale variance. An even sharper contrast between the methods may be drawn by focusing on a single scale such as the R scale, which Jackson and Messick regarded as a relatively pure measure of acquiescence (naysaying). In their college sample, Jackson and Messick (1962b) found that the R scale had a loading of -.80 on the acquiescence factor and a communality of .69: This implies that 64% of the R scale variance is accounted for by the acquiescence factor, 5% by other (presumably content) factors, and 31% by unique variance (specificity and error). In Rorer's college sample, on the other hand, the component analysis indicated 5% of the R scale variance attributable to acquiescence, 69% to content, and 26% to error. Clearly the estimates based on factor analysis and component analysis are far apart, and agree only in their estimates of the size of the error variance.

Jackson and Messick (1965) have pointed out some flaws in the Rorer data which would lead to overestimates of the content variance and underestimates of the acquiescence variance. Because it is not administratively feasible to administer all 551 items of the original and reversed MMPI to subjects at a single sitting, Rorer administered them at two different test sessions. These data were then compared with data for students who had taken the original MMPI twice. Under the component analysis model used by Rorer, the acquiescence component is estimated as a function of the extent to which the test-retest correlations exceed the original-reversed correlations. Because Rorer's test-retest sample was obtained under different conditions than his original-reversed sample, and the differences in testing conditions tended to bias the analysis in the direction of minimizing the discrepancy between the test-retest and original-reversed correlations (e.g., the test-retest interval was four weeks, while the original-reversed interval was only two), Jackson and Messick (1965) concluded that Rorer's acquiescence components were probably underestimated. These flaws in the Rorer study are probably not sufficient to reverse the order of magnitude of the content and acquiescence components, however, so the problem of reconciling the factor analysis and component analysis results is not solved: The factor analysis results imply that the acquiescence component is much larger than the content component, while the component analysis results imply just the reverse. Which results shall we accept, and why do the two methods yield such drastically disparate estimates?

Component analysis uses planned keying variations--i.e., item reversals--to estimate acquiescence, while Jackson and Messick's factor analyses use the existing or "found" keying variations in the MMPI as the basis for interpreting the second factor as acquiescence. The main problem with the use of the existing keying variations to estimate acquiescence is that *true* and *false* keyed scales may be systematically different in content, so that *true* and *false* response discrepancies are a function of both content and acquiescence. When planned item reversals are used and subjects respond to the same items in both a *true* and *false* keyed form, as in the Rorer study, acquiescence appears to be a small proportion of scale variance on the MMPI.

Bock et al. (1969) collected new data for the MMPI. They chose to focus on two MMPI scales, so it would be administratively feasible to have subjects respond to both original and reversed MMPI items on two separate occasions. They selected the 81% *true* keyed *Pt* scale and the 78% *false* keyed *Hy* scale, which are the clinical scales most resembling Welsh's *A* and *R* scale. Like Rorer, Bock et al. found small acquiescence and large content components (.02 and .05 for acquiescence, and .73 and .58 for content, expressed as proportions of original scale variance). Unlike Rorer, however, Bock et al. point out that these small acquiescence components can have a noticeable effect on scale characteristics and behavior, primarily as a consequence of the correlation (.47) between *Pt* scale content and acquiescence. This content-acquiescence correlation implies that 11% ($.11 = 2 \times .47 \times \sqrt{.02 \times .73}$) of the *Pt* scale variance is attributable to content-acquiescence covariance. The covariance component, which indexes the extent of content-acquiescence confounding in the *Pt* scale, is actually larger than the acquiescence component.

The two-factor theory of acquiescence. When Jackson and Messick (1965) reanalyzed Rorer's data for the original-reversed sample, they found that with tests scored in the content direction, the original and reversed subscales loaded on the same factors and in the same direction. This confirmed the finding which Rorer had interpreted as evidence of content consistency for the original and reversed scales. Yet the true and false parts of many scales still loaded in opposite directions on the factor previously identified as acquiescence in the Jackson-Messick analyses. Table 1.1 displays the correlations which Rorer found for the Hs and Hy subscales, as reported by Jackson and Messick (1965). These are correlations for male subjects, and the correlations for female subjects are very similar. The Hs scale is an example of a scale in which the true and false parts have high positive correlations, in both their original and reversed form. Subjects endorsing many Hs true items are making a variety of somatic complaints, and subjects rejecting many Hs false items are, consistently, denying that they are in good health. The keyed direction of Hs items is uniformly undesirable. On the Hy scale, however, we find high original-reversed correlations (.66 for true and .54 for false scales), but low true-false correlations (.01 for original scales and .06 for reversed scales). Rorer had attributed these high original-reversed correlations to content, and they cannot be attributed to acquiescence as traditionally conceived because the reversed-true items are false-keyed in the content direction, and the reversed-false items are true-keyed in the content direction.

Table 1.1

Correlations among Original and Reversed Versions of *True* and *False* Parts of MMPI *Hs* and *Hy* Scales[a]

		Original		Reversed	
		true	*false*	*true*	*false*
Hs Scale (High Content Consistency)					
Original	*true*	1.00			
	false	.61	1.00		
Reversed	*true*	.51	.42	1.00	
	false	.47	.71	.72	1.00
Hy Scale (Low Content Consistency)					
Original	*true*	1.00			
	false	.01	1.00		
Reversed	*true*	.66	.08	1.00	
	false	.13	.54	.06	1.00

[a]Correlations for Rorer's (1965) original-reversed subsample (males only, *N* = 96), as reported by Jackson and Messick (1965). All subscales scored in content direction.

It was the low correlation between the _true_ and _false_ parts of the scales, however, and the finding of a factor which separated the _true_ and _false_ parts of all the scales, which had originally led Jackson and Messick to interpret the factor as acquiescence. Clearly _true_ and _false_ scales correlating .01 cannot be measures of the _same_ content.

The Rorer findings led Jackson and Messick to propose a two-factor theory of acquiescence, which rests in part on an ambiguity in the notion of item reversal. Initially, acquiescence had been defined as the tendency to endorse _like_, _agree_, _yes_, or _true_ on tests with bipolar response options. As in the reversal studies of the F-scale, acquiescence could be detected and measured by writing item reversals which expressed the same content in both _true_ and _false_ keyed form. But there are at least two ways of reversing the sense of an item with the form "I am _X_," say, which may be termed _negation_ and _polar opposition_. A negation reversal has the form "I am _not_ _X_," while a polar opposite reversal has the form "I am _Y_," where _Y_ is an attribute inconsistent with _X_ and at the opposite pole of the content scale. A fourth item type is produced by negation of the polar opposite ("I am _not_ _Y_"), and might be termed a _double negative._ Jackson and Messick suggested that the two kinds of reversal could serve to operationalize two kinds of acquiescence, which I will call _agreement_ and _endorsement_.

Agreement has been called "interpretive aquiescence" (Messick, 1967), "true responding" (Morf, 1968), or "agreement acquiescence" (Bentler et al., 1971), and is operationally defined by differences in scores on true and false keyed scales for the same content. Agreement is thought to be the kind of acquiescence elicited by attitude scales such as the F-scale, and is related to what Damarin and Messick (1965) called "Pattern I acquiescence: Low verbal interpretive skill." Endorsement has been called "trait endorsement" or "trait acceptance" (Jackson & Messick, 1965), "impulsive acquiescence" (Messick, 1967), "item endorsement" (Morf, 1968), or "acceptance acquiescence" (Bentler et al., 1971), and is operationally defined by differences in scores on positively and negatively worded scales for the same content. Endorsement is thought to be the kind of acquiescence elicited by personality inventories such as the MMPI, and is related to what Damarin and Messick (1965) called "Pattern II acquiescence: Speed, tempo and fluency."

In order to separate agreement and endorsement from one another and from content, it is necessary to have a fully crossed 2 x 2 design on the scales for a given kind of content. In the reversal studies of the MMPI, most of the item reversals used negation, so that typically, positively worded true items were given negatively worded and now false-keyed reversals. This leaves open the possibility that a person might respond true to the original and false to the reversal items because of his tendency to endorse items as self-descriptive (rejection of negatively worded items is tantamount to accepting the positively worded originals, according to this theory). Reversal studies of the MMPI sufficed to rule out agreement acquiescence as an explanation for the second factor, but could not rule out endorsement acquiescence since endorsement acquiescence was confounded with content.

Bentler et al. (1972) illustrate negation and polar opposite reversal using the attributes $\underline{X}$ = "happy" and $\underline{Y}$ = "sad." An item "I am happy" appearing on a scale of depression (say) can be reversed by negation ("I am not happy") or by polar opposition ("I am sad"), with a fourth variant produced by double negation ("I am not sad"). Only by having all four kinds of items can effects due to agreement and endorsement be separated from content responding. Table 1.2 displays the four item variants, together with a content key and the two acquiescence keys. The two-factor theory of acquiescence predicts the pattern shown in Figure 1.1(a) for data which have not been scored in the content direction. Usually data are preprocessed by scoring all scales in the nominal content direction (which amounts to reflection of all the false keyed items), yielding the predicted pattern shown in Figure 1.1(b) (after Morf, 1968, p. 11). In data containing all four kinds of scales (PT, PF, NT, NF), the two-factor theory of acquiescence predicts that a plane can be found through the data in which each kind of scale lies in a different quadrant.

Table 1.2

Content, Agreement, and Endorsement Response Patterns

Item Form	Direction of Wording	Direction of Keying	Symbol	Content Key	Acquiescence Keys Agreement	Acquiescence Keys Endorsement
X	positive	true	PT	T	T	T
Y	positive	false	PF	F	T	T
not Y	negative	true	NT	T	T	F
not X	negative	false	NF	F	T	F

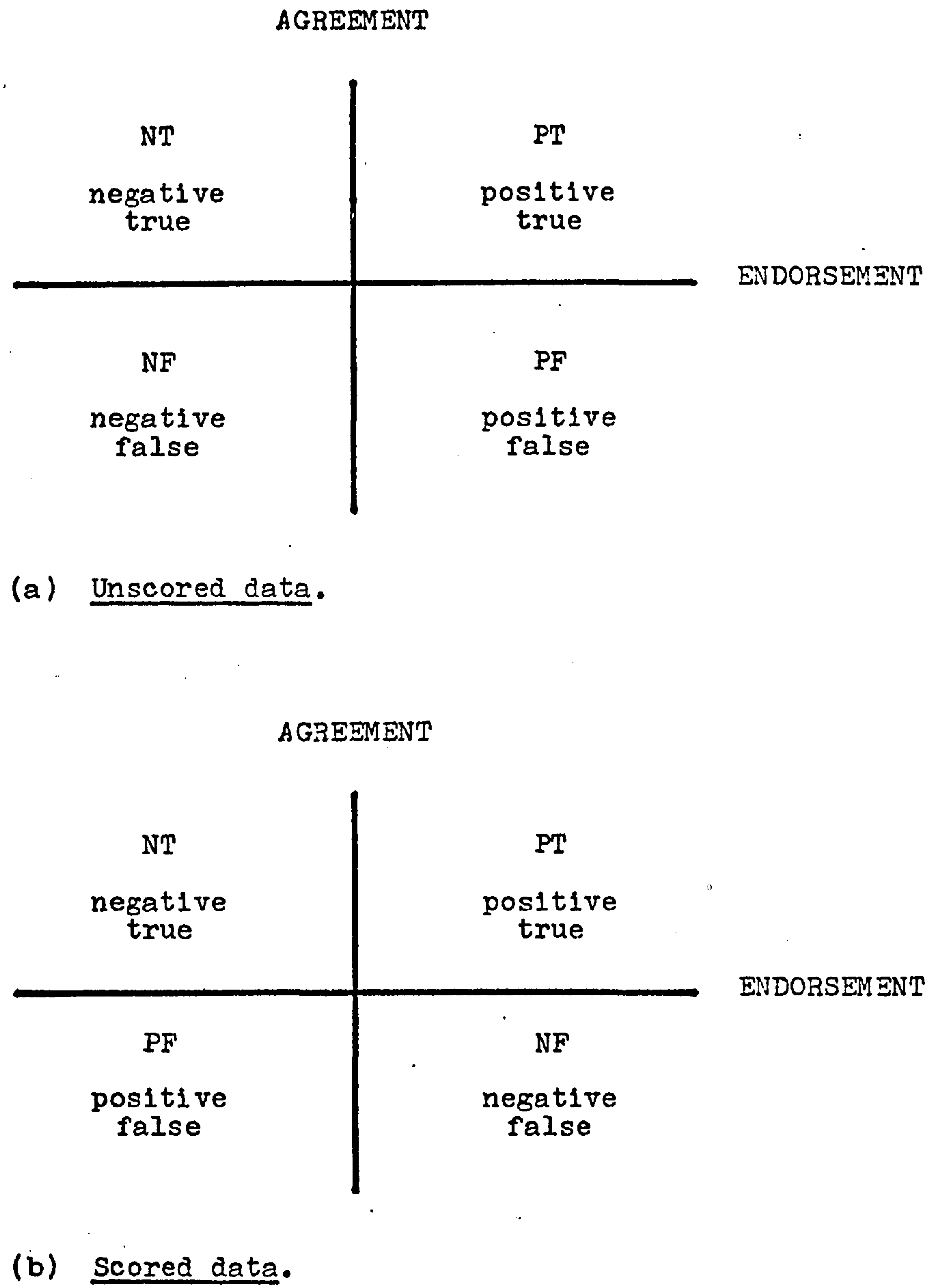


Figure 1.1. Predicted location of scales defined by hypothetical Agreement and Endorsement factors in data which are (a) <u>unscored</u>, or (b) <u>scored</u>, in the content direction.

A content interpretation of the Hs and Hy scales. Before examining the evidence for a two-factor theory of acquiescence, it will be useful to consider a content interpretation of the pattern of correlations found for the Hs and Hy scales. As shown for Rorer's data in Table 1.1, all four versions of the Hs scale tend to have high positive correlations, suggesting that they are all measuring essentially the same thing. An examination of the Hs scale items will show that they almost all involve somatic complaints, as appropriate for a scale of "hypochondriasis." For the Hy scale, however, the true and false original scales correlate only .01, and appear to be measuring somewhat different things. The finding of such low correlations between true and false parts of a scale was part of the evidence suggesting the presence of acquiescence--later specialized to endorsement acquiescence--in the MMPI.

The different behavior of the Hs and Hy scales is especially surprising because they have many items in common. In fact, four of the 11 Hs true items and 16 of the 22 Hs false items also appear on the Hy scale, keyed in the same direction. Table 1.3 lists all of the items appearing on the Hs and Hy scales, grouped in six categories:

A - True keyed and common to Hs and Hy (4 items)
B - True keyed and unique to Hs (7 items)
C - True keyed and unique to Hy (9 items)
D - False keyed and common to Hs and Hy (16 items)
E - False keyed and unique to Hs (6 items)
F - False keyed and unique to Hy (31 items)

Table 1.3

Comparison of Hs and Hy Scale Items

A. True keyed and common to Hs and Hy

23. I am bothered by attacks of nausea and vomiting.

43. My sleep is often fitful and disturbed.

114. Often I feel as if there were a tight band about my head.

189. I feel weak all over much of the time.

B. True keyed and unique to Hs

29. I am bothered by acid stomach several times a week.

62. Parts of my body often have feelings like burning, tingling, crawling or like "going to sleep."

72. I think a great many people exaggerate their misfortunes in order to gain the sympathy and help of others.

108. There often seems to be a fullness in my head or nose most of the time.

125. I have a good deal of stomach trouble.

161. The top of my head sometimes feels tender.

273. I have a numbness in one or more regions of my skin.

C. True keyed and unique to Hy

10. There seems to be a lump in my throat much of the time.

32. I find it hard to keep my mind on a task or a job.

44. Much of the time my head seems to hurt all over.

47. Once a week or oftener I feel suddenly hot all over, without apparent cause.

76. Most of the time I feel blue.

179. I am worried about sex matters.

Table 1.3 (continued)

True keyed and unique to Hy (continued)

186. I frequently notice my hand shakes when I try to do something.

238. I have periods of such great restlessness that I cannot sit long in a chair.

253. I can be friendly with people who do things which I consider wrong.

D. False keyed and common to Hs and Hy

2. I have a good appetite.

3. I wake up fresh and rested most mornings.

7. My hands and feet are usually warm enough.

9. I am about as able to work as I ever was.

51. I am in just as good physical health as most of my friends.

55. I am almost never bothered by pains over the heart or in my chest.

103. I have little or no trouble with my muscles twitching or jumping.

153. During the past few years I have been well most of the time.

163. I do not tire quickly.

175. I seldom or never have dizzy spells.

188. I can read a long while without tiring my eyes.

190. I have very few headaches.

192. I have had no difficulty in keeping my balance while walking.

230. I hardly ever notice my heart pounding and I am very seldom out of breath.

243. I have few or no pains.

274. My eyesight is as good as it has been for years.

Table 1.3 (continued)

E. False keyed and unique to Hs

18. I am very seldom troubled by constipation.

63. I am troubled by discomfort in the pit of my stomach every few days or oftener.

68. I wish I could be as happy as others seem to be.

130. I have never vomited or coughed up blood.

155. I an neither gaining nor losing weight.

281. I do not often notice my ears ringing or buzzing.

F. False keyed and unique to Hy

6. I like to read newspaper articles on crime.

8. My daily life is full of things that keep me interested.

12. I enjoy detective or mystery stories.

26. I feel that it is certainly best to keep my mouth shut when I'm in trouble.

30. At times I feel like swearing.

71. I used to like drop-the-handkerchief.

89. It takes a lot of argument to convince most people of the truth.

93. I think most people would lie to get ahead.

107. I am happy most of the time.

109. Some people are so bossy that I feel like doing the opposite of what they request, even though I know they are right.

124. Most people will use somewhat unfair means to gain profit or an advantage rather than lose it.

128. The sight of blood neither frightens me nor makes me sick.

129. Often I can't understand why I have been so cross and grouchy.

136. I commonly wonder what hidden reason another person may have for doing something nice for me.

Table 1.3 (continued)

<u>False</u> keyed and unique to <u>Hy</u> (continued)

137. I believe that my home life is as pleasant as that of most people I know.

141. My conduct is largely controlled by the customs of those about me.

147. I have often lost out on things because I couldn't make up my mind soon enough.

160. I have never felt better in my life than I do now.

162. I resent having anyone take me in so cleverly that I have had to admit that it was one on me.

170. What others think of me does not bother me.

172. I frequently have to fight against showing that I am bashful.

174. I have never had a fainting spell.

180. I find it hard to make talk when I meet new people.

201. I wish I were not so shy.

213. In walking I am very careful to step over sidewalk cracks.

234. I get mad easily and then get over it soon.

265. It is safer to trust nobody.

267. When in a group of people I have trouble thinking of the right things to talk about.

279. I drink an unusually large amount of water every day.

289. I am always disgusted with the law when a criminal is freed through the arguments of a smart lawyer.

292. I am likely not to speak to people until they speak to me.

The items in the first five categories are overwhelmingly concerned with somatic complaints, which is one of the categories which always appears in attempts to cluster the MMPI items. (For example, 27 of these 42 items appear on Tryon and Bailey's [1970] Body Symptoms scale, and 31 of 42 appear on Wiggins' [1966] Organic Symptoms or Poor Health scales.) In reading through the first five groups of items, there appears to be little difference between the items unique to the *Hs* or *Hy* scales, except perhaps that some of the *Hy* items seem to be symptomatic of tension rather than somatic complaints (10, 32, 76, 179, 186, 238 in group C).

On examining the *false* keyed items unique to the *Hy* scale, in the sixth category, we find a few items involving somatic complaints (160, 174), but most of them appear to be from a different universe of content than the *Hs* and other *Hy* items. Of particular interest for a clinical interpretationof "conversion hysteria" are a number of items which are keyed for denial of problems in talking to people, getting to know them, being in conflict with them, or distrusting their motives (26, 89, 93, 109, 124, 136, 141, 162, 172, 180, 201, 265, 267, 292). Clinically, high *Hs* people have many somatic complaints, and have also been characterized as crabbed, dissatisfied, defeatist and cynical (Cuadra & Reed, 1954). High *Hy* people are also noted for multiple somatic complaints--which are difficult to localize and often transparently useful for purposes of making life miserable for people who are close to them--but they also seem to have a peculiar lack of information about how well they get along with others.

Consider Sullivan's characterization of a more or less typical male hysteric:

> So he has these attacks in and out of season when the provocation is sufficient; and he remains comfortably unaware of their effect on others, partly because he doesn't pay much attention to other people anyway and partly because these are his symptoms, this is his sickness, this is something mysterious and rather awful which abruptly descended upon him at 4:00 A.M. on a particular unforgettable morning and which the doctors haven't succeeded in making much sense of (Sullivan, 1956, p. 208).

Unlike the defeated and embittered high Hs people with whom they share so much apparent pain and suffering, hysterics often give the impression that their life and relationships with others would be perfectly lovely if only it weren't for the current medical problem:

> . . . one discovers sometimes the almost juvenilely simple type of operation set up to profit from the disabling system. The patient will often tell you in the most transparent fashion: "If it were not for this malady then I could do---" and what follows is really a quite grandiose appraisal of one's possibilities. . . . We know that under cover of the hysterical disorder the patient works out dramas that are rather blatantly expressive of what is on his mind, and we marvel sometimes at the prodigies of inattention by which no clue as to what is the source of the difficulty reaches his awareness (ibid., p. 217).

From a content point of view, then, many of the false keyed items which are unique to the Hy scale seem to be getting at the hysteric's repression and denial of what others would regard as his real problems in living.

I mentioned 16 Hy items from group F which are clinically relevant to a diagnosis of hysteria (2 involving somatic complaints, and 14 involving denial of conflict with others)--what about the remaining 15 items from group F? They are quite a mixed bag. There are seven items dealing with aggression and violence, where the keyed *false* response suggests denial of interest in the topic (6, 12, 30, 128, 129, 234, 289). Consider item 6: "I like to read newspaper articles on crime." A *true* response clearly indicates an interest in crime (though not whether the respondent's sympathy lies with the criminal or victim), and the keyed *false* response indicates lack of interest or, possibly, that the respondent has better things to do. To the extent that these items indicate repression of aggressive impulses, they can also be considered clinically relevant to a diagnosis of hysteria. For three more items, the keyed *false* response indicates a lack of interesting things to do, unhappiness, or an unpleasant home life, and is vaguely suggestive of depression (8, 107, 137). It would be futile to try to justify all of the items from a clinical standpoint, because of the inadequacy of the original criterion keying of the scales. The MMPI scales were constructed by choosing items which discriminated about 1500 "normal" persons from rather small groups of neuropsychiatric patients. Fifty persons diagnosed as having conversion hysteria formed the criterion group for the *Hy* scale, and many of the items do not stand up under cross-validation (Kleinmuntz, 1967). I nominate item 71 ("I used to like drop-the-handkerchief") as the item most likely to be on the *Hy* scale by accident.

Thus, there are two main content areas covered by the *Hs* and *Hy* items. The *Hs* scale appears to be an essentially unifactorial measure of somatic complaints, but the longer *Hy* scale contains items from at least two content domains. The 29 *Hy* items in groups A, C and D (20 of which are shared with the *Hs* scale) appear to be measuring the somatic-complaint component, while the 31 *false* *Hy* items in group F appear to be measuring a repression or denial component.

To see how a content interpretation of the *Hs* and *Hy* scales leads to a plausible interpretation of the MMPI factors, examine Figure 1.2, which displays the loadings for selected MMPI scales in the (I,II) plane. These loadings are from the male subsample of Rorer's original-reversed data again, as reported by Jackson and Messick (1965). Double points in Figure 1.2 represent the original and reversed versions of the scales, scored in the content direction. (Projections of the majority of MMPI scales are omitted in this Figure, which shows only scales which have been specifically discussed in this chapter.) The *true* and *false* *Hs* subscales and the *true* *Hy* subscale all load together at the undesirable end of the first factor. These somatic-complaint scales are markedly undesirable in content, like the anxiety and tension items making up the *A* scale, leading to the interpretation of the first factor as Social Desirability. The *false* *Hy* scale loads on the second factor, next to the *R* scale and near the negative pole of the factor identified as Acquiescence.

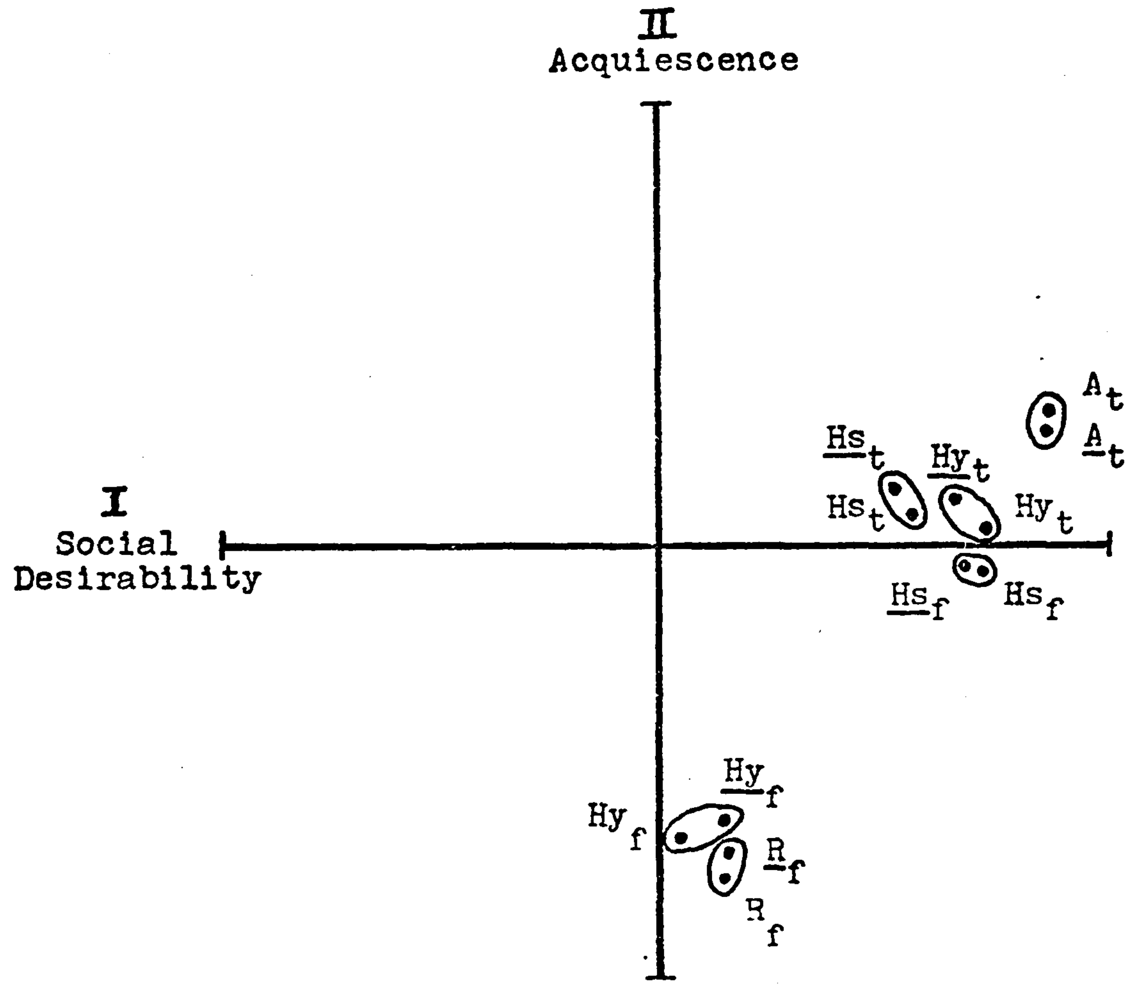


Figure 1.2. Selected loadings in plane of largest two factors for Rorer (1965) data, male subsample, as reported by Jackson and Messick (1965). MMPI reversed scales have been underlined.

Thus, content and style interpretations of the MMPI tend to converge on an identification of the same factors. Social Desirability is the g of psychopathology, and while such undesirable traits as somatic complaints and anxiety complaints are conceptually distinct, they are highly correlated. However, it is the second factor which is most at issue here, and the factor described as acquiescence is marked at its negative pole by scales interpretable from a content point of view as measures of repression. I would argue that the group F Hy scale items meet the definition of "content" as

> . . . response consistencies which reflect a particular set of broader behavioral tendencies, relatively enduring over time, having as their basis some unitary personality trait, need state, attitudinal or belief disposition, or psychopathological syndrome (Jackson & Messick, 1962b, p. 542).

The keyed false responses do not represent repression or denial of just any content, but primarily of content concerned with interpersonal conflict, emotionality, aggression and violence. According to clinical theory, hysterics have difficulty coping with these and deal with them using the mechanism of repression. Vide Shapiro:

> . . . the picture of hysterical neurosis is relatively clear-cut. It was the first neurotic condition to be studied by Freud, and, among neuroses, none has been more definitely or clearly associated with the operation of a specific defense mechanism than has hysteria with repression. The relative simplicity of this association becomes evident if, for example, it is compared with the constellation of defense mechanisms usually considered to be operative in obsessive-compulsive pathology--namely, regression, reaction-formation, isolation of affect, and undoing.

> The mechanism of repression, furthermore, has a simplicity and clarity that is unique among the various defense mechanisms. It was, along with hysterical pathology, the first defense to be studied, and it retains a basic or elemental quality among the mechanisms, which cover a considerable range in both complexity and clarity. The mechanism of repression and its significance in hysterical pathology are, therefore, clear enough and, undoubtedly, real enough (Shapiro, 1965, p. 108).

Shapiro's monograph discusses hysteria as a "neurotic style," and explicitly relates the notion of neurotic style to the work of Gardner, Klein, and their associates on "cognitive style" (cf. Gardner, Holzman, Klein, Linton & Spence, 1959; Gardner, Jackson & Messick, 1960; Shapiro, 1965, pp. 13-15).

At this point, some ambiguity in the notion of "response style" may be noted. In their original paper on content and style in personality assessment, Jackson and Messick (1958) related the notion of style to both "cognitive style" (p. 243) and "response set" (p. 244), with "response style" (p. 251) emerging as the preferred expression and having connotations of both cognitive style and response set. Jackson and Messick indicated that it was difficult to distinguish between content and style in actual behavior:

> One may legitimately ask not only *what* a person says or does (the particular content of his statements and actions) but *how* he acts (his characteristic *mode* or *style* of expression).
>
> What is conceptually a relatively sharp distinction is typically blurred and confounded in a particular concrete act; the what and how are fused in a given goal-directed response (Jackson & Messick, 1958, p. 243).

At the same time, they indicated that it was possible to distinguish between content and style *in data*, by designing assessment techniques "to evoke theoretically important styles of response." Such devices as systematically varying true and false keying, positive and negative wording, absolute and relative phrasing, and so on, would enable researchers to separate the variance attributable to content from the variance attributable to style. Thus, the conceptual distinction between content and style seems appropriate for the notion of "cognitive style," which is confounded with content (and presumably correlated with content in data), while the proposed research model seems appropriate to the original notion of "response set" since it depends on the detection of response habits associated with (from a content point of view) superficial characteristics of item and test formats. For field-independence, the best known cognitive style construct, one would not ordinarily consider partialling out field-independence in order to get better measures of "content." Field-independence is typically conceptualized as a "content" which is measured along with other cognitive "content" variables. For response styles associated with direction of keying or wording, however, it seems natural to consider partialling them out to get better measures of "content." The force of the finding that 25-45% of the MMPI common-factor variance is attributable to acquiescence is that this sort of "massive" response style effect is an indictment of the instrument. That the MMPI is dominated by two large factors accounting for 75% of the common-factor variance is already an undesirable feature for an instrument which is used, for better or worse, for purposes of differential diagnosis; if these two factors are merely attributable to response style, it seems to follow that the MMPI is basically worthless.

There is more congruity between the content and style interpretations of the acquiescence-repression factor than meets the eye. Clinically, repression is a defense mechanism, and an aspect of hysterical neurotic style. Shapiro argues persuasively that it resembles a cognitive style, since the hysteric's faulty recollection is aided by a rather impressionistic and global mode of cognition. It would seem difficult to measure repression as directly as one can measure anxiety or hypochondriasis: People who are anxious or feel sick much of the time are well aware that this is so, and even though it is socially undesirable to have such feelings they can be reported accurately. People who avoid threatening material using the defense mechanism of repression, however, don't <u>know</u> that they do it. Repression takes place outside of awareness, and lacks the strong negative affect associated with anxiety or somatic complaints. Thus, I would rationalize the form of the <u>R</u> scale and the group F <u>Hy</u> scale items by suggesting that most of the items deal with interpersonal conflict, aggression and violence--in rather mild forms--in such a way that the non-keyed <u>true</u> response indicates an interest in the topic, while the keyed <u>false</u> response indicates a lack of interest. On this reading, it is crucial that the items do not imply "I am an aggressive, violent person," which most people can truthfully deny; instead, the items imply something like, "I am interested in aggression and violence." Given the history of the species, it seems rather natural for human beings to express an interest in aggression, violence, and the milder forms of interpersonal conflict, and the systematic denial of such an interest suggests the operation of the defense mechanism of repression. The association between repression and the negative ("nay-saying") pole of the acquiescence factor was accidental, but it makes clinical sense. After the fact, we can interpret the group F _Hy_ items as having a suppressor function: They serve to identify the subgroup of people expressing many somatic complaints who are also using repression as a defense mechanism.

Hathaway and McKinley never conceptualized conversion hysteria, say, as a unifactorial--or even multifactorial--personality dimension which they set out to measure. Instead, they began with a criterion group of people diagnosed as having conversion hysteria, and selected the items which empirically distinguished them from normal controls. Now, if the _Hy_ scale is actually multifactorial, as I have argued, that poses problems for the use of the scale in diagnosis. These problems have little to do with the content-style controversy, but should be mentioned here. An "elevated" _Hy_ score (above 70 after T score conversion) results from about 30 keyed responses, for both males and females (Hathaway & McKinley, 1967, p. 26). Since the scale consists of 60 items, it is possible for the 30 keyed responses to come primarily from the somatic complaint component or primarily from the repression component, and it would clearly make a difference for purposes of interpretation if there were a substantial imbalance either way. MMPI specialists attempt to deal with this problem using profile analysis so that, for example, if _Hs_ and _Hy_ are both elevated it means one thing, while if _Hy_ is elevated and _Hs_ is normal or low it means something else. From a measurement point of view, it would be preferable to attempt to measure the somatic component and repression components separately, rather than to make inferences about them based on profile differences.

As for the status of the controversy over the interpretation of the MMPI second factor, I will mention several points which seem well established, together with several questions which seem unresolved. (1) We have several reversal studies showing that R scale content can be expressed in true as well as false keyed form, and that the resulting (agreement) acquiescence components are small for the R scale (about 5% of scale variance). Such a small agreement component implies only a minor perturbation of the rank order of individuals on R scale content using the unbalanced scale, so that the observed association between content and direction of keying would not, by itself, invalidate the inferences one would like to draw on the basis of R scale content.

¶(2) We have the showing of Bock et al. that small acquiescence components can nevertheless have a strong effect on the correlational behavior of scales if acquiescence is correlated with content. Recall that Bock et al. discussed A and R scale content, and then provided a component analysis of the Pt and Hy scales. An examination of the A and Pt scale items indicates that the Pt scale is a reasonable proxy for the A scale, so that the observed correlation of .47 between Pt content and acquiescence is approximately what would be found for the *A* scale as well. Bock et al. found a negligible correlation (-.09) between *Hy* scale content and acquiescence, but I would judge that this estimate would not necessarily hold for *R* scale content as well, owing to the multifactorial nature of the *Hy* scale. It is entirely possible that acquiescence is uncorrelated with the *Hy* resultant content, but correlated with both components. A simple example will show this. Imagine the following factor structure:

	I	II
Hs	1.0	.0
Hy	.7	.7
R	.0	1.0

This implies that *Hs* and *R* are uncorrelated, while *Hy* is correlated with both. Two factors account for all the variance of the three variables, so we have the luxury of measuring them without error. Now add acquiescence (*Ac*) and a third factor to this structure, and the fact that acquiescence and *Hy* content are uncorrelated is compatible with several very different factor structures, e.g.,

(a)

	I	II	III
Hs	1.0	.0	.0
Hy	.7	.7	.0
R	.0	1.0	.0
Ac	.0	.0	1.0

(b)

	I	II	III
Hs	1.0	.0	.0
Hy	.7	.7	.0
R	.0	1.0	.0
Ac	.5	-.5	.7

(c)

	I	II	III
Hs	1.0	.0	.0
Hy	.7	.7	.0
R	.0	1.0	.0
Ac	.7	-.7	.0

At one extreme, (a) depicts a situation in which acquiescence is uncorrelated with *Hy* and its two components; (b) depicts a situation in which acquiescence is uncorrelated with *Hy* but correlated with both components; and at another extreme, (c) depicts a situation in which acquiescence is entirely predicted by *Hs* and *R*, so that the third factor is not needed, but acquiescence is still uncorrelated with *Hy*. These three situations would have somewhat different implications for an attempt at a substantive interpretation of acquiescence. As it now stands, the literature does not provide direct evidence in the form of a component analysis concerning the relationship between *R* scale content and acquiescence. The reversal studies using the *R* scale do not provide test-retest data for the same subjects, making it impossible to estimate the *R* scale content-acquiescence correlation. On the other hand, the Bock et al. data do not directly address the issue of *R* scale content and acquiescence. It would be feasible to reanalyze the Bock et al. data, performing separate component analyses of the items measuring the somatic component (groups A,C,D, and items 160 and 174 from group F in Table 1.3), and the items measuring what I have interpreted as the repression component (the remaining group F items). A by-product of this analysis would be an estimate of the correlation between the somatic-complaint and repression components.

(3) A third issue raised in the content-style controversy over the second MMPI factor is the Jackson-Messick hypothesis of a second (endorsement) acquiescence factor. My reading of the literature is that the first (agreement) acquiescence factor should not be ruled out as having a bearing on the MMPI factor structure. The endorsement acquiescence hypothesis would, if confirmed, strengthen the response style position in the controversy over the MMPI. To my knowledge, no one has ever prepared sets of polar-opposite original and reversed items for the MMPI (as noted earlier, the existing reversals are primarily of the negation kind), and administered both sets to the same subject. Most of the studies bearing on the two-factor theory of acquiescence have used items drawn from the PRF item pool. The most extensive set of new data was collected by Morf (1968). These data are reanalyzed in Chapter 3, and I will argue there that the data do not provide strong support for the two-factor theory. Neither do they disprove it, however. The next section discusses another line of evidence, other than the evidence based on the MMPI factor structure, which has implications for both the two-factor theory of acquiescence and for a substantive interpretation of acquiescence.

Substantive interpretations of acquiescence. At about the time that the content-style controversy was heating up over the MMPI, Damarin and Messick (1965) completed a review of studies by Cattell and his associates which included variables having a bearing on response style issues. Evidence from a series of these studies (Cattell, 1955; Cattell, Dubin & Saunders, 1954a, 1954b; Cattell & Gruen, 1955; Cattell & Peterson, 1959; Cattell & Scheier, 1959, 1960) suggested the presence of at least two acquiescence patterns, which Damarin and Messick called Pattern I (low verbal interpretive skill) and Pattern II (speed, tempo and fluency). Pattern I acquiescence is evidenced by negative correlations between verbal ability and acquiescence (e.g., Frederiksen & Messick, 1959; Messick & Frederiksen, 1958). This had been noted a number of times in studies using acquiescence measures based on the MMPI. In Cattell's work, the measure Little logical consistency of attitude (MI327), a count of discrepant responses to major and minor parts of syllogisms scattered throughout a questionnaire, serves as a measure of Pattern I acquiescence. Damarin and Messick suggested that Pattern I acquiescence might arise from overgeneralization, "as measured by the unreflective acceptance of sweeping, unqualified beliefs." On this view, the negative relationships between verbal ability and acquiescence arise because people with low verbal ability are more likely to endorse glittering generalities. Pattern II acquiescence, on the other hand, is evidenced by positive correlations between acquiescence and measures of speed, tempo and fluency. This pattern resembles the pattern discovered by Couch and Keniston in their clinical assessment of subjects scoring highest and lowest on the OAS acquiescence measure:

> . . . we asked the subjects about their attitude toward questionnaires. Yeasayers and naysayers described how they approached the task of answering Likert scale items in consistently different ways. The yeasayer responded to the "surface" of the items' meaning, did not reflect on their possible full implications, and thus gave impulsive, immediate answers. This pattern is consistent with the yeasayer's preference for blunt, straightforward, and extreme opinions, which reflect his emotional feelings at the moment. In quite an opposite manner, the naysayer analyzes and dissects each item, considering it from several points of view and attempting to be "logical" in his responses to the highly "interpreted" statements of the questionnaires (Couch & Keniston, 1960, pp.).

Some simplification would result if low verbal interpretive skill and measures of speed, tempo and fluency were highly correlated, but as Damarin and Messick observe:

> There is no evidence that acquiescers of this type have either higher or lower ability than their opposite numbers, the naysayers. Rather, subjects who are high on this pattern seem to acquiesce as part of a generally rapid, slap-dash approach to many of the tasks put before them (Damarin & Messick, 1965, p. 43).

We have, then, two psychological interpretations of acquiescent responding, which Messick (1967) called "interpretive acquiescence" and "impulsive acquiescence." The first is intellectually based, while the second is temperamentally based.

The articulation between the agreement-endorsement distinction arising from the MMPI controversy, and the interpretive-impulsive distinction which has just been made, is not fully clear:

> Further research is required to see if the agreement tendency results primarily from interpretive difficulties and the endorsement tendency primarily from impulsiveness, or whether both intellectual and temperamental processes contribute, perhaps in different proportions, to both agreement and endorsement response tendencies (Messick, 1967, p. 144). [My preferred term "endorsement" has been substituted in this passage for Messick's preferred term "acceptance" for the newer kind of acquiescence; no difference in the interpretation of the two kinds of acquiescence is implied, but the terms "agreement" and "endorsement" facilitate a symbolization of agreement and endorsement components, in later chapters, by α and η].

Pattern I acquiescence had been observed primarily with acquiescence measures derived from the F-scale and other attitude surveys, while the Couch and Keniston OAS measure was derived from items on personality inventories such as the MMPI. It seemed possible that the apparent failure of acquiescence to show up in reversal studies of the MMPI was attributable to the fact that Pattern II acquiescence was more important for the MMPI than Pattern I acquiescence. Noting the ambiguity in the notion of item reversal--polar-opposite versus negation--Jackson and Messick (1965) suggested that negation reversals served to operationalize Pattern I acquiescence or agreement, while polar-opposite reversals would serve to operationalize Pattern II acquiescence or endorsement. Thus the lines were drawn, pairing the interpretive-temperamental acquiescence distinction based on an examination of the correlates of acquiescence, with the agreement-endorsement distinction based on two different ways of operationalizing acquiescence in reversal studies.

In the first study designed with the agreement-endorsement distinction in mind, Jackson and Lay (1968) prepared sets of negation and polar-opposite reversals for five PRF scales. This study provided no evidence that either kind of acquiescence was found in the experimental scales, which Jackson and Lay attributed to the fact that the PRF items used, which had been drawn from the published form of the PRF, were highly content-saturated. No component analysis of the scales was performed, but the factor analysis suggested that acquiescence would not be a sizable component. Also, none of the content factors appeared to be correlated with acquiescence markers based on heterogeneous MMPI items.

The most ambitious attempt to construct faceted scales suitable for measuring both agreement and endorsement is Morf's (1968) dissertation (Morf & Jackson, 1972). Morf used items originally written for four PRF scales, but not included in the published form because of somewhat modest content saturations, and prepared _true_ and _false_ keyed and _positively_ and negatively worded variations on the scales, in a 2 x 2 design. In a factor analysis of the data, Morf found an agreement and endorsement factors showing the predicted pattern in Figure 1.1(b) (cf. Morf & Jackson, 1972, Figure 1, p. 345). The two factors accounted for 30% of the common variance (12% of the total variance), and provided clear support for the two-factor theory of acquiescence.

Implications for Research

No study has yet been made in which agreement and endorsement acquiescence have been measured, using reversal methods, along with measures of the intellectual and temperamental processes thought to underlie them. Thus, it would seem desirable to measure agreement, endorsement, verbal ability and speed simultaneously and study their interrelationships (Messick, 1967). A detailed statement of hypotheses about these relationships is reserved for Chapter 4.

The method of choice for analyzing the data is some form of component structure analysis. Chapter 2 discusses component structure methods as developed up to the time of the Bock et al. (1969) study of MMPI data. Recent developments in maximum likelihood methods for covariance structure analysis provide even better ways of testing structural hypotheses for faceted test data (Jöreskog, 1974), and these methods are also reviewed in Chapter 2.

Chapter 3 applies these methods to a reanalysis of Morf's (1968) data, which provide the principal direct support for the two-factor theory of acquiescence. This reanalysis raises a number of questions about the strength of the support offered by the Morf study, and suggests a number of ways in which that study can be improved.

Chapter 2

Component Analysis and Covariance Structure Analysis

This chapter reviews two kinds of models for analyzing faceted test data. The initial concept of the dissertation was to use component analysis, which had been applied to F-scale data by Chapman and Bock (1958) and to MMPI data by Bock (1964; Bock et al., 1969). The methods seemed generalizable to more complex designs than Bock had considered. In fact, Bock (1960) had indicated that "a monograph reporting the results and the steps in the general method" was in preparation, but no such monograph seems to have appeared. The component analysis model provides an interesting approach to faceted test data, and on stated assumptions it yields computable estimates of the variance components in test data. The model tends to break down when the assumptions are not met, as we shall see, but it can still provide useful information about test data, especially with small datasets.

The first part of the chapter reviews component analysis as applied to $2^{\underline{k}}$ designs, and illustrates the method with several datasets. Component analysis has been largely superseded as a distinct kind of analysis, partly by work on the general multivariate approach to ANOVA (e.g., Bock, 1975), and partly also by the work on maximum likelihood methods of covariance structure analysis. The second part of the chapter reviews covariance structure analysis. This method is illustrated in Chapter 3.

Component Structure Analysis

We begin with $2^{\underline{k}}$ measures of the same content, in a $2 \times 2 \times \ldots \times 2$ design. We will speak of the latent sources of variance underlying the observed measures as "content" and $\underline{k}$ facets or, collectively, as the $\underline{k}+1$ "variance components." The structural model for the observed measures may be written:

$$[2.1] \qquad x = \mu + Bz + e ,$$

where:

- x is a $2^{\underline{k}} \times 1$ observable random vector of test scores, with $\underline{E}(x) = \mu$;
- $z = (\zeta, \alpha, \beta, \gamma \ldots)'$ is a $(\underline{k}+1) \times 1$ latent random vector of scores on the component variates, specialized as a content component (ζ) and $\underline{k}$ components attributable to the facets of the design $(\alpha, \beta, \gamma \ldots)$, with mean vector $\underline{E}(z) = 0$;
- e is a $2^{\underline{k}} \times 1$ latent random vector of error variates, with mean vector $\underline{E}(e) = 0$; and
- B is a $2^{\underline{k}} \times (\underline{k}+1)$ design or pattern matrix of fixed constants.

Covariance matrices will be symbolized by a capital-C with subscripts, so that $\underline{E}(xx') = C_{xx}$, $\underline{E}(ze') = C_{ze}$, etc. In general, we will assume that the error variates are uncorrelated with the latent components $(C_{ze} = C_{ez} = 0)$, and with one another (C_{ee} = diagonal) ; sometimes we will make the more restrictive assumption that the error variances are equal ($C_{ee} = \sigma^2 I$) . Taking crossproducts and expectations of [2.1], we may write the basic expression for the covariance structure of the observed measures as:

[2.2] $$C_{xx} = B C_{zz} B' + C_{ee} ,$$

which expresses the known covariances C_{xx} in terms of the known parameters in B and the unknown covariances C_{zz} and C_{ee} . [2.2] is formally equivalent to the standard factor analysis model with correlated factors, except that the parameters in B are known rather than unknown (Burt, 1947). By solving [2.2] for the component structure of the latent components, a consistent solution will be:

[2.3] $$C_{zz} = B^- C_{xx} B'^- - B^- C_{ee} B'^- ,$$

where B^- is any g-inverse of B such that $B^-B = I$. As it happens, we can always take B to be orthogonal by columns such that $B'B = 2^k I$, which entails a simple form for B with coefficients equal ± 1 , and also a simple form for the g-inverse of B , namely $B^- = 2^{-k} B'$. For example, for one, two and three facet models, the appropriate design matrices are:

[2.4a] $$B_1 = \begin{pmatrix} \zeta & \alpha \\ 1 & 1 \\ 1 & -1 \end{pmatrix} ,$$

$$[2.4b] \qquad B_2 = \begin{pmatrix} \zeta & \alpha & \beta \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} , \quad \text{and}$$

$$[2.4c] \qquad B_3 = \begin{pmatrix} \zeta & \alpha & \beta & \gamma \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix} .$$

In each case, the first column of $B_{\underline{i}}$ in [2.4] corresponds to the content component, which implies that the tests have all been scored in the content direction, and the subsequent columns correspond to successive facets of the design. The designs all imply main-effect models, and the tests may always be ordered in x so that the designs in [2.4] apply. In each case, it can be confirmed that $B'B = 2^{\underline{k}} I$, and taking B^{-} to be the left g-inverse of B, we have immediately:

$$[2.5] \qquad B^{-} = (B'B)^{-1}B' = 2^{-\underline{k}} B' .$$

With this result, [2.3] may be rewritten:

$$[2.6] \qquad C_{zz} = 2^{-2\underline{k}} \left[B' C_{xx} B - B' C_{ee} B \right] .$$

We consider now a general strategy for component structure analysis, implicit in the Chapman and Bock (1958) and Bock et al. (1969) analyses, which involves the order-$\underline{k}$ Hadamard "solution" matrix. The order-$\underline{k}$ Hadamard matrix $H_{\underline{k}}$ may be generated as the Kronecker product $H_1 \otimes H_1 \otimes \ldots \otimes H_1$ with $\underline{k}$ terms, beginning with H_1 below, and for order 1 to 3, we have:

$$[2.7a] \qquad H_1 = 2^{-\frac{1}{2}} \begin{pmatrix} \zeta & \alpha \\ 1 & 1 \\ 1 & -1 \end{pmatrix} ,$$

$$[2.7b] \qquad H_2 = 2^{-1} \begin{pmatrix} \zeta & \alpha & \beta & \alpha\beta \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} , \text{ and}$$

$$[2.7c] \qquad H_3 = 2^{-3/2} \begin{pmatrix} \zeta & \alpha & \beta & \gamma & \alpha\beta & \alpha\gamma & \beta\gamma & \alpha\beta\gamma \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix} .$$

Some general properties of this class of matrices are readily apparent. They are square, not necessarily symmetric matrices of order $2^{\underline{k}} \times 2^{\underline{k}}$. They are orthonormal by columns so that $H'H = I$, which implies $H^{-1} = H'$. Finally, the design effects are in "conventional order," with main effects followed by successively higher degrees of interaction effects (Bock, 1975, pp. 273-277). The first $\underline{k}$+1 columns of [2.7] correspond to the design patterns of [2.4], but note that H_k has more columns than $B_{\underline{k}}$ (for $\underline{k} > 1$). The surplus columns of $H_{\underline{k}}$ have been labeled as interactions involving the facets of the design as, logically, they are. Whether the effect of such interactions on the component structure can be investigated depends partly on the assumptions made about the data, and partly also on the degrees of freedom available for exploring alternate models. In general, we will assume that no interactions are present.

As we shall see when we examine examples of $2^{\underline{k}}$ designs, the observed covariances in C_{xx} may be characterized as a complex function of the latent variates. Considerable simplification results when both sides of [2.2] are postmultiplied by $H_{\underline{k}}$ of the appropriate order, and premultiplied by its transpose. The simplification results from the fact that

$$\left.\begin{array}{llll} [2.8a] & H'_{\underline{k}} B_{\underline{k}} = 2^{\frac{1}{2}} I & \text{for } \underline{k} = 1 \\ [2.8b] & \quad \text{or} \quad = 2^{\underline{k}/2} \begin{pmatrix} I \\ 0 \end{pmatrix} & \text{for } \underline{k} > 1 \end{array}\right\} ,$$

where the $(\underline{k}+1)$-square identity matrix is bordered on the bottom, in the general case [2.8b], by a $(2^{\underline{k}} - \underline{k} - 1) \times (\underline{k} + 1)$ null matrix. Let $H' C_{xx} H = V$, say. Then from [2.2], we can obtain:

$$\left.\begin{array}{llll} [2.9a] & V = 2\,C_{zz} + H' C_{ee} H & \text{for } \underline{k} = 1 \\ [2.9b] & \quad \text{or} \quad = 2^{\underline{k}} \begin{pmatrix} C_{zz} & 0 \\ 0 & 0 \end{pmatrix} + H' C_{ee} H & \text{for } \underline{k} > 1 \end{array}\right\} .$$

The right side of [2.9] may be characterized as a simple function of the component variances in C_{zz}, but a complex function of the error variances in C_{ee} . Still further simplification results if we can assume homogeneous error, with $C_{ee} = \sigma_e^2 I$, which entails $H' C_{ee} H = \sigma_e^2 I$ as well. In particular, with the error variances homogeneous and presumed known, the variance components may be found simply as:

$$[2.10] \qquad \begin{pmatrix} C_{zz} & 0 \\ 0 & 0 \end{pmatrix} = 2^{-\underline{k}} \left(V - \sigma_\epsilon^2 I \right)$$

in the general case. The presence of zero or near-zero elements on the left side of [2.10] serves as a partial check on the assumptions of the model. We may now proceed to work out some of these principles of analysis in a few concrete instances.

<u>A 2-component (1-facet) design.</u> Table 2.1 displays the developments in equations [2.1]-[2.10] for a 2-component design. The design follows the model and concepts of Chapman and Bock's (1958) analysis of the F-scale. In the structural model for individual scores, corresponding to Chapman-Bock equations [1]-[2] and [2.1] of this paper, t and f represent the observed scores on <u>true</u> and <u>false</u> keyed versions of the F-scale. The population means and errors of measurement are represented by μ_t, μ_f, ϵ_t and ϵ_f, and the latent sources of reliable variance are the content and acquiescence scores ζ and α.

The covariance structure of the observed measures, in the second panel of Table 2.1, summarizes the relationships given by Chapman-Bock equations [3], [4] and [6]. It is apparent from the covariance structure that there are five unknown elements (σ_ζ^2, σ_α^2, $\sigma_{\zeta\alpha}$, $\sigma_{\epsilon t}^2$ and $\sigma_{\epsilon f}^2$), but only three known elements (σ_t^2, σ_f^2 and σ_{tf}), so the solution for the elements of C_{zz} is underidentified. The solution is a just-identified one if we assume zero error variances, but that assumption is not very plausible. Assuming that the error variances are equal eliminates one of the unknowns, but still does not yield an identified solution. One defensible strategy is to use estimates of reliability ρ_{tt} and ρ_{ff}, obtained from item analyses of the <u>true</u> and <u>false</u> tests, to estimate the "true score" (substantive) components

Table 2.1

Component Analysis for 2-Component Design

Structural Model

$$x = \mu + B z + e$$

$$\begin{pmatrix} t \\ f \end{pmatrix} = \begin{pmatrix} \mu_t \\ \mu_f \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \zeta \\ \alpha \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \epsilon_f \end{pmatrix}$$

Covariance Structure of Observed Measures

$$C_{xx} = B \; C_{zz} \; B' + E$$

$$\begin{pmatrix} \sigma_t^2 & \sigma_{tf} \\ \sigma_{tf} & \sigma_f^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_\zeta^2 & \sigma_{\zeta\alpha} \\ \sigma_{\zeta\alpha} & \sigma_\alpha^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} + \begin{pmatrix} \sigma_{\epsilon_t}^2 & \\ & \sigma_{\epsilon_f}^2 \end{pmatrix}$$

Solution Operation

$$V = H_1' \; C_{xx} \; H_1$$

$$V = \tfrac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_t^2 & \sigma_{tf} \\ \sigma_{tf} & \sigma_f^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Component Structure of the Latent Variates

$$C_{zz} = 2^{-k} (\; V \; - \; \sigma_\epsilon^2 I \;)$$

$$\begin{pmatrix} \sigma_\zeta^2 & \sigma_{\zeta\alpha} \\ \sigma_{\zeta\alpha} & \sigma_\alpha^2 \end{pmatrix} = \tfrac{1}{2} \left\{ \tfrac{1}{2} \begin{pmatrix} \sigma_t^2 + \sigma_f^2 + 2\sigma_{tf} & \sigma_t^2 - \sigma_f^2 \\ \sigma_t^2 - \sigma_f^2 & \sigma_t^2 + \sigma_f^2 - 2\sigma_{tf} \end{pmatrix} - \begin{pmatrix} \sigma_\epsilon^2 & \\ & \sigma_\epsilon^2 \end{pmatrix} \right\}$$

of variance, and then let the residuals estimate the error variances. That is, we assume that the error variances are known to be:

[2.11a] $\sigma^2_{\epsilon_t} = (1 - \rho_{tt})\sigma^2_t$

[2.11b] $\sigma^2_{\epsilon_f} = (1 - \rho_{ff})\sigma^2_f$

Now the solution is a just-identified one, and the estimates of the variance components in C_{zz} are available by straightforward algebraic methods. If estimates of ρ_{ff} are not available, another defensible strategy is to assume $\sigma^2_{\epsilon_t} = \sigma^2_{\epsilon_f}$ and then use [2.11a] to estimate the common error variance.

The "solution operation" of Table 2.1 is implicit in the Chapman-Bock analysis, and explicit in the Bock et al. (1969) analysis of MMPI data. It amounts to a rotation of the axes of C_{xx} through 45°, and leads to compact formulas for the variance components of C_{zz}, in the bottom panel of Table 2.1. The solution for C_{zz} summarizes the relationships given by Chapman-Bock equations [5], [7] and [8].

Table 2.2 contains a reworking of the Chapman-Bock analyses for six F-scale studies. Studies missing estimates of ρ_{tt} or ρ_{ff} are omitted, and Bass's (1955) study is added to the studies considered by Chapman and Bock. All of the variances reported in the original studies have been rescaled so that $\sigma_t{}^2 = 1.000$, making it easier to compare results across

Table 2.2

Component Analysis for a Panel of 6 Studies

(after Chapman and Bock, 1958)

Statistics	Studies[a]					
	A	B	C	D	E	F
$\underline{N}$ subjects	84	134	144	184	152	346
$\underline{n}$ items/scale	28	10	16	30	10	10
	Data[b]					
σ_f^2	.313	.600	.751	.454	.859	.716
σ_{tf}	.112	.132	-.009	.195	.046	.161
$\underline{v}_{11}$	.768	.932	.867	.922	.976	1.019
$\underline{v}_{22}$	.544	.668	.884	.532	.883	.697
$\underline{v}_{12}$	.344	.200	.125	.273	.070	.142
ρ_{tt}	.81	.69	.53	.71	.52	.60
ρ_{ff}	.50	.41	.41	.42	.43	.42
	Estimates					
$\hat{\sigma}_{\epsilon_t}^2$	.190	.310	.470	.290	.480	.400
$\hat{\sigma}_{\epsilon_f}^2$	.156	.354	.443	.263	.490	.415
$\hat{\sigma}_{\epsilon}^2$	.173	.332	.457	.277	.485	.408
$\hat{\sigma}_{\zeta}^2$	.297	.300	.205	.323	.246	.306
$\hat{\sigma}_{\alpha}^2$	.186	.168	.214	.127	.199	.145
$2\hat{\sigma}_{\zeta\alpha}$	.344	.200	.124	.273	.070	.141
$\hat{\rho}_{\zeta\alpha}$	.732	.445	.297	.673	.159	.337

Table 2.2

(continued)

[a]References to original studies: A - Bass, 1955; B, C - Chapman & Campbell, 1957; D - Chapman & Campbell, unpublished; E, F - Christie et al., 1958.

[b]Published variances were rescaled so that $\sigma_t^2 = 1.000$ for all studies. Elements of V are obtained by transformation of C_{xx}. Reliabilities were obtained by split-half methods in study A, by coefficient-α in studies B, C and D, and by unreported methods in studies E and F.

studies, and σ_f^2 and σ_{tf} (computed as $\sigma_{tf} = r_{tf}\sigma_t\sigma_f$) have been adjusted accordingly. Note that while the calculations are reported with three decimal places, the accuracy of the calculations is limited to two decimal places, at most, owing to the influence of r_{tf}, ρ_{tt} and ρ_{ff} on the results.

The first point to notice in Table 2.2 is that the variances and reliabilities of the false scales are markedly lower than those of the true scales. These are phenomena which the component analysis model attempts to explain, and attributes to acquiescence or, more precisely, to a content-acquiescence correlation. The variances of true and false scales under the model are:

[2.12a] $$\sigma_t^2 = \sigma_\zeta^2 + \sigma_\alpha^2 + 2\sigma_{\zeta\alpha} + \sigma_\epsilon^2,$$ and

[2.12b] $$\sigma_f^2 = \sigma_\zeta^2 + \sigma_\alpha^2 - 2\sigma_{\zeta\alpha} + \sigma_\epsilon^2,$$

so assuming equal error variances, the true scales will have larger variances whenever there is a positive content-acquiescence covariance; and under the same circumstances, the true scales will have larger apparent reliabilities.

The reliability estimates reported in Table 2.2 vary considerably in magnitude. We have already noted the differences between the reliabilities for true and false tests from the same study. One reason for the differences in reliabilities between studies is the fact that they are estimated from scales which vary in length from 10 to 30 items. Using the Spearman-Brown formula to adjust all the true scale reliabilities to a common scale length of, say, 30 items, the values for $\rho_{tt(adj.)}$ range from .68 to .87. Other reasons for the differences in reliabilities between studies include sampling variability from study to study, slightly different scoring methods, and the use of different methods for estimating reliability (coefficient-α and split-plot).

Chapman and Bock chose to complete the solutions for the variance components using two alternate values for the error variance, namely $\sigma_\epsilon^2 = .15\sigma_t^2$ and $\sigma_\epsilon^2 = .30\sigma_t^2$. Table 2.2 uses error variances estimated separately for each individual study. An empirical check on the assumption of equal error variances can be made using [2.11], and as shown in Table 2.2, $\hat{\sigma}^2_{\epsilon_t}$ and $\hat{\sigma}^2_{\epsilon_f}$ are very similar within each study, and do not differ significantly. Accordingly, the error variances were pooled within studies using $\hat{\sigma}_\epsilon^2 = \frac{1}{2}(\hat{\sigma}^2_{\epsilon_t} + \hat{\sigma}^2_{\epsilon_f})$. Finally, using [2.10] and $\hat{\rho}_{\zeta\alpha} = \hat{\sigma}_{\zeta\alpha} / \hat{\sigma}_\zeta \hat{\sigma}_\alpha$, the variance components and correlations between latent variates were obtained, as shown in the last four rows of Table 2.2.

Since the data were scaled to yield true scale variances of 1, the variance components in Table 2.2 are directly interpretable as proportions of true scale variance. A check on the computations (for the first study) is given by:

$$\sigma_t^2 = .297 + .186 + 2 \times .732 \times \sqrt{.297 \times .186} + .173 = 1. ,$$

within rounding error. All six studies show substantial variance components for acquiescence, ranging from 13% to 21% of true scale variance. The acquiescence components tend to be smaller than the content components, which range from 25% to 32% of the true scale variance. The studies agree in showing positive correlations between F-scale content and acquiescence, ranging from .16 to .73. When the content-acquiescence correlation is high (the median value is about .4), a large proportion of true scale variance is attributable to the content-acquiescence covariance, so acquiescence contaminates relationships between the F-scale and outside measures both directly, as a determinant of scores, and indirectly, through its correlation with content. As Chapman and Bock observe,

> Whether the added acquiescence variance in the F scale is "valid" or "invalid" depends on what is being predicted. Prediction of a comparatively acquiescence free measure, such as ratings on Fascism by counselors, would be improved by use of the all positive form. But in predicting scores on another paper-and-pencil test which itself contains extraneous acquiescence variance, the all positive scale would produce too high a correlation (Chapman & Bock, 1958, p. 333).

Bock et al. (1969) also observe that a positive content-acquiescence correlation creates a selection pressure which favors true items during the scale construction process, since true items will be noticeably more highly correlated with a total score than false items, and will result in noticeably more reliable scales than the false items. The contribution of acquiescence to reliability is a spurious one, however, since it degrades the measure of content and artificially inflates the correlations between measures similarly contaminated by acquiescence.

The 2-component model is very limited. Error estimates must be obtained from sources external to the observed covariances, and in the Chapman-Bock model internal consistency and split-half reliabilities are used as (admittedly not very satisfactory) indications of the error in the data. Chapman and Bock suggest that additional parallel-form true and false scales should be used "over an appropriate period" to provide reliability estimates which are theoretically more appropriate for estimating error. Coefficient-α, in particular, is only a lower bound to reliability considered as a ratio of common-factor to total variance of the items comprising a scale, with equality only when the items are strictly unifactorial (e.g., McDonald, 1970). Since the items of the F-scale are necessarily multi-factorial if acquiescence and content are present in the data, coefficient-α must underestimate the reliable variance of the true and false scales. By the procedure we have adopted for the 2-component model, increasing the apparent reliability would decrease the error variance, increase the content and acquiescence components, and decrease the content-acquiescence correlation. Provision of parallel-form measurements would increase the number of tests available for analysis, leading perhaps to larger component models. In general, increasing the number of measures and components provides more flexibility in analysis and better opportunities for estimating components of variance for a given kind of test. Ultimately, measures of a given kind must be related to other measures bearing on the construct validity of the components, so it is useful to carry out a component analysis in tandem with an analysis of related measures. Otherwise one is left with an internal analysis of only one kind of test--F-scales, in the present case--which is not terribly informative.

Some interesting methodological issues can be addressed by using the 2-component model in an illustrative manner, however, because derivations are easier with a 2-component model than with more complex designs. One point to notice is that the model assumes a kind of *tau*-equivalence of all components, which is related to the problem of "adequate reversal" of items discussed throughout the acquiescence literature. The use of pattern matrices B with coefficients equalling ± 1 is tantamount to the assumption that the content component is equal for all measures, the acquiescence component is equal for all measures, and so on for other components of the scales. There are several sets of reversals in the literature, including the "logical" reversals of Bass (1955) who reversed the F-scale items primarily using negation, and the "psychological" reversals of Christie et al. (1958), who reversed the content of F-scale items along with the dogmatic, sweeping style of the items. Jackson et al. (1957), on the other hand, attempted to reverse the F-scale content while retaining the item style of the original items, and their reversals behave differently from the reversals in other studies analyzed by Chapman and Bock. They obtained

$$C_{xx} = \begin{pmatrix} 1.000 & -.239 \\ -.239 & .467 \end{pmatrix},$$

(with scales scored in the content direction, as for the other studies). Based on a split-half reliability of .77 for the false scale, $\hat{\sigma}_\epsilon^2$ is estimated by [2.11b] to be .107, and the variance components are estimated as

$$C_{zz} = \begin{pmatrix} .193 & .142 \\ .142 & .432 \end{pmatrix}, \quad \hat{\rho}_{\zeta\alpha} = .461 .$$

$\hat{\rho}_{\zeta\alpha}$ is in the range found for other studies, but $\hat{\sigma}_\alpha^2$ is over twice the size of $\hat{\sigma}_\zeta^2$, when most of the other studies found $\hat{\sigma}_\alpha^2 \leq \hat{\sigma}_\zeta^2$. Chapman and Bock considered the Christie et al. reversals to be superior, and rejected the Jackson et al. results as being due to inadequate reversal, but the matter bears further study. Christie et al. and Jackson et al. apparently agree that there are really three components of the F-scale--content, acquiescence, and a third variable which may be called "overgeneralization." Representing the third component by β, the implicit model for true scales would be:

$$t = \zeta + \alpha + \beta + \epsilon_t$$

By all accounts, Christie et al. constructed a reversed scale having (after scoring in the content direction) the structure

$$f_1 = \zeta - \alpha + \beta + \epsilon_{f_1} = (\zeta + \beta) - \alpha + \epsilon_{f_1} ,$$

while Jackson et al. constucted a reversed scale

$$f_2 = \zeta - \alpha - \beta + \epsilon_{f_2} = \zeta - (\alpha + \beta) + \epsilon_{f_2} .$$

A 2-component model cannot discriminate among three components, and the Christie et al. reversals evidently confound overgeneralization with content, while the Jackson et al. reversals confound it with acquiescence. Jackson et al. conceptualized overgeneralization as a response style component which would properly be lumped with acquiescence, while Christie et al. conceptualized overgeneralization as a component of content. From the results with the two kinds of reversals, we may conclude that $\sigma_\zeta^2 + \sigma_\beta^2 > \sigma_\alpha^2$ and $\sigma_\alpha^2 + \sigma_\beta^2 > \sigma_\zeta^2$, but it would be preferable to have a 3-component model enabling us to discriminate all three components. Such a model will be discussed later in the chapter.

Another methodological issue which may be addressed using the 2-component model is the issue of a proper metric for the analysis. If the structural model in Table 2.2 holds in the metric of the raw data, and the variance and error components are equal in true and false scales, the variance components will not necessarily be equal in rescaled data: In particular, they cannot be equal in standardized data if there is a content-acquiescence correlation. Examination of [2.12] will show that σ_{ζ}^2, σ_{α}^2 and σ_{ϵ}^2 will not be equal for standardized scales weighted by $1/\sigma_t$ and $1/\sigma_f$. The routine standardization of scales, in factor analytic research, can serve to obscure relationships which are apparent in the metric of the raw data. For example, Bock et al. found that the MMPI Pt scale has a larger content component than the Hy scale, but that the acquiescence components and error variances were approximately equal in absolute terms. This is interesting, since "we would not expect the subject's tendency to acquiesce to vary from scale to scale in an instrument in which items from different scales are intermingled" (Bock et al., 1969, p. 133). If the finding of equal acquiescence components for the two MMPI scales can be replicated, it would imply that we can use simpler models for the joint behavior of the two scales than if acquiescence components have different sizes in the two scales. The finding of equality would be difficult to discern, however, if the data had been standardized prior to analysis. Standardizing scales can have the effect of distributing artifacts of content saturation and content-acquiescence balance throughout a test battery. We can see from the Bock et al. component analysis that acquiescence is a larger proportion of total Hy variance than of total Pt variance, and an analysis of correlations would emphasize that fact, to the neglect of the finding that the acquiescence components and error variances are equal in absolute terms.

Chapman and Bock (1958, p. 332) indicate that the variance components in the 2-component model may be tested for significance using formulas equivalent to:

[2.13a] $(\hat{\sigma}_{\epsilon}^2 + 2\hat{\sigma}_{\zeta}^2)/\hat{\sigma}_{\epsilon}^2 \sim F(N-1, \infty)$ (content)

[2.13b] $(\hat{\sigma}_{\epsilon}^2 + 2\hat{\sigma}_{\alpha}^2)/\hat{\sigma}_{\epsilon}^2 \sim F(N-1, \infty)$ (acquiescence).

These tests assume that the error variances

of the <u>true</u> and <u>false</u> scales are normally distributed, equal, and known. The numerators must be significantly larger than the denominators in order for the variance components to be considered significantly nonzero. Using either $\hat{\sigma}_{\epsilon}^2 = .30\,\sigma_t^2$, which Chapman and Bock considered a conservative estimate of error, or the usually larger estimates of error in Table 2.2, the content and acquiescence components are significantly nonzero ($p < .01$) in all of the studies considered.

<u>A 3-component (2-facet) design</u>. For a 3-component model, we need four tests of the same content, and a 2 x 2 design on the tests will enable us to estimate two method or response style components in addition to the content component. Bock et al. (1969) have provided one example of a 2-facet model, using MMPI data. By using original and reversed scales on two separate occasions, with the same subjects, they were able to estimate variance components for content, acquiescence and trait instability (time 1 vs. time 2, a period of one week), for both the *Pt* and the *Hy* scale.

A variation of the Bock et al. 3-component model will be illustrated here, using F-scale data from Morf's (1968) study. The data will be discussed further in Chapter 3, in connection with a covariance structure analysis of the same data. Morf administered four scales to the same subjects, designated as follows:

RT (<u>Relative</u> phrasing, <u>true</u>-keyed)
AT (<u>Absolute</u> phrasing, <u>true</u>-keyed)
AF (<u>Absolute</u> phrasing, <u>false</u>-keyed)
RF (<u>Relative</u> phrasing, <u>false</u>-keyed)

Thus, the data permit an analysis for components of content (ζ), acquiescence (α), and overgeneralization (β). The top panel of Table 2.3 displays the structural model for the data. The scales have been reordered slightly so that, after scoring in

Table 2.3

Component Analysis for a 3-Component Design (F-scale)

Structural Model

$$\mathbf{x} = \mu + B \quad z + e$$

$$\begin{matrix} RT \\ AT \\ AF \\ RF \end{matrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \mu + \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \zeta \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

Covariance Structure of Observed Measures

$$C_{xx} = B\, C_{zz}\, B' + C_{ee} = \begin{pmatrix} 1.0000 & & & \text{sym.} \\ .6035 & .9764 & & \\ .0113 & -.0307 & .5113 & \\ -.0019 & -.0293 & .2280 & .6168 \end{pmatrix}$$

Solution Operation

$$V = H' C_{xx} H = 4 \begin{pmatrix} C_{zz} & 0 \\ 0 & 0 \end{pmatrix} + H' C_{ee} H$$

$$= \begin{pmatrix} 1.1666 & & & \text{sym.} \\ .3998 & 1.2172 & & \\ -.0002 & .0179 & .3677 & \\ .0467 & -.0408 & .0243 & .3531 \end{pmatrix}$$

the content direction, the appropriate design matrix has the form of [2.4b]. The observed covariances are given in the second panel of Table 2.3. For convenience, the covariance matrix has been scaled so that the largest variance (for the RT scale) is one. As in [2.2], the expected values of the observed covariances may be found as $B C_{zz} B' + C_{ee}$, where

$$C_{zz} = \begin{pmatrix} \sigma_\zeta^2 & & \text{sym} \\ \sigma_{\zeta\alpha} & \sigma_\alpha^2 & \\ \sigma_{\zeta\beta} & \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix} , \quad \text{and}$$

$$C_{ee} = \text{diag}(\ \sigma_{\epsilon_1}^2 ,\ \sigma_{\epsilon_2}^2 ,\ \sigma_{\epsilon_3}^2 ,\ \sigma_{\epsilon_4}^2\ ) .$$

Finally, using the solution matrix H of [2.7b], the obtained value for V of [2.9b] is given in the bottom panel of Table 2.3.

Before discussing the properties of V, it may be noted that with the assumption of unequal error variances, the solution for the 3-component model is a just-identified one, since there are 10 observed values in C_{xx} and 10 unknown values in C_{zz} and C_{ee}. A simultaneous solution for the unknowns may be obtained by stringing out the elements of C_{xx} into a vector u, say, and stringing out the elements of C_{zz} and C_{ee} into another vector y. Then the unknowns may be found by solving the system of equations

$$[2.14] \qquad u = Ly ,$$

where L is a 10 x 10 matrix of coefficients expressing the expected values of elements of u in terms of the expected values of elements of y . The system [2.14] is shown in detail in Table 2.4. This is a consistent system of equations having the unique solution given in Table 2.5.

What can we say about the solution in Table 2.5? First, it is of interest to test the error variances for homogeneity, using Bartlett's test, and we obtain $\chi^2(3) = 16.02$, $p < .005$. Apparently the AF measure has a smaller error variance than the others (it is also the measure with the smallest observed variance). Second, we note that $\sigma_\alpha^2 = .216$ is slightly larger than $\sigma_\zeta^2 = .203$, and that both are in the range found for these components in the 2-component analyses in Table 2.2. The content-acquiescence correlation may be estimated as .448 , which is positive and moderately large, as we have found earlier. The overgeneralization component $\sigma_\beta^2 = .004$ is rather small, however, and is not (as we will see in a moment) significantly different from zero. As discussed in Chapter 3, there are certain problems with the manner in which these F-scale reversals were obtained, making it difficult to interpret the results unequivocally in terms of response style hypotheses. Taking the data at face value, however, we would have to conclude that the hypothesis of an overgeneralization component for the data is not confirmed.

Table 2.4

Expected Values of C_{xx} for a 3-Component Design

$$
\begin{array}{ccccc}
u & = & L & & y
\end{array}
$$

$$
\begin{pmatrix} \sigma_{11}^2 \\ \sigma_{22}^2 \\ \sigma_{33}^2 \\ \sigma_{44}^2 \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{32} \\ \sigma_{41} \\ \sigma_{42} \\ \sigma_{43} \end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 1 & 2 & 2 & 2 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 2 & -2 & -2 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & -2 & 2 & -2 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & -2 & -2 & 2 & 0 & 0 & 0 & 1 \\
1 & 1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & -1 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
1 & -1 & -1 & 0 & 0 & -2 & 0 & 0 & 0 & 0 \\
1 & -1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & -1 & -2 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix} \sigma_{\zeta}^2 \\ \sigma_{\alpha}^2 \\ \sigma_{\beta}^2 \\ \sigma_{\zeta\alpha} \\ \sigma_{\zeta\beta} \\ \sigma_{\alpha\beta} \\ \sigma_{\epsilon_1}^2 \\ \sigma_{\epsilon_2}^2 \\ \sigma_{\epsilon_3}^2 \\ \sigma_{\epsilon_4}^2 \end{pmatrix}
$$

Table 2.5

Solution for Variance Components from Table 2.4

<u>Estimates of</u> C_{zz} <u>(Correlations above Diagonal)</u>

$$C_{zz} = \begin{pmatrix} \sigma_\zeta^2 & & \\ \sigma_{\zeta\alpha} & \sigma_\alpha^2 & \\ \sigma_{\zeta\beta} & \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} .2034 & (.4480) & (.3670) \\ .0939 & .2160 & (-.2547) \\ .0101 & -.0072 & .0036 \end{pmatrix}$$

<u>Estimates of</u> C_{ee}

$$C_{ee} = \text{diag}(\ .3833,\ .3715,\ .2413,\ .4162\ )$$

	RT	AT	AF	RF
C_{ee} = diag	.3833	.3715	.2413	.4162

The expected values of V as a function of the latent components are given in Table 2.6, under the assumptions of homogeneous and heterogeneous error. It can be seen that the expected values of V are simpler than the expected values of C_{xx} in Table 2.4, especially under the assumption of homogeneous error. Expressions for the variance components are readily derived from Table 2.6. Whether error is homogeneous or not, the variance components are estimated by

$$\hat{\sigma}_{\zeta}^2 = \tfrac{1}{4}(\underline{v}_{11} - \underline{v}_{44}) ,$$

$$\hat{\sigma}_{\alpha}^2 = \tfrac{1}{4}(\underline{v}_{22} - \underline{v}_{44}) , \quad \text{and}$$

$$\hat{\sigma}_{\beta}^2 = \tfrac{1}{4}(\underline{v}_{33} - \underline{v}_{44}) ,$$

yielding (within rounding error) the results in Table 2.5. Significance tests for the hypotheses that the variance components are zero are given in the bottom panel of Table 2.6, and assume multivariate normality for the observed measures (Bock et al., 1969, p. 133). For the present data, the critical $\underline{F}_{.05;193,193} = 1.26$, and we find:

$\underline{F}(\text{content}) = 3.30$, $\underline{p} < .0001$

$\underline{F}(\text{acquiescence}) = 3.45$, $\underline{p} < .0001$

$\underline{F}(\text{overgeneralization}) = 1.04$, n.s.

As mentioned earlier, the overgeneralization component is evidently null.

Table 2.6

Expected Values of V for a 3-Component Design

Element	Assuming $C_{ee} = \sigma^2 I$	Assuming C_{ee} = diag.
$\underline{v}_{11}$	$4\sigma_\zeta^2 + \sigma_\epsilon^2$	$4\sigma_\zeta^2 + \frac{1}{4}\Sigma\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{22}$	$4\sigma_\alpha^2 + \sigma_\epsilon^2$	$4\sigma_\alpha^2 + \frac{1}{4}\Sigma\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{33}$	$4\sigma_\beta^2 + \sigma_\epsilon^2$	$4\sigma_\beta^2 + \frac{1}{4}\Sigma\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{44}$	σ_ϵ^2	$\frac{1}{4}\Sigma\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{21}$	$4\sigma_{\zeta\alpha}$	$4\sigma_{\zeta\alpha} + \frac{1}{2}\Sigma\underline{h}_{2\underline{i}}\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{31}$	$4\sigma_{\zeta\beta}$	$4\sigma_{\zeta\beta} + \frac{1}{2}\Sigma\underline{h}_{3\underline{i}}\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{32}$	$4\sigma_{\alpha\beta}$	$4\sigma_{\alpha\beta} + \frac{1}{2}\Sigma\underline{h}_{4\underline{i}}\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{41}$	0	$\frac{1}{2}\Sigma\underline{h}_{4\underline{i}}\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{42}$	0	$\frac{1}{2}\Sigma\underline{h}_{3\underline{i}}\sigma_{\epsilon_{\underline{i}}}^2$
$\underline{v}_{43}$	0	$\frac{1}{2}\Sigma\underline{h}_{2\underline{i}}\sigma_{\epsilon_{\underline{i}}}^2$

Significance Tests for Components

Content:	$\underline{v}_{11}/\underline{v}_{44}$	$\sim \underline{F}(\underline{N}-1,\underline{N}-1)$
Acquiescence:	$\underline{v}_{22}/\underline{v}_{44}$	$\sim \underline{F}(\underline{N}-1,\underline{N}-1)$
Overgeneralization:	$\underline{v}_{33}/\underline{v}_{44}$	$\sim \underline{F}(\underline{N}-1,\underline{N}-1)$

Bock et al. considered the model assuming homogeneous error, and in that case, component covariances are estimated by:

$$\hat{\sigma}_{\zeta\alpha} = \tfrac{1}{4}\underline{v}_{21} ,$$
$$\hat{\sigma}_{\zeta\beta} = \tfrac{1}{4}\underline{v}_{31} , \text{ and}$$
$$\hat{\sigma}_{\alpha\beta} = \tfrac{1}{4}\underline{v}_{32} .$$

Population values of the component correlations are estimated by $\hat{\rho}_{\zeta\alpha} = \hat{\sigma}_{\zeta\alpha} / \hat{\sigma}_{\zeta}\hat{\sigma}_{\alpha}$, etc., but the sample values are estimated by $\underline{r}_{\zeta\alpha} = \underline{v}_{21}/\sqrt{\underline{v}_{11}\underline{v}_{22}}$, etc. $\hat{\rho}_{\zeta\alpha}$ resembles a correlation corrected for attenuation, while $\underline{r}_{\zeta\alpha}$ represents the correlation which would be attained by scoring the scales for content and acquiescence using

$$\hat{\zeta} = x_1 + x_2 + x_3 + x_4 \text{ and}$$
$$\hat{\alpha} = x_1 + x_2 - x_3 - x_4 ,$$

and correlating them. Assuming multivariate normality and homogeneous error, Bock et al. indicate that $\underline{r}_{\zeta\alpha}$ may be tested for significance in the usual manner, on $\underline{N}$-2 degrees of freedom, as a test of $\underline{H}_0$: $\sigma_{\zeta\alpha} = 0$.

With heterogeneous error, the off-diagonal elements of V are biased by a weighted function of the error components. The bias has the form $\frac{1}{2}\sum_{i}\underline{h}_{ji}\,\sigma^2_{\epsilon_i}$, where $\underline{h}_{ji}$ is the $\underline{ji}$-th element of H in [2.7b]. With summation over $\underline{i}$, the bias is essentially a crossproduct of the vector of error variances and a particular column of H . From the pattern of the bias terms in Table 2.6, it is clear that the component covariances may be estimated by:

$$\hat{\sigma}_{\zeta\alpha} = \tfrac{1}{4}(\underline{v}_{21} - \underline{v}_{43}) ,$$

$$\hat{\sigma}_{\zeta\beta} = \tfrac{1}{4}(\underline{v}_{31} - \underline{v}_{42}) , \quad \text{and}$$

$$\hat{\sigma}_{\alpha\beta} = \tfrac{1}{4}(\underline{v}_{32} - \underline{v}_{41}) ,$$

yielding (within rounding error) the values in Table 2.5. Given that the overgeneralization component is null, the component covariances involving overgeneralization in Table 2.5 cannot be meaningful, even though their (population) correlation estimates appear to be nontrivial. The matter requires further study, but I would speculate that the hypothesis $\underline{H}_0$: $\sigma_{\zeta\alpha} = 0$ (for example) may be evaluated by computing a sample correlation corrected for bias due to heterogeneous error by means of

[2.15] $$\underline{r}^{*}_{\zeta\alpha} = 4\hat{\sigma}_{\zeta\alpha} / \sqrt{\underline{v}_{11}\,\underline{v}_{22}} ,$$

and evaluating it on $\underline{N}$-2 degrees of freedom. Using [2.15], we obtain:

$$\underline{r}^{*}_{\zeta\alpha} = .3152 , \quad \underline{t}(192) = 4.60 , \quad \underline{p} < .0001 ,$$

$$\underline{r}^{*}_{\zeta\beta} = .0617 , \quad \underline{t}(192) = .86 , \quad \text{n.s.} , \quad \text{and}$$

$$\underline{r}^{*}_{\alpha\beta} = .0430 , \quad \underline{t}(192) = .60 , \quad \text{n.s.}$$

Thus, while the estimates of $\hat{\rho}_{\zeta\beta}$ and $\hat{\rho}_{\alpha\beta}$ in Table 2.6 appear to be nontrivial, the associated component covariances may be regarded as null, and so these component correlations should be set to zero. $\hat{\rho}_{\zeta\alpha}$, on the other hand, is clearly nonzero.

For the sake of completeness, formulas for the error variances in terms of the elements of V may be derived for the case of heterogeneous error. These turn out to be:

$$\sigma^2_{\epsilon_1} = \underline{v}_{44} + \underline{v}_{43} + \underline{v}_{42} + \underline{v}_{41} ,$$

$$\sigma^2_{\epsilon_2} = \underline{v}_{44} + \underline{v}_{43} - \underline{v}_{42} - \underline{v}_{41} ,$$

$$\sigma^2_{\epsilon_3} = \underline{v}_{44} - \underline{v}_{43} + \underline{v}_{42} - \underline{v}_{41} , \text{ and}$$

$$\sigma^2_{\epsilon_4} = \underline{v}_{44} - \underline{v}_{43} - \underline{v}_{42} + \underline{v}_{41} .$$

The sign pattern of H can be seen here as elsewhere in the data.

<u>Evaluation of component analysis as a method</u>. In principle, component analysis may be extended to any 2^k design and to designs in which the facets have more than two levels. The steps in the method are basically: (a) Write the structural model expressing the observed measures as a function of the latent variates, as in [2.1]. (b) Obtain the covariance structure for both sides of the structural model, which expresses the observed covariances in terms of the latent components, as in [2.2]. (c) Pre- and post-multiplication of the covariance structure by the appropriate solution matrix will result in a tidy simplification of the right side of equation [2.2], enabling one to express the latent components as a function of the observed covariances.

In practice, there are usually a number of ways to proceed with the analysis. For the simple case of the 2 x 2 design, Bock (1960) provided one model permitting the estimation of four components, which requires $H' C_{XX} H$ to be diagonal, and another model (Bock et al., 1969) permitting estimation of three components, which requires homogeneous error. The last section of the chapter provided yet another model for the 2 x 2 design which allows heterogeneous error. All of the models involve relatively cumbersome algebra, and a certain amount of "fiddling" with the data and model before a researcher can be reasonably certain that the model fits adequately. It would seem difficult to specify rules of analysis, for the general case, with enough precision that a general computer program could be written to perform the analysis. The analysis will often involve an artistic flair for making the minimal number of assumptions required for a just-identified model which is suitable for analysis. In a just-identified model, the matrix $V = H' C_{xx} H$ contains precisely as many nonzero expected values as there are unknown variances and covariances.

If the model is underidentified, there are fewer nonzero elements of V than there are unknowns, and no solution is possible unless we adopt the strategy adopted for the 2-component model and obtain independent estimates for some of the unknowns. If the model is overidentified, however, other problems arise. For example, we concluded that the overgeneralization component variance and covariances were essentially zero in the Morf data. Implicitly, I then ignored the third row of V , retaining the estimates of the error variances which had been based on the fourth row of V . If σ_β^2 is really zero, however, the elements of the third row of V are as much a function of the error variances as the elements of the fourth row, so both rows should be used to estimate the error variances. This yields an overidentified model, since there are now more equations than unknowns, and in sample data the system of equations will be inconsistent with probability 1. With an over-identified model, the most direct approach to a solution would be to go back to the equations [2.14] displayed in Table 2.4, striking out the columns of L and rows of

y which no longer apply. The least-squares estimates of the unknowns may then be obtained using $\hat{y} = (L'L)^{-1}L'u$ (with L missing columns 3, 5 and 6 and no longer square, as indicated). Alternately, elements of V may be expressed as a function of their expected values, and least-square estimates of the unknowns can be obtained from the resulting equations. With an overidentified model, some of the simplicity of analysis brought about by the solution operation $H' C_{xx} H$ tends to be lost. Moreover, it is no longer clear that appropriate significance tests for the components can be obtained as a simple function of the elements of V.

One of the advantages of component analysis is that, under multivariate normality assumptions, significance tests are available for the individual variance components and for aspects of fit to the model (e.g., that certain elements of V have expected values of zero in the population). Despite the completely worked small examples of component analysis provided by Bock, however, it is not clear how significance tests should be performed in more complex designs. Rules have been formulated for component analysis models involving balanced designs with uncorrelated components and homogeneous error (e.g., Searle, 1971, ch. 9), but correlated components, heterogeneous error variances and overidentified models all pose problems for a statistically rigorous development of component analysis models--and faceted test data can involve all three of these complications. Even small models do not provide an overall test of adequacy for the model, so that one must test the implications of the model in a piecemeal and often ad hoc manner, using whatever degrees of freedom are available for such tests.

For a variety of reasons, then, component analysis is not completely satisfactory as a method for analyzing faceted test data. Conclusions about the data can be reached, of course, but they require tricky and often tedious detective work. Nevertheless, component analysis provides an interesting heuristic approach to the analysis of test data, and it can be a useful and informative method for small data sets. Maximum likelihood methods for analyzing covariance structures proved a more versatile and powerful approach, however, and these are discussed later in the chapter.

<u>Relationships between component analysis and analysis of variance</u>. It may be useful here to indicate the relationships between component analysis, on the one hand, and univariate and multivariate repeated-measures ANOVA (Bock, 1975, pp. 456-464; Winer, 1972, pp. 335-347, 496-497). Consider a Subjects-by-Treatments-by-Treatments ANOVA layout, with two repeated measures factors, each having two levels. The data layout corresponds to the 2 x 2 component analysis design discussed earlier. The standard ANOVA treatment of the data treats it as a special case of a 3-way layout, with a random subjects factor crossed by two fixed treatment factors, and the analysis partitions the variance as shown in Table 2.7. The usual analysis yields the two treatment main effects and their interaction, with a separate error term for each. The error terms are formally subject-by-treatment interactions, and the hope is usually that they can be pooled for the tests of the treatments. (If the subject-by-treatment interactions are null, the three error terms each estimate the replication error σ_ε^2, which are not usually estimated unless the entire design is repeated using the same subjects.)

Note that in the expressions for the expected mean-squares, at the right of Table 2.7, the "treatment components" σ_a^2, σ_b^2 and σ_{ab}^2 are different in character from the "subject-by-treatment" components $\sigma_{\pi a}^2$, $\sigma_{\pi b}^2$ and $\sigma_{\pi ab}^2$: The "treatment components" (for want of a better term) are a function of the variance of the treatment means around the grand mean, while the "subject-by-treatment" components are a function of individual dif-

Table 2.7

Expected Mean-Squares in the

Subjects-by-Treatments-by-Treatments ANOVA[a]

Source	*df*	Mean Square	*E*(MS)
Grand mean	1	MS_G	-
Subjects	*N*-1	MS_S	$\sigma_\varepsilon^2 + 4\sigma_\pi^2$
Treatment A	1	MS_A	$\sigma_\varepsilon^2 + 2\sigma_{\pi a}^2 + 2N\sigma_a^2$
Error (A)	*N*-1	MS_{SA}	$\sigma_\varepsilon^2 + 2\sigma_{\pi a}^2$
Treatment B	1	MS_B	$\sigma_\varepsilon^2 + 2\sigma_{\pi b}^2 + 2N\sigma_b^2$
Error (B)	*N*-1	MS_{SB}	$\sigma_\varepsilon^2 + 2\sigma_{\pi b}^2$
AB Interaction	1	MS_{AB}	$\sigma_\varepsilon^2 + \sigma_{\pi ab}^2 + N\sigma_{ab}^2$
Error (AB)	*N*-1	MS_{SAB}	$\sigma_\varepsilon^2 + \sigma_{\pi ab}^2$
Replication error	-	None	σ_ε^2

[a]After Winer, 1972, p. 346; specialized for the case of the 2 x 2 design.

ferences variance around the treatment means. The "treatment components" are a quadratic function of the fixed effects, and are only analogous to variances, while the "subject-by-treatment" components are true variance components (Searle, 1971, p. 388). It is only the latter which are of interest in a component analysis.

In a multivariate ANOVA treatment of the same data, we can let the data be represented by the $N \times p$ matrix $Y = \{y_{ij}\}$, with a $p \times 1$ mean vector

$$\bar{y} = \left\{ \bar{y}_j = \frac{1}{N} \sum_{i=1}^{N} y_{ij} \right\},$$

and a $p \times p$ covariance matrix

$$S = \frac{1}{N-1} (Y'Y - N\bar{y}\bar{y}') .$$

As shown by Bock (1975), pre- and post-multiplication of the quadratic forms $N\bar{y}\bar{y}'$ and S by an orthogonal $p \times p$ solution matrix P, say, yields all the mean squares required for the univariate ANOVA in Table 2.7. For the case of a 2×2 design, the number of repeated measures is $p = 4$, and the solution matrix P is the Hadamard matrix H_2 of [2.7b]. Adapting Bock's notation, we have:

$$MSM^* = N P'\bar{y}\bar{y}'P , \text{ and}$$

$$MSE^* = P'(Y'Y - N\bar{y}\bar{y}')P/(N-1) = P'SP ,$$

for the transformed mean-square matrices for the mean vector and error, respectively. MSM^* and MSE^* are each $p \times p$ matrices, and their diagonal elements may be shown to be precisely:

$$\text{diag}[MSM^*] = [\; MS_G \,,\; MS_A \,,\; MS_B \,,\; MS_{AB} \;] \text{ and}$$

$$\text{diag}[MSE^*] = [\; MS_S \,,\; MS_{SA} \,,\; MS_{SB} \,,\; MS_{SAB} \;] .$$

Thus, the univariate F-ratios for the treatment effects are obtained by taking the last $p - 1$ diagonal elements of MSM* as numerators, and the corresponding diagonal elements of MSE* as denominators. Moreover, if the ANOVA assumptions about the population covariance structure Σ are met, then P'SP will be nearly diagonal. Bartlett's sphericity test may be used to test whether $P'\Sigma P$ is diagonal in the population (cf. Bock, 1975, p. 462). Besides providing the univariate results in a simple computational layout, the multivariate treatment of the data leads directly to the appropriate multivariate test of treatment differences in the case where the univariate assumptions are not met.

For present purposes, we are not interested in tests of differences between means. As indicated in Chapter 1, the means for faceted test data are essentially arbitrary. In component analysis, we are basically interested in the covariance structure S , and it should be apparent that the component analysis "solution operation" of [2.9] is identical with the operation used to obtain MSE* in the multivariate treatment of repeated measures. That is, $V = H' C_{xx} H$ and $MSE^* = P' S P$ differ only in notation, not in form or effect. Another notational difference between the component analysis and the ANOVA handling of the data can be found in the expressions for the expected values of the variance components. Comparing the expressions in Table 2.6 with those in Table 2.7, we can set up the following equivalences:

$$\sigma_{\epsilon}^2 + 4\sigma_{\zeta}^2 \equiv \sigma_{\varepsilon}^2 + 4\sigma_{\pi}^2$$

$$\sigma_{\epsilon}^2 + 4\sigma_{\alpha}^2 \equiv \sigma_{\varepsilon}^2 + 2\sigma_{\pi a}^2$$

$$\sigma_{\epsilon}^2 + 4\sigma_{\beta}^2 \equiv \sigma_{\varepsilon}^2 + 2\sigma_{\pi b}^2$$

$$\sigma_{\epsilon}^2 \equiv \sigma_{\varepsilon}^2 + \sigma_{\pi ab}^2$$

As this may suggest, there are two alternate notational systems for variance components. Not only are the subscripts different, but the coefficients of the variance components are different in the two cases. There is no term corresponding to $\sigma_{\pi ab}^2$ on the left side of the equivalences because the component analysis handling of the data assumed no α-by-β interaction variance component, which would imply individual differences associated with the $\alpha\beta$ product. On the other hand, the ANOVA handling of the data allows no component covariances, estimated from the off-diagonal elements of V, since it requires the components to be uncorrelated. With test data, it seems more important to allow the underlying components to be correlated, than to allow for component interactions, and we have already found significant off-diagonal elements of V implying a $\sigma_{\zeta\alpha}^2$ term or content-acquiescence covariance in several analyses. (An $\alpha\beta$ component interaction has no necessary relationship to an AB interaction involving the test or "treatment" means, by the way, except that if the former is present it leads to a $\sigma^2_{\pi ab}$ term in the expected values for the AB and SAB mean squares. An $\alpha\beta$ component interaction implies that the α-on-β regression is different for different levels of β, which is a somewhat more esoteric phenomenon than an AB interaction of the means.)

With the $\sigma^2_{\pi ab}$ term dropped from the right side of the equivalences, it is easy to trace the correspondences between the two notational systems. σ^2_{ϵ} and σ^2_{ε}, sometimes called "replication error," are equivalent in interpretation and effect. If the subjects were brainwashed and the same tests or "treatments" were administered, and this were repeated a number of times, we would expect slightly different results each time, even though the sample had not changed: This is replication error. The content component σ^2_{ζ} and the subjects or "persons" component σ^2_{π} are evidently equivalent in the two notational systems, and refer in any case to individual differences in the common content of the tests or, in ANOVA terms, to the common content of the dependent variable. The terms $4\sigma^2_{\alpha}$ and $2\sigma^2_{\pi a}$ are also equivalent, as are the terms $4\sigma^2_{\beta}$ and $2\sigma^2_{\pi b}$, and are interpreted as individual differences variance associated with the two design facets. The differences in the coefficients in these pairs of expressions appear to be a definitional one. In the component analysis model, the acquiescence effects $\alpha_{\underline{i}}$ for each person are directly defined as being distributed $\underline{N}(0,\sigma^2_{\alpha})$ (Bock et al., 1969, p. 131). A "variance due to interaction in the population" such as $\sigma^2_{\pi a}$, however, is being defined in terms of the means in a 2-way S x A table which has been collapsed over the B factor of the design; i.e.,

$$\sigma^2_{\pi a} = \sum_{\underline{i}=1}^{\underline{N}} \sum_{\underline{j}=1}^{2} (\pi a_{\underline{ij}})^2 / (\underline{N} - 1)$$

(after Winer, 1972, pp. 318-319). Computationally, σ^2_{α} is defined as a mean sum of squares over $\underline{N}$ persons, while $\sigma^2_{\pi a}$ is defined as a mean sum of squares over $2\underline{N}$ observations, with a $\pi a_{\underline{i}1}$ and $\pi a_{\underline{i}2}$ effect both being counted for each person. This leads, in the 2 x 2 design, to the relationships $\hat{\sigma}^2_{\pi a} = 2\hat{\sigma}^2_{\alpha}$ and $\hat{\sigma}^2_{\pi b} = 2\hat{\sigma}^2_{\beta}$.

The component analysis and ANOVA expressions for the expected mean squares differ conceptually as well as computationally. The ANOVA expectations are based on the Cornfield and Tukey (1956) "pigeonhole model" for deriving expected values for mean squares, and involve an assumption that the π main effect and the πa , πb and πab (if any) interactions are all uncorrelated. The expected values under the component analysis model, on the other hand, are more akin to Scheffé's (1959, ch. 8) reformulation of the mixed-effect ANOVA, which is aimed at relaxing this assumption. There are two quite different mixed-effect models for ANOVA which lead to the same tests of significance for the treatment effects (when $P'\Sigma P$ is diagonal), but entail different definitions for the variance components and different ways of conceptualizing the covariance structure underlying the analysis. Hocking (1973), Scheffé (1959) and Searle (1971, pp. 400-404) have discussed the differences in the two kinds of models in the context of the 2-way mixed-effects layout, but I know of no discussion of these differences in the context of a 3-way mixed-effects layout such as the 2 x 2 component analysis design which has been discussed in this section. Of the two kinds of models, the one which leads to the component analysis values for the expected mean squares has a more direct extension to maximum-likelihood methods for covariance structure analysis, and may be preferred for present purposes.

Covariance Structure Analysis

Mulaik (1972, ch. 15) has provided a readable expository review of covariance structure analysis (CSA) using maximum likelihood (ML) methods. He attributes the recent elaboration of these methods to both Bock and Bargmann (1966) and Jöreskog (1966, 1967, 1969, 1970, 1974; Jöreskog & Lawley, 1968), who were working independently and from somewhat different starting points. Bock and Bargmann were working on CSA as a general method of variance component estimation, while Jöreskog was working on the extension of ML factor analysis (Lawley, 1940) to tests of simple structure hypotheses. Jöreskog (1966) had formulated tests of simple structure hypotheses in terms of the factor analysis model

$$[2.16] \quad S \cong BCB' + E \, ,$$

where S is the $n \times n$ matrix of obtained covariances, B is the $n \times r$ factor pattern, C is the $r \times r$ matrix of correlations among the factors, and E is the $n \times n$ diagonal matrix of unique variances. [2.16] resembles the component structure model [2.2], but instead of being a matrix of fixed coefficients, B contains a "simple structure" factor pattern in which some of the coefficients are fixed at a value of zero, but the others are free to vary. The methods which Jöreskog developed to obtain a ML solution for [2.16], with a mixture of fixed and free coefficients in B and/or C, has extremely wide generality: In particular, it has applications to many kinds of CSA problems and to complex path analysis models which had previously been difficult to handle satisfactorily.

Jöreskog's (1974) review shows how an amazing variety of parametric estimation problems can be brought within the framework of his latest and most elaborate version of the CSA model. In addition to the theoretical work leading to general methods for solving restricted ML equations, Jöreskog has made important contributions to the numerical analysis of actual data. The ML equations do not usually have "computable" solutions which may be obtained in a fixed number of steps. Instead, the solutions must be obtained by means of numerical analysis (e.g., Householder, 1953), which involves iteration from an initial trial solution to a solution meeting some convergence criterion. With his co-workers, Jöreskog has provided a series of computer programs which implement the ML methods for a variety of models, in particular: a subroutine package for minimizing a function of several variables (Gruvaeus & Jöreskog, 1970), UFABY3 (Jöreskog & van Thillo, 1971), ACOVS and ACOVSF (Jöreskog, Gruvaeus, & van Thillo, 1970), ACOVSM (Jöreskog, van Thillo, & Gruvaeus, 1971), SIFASP (van Thillo & Jöreskog, 1970), and LISREL (Jöreskog & van Thillo, 1972). The next few sections of this chapter will explicate the general CSA model as embodied in the ACOVSF program.

The ACOVSF model. We begin with a very general structural model:

[2.17] $$x - \mu = A(Bz + d) + e , \quad \text{where:}$$

- x is an $\underline{n} \times 1$ observable random vector such that $\underline{E}(x) = \mu$ and $\underline{E}(xx') = \Sigma$ in the population, and $\bar{x}$ and S respectively are the obtained mean vector and covariance matrix;
- A is an $\underline{n} \times \underline{m}$ matrix of coefficients with rank $\underline{m} \leq \underline{n}$;
- B is an $\underline{m} \times \underline{r}$ matrix of coefficients with rank $\underline{r} \leq \underline{m}$;
- z is an $\underline{r} \times 1$ latent random vector such that $\underline{E}(z) = 0$ and $\underline{E}(zz') = C$ is symmetric with rank $\underline{r}$;
- d is an $\underline{m} \times 1$ latent random vector of disturbances such that $\underline{E}(d) = 0$ and $\underline{E}(dd') = D$ is diagonal; and
- e is an $\underline{n} \times 1$ latent random vector of disturbances such that $\underline{E}(e) = 0$ and $\underline{E}(ee') = E$ is diagonal.

The latent variates z , d and e are assumed to be uncorrelated, so that $\underline{E}(zd') = \underline{E}(ze') = \underline{E}(de') = 0$, and taking crossproducts and expectations of both sides of [2.17] yields:

[2.18] $$S \cong A(BCB' + D)A' + E .$$

Since the dimensions must meet $\underline{r} \leq \underline{m} \leq \underline{n}$, the "outer" matrices $A(.)A'$ and E have a size $\underline{n} \times \underline{n}$ given by the data, *and the* "inner" smallest matrix C has a size $\underline{r} \times \underline{r}$ which determines the rank of the model. Notice that the covariance structure

[2.18] contains one factor model nested inside of another, with $Q = BCB' + D$, say, representing the second-order factors, and $AQA' + E$ representing the first-order factors.

With no restrictions on the right side of [2.18], an infinite number of models will fit the data, and a sufficient number of restrictions must be placed on the constituent matrices A, B, C, D and E to yield a unique solution. Jöreskog's model is so general that any element of the constituent matrices may be specified as: (a) free; (b) fixed at any real value; or (c) constrained to equal another free element. Almost any covariance structure model can be expressed through an appropriate choice of free, fixed and constrained coefficients. In some applications, $A = I$, $B = I$, $C = I$, $D = 0$, and/or $E = 0$. When $A = I$ and $D = 0$, for example, the model reduces to the model [2.16] and when, in addition, $C = I$, the model reduces to the model for factor analysis in an orthogonal rotation. Other possibilities are given by Jöreskog (1974).

<u>ML estimation of free parameters</u>. For given estimates of the constituent matrices, some of whose elements are fixed by hypothesis, we will agree to write

$$[2.19] \qquad \hat{\Sigma} = \hat{A}(\hat{B}\hat{C}\hat{B}' + \hat{D})\hat{A}' + \hat{E} ,$$

where the carets indicate a particular set of estimates for all of the free parameters of the model. We seek estimates of these parameters which maximize the likelihood (or log-likelihood) of the observed S given $\hat{\Sigma}$. If Σ is multivariate-normal, the log-likelihood of the observed S is

$$[2.20] \qquad \underline{L}_{\Omega} = -\tfrac{1}{2}\underline{Nn}\ln(2\pi) - \tfrac{1}{2}(\underline{N}-1)[\ln|S| + \underline{n}] ,$$

where Ω is the set of all $\underline{n} \times \underline{n}$ positive definite matrices, while

$$[2.21] \qquad \underline{L}_{\omega} = -\tfrac{1}{2}\underline{Nn}\ln(2\pi) - \tfrac{1}{2}(\underline{N}-1)[\ln|\hat{\Sigma}| + \mathrm{tr}(S\hat{\Sigma}^{-1})] ,$$

where ω is the set of all $\hat{\Sigma}$ meeting [2.19] given the restrictions (Jöreskog, 1967). Instead of maximizing [2.21], it is convenient to minimize

$$[2.22] \qquad \begin{aligned} \underline{F} &= -2(\underline{L}_{\omega} - \underline{L}_{\Omega}) \\ &= (\underline{N}-1)[\ln|\hat{\Sigma}| - \ln|S| + \mathrm{tr}(S\hat{\Sigma}^{-1}) - \underline{n}] . \end{aligned}$$

$\underline{F}$ is minus 2 times the logarithm of the likelihood ratio for testing ω against the general alternative Ω, and in large samples $\underline{F}$ is distributed as χ^2 with degrees of freedom equal to the difference in the number of parameters estimated under the two hypotheses $\underline{H}_\omega$ and $\underline{H}_\Omega$ (Jöreskog, 1970). The number of parameters estimated under $\underline{H}_\Omega$ is $\frac{1}{2}\underline{n}(\underline{n}+1)$ for the unique elements of S plus $\underline{n}$ for the mean vector, while the number of free parameters estimated under $\underline{H}_\omega$ is $\underline{p}$ (say) plus $\underline{n}$ for the mean vector, so the test statistic is:

[2.23] $$\underline{F} \sim \chi^2 [\tfrac{1}{2}\underline{n}(\underline{n}+1) - \underline{p}] ,$$

at some selected α level. If $\underline{F} < \chi^2_\alpha$, we can say that the restricted model "fits the data." If the model does not fit, it may be because too low a dimensionality $\underline{r}$ has been specified, or because too many restrictions have been imposed on the model.

To minimize [2.22], we need the partial derivatives of $\underline{F}$ with respect to each of the constituent matrices. According to Jöreskog (1970), these turn out to be (omitting a factor of $\underline{N}$-1):

[2.24a] $\partial\underline{F}/\partial A = 2PAQ$;

[2.24b] $\partial\underline{F}/\partial B = 2A'PABC$;

[2.24c] $\partial\underline{F}/\partial C = 2B'A'PAB - \text{diag}(B'A'PAB)$;

[2.24d] $\partial\underline{F}/\partial D = 2\,\text{diag}(A'PA)D$; and

[2.24e] $\partial\underline{F}/\partial E = 2\,\text{diag}(P)E$; where

$P = \hat{\Sigma}^{-1}(\hat{\Sigma} - S)\hat{\Sigma}^{-1}$ and $Q = BCB' + D$. The matrices of partial derivatives are each matrices of the same size as the matrix of the denominator (i.e., $\partial \underline{F}/\partial A$ is $\underline{n} \times \underline{m}$, like A, etc.), and each has zero values wherever the matrix of the denominator is free, and equalized values wherever the matrix of the denominator is constrained.

The most direct approach to minimization of a differentiable function of several variables is to set the partial derivatives to zero and solve the resulting equations algebraically. This can seldom be accomplished for irregular CSA models of the kind considered by Jöreskog, so various numerical techniques must be used. The ensuing account relies on discussions of the principles of numerical analysis sketched by Bock and Bargmann (1966), Bramble (1971), Gruvaeus and Jöreskog (1970), Jöreskog (1966, 1967) and Mulaik (1972).

The one-to-one matching of the elements of the parameter matrices and the elements of the partial derivative matrices in [2.24] permits a useful device: Let the free parameters be strung out in a vector θ with $\underline{p}$ elements, and then the corresponding elements of the partial derivative matrices may be strung out in another vector g called the "gradient," which has the same size as θ and the same arrangement of elements. Some "bookkeeping" will be needed to keep track of the fixed and constrained parameters, but the free parameters will be available in a tidy form. $\underline{F}$ of [2.22] may now be understood as $\underline{F}(\theta^{\underline{i}})$, the value of $\underline{F}$ for a given set of parameters $\theta^{\underline{i}}$ at the $\underline{i}$-th iteration, and the gradient $g(\theta^{\underline{i}})$ will contain the partial derivatives evaluated with the same parameters. The ML estimates of the parameters are taken as the values $\hat{\theta}$ for which $\underline{F}(\hat{\theta})$ is a minimum, at which point $g(\hat{\theta}) = 0$. The minimization problem may now be expressed by saying that, beginning with trial values of $\theta^{\underline{i}}$ for $\underline{i} = 1$, we seek values of $\theta^{\underline{i}+1}$ which make $\underline{F}(\theta^{\underline{i}+1}) < \underline{F}(\theta^{\underline{i}})$ and $g \rightarrow 0$.

Figure 2.1 illustrates the principles involved in minimization. $\underline{F}(\theta)$ is a continuous $\underline{p}$-dimensional manifold weaving through Ω-space, illustrated in Figure 2.1 by a wavy line representing the trace of $\underline{F}(\theta)$ on a plane passing through the minimum in Ω-space. Suppose we start at the point $\underline{F}(\theta^1)$ which happens to lie in this plane, and we want to find $\underline{F}(\hat{\theta})$. Each point $\underline{F}(\theta^{\underline{i}})$ in the plane will be a point on the wavy line, and each gradient $g(\theta^{\underline{i}})$ evaluated at the same point may be represented geometrically as the slope of the line tangent to $\underline{F}(\theta)$ at the same point. (The tangent lines are illustrated as lying in the plane, but they may intersect the plane instead.) Clearly, one end of the gradient always points in the general direction of the minimum. The method of <u>steepest descent</u> treats $\underline{F}(\theta)$ as linear, and from trial values $\theta^{\underline{i}}$ we take new values

$$[2.25] \qquad \theta^{\underline{i}+1} = \theta^{\underline{i}} + \underline{c}_{\underline{i}}\, g(\theta^{\underline{i}}) \,,$$

where $\underline{c}_{\underline{i}}$ is a suitable small constant; in other words, we take a short step in the direction of the gradient, or "downhill."

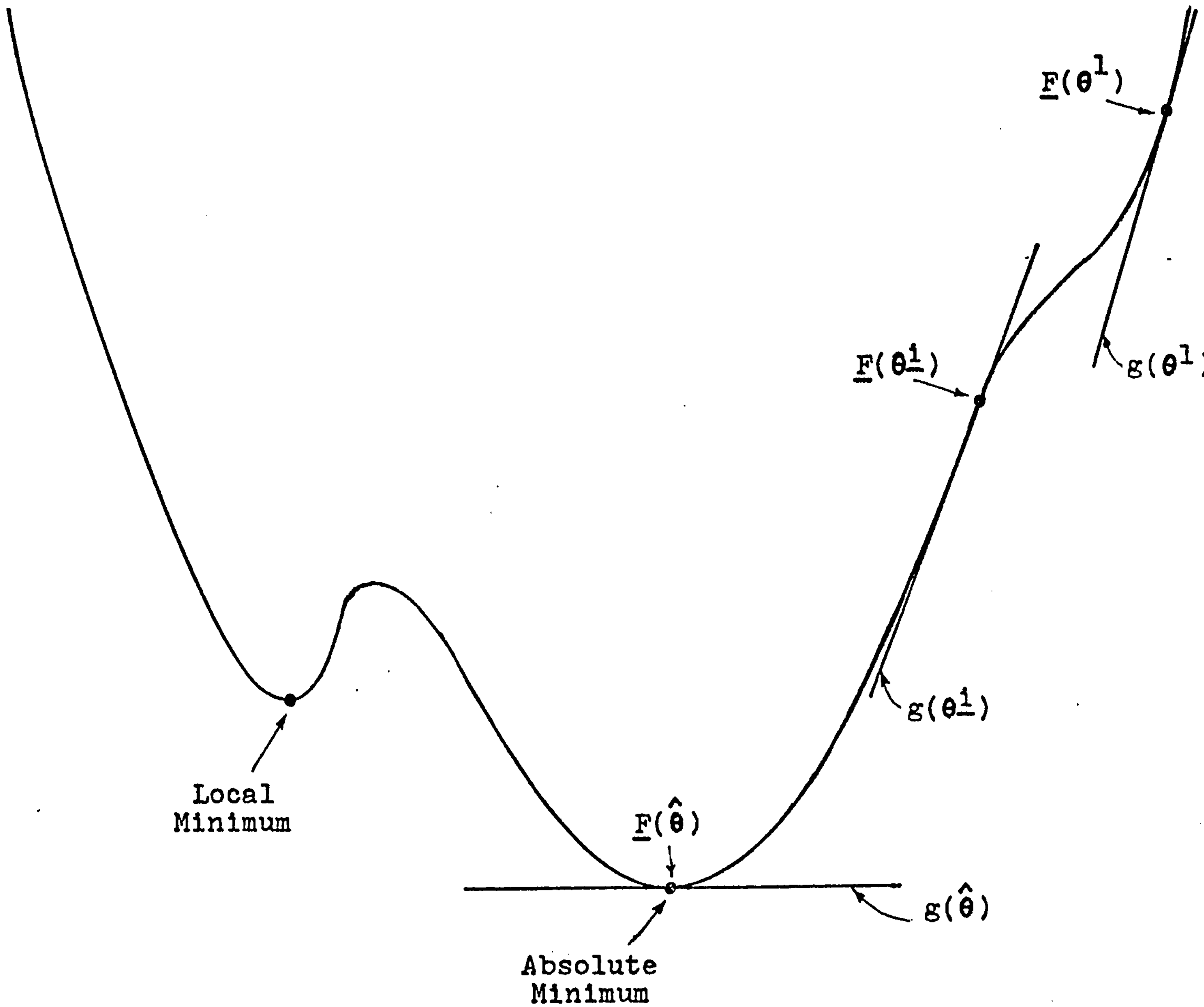


Figure 2.1. Minimization of $\underline{F}(\theta^{\underline{i}})$ using gradient vector $g(\theta^{\underline{i}}) = \partial\underline{F}/\partial\theta$ evaluated at $\theta^{\underline{i}}$.

Iterations using [2.25] continue until all values of g are zero, within some tolerance ϵ, which is taken to be the minimum (more properly, a minimum) of the function.

The method of steepest descent always converges, in the sense that each $\underline{F}(\theta^{\underline{i}+1}) < \underline{F}(\theta^{\underline{i}})$ and the process reaches a point where $g(\theta^{\underline{i}}) \cong 0$. The method is very slow, however, particularly as the trial values near the minimum. Minimization may be speeded up somewhat by interpolating steps having the form

$$[2.26] \qquad \theta^{\underline{i}+1} = \theta^{\underline{i}} + \underline{c}_{\underline{i}}(\theta^{\underline{i}} - \theta^{\underline{i}-2}) ,$$

at every third iteration. This is the method of resultant descents, which uses the resultant of gradients at $g(\theta^{\underline{i}})$ and $g(\theta^{\underline{i}-1})$ to find the direction at step $\underline{i}+1$ (Jöreskog, 1966).

A problem with all of the minimization methods to be discussed is that there is no guarantee that the obtained minimum $\underline{F}(\hat{\theta})$ will be an absolute minimum over the entire parameter space. If the initial trial value had been taken on the left side of the wavy line in Figure 2.1, the minimization process might stop at the point flagged as a "local minimum." A good starting point will speed up convergence, and may determine which local minimum is reached, if there is more than one. Bock and Bargmann (1966) seem to feel that it is not difficult to find a good starting point, and Jöreskog recommends that several starting points be used, to see if they all converge to the same minimum. Chanda (1954) has explored the conditions under which the ML criterion is unimodal. If Σ is strictly multivariate-normal, then the probability of finding local minima and saddle points diminishes as the sample size increases, but the problem needs further exploration.

Jöreskog has concerned himself with another kind of minimization misbehavior, called "boundary problems," which are said to occur when the minimization process leads to a region of parameter space where the parameters take on unrealistic values. This can happen if the elements of E or sometimes D begin to take on negative values. Jöreskog generally handles this problem by requiring elements of E to be greater than or equal to .005 through the iterations. Sometimes the condition will correct itself, but if the minimum is reached with an element of E on the "boundary," with a value of .005, the solution is said to be an "improper" one. Jöreskog recommends using new starting values if this occurs.

Bock and Bargmann adopted the Newton-Raphson method of minimization. In addition to the gradient g, which is the derivative of $\underline{F}$ with respect to the parameters (i.e., $g = \partial\underline{F}/\partial\theta$), this method requires analytic expressions for the second derivatives of $\underline{F}$ with respect to the parameters, which may be represented as

$$[2.27] \qquad H = \partial^2\underline{F}/\partial\theta_{\underline{j}}\partial\theta_{\underline{k}} = \partial g/\partial\theta .$$

H is called the "Hessian" matrix, and has an order of $\underline{p} \times \underline{p}$.

For the Newton-Raphson method, let $\underline{H}^i$ indicate the Hessian matrix evaluated at the trial point $\underline{\theta}^i$; then the next trial point is found as

$$[2.28] \qquad \underline{\theta}^{i+1} = \underline{\theta}^i - (\underline{H}^i)^{-1} g(\underline{\theta}^i) \ .$$

In this method, $\underline{F}(\theta)$ is implicitly treated as a quadratic function, since [2.28] is derived from an order-two Taylor series expansion. Bock and Bargmann indicate that the minimization process converges rapidly, if it converges at all. The negative of the expected value of the Hessian at point $\underline{i}$ is the information matrix $\underline{J}^i = -\underline{E}(\underline{H}^i)$ of the parameters at the same point. $\underline{E}(\underline{H}^i)$ usually has a simpler expression than $\underline{H}^i$ itself, and a simplification of the Newton-Raphson method known as Fisher's <u>scoring method</u> is obtained by taking the trial points as

$$[2.29] \qquad \underline{\theta}^{i+1} = \underline{\theta}^i + (\underline{J}^i)^{-1} g(\underline{\theta}^i) \ ,$$

using the information matrix rather than the Hessian. At the minimum, $J = -H$, and J^{-1} contains the variance-covariance matrix of the ML estimates.

A drawback of the Newton-Raphson and scoring methods is that they require the computation of H^{-1} or J^{-1} at each iteration, which can be a formidable task. In a possibly typical unrestricted factor analysis model involving 40 variables and 10 factors, the number of free parameters $\underline{p}$ will be on the order of 400, so H and J are on the order of 160,000. One variant of the

Newton-Raphson method evaluates H^{-1} only once, or only periodically, to save the labor of repeated inversions. Apart from the difficulty of obtaining many inversions, it may be difficult to obtain even one accurate inversion of this size, owing to rounding error. For large problems, H may become singular or nearly so during the process of iteration. Nevertheless, Bock and Bargmann have used the method successfully with several small problems, inverting H at each iteration.

The final minimization method is usually known in the literature as the Fletcher-Powell method. It was originally developed by Davidon (1959), improved by Fletcher and Powell (1963) who added an elegant proof of convergence, and improved further by Jöreskog (1967) who applied it to the problem of ML estimation. In this method, expressions for the second derivatives of $\underline{F}$ never have to be obtained analytically, and a $\underline{p} \times \underline{p}$ matrix never has to be inverted; yet it yields iteratively improved matrices $W^{\underline{i}}$, say, such that $W^{\underline{i}} \rightarrow J^{-1}$, the variance-covariance matrix of the parameters evaluated at the minimum, as $\underline{F}(\theta^{\underline{i}}) \rightarrow \underline{F}(\hat{\theta})$. The iterative equation is

$$[2.30] \qquad \theta^{\underline{i}+1} = \theta^{\underline{i}} - \beta W^{\underline{i}} g(\theta^{\underline{i}}) ,$$

where β is a positive constant determined separately for each iteration. The new gradient $g(\theta^{\underline{i}+1})$ may in principle be differentiated with respect to β , and for a given β will have a slope

$$[2.31] \qquad \underline{s}(\beta) = \partial[g(\theta^{\underline{i}+1})]/\partial\beta = [-W^{i}g(\theta^{\underline{i}})]'g(\theta^{\underline{i}+1}) ,$$

which is a scalar. In particular, for $\beta = 0$, the slope $\underline{s}(\beta=0) = [-W^{\underline{i}}g(\theta^{\underline{i}})]'g(\theta^{\underline{i}})$, which may be taken (since its sign is arbitrary) as negative. (And if $\underline{s}(\beta=0) = 0$, the minimization is complete.) The function [2.31] has at least one minimum at a point with $\underline{s}(\beta=\beta_{min}) = 0$, for some $\beta_{min} > 0$, and any number of points β^* for which $\underline{s}(\beta=\beta^*)$ is positive and $\beta^* > \beta_{min}$. Since [2.31] would be difficult to evaluate directly, the Fletcher-Powell method estimates β_{min} for use as the constant β in [2.30] by finding a value β^* and interpolating for an estimate of β_{min} . That is, we know that

$$[2.32] \qquad 0 < \beta_{min} < \beta^* , \text{ for which}$$

$$[2.33] \qquad \underline{s}(\beta=0) < \underline{s}(\beta=\beta_{min}) = 0 < \underline{s}(\beta=\beta^*) ,$$

and the two points 0 and β^* and their slopes enable us to fit a third-degree polynomial to [2.31]. The minimum value for the polynomial in the range [2.32] will approximate β_{min}, and is then taken to be the constant β for iteration $\underline{i}$. A suitable value for β^* at iteration $\underline{i}$ may be obtained by extrapolation. $\beta = 1$ may be taken as the first trial value: If $\underline{s}(\beta=1) > 0$, we may take $\beta^* = 1$ and interpolate for β_{min} ; otherwise, we may extrapoate linearly for β_{min} using

$$\beta_{min} = -\underline{s}(\beta=0) / [\underline{s}(\beta=1) - \underline{s}(\beta=0)] .$$

In practice, Jöreskog recommends several interpolations and/or extrapolations at each iteration, to allow for an irregular form of [2.31] and a narrower bracketing of β_{min} than [2.32] implies (Gruvaeus & Joreskog, 1970). Finally, once β is determined, $\underline{\theta}^{i+1}$ is obtained using [2.30], along with the new gradient $g(\underline{\theta}^{i+1})$ and a new value for $\underline{W}^{i+1}$. The latter is obtained by means of:

[2.34a] $\underline{t}^{i} = \beta \underline{W}^{i} g(\underline{\theta}^{i})$,

[2.34b] $\underline{u}^{i} = g(\underline{\theta}^{i+1}) - g(\underline{\theta}^{i})$, and

[2.34c] $\underline{v}^{i} = \underline{W}^{i} \underline{u}^{i}$; and finally,

[2.34d] $\underline{W}^{i+1} = \underline{W}^{i} + \underline{t}^{i}(\underline{t}^{i})' / (\underline{t}^{i})'\underline{u}^{i} + \underline{v}^{i}(\underline{v}^{i})' / (\underline{u}^{i})'\underline{v}^{i}$.

In principle, there is no reason why two different minimization methods cannot both be used for a particular problem. Jöreskog typically starts off with a couple of steepest descent iterations, which approach a minimum very rapidly at first, and then switches to Fletcher-Powell iterations, to refine the estimates of the ML parameters closer to the minimum.

Implications of Content-Acquiescence Covariance

I have alluded several times to the principle that acquiescence can have a large effect on the behavior of scales when content and acquiescence are correlated, and this section illustrates the principle with a short example. Recall that Bock et al. cited the Couch and Keniston correlation of $r_{OAS,A} = .50$, between the *OAS* acquiescence measure and the MMPI *A* scale, as a piece of evidence for the influence of acquiescence on the MMPI factor structure. Block and Rorer had attributed this result to the contamination of the *OAS* items by *A* scale content, but another explanation is possible.

With the aid of a path model in Figure 2.2, we can trace the effect of a content-acquiescence covariance on the correlation between a balanced measure of content and a content-free measure of acquiescence. As shown in Figure 2.2, the latent sources of variance are ζ and α, and the observed measures are the *true* measure of content t, the *false* measure f, a balanced composite measure $b = \frac{1}{2}(t + f)$, and the content-free measure of acquiescence a. The path coefficients depicted are the ones resulting from a component or covariance structure analysis in which the *true* and *false* measures of content have components $\sigma_\zeta^2 = .64$, $\sigma_\alpha^2 = .04$, and $\sigma_{\zeta\alpha} = .16$ (since $\rho_{\zeta\alpha} = .50$). The *true* and *false* measures are represented as having error variances of .16, and the independent measure of aquiescence is represented as having an error variance of .15 and a reliability of .85. These are hypothetical coefficients,

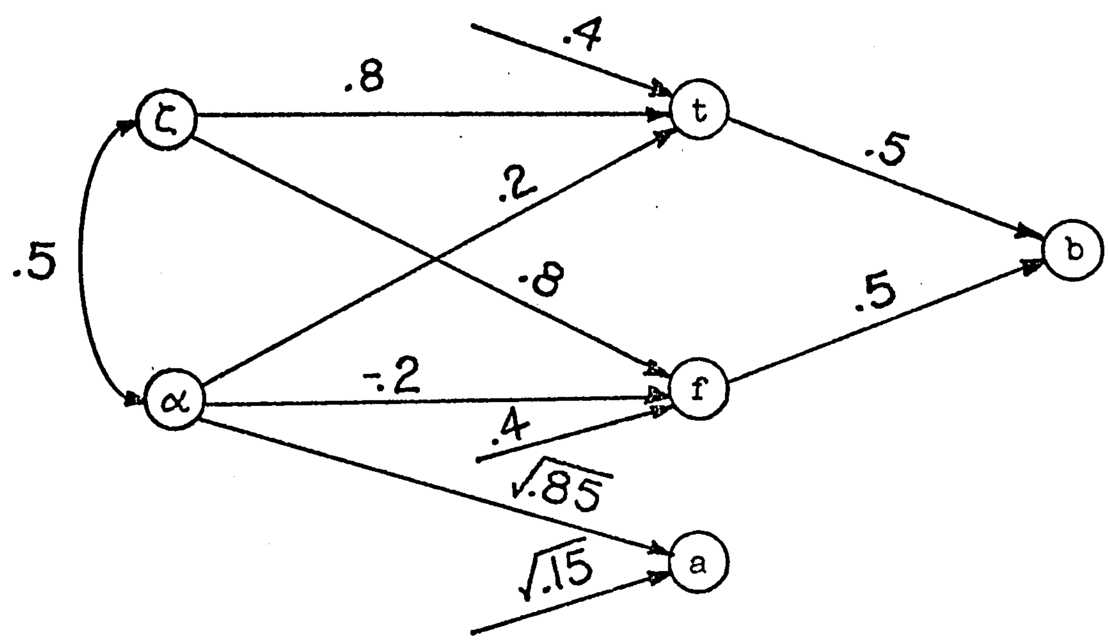


Figure 2.2. Path model showing the relationship between the latent variates ζ and α and observed measures representing <u>true</u> (t), <u>false</u> (f), and <u>balanced</u> (b) measures of content, together with a content-free measure of acquiescence (a).

with values which simplify computations, but resemble the coefficients obtained with the actual A and OAS scales.

Routine use of the conventions for path analysis (e.g., Duncan, 1966) enable us to derive the following implications of the model in Figure 2.2:

$$\sigma_t^2 = .8^2 + .2^2 + 2 \times .5 \times .8 \times .2 + .4^2 = 1.00$$

$$\sigma_f^2 = .8^2 + .2^2 - 2 \times .5 \times .8 \times .2 + .4^2 = .68$$

$$\sigma_b^2 = \sigma_\zeta^2 + \tfrac{1}{2}\sigma_\epsilon^2 = .8^2 + \tfrac{1}{2}(.4^2) = .72$$

$$\sigma_{ta} = (.8)(.5)\sqrt{.85} + (.2)\sqrt{.85} = .533$$

$$\sigma_{fa} = (.8)(.5)\sqrt{.85} - (.2)\sqrt{.85} = .210$$

$$\sigma_{ba} = (.8)(.5)\sqrt{.85} = .372$$

$$\sigma_a^2 = .85 + .15 = 1.00$$

From these summary statistics, correlations between the content measures and the acquiescence measure may be obtained as:

$$r_{ta} = \sigma_{ta} / \sigma_t \sigma_a = .533$$

$$r_{fa} = \sigma_{fa} / \sigma_f \sigma_a = .255$$

$$r_{ba} = \sigma_{ba} / \sigma_b \sigma_a = .436$$

The derived result for the true scale, $r_{ta} = .533$, closely resembles the obtained result $r_{OAS,A} = .50$, showing that the obtained result can be explained without resorting to the assumption that the OAS is contaminated by first-factor content.

Notice that the derived result for the balanced scale, $\underline{r}_{ba}$ = .436 , is also quite high. This result faithfully mirrors the parameter estimate for the content-acquiescence correlation, $\rho_{\zeta\alpha} = .50$, which can be seen by correcting $\underline{r}_{ba}$ for attenuation due to error of measurement for the b and a measures. The reliability of the a measure is .85, and the reliability of the b measure is

$$\rho_{bb} = \sigma_{\zeta}^{2} / \sigma_{b}^{2} = .64 / .72 = .89 ,$$

which yields the attenuation-corrected correlation

$$\rho^{*}_{ba} = .436 / \sqrt{(.85)(.89)} = .50 = \rho_{\zeta\alpha} .$$

ρ^{*}_{ba} , in other words, is just the population correlation between acquiescence and a balanced measure of content, when both are measured without error, and this is equivalent to the value of $\rho_{\zeta\alpha}$ with which we began.

Simply-Patterned ACOVS Models

Wiley, Schmidt and Bramble (1973) have defined a class of covariance structure models which I will call "simply-patterned." The class of models includes component analysis as a special case, and has the general form

$$[2.35] \qquad \hat{\Sigma} = A(BCB')A + E \, ,$$

where A and E are $\underline{n} \times \underline{n}$ diagonal matrices containing scaling factors and error variances, respectively; C is an $\underline{r} \times \underline{r}$ symmetric matrix containing component covariances; and B is a fixed $\underline{n} \times \underline{r}$ pattern matrix consisting of $\underline{r}$ columns from a Hadamard matrix of size $\underline{n}$. Thus, the model is suitable for any data having a $2^{\underline{k}}$ component analysis design. This section works through the principles of covariance structure analysis for the simply-patterned class of models, for the case of the 2^3 design, touching on the problems of determining the rank of the model, testing a range of models of a given rank, deciding when to stop fitting, and identification of parameters.

Testing a range of models. For the 2^3 design, the rank of the observed data is $\underline{p} = 8$, and the class of simply-patterned models may be defined in terms of the columns of the matrix H_3 in equation [2.7c]. In principle, any or all of the columns of H_3 may be used--the definition of the class of models is that broad. In practice, we wish to fit a model of low rank, and are likely to be most interested in the columns of H_3 which represent the main effects of the design.

The first problem to be faced is that of deciding the rank $\underline{r}$ of the model, and the probable components which can be detected in the data, but the discussion of this problem will be deferred. Once a particular pattern $B_{\underline{r}}$, say, has been chosen, the next problem is that of testing a "range" of models to determine whether we can fit models incorporating the pattern $B_{\underline{r}}$, and with or without restrictions on the other matrices in the model. Table 2.8 lists the variants of the elementary matrices used to define a range of models. The most restricted model we shall consider may be written

$$[2.36] \qquad \hat{\Sigma} = A_1(B_{\underline{r}} C_1 B'_{\underline{r}})A_1 + E_1 ,$$

implying no scaling factors ($A_1 = I$), uncorrelated components (C_1 = diagonal), and homogeneous error variances ($E_1 = \sigma^2 I$). The least restricted model is

$$[2.37] \qquad \hat{\Sigma} = A_3(B_{\underline{r}} C_2 B'_{\underline{r}})A_3 + E_2 ,$$

which implies that the scaling factors are estimated from the data (except for one fixed parameter needed to assure identification of the model), correlated components, and heterogeneous error variances. For ease of recall, the lower-valued subscripts in [2.36-2.37] and Table 2.8 refer to more restricted forms of the elementary matrices.

Thus, [2.36] and [2.37] define the outer limits of a range of models to be tested with a given pattern matrix. Conceptually, the range has an illimitable number of fine gradations between the most restricted model [2.36] and the least restricted model

Table 2.8

Elementary Matrices Used for Fitting Simply-Patterned ACOVS

Models $\hat{\Sigma} = A(BCB')A + E$

Scaling Factors: A ($\underline{p}$ x $\underline{p}$)

$A_1 = I$

$A_2 = \text{diag}(1, 1, \underline{a}_3, \underline{a}_4, \ldots \underline{a}_p)$

$A_3 = \text{diag}(1, \underline{a}_2, \underline{a}_3, \ldots \underline{a}_p)$

$A_4 = \text{diag}(\underline{a}_1, \underline{a}_2, \ldots \underline{a}_p)$

Pattern Matrices: B ($\underline{p}$ x $\underline{r}$)

$B_1' = (1, 1, 1, \ldots 1)$

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

$$B_3' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

$$B_4' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

Component Covariances: C ($\underline{r}$ x $\underline{r}$)

$C_1 = \text{diag}(\underline{c}_1, \underline{c}_2, \ldots \underline{c}_r)$

C_2 = general symmetric

Error Variances: E ($\underline{p}$ x $\underline{p}$)

$E_1 = \sigma^2 I$ (homogeneous error)

$E_2 = \text{diag}(\underline{e}_1, \underline{e}_2, \ldots \underline{e}_p)$

[2.37]: The class of models given by [2.35] defines a vector space, even with $B_{\underline{r}}$ determined in advance, and by definition, a vector space permits an infinite number of subdivisions. In practice, the system of eight models generated by crossing A_1 vs. A_3 with C_1 vs. C_2 and E_1 vs. E_2 will provide a satisfactory framework for interpreting the data and deciding what sorts of restrictions can be applied without seriously degrading the fit of the data to the model.

Wiley et al. discuss this 2 x 2 x 2 system of models, and also a set of models having the form

$$[2.38] \qquad \hat{\Sigma} = A(BCB' + \underline{k}I)A + E \; ,$$

where $\underline{k}$ is a fixed constant. The most restricted form of [2.38] may be obtained by reparameterization of [2.37], by allowing all eight scaling factors to be unconstrained. [2.38] spans the same space as [2.35], and may be considered a member of the class of simply-patterned models. [2.38] and its variants were not found to be interpretively useful for the data considered in this dissertation, however, and will not be discussed further.

There are also models more restricted than [2.36] which are still simply-patterned--for example, we can constrain C to be homogeneous as well as diagonal, or constrain C to equal the identity matrix, but these are not generally useful.

Deciding when to stop fitting. A useful theoretical distinction can be made between confirmatory and exploratory factor analysis (Jöreskog, 1969), which I will parody here. In a purely confirmatory analysis, the researcher has a strong hypothesis about the covariance structure of the data--ideally a single structural model which he wishes to test for goodness of fit. Assuming a large sample from a multivariate normal distribution, equation [2.22] has a χ^2 distribution in this situation. A single test of the hypothesis is made, and the researcher accepts or rejects it and writes up the results. This never happens, of course, but if it were to happen, the probability statement based on the obtained χ^2 would have a rigorous statistical meaning. In a purely exploratory analysis, on the other hand, the researcher has only weak hypotheses about the data--perhaps some notion of the kinds of models which might be suitable--but basically wants the data to tell him what sorts of effects and relationships are present in his collection of variables. If the phenomena observed in the data seem plausible, in the light of psychological theory or whatever, the researcher can (and should) follow up his exploratory analysis with a confirmatory one. In the exploratory situation, the obtained χ^2 values for the models he tests do not have any rigorous statistical meaning, owing to multiple nonindependent tests of the same data, capitalization on chance, and so on, but differences in the obtained χ^2 values can serve as an index of the size of effects in the sample, and will enable him to form a judgment whether some effects are more important and more likely to replicate than others.

Most real research situations involve a rather tangled mixture of confirmatory and exploratory hypotheses. For our component analysis problem, the analysis may be regarded as purely confirmatory with respect to the hypothesis that the measures have a common content component, partly confirmatory and partly exploratory with respect to the hypothesis that components associated with the design facets are present in the data, and mostly exploratory with respect to the hypotheses that the components are correlated or not, that the error variances are homogeneous or not, and that the components are present in the metric of the data as given. Unfortunately, it is difficult to formulate rigorous tests of the confirmatory part of the analysis without making a number of decisions about these other issues: All too often "decisions" about the metric of the data, independence of the components and homogeneity of error get buried as statistical "assumptions." Many kinds of analyses are greatly simplified if such assumptions can be made, but it may be as important to test the assumptions as to test the major hypotheses of the study. One of the great strengths of ACOVS methods is that they bring tests of some of the assumptions into the open, where they can be examined along with the tests of the major hypotheses. But then it may be necessary to regard the data analytic enterprise as inherently exploratory with any given dataset, and only confirmatory to the extent that new or similar datasets give similar results. Jöreskog has written, "In practice, the [confirmatory-exploratory] distinction is not always clear cut" [1974, p. 1]. I would go a step further, and assert that the distinction is never clear cut, and cannot be.

A recent statement about the interpretation of χ^2 tests in maximum-likelihood estimation is worth quoting in full:

> The χ^2 test [for a given model] is a test of the specified model against the most general alternative that Σ is any positive definite matrix.
>
> Suppose $\underline{H}_0$ represents one model under given specifications of fixed, free, and constrained parameters. Then it is possible, in large samples, to test the model $\underline{H}_0$ against any more general model $\underline{H}_1$, by estimating each of them separately and comparing their χ^2 goodness-of-fit values. The difference in χ^2 is asymptotically a χ^2 with degrees of freedom equal to the corresponding difference in degrees of freedom. In many situations, it is possible to set up a sequence of hypotheses such that each one is a special case of the preceding one and to test these hypotheses sequentially.
>
> The values of χ^2 should be interpreted very cautiously. In most empricial work many of the hypotheses may not be realistic. If a sufficiently large sample were obtained, the test statistic would, no doubt, indicate that any such hypothesis is statistically untenable. The hypothesis should rather be that [the specified model] represents a reasonable approximation to the population variance-covariance matrix. From this point of view the statistical problem is not one of testing a given hypothesis (which a priori may be considered false), but rather one of fitting various models with different numbers of parameters and of deciding when to stop fitting. In other words, the problem is to extract as much information as possible out of a sample of given size without going so far that the result is affected to a large extent by 'noise.' it is reasonable and likely that more information can be extracted from a large sample than from a small sample. In such a problem the differences between χ^2 values matter rather than the χ^2 values themselves. . . . A large drop in χ^2 [when more parameters are introduced], compared to the difference in degrees of freedom, indicates that the changes made in the model represent a real improvement. On the other hand, a drop in χ^2 close to the difference in degrees of freedom indicates that the improvement in fit is obtained by 'capitalizing on chance,' and the added parameters may not have real significance and meaning [Jöreskog, 1974, p. 4; my italics].

The force of saying that we are using "large-sample" tests for goodness of fit, and that χ^2 differences are "asymptotically" χ^2 , is that we assume that the observed values not only estimate the population values, but estimate them rather well. Paradoxically, if our sample is large enough to really represent the population, we are almost certain to find that the hypothesis of fit is "statistically untenable" or false. Thus, we hope to find a "reasonable approximation" of the population covariance matrix with, hopefully, as few parameters as possible.

There is a strong tug toward fitting the most parsimonious possible model for the data, but, clearly, an open-ended exploratory analysis will yield an "acceptable fit" to models which are too parsimonious to replicate--let alone represent the true population state of affairs. One must be diffident about presenting the results for the most restricted model located in the data, because if it fits, any number of less restricted models will also fit. I like to think of the fitting process as one of informally mapping the region defined by models having the form of [2.35] which fit the data, aiming for generalizations of the following kind: "X and Y are the most important components in the data; the other components do not materially improve the fit." "Correlated components are needed for acceptable fit." "Heterogeneous error variances are not required for acceptable fit." And so on. The emphasis should be one of identifying features of the models which improve the fit materially, rather than on finding a particular model which "fits best."

Determination of rank and identification of the probable components. There are practical as well as theoretical considerations involved in the decision to stop fitting models to the data. In principle, we have 8 choices for models using a single column of H_3, 28 choices for models using 2 columns, 56 choices for models using 3 columns, and so on. For each choice of B, there are (ignoring the fact that there will be some redundant and underidentified models) eight distinct models in the range, plus any number of other simply-patterned models which could be fitted. Control card preparation for covariance structure analysis can be somewhat onerous, however. I use a version of the ACOVSF program (Jöreskog et al., 1970), and a typical set of control cards for testing a range of eight models contains about 300 lines, and generates over 1000 lines of output. The enormous flexibility of the program is purchased at a certain cost! It is useful, therefore, to make a preliminary determination of the rank of the data and the probable components present in the data before fitting ACOVS models with the ACOVSF program. (Incidently, there is an ACOVS computer program, and the ACOVSF computer program is a version of it which uses estimates of second derivatives in the computations and also yields standard errors for the parameters. I use "ACOVS" as an acronym for "analysis of covariance structure"; unless otherwise indicated, the computer program used is a linear descendent of ACOVSF.)

I like to look at several preliminary tests on the dimensionality of an observed matrix S thought to be suitable for component analysis. It is very comforting when the tests all agree, as they sometimes do. The tests are sensitive to different information about the data, however, and it can be useful to see which tests agree and which do not.

The easiest test to apply is the roots-of-R-greater-than-one test, which is the Guttman-Kaiser rule for the number of factors in factor analysis, and indicates the number of principal components having a positive coefficient-α (Kaiser, 1961). It may be considered a lower bound on the number of variance components present in the data.

The next few tests are likelihood-ratio tests on a covariance structure which are available in a closed form. They can all be expressed as tests for equality of s variances (or eigenvalues, which may be interpeted as variances in multivariate work), and have the form:

$$[2.39] \qquad \chi^2 = -K \ln W = -K \ln[\prod_s \phi_i / (\sum_s \phi_i / s)^s] ,$$

where W is the likelihood ratio, obtained as the product of the variances ϕ_i ($i = 1, s$) divided by the s-th power of their mean, and K is a scale factor which improves the large-sample approximation $-(N-1)\ln W \sim \chi^2$. K is always less than $N-1$ and is on the order of $N - s/3$ for several of the tests below (Box, 1949), but when $N >> s$, K is negligibly different from $N-1$.

Among the classical likelihood-ratio tests which may be used are the Wilks (1932), Mauchly (1940), and Bartlett (1950, 1951a, 1951b) "sphericity" tests on R and S . Wilks provided the classical test of the hypothesis $\underline{H}_0$: $P = I$ that a correlation matrix is equal to the identity matrix. If the Wilks test cannot be rejected, there is no point in pursuing a component analysis of the data. Mauchly (1940) provided a more restrictive test of the hypothesis $\underline{H}_0$: $\Sigma = \sigma^2 I$ of sphericity and equality of a variances for a covariance matrix in a given metric. Bartlett extended Wilks's test to the hypothesis $\underline{H}_0$: $P = F_{\underline{r}}F'_{\underline{r}} + \sigma^2 I$, where $F_{\underline{r}}$ contains the correlations of the variables with the first $\underline{r}$ principal components of R. Bartlett's test may be expressed as a test for equality of the last $\underline{s} = \underline{p} - \underline{r}$ eigenvalues of R, and is interpreted as a test for "significance" of the first $\underline{r}$ principal components. The $\phi_{\underline{i}}$ of equation [2.39] are the $\underline{s}$ smallest eigenvalues, and $\underline{K}$ is taken as $\underline{N} - 1 - (2\underline{s}^2+\underline{s}+2)/6\underline{s}$. The resulting χ^2 is interpreted with $\frac{1}{2}(\underline{s}^2+\underline{s}-2)$ degrees of freedom.

All of these classical tests can be performed using ACOVSF, and the ACOVS significance test is constructed on exactly the same likelihood-ratio principles as the classical tests. In fact, the ACOVS fitting function [2.22] may be expressed in the form [2.39] by taking the $\phi_{\underline{i}}$ to be the eigenvalues of $S\hat{\Sigma}^{-1}$ at the minimum. The ACOVSF program does not take advantage of the scaling factor $\underline{K}$, which is only available when there is a closed-form estimate $\hat{\Sigma}$, but ACOVSF can find $\hat{\Sigma}$ when closed-form estimates of Σ do not exist or would be extremely tedious to obtain.

It is useful to have a preprocessor program which computes some of the classical tests on dimensionality, and also computes the matrix $V = H_3' S H$ of equation [2.9]. Much can be learned about the components of S by examining V. If we let θ_j^2 represent the component corresponding to the j-th column of H, it can be shown that the diagonal elements of V have the expected values

$$[2.40] \qquad \underline{E}(\underline{v}_{jj}) = \underline{p}\,\theta_j^2 + \overline{\sigma^2} ,$$

where $\overline{\sigma^2}$ is the mean of the error variances (see the discussion, earlier in the chapter, in connection with Table 2.6). θ_1^2 is the content component σ_ζ^2, θ_2^2 is the component σ_α^2 corresponding to the α facet of the design, and so on. It

that the diagonal elements for th

will often be apparent by inspection that the/content component and one or more of the other components are considerably larger than the others; if so, these are the "probable components" present in the data. We aim to fit $\underline{r}$ fewer than $\underline{p}$ components, and by formally testing for equality of the smaller diagonal elements of V --which estimate $\overline{\sigma^2}$ if the corresponding variance components are null--we can obtain a preliminary estimate of $\underline{r}$. The rationale for looking at V in order to make decisions about the components of S has been explored earlier in the chapter, and comes from the work of Bock and his associates (particularly Bock, 1960; Bock & Bargmann, 1966; Bock et al., 1969; see also Bock, 1975, for an application of the same principles in testing the assumptions for a repeated-measures ANOVA).

Somewhat speculatively, I have adopted the following procedure for testing V , and found it useful. (Examples will be given in the results section of the dissertation.)

1. Obtain $V = H'SH$.

2. Reorder V so that the last $p-1$ diagonal elements are in descending order by size, which is an implicit reordering of the columns of H , reserving the first column of H and the leading diagonal element of V for the content component (content should be one of the larger components in the data, but it doesn't have to be).

3. For $s = p, p-1, \ldots 2$, obtain the eigenvalues $\phi_i^{(s)}$, $i = 1, s$ of V_s , where V_s is the lower-right $s \times s$ submatrix of the reordered V .

4. Test V_s for sphericity at each step using [2.39], where $M = N - 1 - (2s^2+s+2)/6s$, and

$$[2.41] \qquad -M \ln W \sim \chi^2[\tfrac{1}{2}(s^2+s-2)] \ .$$

This is simply an application of the classical Mauchly (1940) test for the successive submatrices of V .

For $s = p$ and $p-1$, [2.41] corresponds to well-known tests on S . With data thought to be suitable for component analysis, it should be possible to comfortably reject the hypothesis of sphericity with $s \geq p-1$, and fail to reject for some $s > 2$. The value of s for which we first fail to reject sets an upper limit on the number of components which can be detected in the data, in the metric of the data as given, and assuming uncorrelated and (strictly speaking) homogeneous error variances for the measures.

For $\underline{s} = \underline{p}$, [2.41] is indentical to Mauchly's test. I have described it as a test on V, but since V and S differ only by a similarity transformation and therefore have the same eigenvalues (Graybill, 1969, p. 45), and Mauchly's test for sphericity of S depends only on the eigenvalues of S, the same test can be made using the eigenvalues of V. If we fail to reject the hypothesis of sphericity for $\underline{s} = \underline{p}$, it is a very bad omen: It implies that our measures of the same ostensible content are homoscedastic and uncorrelated.

For $\underline{s} = \underline{p}-1$, [2.41] is identical to Huynh and Feldt's (1970) test of the hypothesis $\underline{H}_0: \hat{\Sigma} = aj' + ja' + \sigma^2 I$ (where a is a real-valued $\underline{p} \times 1$ vector, and j is the $\underline{p} \times 1$ unit vector), which is a test of fit to the "Type H" pattern required for pooling error terms in the one-sample repeated-measures ANOVA. If we fail to reject this hypothesis, it implies that only one random component σ_ζ^2 is needed, and none of the components associated with the design facets are detectable; however, the error variances may be regarded as possibly correlated with individual differences on the content component. If mean differences on the observed measures are meaningful, failure to rejct the hypothesis of sphericity at this step will justify an analysis of mean differences using mixed-model assumptions.

For $\underline{s} \leq \underline{p}-2$, [2.41] is testing the hypothesis that V has the form:

$$[2.42] \quad \underline{E}(V) = \underline{p}\begin{pmatrix} C_{zz} & C_{ze} \\ C_{ez} & 0 \end{pmatrix} + \sigma^2 I ,$$

where 0 is the $\underline{s} \times \underline{s}$ null matrix. Comparing this with [2.9b], [2.42] implies the strong assumption that $C_{ee} = \sigma^2 I$, but contrary to the development of the component analysis model in the early part of the chapter, it allows $C_{ze} \neq 0$. Other assumptions are possible, of course, and may be tested using ACOVS methods. As the purpose of the tests on V is to gain an idea of the dimensionality and probable components in the data, the matter need not be pursued further here.

As a final test on dimensionality, it is useful to perform an unrestricted maximum-likelihood factor analysis (UMLFA) on the data, using a program such as UFABY3 (Jöreskog & van Thillo, 1971), testing for goodness of fit using successive numbers of factors. This also sets a lower limit on the rank of the simply-patterned component model. Depending on the assumptions made for the component model, it may not be possible to fit a component model with a rank as low as the first UMLFA model which fits the data.

Final remarks on dimensionality. The search procedure for making preliminary checks on dimensionality and the probable components is highly exploratory, of course, and together with the subsequent testing of ACOVS models it is highly conducive to Type I error. The significance tests associated with the various maximum—likelihood criteria used throughout the process have hardly a shred of statistical meaning left at the end. The significance tests are being used as a kind of yardstick against which the merits of models can be compared, for purposes of separating the plausible wheat from the implausible chaff.

In evaluating alternate models, the size of various efects may be as important as the significance levels associated with them. Large first differences in the roots of S and R , or in the diagonal values of V , tend to be accompanied by large χ^2 differences in the significance tests for dimensionality. The sizes of the estimated variance components also tend to speak for themselves. Suppose the following estimates are obtained: $\hat{\sigma}_\xi^2 = .5$, $\hat{\sigma}_\alpha^2 = .3$, and $\hat{\sigma}_\beta^2 = \hat{\sigma}_\gamma^2 = .02$; even if the β and γ components are real, they are an order of magnitude smaller than the first two components, and (depending on the size of the error estimate) probably cannot be measured very well. In the datasets I have looked at, it is comforting to discover that the variance component estimates do not change very much across a wide variety of trial estimates on the other parameters, especially for the larger components. Often the rough estimates of components based on the elements of V , such as the ones used earlier in the chapter, will be within 5 or 10% of the estimates obtained after testing many ACOVS models.

Characterizations of particular simply-patterned ACOVS models. In this section, some general features of analysis for 2-component and 3-component models are discussed. To be thorough, we can begin with 0-component and 1-component models. The 0-component models of interest are:

$$\hat{\Sigma} = I \text{ (for correlations), and}$$
$$\hat{\Sigma} = \sigma^2 I ,$$

which may be tested in closed form for goodness of fit by the Wilks (1932) and Mauchly (1940) likelihood-ratio tests mentioned earlier. The most interesting 1-component simply-patterned models are:

$$\hat{\Sigma} = \sigma_\zeta^2 jj' + \sigma^2 I ,$$
$$\hat{\Sigma} = \sigma_\zeta^2 jj' + E ,$$
$$\hat{\Sigma} = ff' + \sigma^2 I , \text{ and}$$
$$\hat{\Sigma} = ff' + E ,$$

where f is a $p \times 1$ free vector which may be expressed in terms of the elementary matrices in Table 2.8 as $f = A_4B_1$. The first of the 1-component models is the compound symmetry model, for which Wilks (1946) provided the closed-form test of fit, and the last is the unrestricted 1-factor model, which may be tested by UMLFA methods.

With data suitable for component analysis, we should be able to reject all of the 0-component and 1-component models, as mentioned earlier, but we can use the 1-component models to illustrate the "identification" problem in maximum-likelihood estimation. The general model $\hat{\Sigma} = A(BCB')A + E$ is said to "unidentified" if the free parameters of the model do not have unique values at the minimum $\hat{\Sigma}$. For example, if $A = A_4$ and $C = C_2$, which are the least-restricted forms of these matrices which we are considering, the general model will be unidentified even if there is a unique value of $\hat{\Sigma}$ for which the maximum-likelihood equations are minimized. For any estimate of $\hat{\Sigma}$, it will always be possible to divide A_4 by any arbitrary constant $\underline{k}$ and multiply C_2 by $\underline{k}^2$, and still obtain the same $\hat{\Sigma}$. Thus, the least-restricted elementary matrices A_4 and C_2 can never be used simultaneously. In the 1-component models above, the term $\sigma_\zeta^2 jj'$ may be specified as $I(B_1 C_2 B_1')I$, and the term ff' may be specified as $A_2(B_1 I B_1')A_2$, which sets A to fixed values in the first case and C to fixed values in the second.

The range of 2-component models introduces some complications. The pattern matrix for 2-component models may always be expressed (by rearranging the observed measures if necessary) as

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix} = (L L L L),$$

where

$$L = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

is the "building block," so to speak, for the pattern B_2. If the two components are content and agreement, this specification of the model assumes that the observed measures are alternately <u>true</u> and <u>false</u> keyed. It follows from the general model that

$$\hat{\Sigma} = A \begin{pmatrix} LCL & LCL & LCL & LCL \\ LCL & LCL & LCL & LCL \\ LCL & LCL & LCL & LCL \\ LCL & LCL & LCL & LCL \end{pmatrix} A + E = A(J \otimes LCL)A + E ,$$

where $J = jj'$ is the 4×4 unit matrix. It is useful to distinguish between the "common" or "component" part of the model, which is $\hat{\Sigma} - E = A(J \otimes LCL)A$, and the error part, which is simply E. We will allow E to be either homogeneous or heterogeneous, depending on which gives the better fit. Now consider the common part of the model. The meaning of the scaling factors in A is that, in some metric for the data, we require the component-space to have the form

$$A^{-1}(\hat{\Sigma} - E)A^{-1} = J \otimes LCL ,$$

in which each of the 2×2 blocks LCL is identical with every other such block. If $A = I$, then we are constraining component-space to have the form $J \otimes LCL$ in the given metric for the data, and in general, we will prefer to have A known, for ease of interpretation. If the "proper" metric for the data is not known, however, optimal scaling is provided by letting the elements of A be free, subject only to the minimal restrictions needed for identifiability.

Let $G = LCL$, and since $L^{-1} = \frac{1}{2}L$, we can write $C = \frac{1}{4}LCL$. Since $LCL = IGI$, we can rewrite the simply patterned model as

$$\hat{\Sigma} = A(J \otimes IGI)A + E = A(IIII)'G(IIII)A + E .$$

Thus, any 2-component model based on $B_2' = (L\,L\,L\,L)$ can be re-expressed in terms of the pattern $B_2^{*\prime} = (I\,I\,I\,I)$, since $B_2 C B_2' = B_2^* G B_2^{*\prime}$, and vice versa. Note that

$$C = \begin{pmatrix} \hat{\sigma}_\zeta^2 & \\ \hat{\sigma}_{\zeta\alpha} & \hat{\sigma}_\alpha^2 \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} \hat{\sigma}_{\underline{t}}^2 & \\ \hat{\sigma}_{\underline{tf}} & \hat{\sigma}_{\underline{f}}^2 \end{pmatrix}.$$

In component-space and in the metric A, C contains the content and agreement covariance matrix, and G contains the true and false covariance matrix. But while models based on B_2 and C can be freely converted to models based on B_2^* and G, and vice versa, placing restrictions on C will not generally have the same effect as placing restrictions on G. If C is required to be diagonal, for example, we find

$$G = LCL = \begin{pmatrix} \hat{\sigma}_\zeta^2 + \hat{\sigma}_\alpha^2 & \\ \hat{\sigma}_\zeta^2 - \hat{\sigma}_\alpha^2 & \hat{\sigma}_\zeta^2 + \hat{\sigma}_\alpha^2 \end{pmatrix},$$

so that G is not diagonal, but does have the implicit restriction $\hat{\sigma}_{\underline{t}}^2 = \hat{\sigma}_{\underline{f}}^2 = \hat{\sigma}_\zeta^2 + \hat{\sigma}_\alpha^2$. Similarly, if C is required to have equal diagonal elements, making it proportional to a correlation matrix, G is implicitly required to be diagonal but not necessarily homogeneous. I mention these equivalences, because we may occasionally want to consider restrictions on C other than $C = C_1$ and $C = C_2$, and it is useful to know what/certain restrictions on the components imply for the observed covariances. Whatever the metric of the data, and whatever restrictions are on C, the general 2-component model implies that any true-false pair of measures has a constant correlation

$$\hat{\rho}_{\underline{tf}} = g_{12}/\sqrt{g_{11}g_{22}}$$

in component-space. The true-false pairs cannot have a constant correlation in test-space under the model, however, unless the error variances are homogeneous. If the error variances are homogeneous,

$$\hat{\underline{r}}_{\underline{tf}} = 8g_{12}/\sqrt{(8g_{11}+\hat{\sigma}^2)(8g_{22}+\hat{\sigma}^2)}$$

estimates the constant correlation in test-space. (These examples are expressed in terms of the 2-component content and agreement model, but can be modified for other pairs of components.)

In testing a range of 2-component models, I will use the crossed combinations: A_1 vs. A_2 , C_1 vs. C_2 , and E_1 vs. E_2 from Table 2.8. Note that A_2 has fixed 1's for the two leading diagonal elements. As a general rule, at least one diagonal element of A must be fixed whenever C is completely free, as mentioned earlier. In the 2-component model, however, one additional restriction is needed for identification whenever $C = C_2$. The nature of the identification problem is a little difficult to pin down, but it turns out that if A_3 and C_2 are paired, we can obtain the same estimate $\hat{\Sigma}$ by multiplying g_{12} by an arbitrary $\underline{k}$ and g_{22} by $\underline{k}^2$, and the transformation is absorbed by dividing $\underline{a}_2$, $\underline{a}_4$, $\underline{a}_6$ and $\underline{a}_8$ by the same value $\underline{k}$. The additional restriction can be placed on C , if desired. The approach adopted here has the effect of expressing the variance components in the metric of the first true-false pair of observed measures.

3-component and 4-component simply-patterned models are obtained by taking B equal to B_3 or B_4 from Table 2.8. It may be necessary to permute the rows and columns of S so that B will have the particular columns given in Table 2.8 or, equivalently, the rows of B can be permuted to match the components of the observed S . For 3-component and 4-component models, we can write [2.35] as

$$\hat{\Sigma} = A \begin{pmatrix} K \\ K \end{pmatrix} C \, (K' \; K') \, A + E$$

$$= A \begin{pmatrix} KCK' & KCK' \\ KCK' & KCK' \end{pmatrix} A + E$$

$$= A \, (J \otimes KCK') \, A + E ,$$

where J is a 2 x 2 unit matrix, and K represents the 4 x 3 upper half of B_3 or the 4 x 4 upper half of B_4 . Thus, for these models, the component part can be partitioned into four identical submatrices (in the metric of A). When C is diagonal, it can be shown by carrying out the multiplication KCK' that each of the submatrices has the Bargmann (1957) "equipredictability" pattern. For identification, at least one element of A must be fixed whenever $C = C_2$, and it is convenient to set the scaling matrix to $A = A_3$ if scaling factors are to be used.

Chapter 3

Reanalysis of Morf's (1968) Data

Drs. Morf and Jackson graciously provided data collected for Morf's (1968) dissertation, giving me the opportunity to test some covariance structure analyses of the data. This chapter reports the results of the reanalysis, discusses some of the theoretical and methodological problems encountered, and lays the groundwork for the design and hypotheses of the present dissertation.

As discussed earlier, Morf's study is the most ambitious attempt to construct faceted scales suitable for measuring both the agreement and endorsement types of acquiescence. For each of four PRF content areas--Exhibition, Play, Succorance and Understanding--Morf constructed four short scales in a 2 x 2 design, for measuring content and the facets of agreement and endorsement. Additionally, each of the resulting items was translated from self-descriptive ("I am _X_") to attitude ("_X_ is good," or "I like _X_") format, which provides a potential basis for measuring a third response style component which I will call "form." Thus, the heart of the data base was a completely faceted structure, with 4 content areas x 2 agreement x 2 endorsement x 2 form facets, yielding a total of 32 variables. Nineteen other variables, some faceted and some not, were also included, and involved scales for: heterogeneous

MMPI and PRF content; PRF desirability; the F-scale variations of Jackson and Lay (1968); social desirability scales for an adjective checklist; and sex.

In Morf's analysis of the data, the entire correlation matrix (51 variables) was factored and targeted for eight factors:

Stylistic factors:

1. true responding (agreement)
2. item endorsement (endorsement)
3. adjective endorsement
4. desirability

Content factors:

5. Exhibition
6. Play
7. Succorance
8. Understanding

Preliminary analyses had indicated that a factor for the response style of form would not emerge, but the fit of the data to the above eight factors appears to be very good. Morf used Bentler's (1971) Clustran procedure, which yields a best fit to oblique hypothesis vectors, followed by rotation to the orthogonal matrix which best fits the oblique loadings.

Factoring the matrix of intercorrelations for the oblique solution yields some interpretive bonuses among the second-order factors. At the second-order level, the item endorsement and adjective endorsement factors merge, suggesting that similar impulsive response patterns might be involved in both endorsement factors. The true responding (agreement) and understanding factors also merge into a second-order factor, loading in opposite directions, suggesting that agreement is negatively related to the dimension of intellectual curiosity which the Understanding scale was intended to measure. These results lend support to the hypothesis that agreement is related to intelligence, and endorsement to impulsiveness (Messick, 1967).

Theoretical Issues

It is well known that factor analysis suffers from two major kinds of indeterminacy--the indeterminacy of the communalities and unique variances of the measures, and the indeterminacy of rotation. Indeterminacy of the communalities is closely related to the number-of-factors problem, since it is relatively simple to solve for the unknown communalities when a decision about the number of factors has been reached. Once the number-of-factors issues is settled, indeterminacy of rotation poses problems which can bedevil attempts to make substantive sense of the data, since we want to identify a convenient set of axes for common-factor space with the theoretical constructs thought to underlie the data.

The next two sections of the chapter briefly evaluate Morf's analysis from the standpoint of the issues raised by the two kinds of indeterminacy. A third section discusses the problem of "adequate reversal" of items and scales.

<u>Indeterminacy of communalities and the issue of significance tests</u>. Morf's conclusions about his data do not rely on formal significance tests, apart from a few χ^2 tests on the sign patterns obtained from his targeted factor loadings. In this, he was following standard factor-analytic practice. Since the early days of factor analysis, the lack of significance tests has been one of the most frequent criticisms of the method. Although maximum-likelihood tests for the number of factors in factor analysis, and for the dimensionality of a correlation matrix, have been available since Lawley's (1940) and Bartlett's (1950, 1951) classic papers, my impression is that over 90% of factor-analytic results are still being reported without preliminary statistical checks on the dimensionality of the data being analyzed.

A researcher must decide the important number-of-factors question before he can solve for the unknown communalities of the measures. The most widely used rule-of-thumb for deciding the number-of-factors question is probably still the Guttman-Kaiser rule: Retain as many factors as the correlation matrix has eigenvalues greater than one. Kaiser's (1960) state-of-the-art paper provides a rationale for the rule in terms of: (a) algebraic criteria of necessity; (b) psychometric criteria of reliability; and (c) psychological criteria of meaningfulness. Point (a) refers to the fact that the rule corresponds to Guttman's (1954) weak lower bound for the number of factors (Guttman's strong lower bound yields too many factors). Point (b) refers to the principle that principal components must have eigenvalues greater than one in order to have a positive coefficient-α; apparently Kaiser never published a formal proof of this principle, but McDonald (1970) has provided a proof under certain restrictive assumptions. Finally, Kaiser (1960, p. 144) considers point (c) decisive: The rule leads "almost invariably . . . to the number of factors which practicing psychologists were able to interpret." He summarily dismisses the "statistically correct but scientifically issue-confusing significance tests of Lawley and Company," and provides words of comfort for "those of you who have been browbeaten by the imprecations of second-rate statisticians into thinking that a significance test for the number of factors is essential to the proper application of factor analysis" (Kaiser, 1960, pp. 143-144). A decade later, however, in another state-of-the-art paper, Kaiser concluded that "the most important future work . . . should continue to concentrate on the number-of-factors question" (Kaiser, 1970, p. 414), so perhaps the issue is not yet settled. The later paper also provides an unusual glimpse of a factor analyst at work, deciding the number-of-factors issue in a doubtful case: The process involves "root-staring, plotting the value of eigenvalues against their ordinal numbers, consulting my tummy for a nice answer upon staring at this plot, and then seeing if I could make up some . . . rational, simple rule of behavior which explicates objectively what my tummy tells me" (p. 407). In his inimitable manner, Kaiser was asserting that the "right" factor solution tends to be esthetically pleasing, and that the best factor analysts have always been guided by experience, intuition and esthetic sense--not necessarily in that order. Naturally, psychologists whose tummies react badly to eigenvalue plots are made somewhat uneasy by this state of affairs; they might be comforted if some formal significance tests could be found to correlate with feelings in the tummy.

The issue of significance tests in factor analysis must be considered historically, in terms of available techniques. Although Lawley established a theoretical basis for a test of numbers of factors in 1940, it involved "matrix inversion, and eigenvector-eigenvalue problems of the worst sort," as Kaiser (1960) put it. The convergence properties of early maximum-likelihood solutions were rather poor. As McDonald (1970) made clear, there is no necessary connection between the theoretical basis of factor analysis models and the properties of iterative algorithms needed to obtain solutions with sample data: The models yield matrix functions to be maximized or minimized, but the practical problems involved in actually finding the stationary points of the functions, within the limitations of accuracy for digital computers, can be severe. It is hardly helpful to have an elegant statistical model for factor analysis, if one's data are too extensive to analyze by hand, and the available computer programs behave somewhat unpredictably. Jöreskog's (1966, 1967) computational breakthrough, which involves the use of a minimization algorithm published by Fletcher and Powell (1963), has been acknowledged as providing the first reasonably reliable practical method of maximum likelihood factor analysis (e.g., McDonald, 1970; Hakstian & Muller, 1973).

Maximum likelihood tests for dimensionality need not be decisive, of course. They depend for power on the number of subjects, measures and factors, the selected alpha level, and the (unknown) actual size of relationships between measures in the population. A researcher may decide that he has too much power, and choose not to interpret the trivial factors. Alternately, he may decide that he has too little power, and either eliminate some of his measures from the analysis, or collect data from more subjects. The power functions of multivariate procedures are technically very formidable, and they depend in any case on unknown population parameters, so that a researcher must still rely on rules-of-thumb and "experience" for deciding whether he has too much, too little, or just the right amount of power (Jöreskog, 1969). A reasonable rule would be that data at hand should pass a statistical test for dimensionality as an upper bound for the number of interpretable factors (Gorsuch, 1974, p. 141).

Montanelli (1974) and Browne (1968) have indicated that maximum likelihood methods sometimes "underestimate" the correct number of factors when *N*s are small, but the sense of "underestimates" seems misleading in this context. One would hardly complain that a *t* test underestimates the mean difference because a sample mean difference was found to be not significant in a study with a known population mean difference of $.5\sigma$ and *n* = 32 in each group; with $\alpha = .05$, the power of the *t* test can be readily calculated as .50, so that there was only a 50-50 chance of finding a significant mean difference from the outset (Cohen, 1969, p. 34). Instead of concluding that maximum likelihood sometimes underestimates the number of factors, we should conclude that sample sizes are sometimes much too small. If the *N* is too small to use maximum likelihood methods, then it is probably too small to use other methods as well. Hakstian and Muller (1973) compared maximum likelihood and other decision rules for numbers of factors, for 17 published data sets involving 6-34 measures and 100-421 subjects. Maximum likelihood yielded the same number of factors or more factors than the roots-greater-than-one rule, for 15 of the data sets, and also matched the number of factors extracted in the original studies more often than the roots-greater-than-one rule; so it is probably not the case that the use of maximum likelihood will result in fewer factors being extracted, or that the published literature is full of studies having inadequate sample sizes for the number of factors extracted (cf. Armstrong & Soelberg, 1968; Tobias & Carlson, 1969).

Morf extracted eight factors from his data, essentially because that was the number hypothesized. By several common tests for numbers of factors, the data were underfactored. The tests were applied to correlations involving 46 of the 51 variables included by Morf (three adjective checklist measures were not included in the data sent to me; an infrequency measure was omitted because of its low variability; sex was also omitted). Bartlett's test for dimensionality of the correlation matrix was the winner, with 19 significant dimensions before the residual roots stabilized at a non-significant value ($\chi^2(351) = 399.07$; $p < .10$). This is undoubtedly an overestimate, and on theoretical grounds we would expect Bartlett's test to overestimate the dimensionality of common-factor space. The roots greater-than-one rule was next with 14 factors, while the maximum likelihood test (Jöreskog & van Thillo, 1971) yielded 12 factors ($\chi^2(549) = 600.45$, $p = .063$) before a non-significant chi square was found. Since the roots-greater-than-one rule provides a theoretical lower limit, and the maximum likelihood test a statistical upper limit, it would seem that the appropriate number of factors for the data is in the neighborhood of 13.

Overfactoring and underfactoring lead to "fission" and "fusion," respectively, to use Cattell's expressions: If too many factors are extracted, the "true" factors will split apart in nonreplicable ways; while if too few are extracted, some of the "true" factors must fuse in order to fit the dimensionality allowed. There are indications that it is usually better to overfactor (slightly) than to underfactor (Gorsuch, 1974, pp. 156-160). In examining the solutions obtained by extracting successively more factors from data having a known factor structure, the known structure can usually be identified when a few too many factors are extracted, since the surplus factors can be isolated as specifics; but as still more factors are extracted, the known structure will become fragmented.

The question of the correct number of factors is particularly important for confirmatory factor analysis where, as in the case of the Morf data, h factors are predicted but the data will support $r > h$ factors. Morf chose to extract h factors and rotate them to a target matrix, but extracting too few factors has the effect of injecting unwanted variance into the solution and omitting some of the wanted variance. On the other hand, if he had extracted r factors, it would have been perhaps too easy to fit the predicted factors, since he would have been fitting an 8-dimensional set of hypotheses in a space of higher dimensionality.

One of the aims of this dissertation is that of applying maximum likelihood methods--which were not readily available when Morf analyzed his data--to tests of response style hypotheses. Jöreskog's (1970) general approach to the analysis of covariance structures provides tests not only for dimensionality of the data, but also for the fit of hypothesized factor structures. It provides a statistical basis for addressing the tricky problem of rotational indeterminacy, which is the topic of the next section.

Indeterminacy of rotation and the issue of procrustes methods. Block (1971, 1972) has criticized the Morf analysis for its use of procrustes rotation methods, and monte carlo studies of these methods indicate that they can yield superficially "good" solutions with random data (Horn, 1967; Humphreys, Ilgen, McGrath, & Montanelli, 1969). Humphreys et al. showed, in particular, that capitalization on chance in the rotation process tends to decrease with high numbers of subjects, with low factor/variable or high marker/factor ratios (which are confounded in their study), and to some extent with high numbers of variables. The elaborate faceting of the Morf data, which yields a high marker/factor ratio for the acquiescence factors, provides a safeguard against abuse of procrustes methods (Bentler et al, 1972). One index of the extent to which a targeted rotation results from a valid underlying structure is the number of "high" (over .3 or .4, say) loadings obtained. If the underlying structure is very different from the targeted structure, the number of high loadings is depressed by the stretching and squeezing of test vectors to fit their procrustean bed. For the Morf data, 17.2% of the targeted loadings exceed .3, and 10.3% exceed .40. This compares very favorably to Guilford and Hoepfner's (1971) results using targeted rotations. They report only 7.0% of targeted loadings over .3, and 4.6% over .4. Guilford and Hoefner also adjust their target matrix to fit their data, inserting small positive or negative loadings in the target in order to improve the fit (cf. Guilford & Hoepfner, 1971, pp. 53, 56; Horn & Knapp, 1973, p. 41).

The existing monte carlo work on procrustes rotation suggests that extreme caution needs to be used in interpreting the results of procrustes methods, but provides only crude guidelines for the practical researcher. The random data used in the Humphreys et al. study do not really resemble data collected in actual research, and more monte carlo work needs to be done, perhaps using data with different kinds of known structure, such as the Tucker et al. (1969) series of simulated correlation matrices. Jackson and Morf (1973, 1974) have provided two auxiliary studies bearing on the adequacy of the Morf analysis. The first study capitalizes on the fact that the form facet (self-descriptive vs. attitude scales) did not yield a factor, which enables them to repeat the analysis using more or less parallel subsets of the data. This study shows that the facet design of the Morf data can be found in both sets of measures. This is not too surprising, since each of the parallel subsets has a higher factor/measure and lower marker/factor ratio than both subsets combined, making it easier to fit an arbitrary target. However, if the fit in each subset is entirely due to capitalization on chance, we would not expect factors obtained in one subset to correlate with the corresponding factors in the second subset, and Jackson and Morf (1973) found inter-battery correlations of .73 and .54, respectively, for the crucial agreement and endorsement acquiescence factors. Block had suggested that the first auxiliary study "is logically faulty and does not address itself to the question, Will the Morf and Jackson factors fit a randomly constructed target matrix as well as the factors fit the hypothesized target matrix?" (Block, 1972, p. 10, referring to a prepublication copy of the Jackson and Morf, 1973, study). The second auxiliary study indicates that the answer to Block's question is _No_ (Jackson and Morf, 1974).

In a sense, the use of procrustes methods is not the real problem. Data should never be collected for a factor analytic study without _some_ hypotheses about the nature of the factors expected, and if very specific hypotheses can be made for the structure of the data, it seems only natural to analyze the data in ways which maximize fit to the hypothesized structure. In principle, we do precisely that in using analysis of variance to test hypotheses about means; the major difference is that the standard methods of ANOVA indicate whether, on stated assumptions, the observed means could be obtained by random sampling from a population in which the means are really equal. Accordingly, Jöreskog's (1970) hypothesis-testing methods provide a welcome addition to the techniques available for factor analysis, and to methods available for testing response style hypotheses.

Maximum likelihood methods do not completely solve the problem of rotational indeterminacy, because if the predicted pattern B and the estimates $\hat{\Sigma} = BCB' + E$ provide an acceptable fit to the obtained covariance matrix, then any pattern $B^* = BT$ with covariances $C^* = T^{-1}CT'^{-1}$, where T is any nonsingular matrix, will fit the data equally well (Jöreskog, 1970). Since T^{-1} exists, $(BT)(T^{-1}CT'^{-1})(BT)$ or $B^*C^*B^{*'}$ reduces to BCB'. However, maximum likelihood methods can be helpful when we need to choose between two solutions $B_1CB_1' + E$ and $B_2CB_2' + E$, where B_1 and B_2 have the same full-column rank, but $B_2 \neq B_1$. This principle will be illustrated later in the chapter, after a discussion of the theoretical issue of adequate reversal.

Adequacy of reversal and counterbalancing. Another criticism of the Morf study is that the items used were inadequately reversed, since some of the reversals may be logical contraries rather than logical contradictories (Block, 1971; Samelson, 1972). As discussed in Chapter 1, this criticism appears to be partly a misunderstanding of the factor analytic and covariance structure approaches to the study of acquiescence. Reversals need not be contradictory (though they must be contrary) along the content dimension, provided that the reversals affect the fixed means of the scales and not the rank order of individuals along the content dimension. The analysis of the data is orthogonal to the scale means, and is primarily responsive to the rank order of subjects on the scales.

Morf's results may have been affected by a different kind reversal problem, however, because the content stems used to define the response style facets are distributed nonrandomly across the facets of his design. When we examine the items defining the agreement facet for the four PRF content areas, for example, the true keyed items are based on one set of stems, while the false keyed items are based on another. For the most part, the true and false keyed items are not reversals at all, but entirely different items. The same is true for the positively and negatively phrased items defining the endorsement facet.

Table 3.1 displays the manner in which the content stems are matched with the design facets for the PRF content areas. Morf considered it unwise to administer eight variants of the same stem to a single subject, for obvious reasons, and most of the stems appear in either two or four variants. For example, the first stem for Exhibition is concerned with having a "flashy car," and the two variants were scored for the ESPT and EAPT scales; similarly, the first stem for Play is concerned with "children's games," and the two variants are scored for the PSPT and PAPT scales. Morf's complete test instrument, which contained a number of subscales not discussed here, contained a total of 560 items. These items were divided into two separate questionnaires, indicated by the letters A and B in Table 3.1, and administered at two different test sessions. So far as possible, items based on the same stem were not administered at the same test session. Some other features of the PRF scales which may be noted from Table 3.1 are that the number of items per scale varied from 4 to 8; the content areas used differing numbers of stems, ranging from 18 for Exhibition to 24 for Understanding; finally, each content area used a constant number of 24 self-descriptive and 24 attitude items.

Morf (1968, pp. 41-42) wrote the self-descriptive items first, aiming for the pattern shown for the Understanding scales in Table 3.1: Approximately six items were obtained

Table 3.1

Patterns of Stem Overlap for Morf's PRF Content Areas[a]

	Exhibition								Play								Succorance								Understanding							
	S				A				S				A				S				A				S				A			
	P		N		P		N		P		N		P		N		P		N		P		N		P		N		P		N	
Stem	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F
1.	B				A				B				A				B	A			A	B			B				A			
2.	A				B				B				A				A			B	B			A	A				B			
3.	A				B				A				B				A				B				A				B			
4.	B				A				A				B				B				A				A				B			
5.	B				A				B				A				A				B				B				A			
6.	A				B				A				B				B				A				B				A			
7.		A	B			B	A			B	A			A	B			A	B			B	A			B				A		
8.		A	B				A	B		A	B			B	A			B	A			A	B			A				B		
9.		B	A			A	B			B	A							A				B				B				A		
10.		B	A			A	B							A	B			B				A				B				A		
11.		B	A			A	B			A				B				B				A				A				B		
12.			B	B		A	A			A				B					B				A			A				B		
13.				A		B								A					A				B				A				B	
14.				A		B				B	B				A				B				A				A				B	
15.				B		A					A				B				A		B						B				A	
16.				A				B			B				A					B				A			B				A	
17.				A				B				A				B				B				A			B				A	
18.				B				A				B				A				A				B				B				A
19.												B				A				A				B				A				B
20.												A				B				A				B				B				A
21.												B				A												B				A
22.												A				B												A				B
23.																												A				B
24.																												A				B
Items	6	5	6	7	6	8	6	4	6	6	6	6	6	6	6	6	6	6	6	6	7	6	5	6	6	6	5	7	6	6	5	7

[a]A indicates item appears on scale in questionnaire A, administered at test session one, while B indicates item appears on scale in questionnaire B, administered at test session two.

or written for each of the four self-descriptive scales, with no stem being used twice; then each self-descriptive item was rewritten as an attitude item, expressing an attitude toward the "external referent" contained in the self-descriptive item. This resulted in 24 attitude items, each having a reversed self-descriptive item based on the same stem. Morf was less successful in obtaining this pattern for the other three content areas, where the same stem was sometimes used three or four times. Each self-descriptive item was then randomly assigned to the _A_ or _B_ questionnaire, and its counterpart was assigned to the other questionnaire.

Morf's item stems may be characterized as counterbalanced with respect to test session and the facet of form, but _not_ counterbalanced with respect to the facets of agreement and endorsement. For example, each self-descriptive item for Understanding has a corresponding attitude item based on the same stem, and each questionnaire _A_ item has a corresponding questionnaire _B_ item. But the true keyed items are based on one set of stems (1-6 and 13-17) while the false keyed items are based on another (7-12 and 18-24), and the positively phrased items are based on one set of stems (1-12) while the negatively phrased items are based on another (13-24). The picture is less clear for the other three content areas, but it is always the case that the stems are not counterbalanced with respect to the facets of agreement and endorsement.

Because of the lack of counterbalancing in Morf's item pool, the design facets of agreement and endorsement may be confounded with variance attributable to differing groups of stems within each content area. With a rough content analysis of the Exhibition stems, for example, we can discern one fairly homogeneous group of stems concerned with artistic or athletic performance (stems 3-7,11,16 in Table 3.1), while the remainder are concerned with being noticed in ordinary social situations (being the center of attention for unspecified reasons, standing out at parties, telling jokes, saying "something shocking," and so on). The performance items are appropriate for the construct of Exhibition, but they may also be tapping achievement needs or aesthetic interests which are not measured by any of the other items. If all or most of the performance items were included in a few cells of the 2 x 2 x 2 design (in Table 3.1, they are predominantly positively phrased and true keyed), it could yield a spurious agreement or endorsement effect which is really due to the content imbalance.

Counterbalancing the design facets across the content stems is a way of lessening the risk of inadvertent reversal failure in the scales, and the last section of this chapter discusses a balanced incomplete block design for the stems which will be used in this dissertation. We turn now to the reanalysis of the Morf data. As we shall see, at least some of Morf's scales are predicted better by patterns of stem overlap than by response style hypotheses.

Reanalysis of Morf Data--PRF

Preliminary remarks. Subject #29 was missing a card in the raw data furnished to me and was dropped, along with subjects #174 and 175, who both had three scored responses on the Infrequency scale. I also found several errors in the scoring keys. Item 352 rather than 552 belongs on the ESPF scale, and items 7 and 334 seem to belong on the HPPT scale. There were a few missing responses here and there. I scored these as 1.5 (intermediate between 1 and 2) instead of, say, plugging in the applicable item means. These may account for minor discrepancies between my results and Morf's results.

I also chose to analyze covariances between the scales rather than correlations, on the grounds that response style and error components should be more or less equal from scale to scale in the metric of the raw data, but would then not be equal in the metric of standard scores for each scale. Moreover, a sign pattern for response style--the pattern (1,-1,1,-1,1,-1,1,-1) for agreement, for example--applies properly to the raw scores since, on theoretical grounds, agreement cannot be expected to be a constant proportion of scale variance if agreement is present in the data. The fact that the scales vary in numbers of items, as shown in Table 3.1, then presented a problem, since the scales had differing variances as a function of number of items. The scales were initially scored for numbers of responses in the keyed (content) direction, and these raw scores were rescaled by the scale lengths, making them equivalent to _proportions_ of items in the content direction. All scales were then rescaled a second time by the standard deviation of the scale having the largest variance (SSNF), so that the final scales have variances less than or equal to one. The resulting variance-covariance matrices are shown in Table 3.2. The rescaling was double-checked by performing the equivalent operations on the variance-covariance matrix of the raw scores: Let S represent the variance-covariance matrix of the raw scores; then each matrix reported in Table 3.2 may be found as $S^* = \underline{k}(N S N)$, where N is a diagonal matrix containing reciprocals of the scale lengths, and $\underline{k} = 3.9279$, the reciprocal of the standard deviation of the SSNF scale--i.e., the reciprocal of the square root of the largest diagonal element of $N S N$. The net effect of the rescaling is to yield variance-covariance matrices having elements of about the same order as correlations, but in fact somewhat smaller than the actual correlations--a feature of the data which should be noted in examining the results.

Table 3.2

Variance-covariance Matrices for PRF Scales[a]

	SPT	SPF	SNT	SNF	APT	APF	ANT	ANF
				Exhibition				
SPT	6607							
SPF	1820	9545						
SNT	2421	5206	8081					
SNF	1587	2682	2769	8437				
APT	2515	0791	0773	0735	3604			
APF	0530	2862	1441	1970	0093	5446		
ANT	0693	2328	1225	-0293	0802	0156	6344	
ANF	0252	2321	1186	2932	0099	1043	-0118	6848
				Play				
SPT	7188							
SPF	0878	9200						
SNT	2264	2883	6463					
SNF	1834	2288	1216	6856				
APT	2650	2322	2512	1464	7936			
APF	-0236	3373	1294	1504	0296	7146		
ANT	1917	2636	3463	0869	2074	1575	6704	
ANF	1778	2545	0945	3691	2120	1988	0778	7745
				Succorance				
SPT	7563							
SPF	1298	7464						
SNT	2315	1796	7022					
SNF	3789	1457	2989	10000				
APT	2419	0709	1301	1849	3962			
APF	1111	4181	1572	0883	0272	7271		
ANT	1255	0980	2758	1260	0474	1341	7573	
ANF	2393	1726	1045	2752	0681	1235	1240	5860
				Understanding				
SPT	8705							
SPF	2608	7832						
SNT	2048	1836	8887					
SNF	2929	2986	0653	7372				
APT	3555	1048	2452	0675	8409			
APF	1851	4645	1694	2271	0639	6879		
ANT	0758	0374	2292	-0079	0803	0505	7228	
ANF	1807	2394	0618	2843	1177	2258	-0025	4342

[a]To 4 decimal places (decimal omitted).

<u>Unrestricted Maximum Likelihood Factor Analyses (UMLFA)</u>. Significance tests from a UMLFA for each of the PRF content areas, using the program UFABY3 (Jöreskog & van Thillo, 1972), are reported in Table 3.3. If we adopt a .10 significance criterion for goodness of fit, we are led to accept a dimensionality of three for the Exhibition and Play scales, and a dimensionality of four for the Succorance and Understanding scales. (Since the hypothesis is one of goodness of fit, and we seek "non-significant" results in order to accept the null hypothesis that the model fits, use of a higher significance level than the usual .05 is a conservative strategy.) For the areas of Exhibition and Play, the 3-factor solutions are clearly appropriate: The 3-factor fits are adequate (p = .310 and p = .652); the 3-factor solutions yield significant improvement over the corresponding 2-factor solutions (p < .001 and p < .001); and the 3-factor solutions do not differ significantly from the corresponding 4-factor solutions (p = .253 and p = .522). For the areas of Succorance and Understanding, the results are less clear-cut: The 3-factor solutions are almost adequate (p = .073 and p = .063), and are significantly improved over the corresponding 2-factor solutions (p = .003 and p < .001), but the corresponding 4-factor solutions show continued improvement in fit over the 3-factor solutions (p = .084 and p = .030). The decision reached was to accept 3-factor solutions for Exhibition and Play, and 4-factor solutions for

Table 3.3

Significance Tests from UMLFA of PRF Scales in Morf Data

Factors	Goodness of Fit			Improvement in Fit		
	Chi Square	df	p	χ^2	df	p
Exhibition						
1	108.48	20	<.001			
2	47.05	13	<.001	61.43	7	<.001
3	8.26	7	.310	38.79	6	<.001
4	1.67	2	.433	6.59	5	.253
Play						
1	108.53	20	<.001			
2	42.77	13	<.001	65.76	7	<.001
3	5.07	7	.652	37.70	6	<.001
4	.86	2	.649	4.21	5	.522
Succorance						
1	102.43	20	<.001			
2	33.24	13	.002	69.19	7	<.001
3	12.98	7	.073	20.26	6	.003
4	3.25	2	.196	9.72	5	.084
Understanding						
1	94.54	20	<.001			
2	47.27	13	<.001	47.27	7	<.001
3	13.42	7	.063	33.85	6	<.001
4	1.05	2	.591	12.37	5	.030

Succorance and Understanding. The test of dimensionality provided by UMLFA is scale-free, so the same chi square values and probability values would have been obtained if correlations or some other linear rescaling of the data had been used.

Table 3.4 displays the varimax factor patterns which result, for each of the content areas, based on the decisions adopted for the test of dimensionality. The factor patterns closely resemble the patterns predicted on the basis of stem overlap, as indicated by the column labeled "Stem pattern" in Table 3.4, a point which requires some discussion.

Let us say that two scales have "stem overlap" when they share three or more items having the same stem. We can illustrate this concept using the pattern of stem overlap for the Exhibition area in Table 3.1. Exhibition scales 1 and 5 (ESPT and EAPT) have stem overlap by this definition, because they have six items sharing the same stem. Similarly, scales 2, 3, 6 and 7 (ESPF, ESNT, EAPF and EANT) form a set having stem overlap, and scales 4 and 8 (ESNF and EANF) form another set having stem overlap. Additionally, scale 6 has stem overlap with scale 4, though scale 4 does not have stem overlap with 2, 3 and 7. If stem overlap causes scales to covary, as it should, these considerations would lead us to predict the following factor pattern for the eight Exhibition scales:

Table 3.4

Varimax Loadings and Stem Patterns from UMLFA of PRF Scales

Scale	Varimax loadings[a] 1	2	3	4	Stem pattern 1	2	3	4
Exhibition								
SPT	764*	121	048		x	0	0	
SPF	072	916*	325		0	x	0	
SNT	225	449*	287		0	x	0	
SNF	158	-039	901*		0	0	x	
APT	319*	051	028		x	0	0	
APF	015	231*	226		0	x	x	
ANT	058	261*	-032		0	x	0	
ANF	-011	136	333*		0	0	x	
Play								
SPT	444*	-148	272		x	0	0	
SPF	291	536*	216		0	x	0	
SNT	575*	213	024		0	x	0	
SNF	124	146	499*		0	0	x	
APT	431*	012	252		x	0	0	
APF	021	560*	138		0	x	0	
ANT	504*	249	-012		0	x	0	
ANF	071	184	665*		0	0	x	
Succorance								
SPT	577*	056	135	266	x	0	0	0
SPF	091	850*	069	076	0	x	0	0
SNT	209	125	802*	-016	0	0	x	0
SNF	432*	074	255	334	0	0	0	x
APT	400*	036	053	003	x	0	0	0
APF	034	470*	116	101	0	x	0	0
ANT	055	071	322*	145	0	0	x	0
ANF	141	133	083	533*	0	0	0	x
Understanding								
SPT	334	182	182	343*	x	0	0	0
SPF	060	613*	115	247	0	x	0	0
SNT	161	147	600*	-009	0	0	x	0
SNF	018	189	-009	744*	0	0	0	x
APT	898*	028	155	064	x	0	0	0
APF	023	684*	098	131	0	x	0	0
SNT	026	013	371*	-011	0	0	x	0
SNF	104	266	-018	310*	0	0	0	x

[a]Highest loading for each scale is starred.

$$[3.1] \qquad \begin{pmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & x & 0 \\ 0 & 0 & x \\ x & 0 & 0 \\ 0 & x & x \\ 0 & x & 0 \\ 0 & 0 & x \end{pmatrix},$$

and this is the "stem pattern" shown for Exhibition in Table 3.4. The x's represent positive, nonzero loadings, and the 0's represent zero or near-zero loadings. Scale 6 is indicated as having stem overlap with factors II and III. This pattern is a "simple structure" pattern, and is close to the pattern which results from UMLFA on the eight Exhibition scales, followed by varimax rotation.

Stem patterns for the other three content areas are readily derived from Table 3.1, and are shown in Table 3.4. Examining the results, we can see that the varimax pattern resembles the stem pattern for Exhibition. For Succorance and Understanding, the varimax pattern also resembles the stem pattern, except that one scale in each area (SSNF and USPT) has slightly higher loadings on a non-predicted factor than on the predicted one; in each case, however, a minor additional rotation in the (I,IV) plane will yield the predicted stem pattern. For the area of Play, the varimax and stem patterns do not resemble one another: The stem overlap hypothesis predicts that scales 2, 3, 6 and 7 will load together; instead, scales 1, 3, 5 and 7 load together, while scales 2 and 6 load on a separate factor. Moreover, further rotation will not bring the varimax and stem patterns for Play into good alignment. As we shall see, the varimax pattern for Play is more consistent with a response style interpretation of the factors than with a stem overlap interpretation, but the reverse is true for the other three content areas.

Procrustes rotation to stem and style patterns. Stem patterns for the PRF content areas were discussed in the previous section, but we need to specify style patterns in order to compare the fit of stem and style patterns to the data. For a 4-factor solution, the response style hypothesis that the underlying factors are content, agreement, endorsement and form, may be tested by rotating the obtained factors to the following target matrix:

$$[3.2] \qquad \begin{array}{cccc} C & A & E & F \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix} .$$

From a number of results indicating that the form component is not needed for the areas of Exhibition and Play, the first three columns of [3.2] may be used to represent the response style hypothesis that (for a 3-factor solution) the underlying factors are content, agreement and endorsement.

Some additional discussion of [3.2] may be useful. The portion of the target matrix representing the content, acquiescence and endorsement components is obtained by

$$[3.2] \quad \begin{array}{l} \\ \text{SPT} \\ \text{SPF} \\ \text{SNT} \\ \text{SNF} \\ \text{APT} \\ \text{APF} \\ \text{ANT} \\ \text{ANF} \end{array} \overset{\text{UNSCORED}}{\begin{pmatrix} \text{C} & \text{A} & \text{E} \\ 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{pmatrix}} \otimes \overset{\text{KEY}}{\begin{pmatrix} \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}} = \overset{\text{SCORED}}{\begin{pmatrix} \text{C} & \text{A} & \text{E} \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}},$$

where $\otimes$ represents a Hadamard or element-by-element product of vectors (Rao, 1973, p. 30). The order of the measures is the one used in Table 3.1 and throughout. With the data in unscored form, as indicated by the matrix on the left (with _true_ responses coded 2 and _false_ responses coded 1, say): (a) content is represented by the pattern given by the key, which is also the first column of the matrix on the left; (b) acquiescence is measured by counting _true_ responses, as represented by the second column; and (c) endorsement is measured by counting _true_ responses to positively phrased items and _false_ responses to negatively phrased items, as represented by the third column. The data were systematically scored in the content direction, however, so the matrix on the right must be multiplied through by the scoring key to obtain the appropriate target matrix for content-scored data, on the right. This is the reason for the somewhat contraintuitive pattern used to represent endorsement in [3.2].

The form component may be construed in one of two ways, which can be labeled F1 and F2. The pattern (1 1 1 1-1-1-1-1) for unscored data represents a differential tendency to respond true to self-descriptive items, regardless of content (F1), while the same pattern for scored data represents a differential tendency to endorse the content of items presented in self-descriptive form (F2). F1 resembles a "response style," while F2 would be interpreted along content lines (i.e., items implying "I am exhibitionistic" may be endorsed more often than items implying "It's good for people to be exhibitionistic.") If the F1 interpretation of form is adopted, then the form component should be represented by

$$\underset{\text{F2}}{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}} \otimes \underset{\text{KEY}}{\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}} = \underset{\text{F1}}{\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}} ,$$

and if the F2 interpretation is adopted, then the form component should be represented by the F2 pattern on the left. Both the F1 and F2 pattern were used, during various phases of the reanalysis, but the F1 pattern invariably yielded smaller variance components than the F2 pattern. Accordingly, the F2 pattern was retained, as shown in [3.2], to represent the form component. The F2 pattern never accounts for as much variance as the content, acquiescence and endorsement patterns, but it accounts for more than the F1 pattern.

To compare the fit to alternate stem and style factor patterns, orthogonal procrustes rotations using Schönemann's (1966) procedure were obtained for the varimax patterns in Table 3.4. The procrustes problem may be sketched as follows: Let A be the obtained factors in any orthogonal rotation, and let B be a rational target matrix corresponding to the stem or style hypotheses for the data. We want to rotate A to a least-squares approximation of B using

[3.3] $$\hat{B} = AT ,$$

where T is an orthonormal matrix such that $T'T = TT' = I$. Another way of stating the problem is that we want to minimize $tr(E'E)$, where $E = AT - B$, and Schönemann (1966) has given a completely general solution. Suppose now that we have alternate rational target matrices B_1 for stem hypotheses and B_2 for style hypotheses. For Exhibition, for example, the rational targets are

[3.4] $$B_1' = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

from Table 3.4, and

[3.5] $$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

by taking the first three columns of Equation 3.2. It is a simple matter to obtain the least-square approximations

AT_1 and AT_2 to the rational targets B_1 and B_2, respectively, but we are still in a poor position to say that one fit is better than the other. The functions minimized, say:

[3.6] $\quad tr(E_1'E_1)$, where $E_1 = AT_1 - B_1$, and

[3.7] $\quad tr(E_2'E_2)$, where $E_2 = AT_2 - B_2$,

are partly a function of the targets, and $tr(B_1'B_1) \neq tr(B_2'B_2)$ in general. For a given target B, however, we may "condition" the target to the observed communalities of A, as follows: Let $D_A = diag(AA')$ and $D_B = diag(BB')$; then we may define a "conditioned" target $B^* = D_A^{\frac{1}{2}} D_B^{-\frac{1}{2}} B$, which will have the same row sums of squares as A. Then, if AT fits B^* exactly, $tr(E'E)$ will be zero, and as $tr(E'E)$ gets larger, we can say that the fit is poorer; or if we find $tr[(AT_1^* - B_1^*)'(AT_1^* - B_1^*)] > tr[(AT_2^* - B_2^*)'(AT_2^* - B_2^*)]$, we can say that the obtained A fits B_2^* better, in a least-squares sense, than it fits B_1^*. (If B_1^* can be rotated orthonormally to B_2^*, of course, A will fit them equally well). I know of no significance test which applies to this procedure, but at least it enables us to quantify the differences in fit to the stem and style patterns.

The results of the procedure sketched in the previous paragraph are given in Table 3.5 for stem patterns and 3.6 for style patterns. Recapitulating the procedure, we have varimax

Table 3.5

Orthogonal Procrustes Rotations to Stem Patterns for Morf Data

Scale	Loadings[a] 1	2	3	4	h^2	
Exhibition						
SPT	755*	160	072		601	
SPF	024	964*	137		948	
SNT	191	505*	205		334	tr(E'E) = .231
SNF	092	149	899*		839	
APT	315*	069	038		105	$\max\|e_{ij}\|$ = .205
APF	-007	272*	175*		104	
ANT	054	251*	-079		072	
ANF	-040	200	298*		130	
Play						
SPT	503*	130	152		293	
SPF	-013	599*	243		418	
SNT	342	507*	-053		377	tr(E'E) = .642
SNF	144	178	483*		285	
APT	398*	253	162		249	$\max\|e_{ij}\|$ = .342
APF	-260	464*	224		333	
ANT	258	496*	-067		317	
ANF	124	174	660*		481	
Succorance						
SPT	525*	065	137	355	425	
SPF	064	851*	077	085	741	
SNT	141	116	811*	108	702	tr(E'E) = .567
SNF	363	080	239	417*	369	
APT	390*	040	078	064	164	$\max\|e_{ij}\|$ = .363
APF	006	470*	111	113	700	
ANT	009	068	304*	188	132	
ANF	064	140	025	551*	328	
Understanding						
SPT	385*	180	159	300	296	
SPF	106	615*	094	238	455	
SNT	216	166	580*	012	410	tr(E'E) = .516
SNF	097	183	010	740*	591	
APT	910*	015	080	-033	835	$\max\|e_{ij}\|$ = .300
APF	057	685*	075	128	495	
ANT	055	026	366*	-030	139	
ANF	139	261	-026	301*	179	

[a]Targeted loadings are starred.

Table 3.6

Orthogonal Procrustes Rotations to Style Patterns for Morf Data

Scale	Loadings 1	2	3	4	h^2	
Exhibition						
SPT	573	475	217		601	
SPF	746	-256	-571		948	
SNT	550	-091	-150		334	tr(E'E) = .539
SNF	546	-549	490		839	
APT	244	192	094		105	$\max\lvert e_{ij}\rvert$ = .424
APF	262	-172	-082		105	
ANT	172	043	-202		072	
ANF	246	-262	032		130	
Play						
SPT	400	277	236		293	
SPF	558	-186	-268		418	
SNT	506	278	-209		377	tr(E'E) = .299
SNF	435	-212	226		285	
APT	449	188	111		249	$\max\lvert e_{ij}\rvert$ = .187
APF	336	-356	-304		333	
ANT	452	225	-249		317	
ANF	511	-343	320		481	
Succorance						
SPT	458	047	367	280	425	
SPF	431	-559	-367	330	741	
SNT	571	468	-321	233	702	tr(E'E) = 1.000
SNF	532	081	240	147	369	
APT	175	053	202	299	164	$\max\lvert e_{ij}\rvert$ = .502
APF	317	-274	-225	139	246	
ANT	325	132	-095	-009	132	
ANF	489	-157	195	-165	328	
Understanding						
SPT	507	046	175	076	296	
SPF	561	-327	-181	001	455	
SNT	441	342	-256	184	410	tr(E'E) = .740
SNF	436	-349	327	414	591	
APT	581	458	389	-370	835	$\max\lvert e_{ij}\rvert$ = .336
APF	530	-355	-287	-082	495	
ANT	167	237	-182	147	139	
ANF	342	-217	104	060	179	

pattern matrices A for each of the PRF content areas, and rational target matrices B_1 and B_2 given by the stem and style hypotheses for the data; we "condition" B_1 and B_2 to the obtained communalities of A, to get B_1^* and B_2^*, and then rotate A orthonormally to least-square approximations of B_1^* and B_2^*. In Tables 3.5 and 3.6, two indices of goodness of fit are printed:

[3.8] $\quad tr(E'E)$, where $E = AT - B^*$,

which is the function minimized by Schönemann's procedure; and

[3.9] $\quad \underline{e} = \max_{i,j} |\underline{e}_{ij}|$,

the absolute value of the largest element of E.

The results in Tables 3.5 and 3.6 may be briefly summarized. By usual standards of eyeballing factor patterns, all of the solutions "look" good. By the criteria of equations [3.8] and [3.9], however, the stem hypotheses provide a better fit for the content areas of Exhibition, Succorance and Understanding, while the style hypotheses provide a better fit for the content area of Play. If we allow correlated factors, we can achieve better fits to both the style and stem patterns, and neither the stem nor the style hypotheses require that the underlying factors be orthogonal. Analyses using correlated factors will be reported in a later section, using maximum likelihood methods, but they essentially confirm the results obtained using procrustes rotations: Stem hypotheses fit better for three content areas, and style hypotheses fit better for the fourth.

It may be noted that the stem and style hypotheses are distinct, since $B_2 \neq B_1T$ (with no restrictions on T) for each of the four content areas. Standard least-square theory yields a method for checking the equivalence of the stem and style hypotheses, since it is well known that

$$[3.10] \qquad \hat{B}_2 = B_1T$$

provides a least-square approximation of B_2, where B_1 and B_2 have equal full-column rank, and

$$[3.11] \qquad T = (B_1'B_1)^{-1}B_1'B_2 \; ;$$

and if [3.10] holds exactly, then T defined by [3.11] is the matrix of transformation. As an example, we may show that the rational targets of [3.4] and [3.5] are not equivalent. Application of [3.10] and [3.11] yield

$$T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 1/3 \end{pmatrix} \text{ and}$$

$$\hat{B}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 0 & 0 & -1 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1/3 & 1 & -2/3 & -1 & 1/3 \end{pmatrix},$$

which is clearly not equivalent to the style pattern of [3.5]. Thus, even though the stem and style patterns have the same dimensionality or rank, they are not equivalent in the sense that one is merely an arbitrary rotation of the other.

<u>Analyses of covariance structure--Specification of models</u>. Analyses of covariance structure for the PRF scales were made in terms of a general model having the form:

$$[3.12] \quad \hat{\Sigma} = A B C B' A' + E ,$$

where $\hat{\Sigma}$ estimates the population variance-covariance matrix providing the best fit to the obtained matrix S , given certain restrictions on A , B , C and E . The analyses were performed using the program ACOVSF, which minimizes a transform of the maximum likelihood function under multivariate normality assumptions, namely,

$$[3.13] \quad \underline{M} = \ln|\hat{\Sigma}| + \mathrm{tr}(S\hat{\Sigma}^{-1}) - \ln|S|$$

(Joreskog,Gruvaeus, & van Thillo, 1970). In large samples,

$$[3.14] \quad (\underline{N} - 1)(\underline{M} - \underline{n}) \sim \chi^2 , \quad \text{with } \underline{df} = \tfrac{1}{2}\underline{n}(\underline{n} + 1) - \underline{m} ,$$

where $\underline{N}$ is the number of subjects, $\underline{n}$ is the number of measures, and $\underline{m}$ is the number of parameters to be estimated. For our purposes, each of the elementary matrices of the general model [3.12] can have one of two variants, distinguished by the subscripts $_1$ and $_2$:

B is a fixed $\underline{n} \times \underline{r}$ matrix, where $\underline{r}$ is the number of hypothesized components or factors, with coefficients representing either a stem (B_1) or a style (B_2) hypothesis for the data.

C is an $\underline{r} \times \underline{r}$ matrix representing covariances among the hypothesized components, and may be either a diagonal matrix C_1 or a general symmetric matrix C_2 .

E is an $\underline{n} \times \underline{n}$ diagonal matrix representing the unique variances of the scales, which may be either a homogeneous matrix $E_1 = \sigma^2 I$ or a heterogeneous matrix E_2 .

A is a diagonal $\underline{n} \times \underline{n}$ matrix representing scaling factors for the data, and may be either an identity matrix $A_1 = I$ or a restricted diagonal matrix $A_2 = \begin{pmatrix} I & 0 \\ 0 & D \end{pmatrix}$, where D is a general diagonal matrix, and I , 0 and D are each $\frac{1}{2}\underline{n} \times \frac{1}{2}\underline{n}$.

For ease of recall, the subscript $_1$ designates a more restricted matrix (or stem hypotheses, in the case of B_1) , and the subscript $_2$ designates the less restricted matrix (or style hypotheses, in the case of B_2) .

The explanation for the scaling factors in A_2 is that A cannot be a free diagonal matrix while C is either a free diagonal or free general matrix; otherwise, elements of A could be multiplied by a constant and elements of C could be divided by the square of the constant, and the same estimate of $\hat{\Sigma}$ would be obtained. Additional restrictions may be imposed by setting certain of the elements of A to one, the exact number depending on the number of measures being analyzed. For our purposes, $\underline{n} = 8$, and it is convenient to work with scaling factors having the form of A_2 , where the first four elements are set to one and the last four elements are free to vary. This implies that, with B partitioned as $\begin{pmatrix} B_x \\ B_y \end{pmatrix}$, say, the estimate of the population covariance matrix is partitioned into 4 x 4 submatrices as follows:

$$[3.15] \qquad \hat{\Sigma} = \begin{pmatrix} B_x C B_x' & B_x C B_y' D \\ D B_y C B_x' & D B_y C B_y' D \end{pmatrix} + E .$$

Since the data are arranged with the four self-descriptive scales followed by the four attitude scales, the scaling factors in D will indicate the extent to which the attitude scales have different variances than the self-descriptive scales. Any four elements of A could be fixed, but these four happen to be interpretively useful.

By systematically varying which member of the four pairs of elementary matrices is used, a 2 x 2 x 2 x 2 system of analyses is generated. The development here resembles that of Wiley, Schmidt and Bramble (1973): For a given choice of B , which determines whether stem or style hypotheses are being used, there are eight possible models, as shown schematically in Figure 3.1. Other things being equal, we should prefer a more restricted model to a less restricted one.

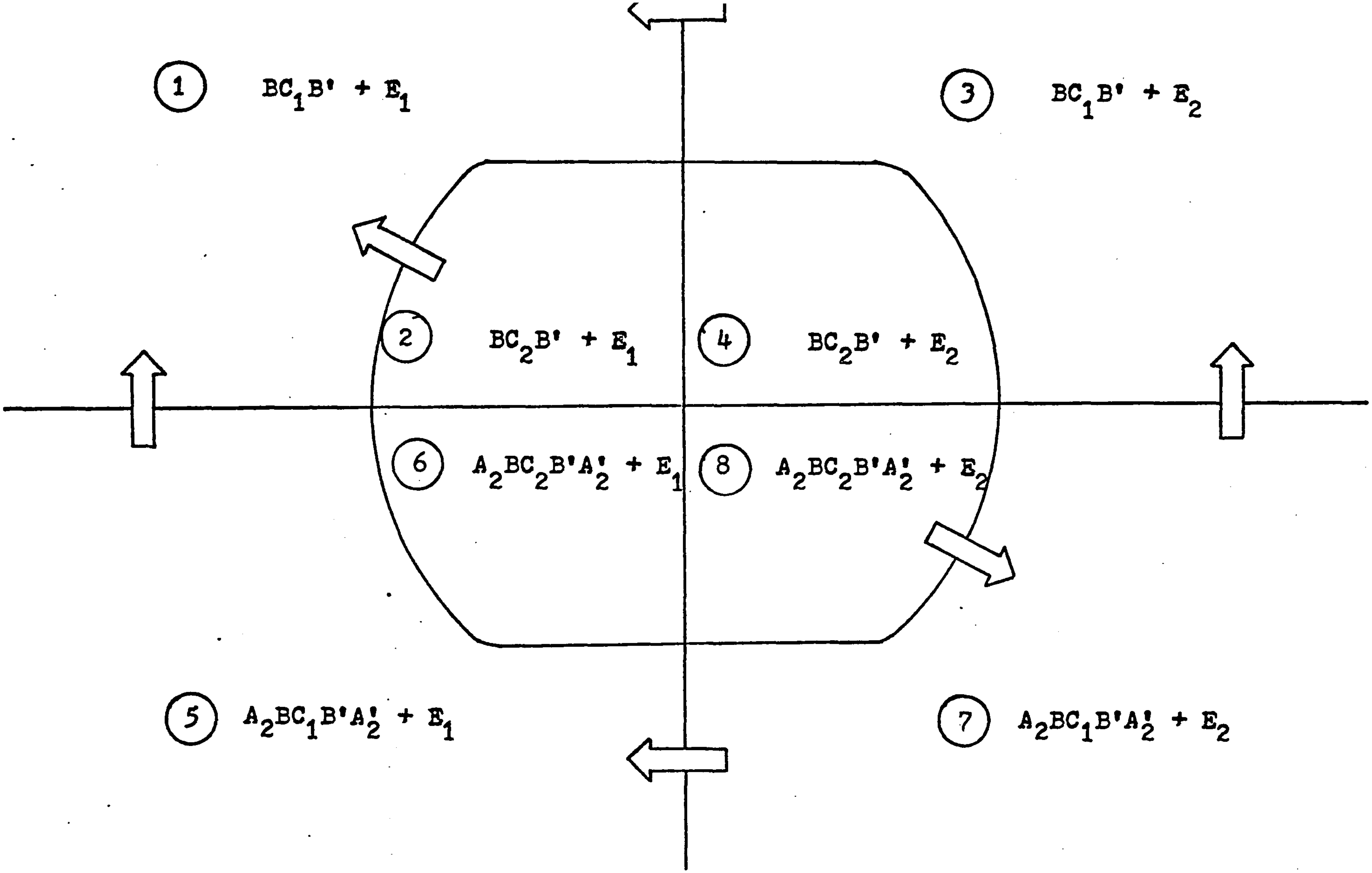


Figure 3.1. 2 x 2 x 2 System of models permitted with a given pattern B . Arrows point in the direction of more restricted models.

In terms of Figure 3.1, the principle of parsimony dictates that we should prefer models which lie in the direction of the arrows, which point toward more restricted models:

(a) toward the left side of the figure (homogeneous E);

(b) toward the top of the figure (A = I) ; and

(c) toward the outer ring of models (diagonal C) .

If the <u>most</u> restricted model (1) fits, we may accept it unless there is significant improvement by using one of the models (2), (3) or (5) which is formed by varying one of the elementary matrices of the general model [3.12]. Similarly, if the <u>least</u> restricted model (8) fits, we should not accept it unless it differs significantly from each of the more restricted adjacent models (4), (6) and (7). The significance test for differences between adjacent models can be made by evaluating the chi square difference between models, on the difference in degrees of freedom (Bock & Bargmann, 1966). It should be noted that the 2 x 2 x 2 x 2 system of models does not exhaust the possible models which may be plausible for the data, and that, in any case, the signficance tests for differences between models may not lead to a unique model which "fits best." However, the 2 x 2 x 2 x 2 system of models <u>does</u> provide a framework for analysis which may narrow the range of acceptable models.

Since alternate stem and style models with the same degree of restriction--i.e., with differing B , but the same A , C and E --will not differ in degrees of freedom, the significance test for differences between models cannot apply. Often it will be apparent by inspection whether the stem or style model fits better. In the absence of a rule of thumb for this situation, I would suggest that alternate models having the same degrees of freedom may be judged "significantly" different if the chi square difference is significant on one degree of freedom.

<u>ACOVSF solutions for PRF scales</u>. Analyses of covariance structure, using the range of models shown in Figure 3.1, are displayed in Tables 3.7 to 3.10. None of the models fit particularly well in Table 3.7 (Exhibition), which poses some special problems. I will defer discussion of the Exhibition results until last.

The results for Table 3.8 (Play) indicate that the style hypotheses yield uniformly better fits to the data than the corresponding stem hypotheses. The most restricted style model (1) fits adequately, with $\chi^2(32) = 43.34$, $p = .0871$. Model (1) does not differ significantly at the .05 level from models (2) or (5), so the fit is not improved by introducing correlated components or scaling factors. It does differ significantly from model (3), however, ($\chi^2(7) = 16.68$, $p = .0196$), so the fit is improved by introducing heterogeneous error variances. The parameter estimates under model (3) are:

$$C = \text{diag}(.207, .066, .055)$$

$$E^{\frac{1}{2}} = \text{diag}(.651, .729, .540, .604, .716, .667, .580, .632) ,$$

with the diagonal elements of C representing the facets of content, agreement and endorsement, in that order. Note that elements of $E^{\frac{1}{2}}$ are reported; these are the square roots of the error variances under the model. Under model (1), the parameter estimates are:

$$C = \text{diag}(.206, .065, .055) , \quad E^{\frac{1}{2}} = .644\, I ,$$

so the component variances in C are very similar under both models.

Table 3.7

ACOVSF Results for EXHIBITION Scales

Model[a]	Parameters C	E	A	χ^2	df	p
Stem pattern B_1						
1	1	1	1	230.88	32	<.0001
2	2	1	1	214.76	29	"
3	1	2	1	158.75	25	"
4	2	2	1	148.56	22	"
5	1	1	2	136.03	28	"
6	2	1	2	92.86	25	"
7	1	2	2	89.33	21	"
8	2	2	2	41.13	18	.0015
Style pattern B_2						
1	1	1	1	179.42	32	<.0001
2	2	1	1	167.80	29	"
3	1	2	1	131.92	25	"
4	2	2	1	125.60	22	"
5	1	1	2	123.61	28	"
6	2	1	2	108.22	25	"
7	1	2	2	86.84	21	"
8	2	2	2	63.02	18	"

[a]Note: Models have the general form: $\hat{\Sigma} = A_i B_j C_k B_j' A_i' + E_m$, where: $A_1 = I$ and $A_2 = \begin{pmatrix} I & 0 \\ 0 & D \end{pmatrix}$ with D diagonal and free; B_1 is a fixed stem pattern and B_2 is a fixed style pattern, below; C_1 is diagonal and free and C_2 is general and free; and $E_1 = \sigma^2 I$ and E_2 is diagonal and free.

$$B_1' = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad \text{(stems)}$$

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix} \quad \begin{matrix} \text{(content)} \\ \text{(agreement)} \\ \text{(endorsement)} \end{matrix}$$

Table 3.8

ACOVSF Results for PLAY Scales

Model[a]	Parameters C	E	A	χ^2	df	p
Stem pattern B_1						
1	1	1	1	149.60	32	<.0001
2	2	1	1	89.62	29	"
3	1	2	1	132.16	25	"
4	2	2	1	70.99	22	"
5	1	1	2	144.53	28	"
6	2	1	2	82.30	25	"
7	1	2	2	124.23	21	"
8	2	2	2	59.19	18	"
Style pattern B_2						
1	1	1	1	43.34	32	.0871
2	2	1	1	42.72	29	.0483
3	1	2	1	26.66	25	.3732
4	2	2	1	26.33	22	.2379
5	1	1	2	40.32	28	.0620
6	2	1	2	37.20	25	.0552
7	1	2	2	21.40	21	.4348
8	2	2	2	19.39	18	.3685

[a]Note: For parameter matrices C, E and A , see note to Table 3.7.

$$B_1' = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{matrix}\text{(stems)}\end{matrix}$$

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix} \quad \begin{matrix}\text{(content)} \\ \text{(agreement)} \\ \text{(endorsement)}\end{matrix}$$

Table 3.9

ACOVSF Results for SUCCORANCE Scales

Model[a]	Parameters			χ^2	df	p
	C	E	A			
Stem pattern B_1						
1	1	1	1	180.44	31	<.0001
2	2	1	1	87.05	25	"
3	1	2	1	124.07	24	"
4	2	2	1	48.46	18	.0001
5	1	1	2	141.88	27	<.0001
6	2	1	2	44.86	21	.0018
7	1	2	2	124.07	20	.0001
8	2	2	2	17.28	14	.2418
Style pattern B_2						
1	1	1	1	121.18	31	<.0001
2	2	1	1	81.50	25	"
3	1	2	1	81.68	24	"
4	2	2	1	48.76	18	.0001
5	1	1	2	88.48	27	<.0001
6	2	1	2	69.85	21	"
7	1	2	2	54.44	20	"
8	2	2	2	38.61	14	.0004

[a]Note: For parameter matrices C, E and A, see note to Table 3.7.

$$B_1' = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{(stems)}$$

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix} \quad \begin{matrix} \text{(content)} \\ \text{(agreement)} \\ \text{(endorsement)} \\ \text{(form)} \end{matrix}$$

Table 3.10

ACOVSF Results for UNDERSTANDING Scales

Model[a]	Parameters C	E	A	χ^2	df	p
			Stem pattern B_1			
1	1	1	1	187.81	31	<.0001
2	2	1	1	91.97	25	"
3	1	2	1	127.06	24	"
4	2	2	1	31.80	18	.0232
5	1	1	2	172.91	27	<.0001
6	2	1	2	77.16	21	"
7	1	2	2	127.06	20	"
8	2	2	2	17.98	14	.2079
			Style pattern B_2			
1	1	1	1	124.66	31	<.0001
2	2	1	1	102.00	25	"
3	1	2	1	61.83	24	"
4	2	2	1	45.85	18	.0003
5	1	1	2	99.52	27	<.0001
6	2	1	2	87.49	21	"
7	1	2	2	36.01	20	.0153
8	2	2	2	24.57	14	.0391

[a]Note: For parameter matrices C , E and A , see note to Table 3.7. The stem and style patterns are the same as for Table 3.9:

$$B_1' = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{(stems)}$$

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix} \quad \begin{matrix} \text{(content)} \\ \text{(agreement)} \\ \text{(endorsement)} \\ \text{(form)} \end{matrix}$$

The ACOVSF program also provides standard errors (s.e.'s) for the maximum likelihood estimates, and for model (3) it gave:

$$\underline{\text{s.e.}}(C) = (.026, .012, .011)$$
$$\underline{\text{s.e.}}(E^{\frac{1}{2}}) = (.044, .046, .040, .042, .045, .044, .040, .043) .$$

Estimates divided by their s.e.'s are approximately distributed as t , and if the estimates exceed twice their s.e.'s, they may be regarded as individually "significant"--that is, reliably nonzero (Jöreskog et al., 1970). All of the estimates for style model (1) and (3) meet this test. 2-facet solutions (not reported here) were also obtained using content, agreement and form, and also content, endorsement and form, but they were not as good as the solutions using content, agreement and endorsement.

By contrast with the style models for Play, none of the stem models are even close to being acceptable. Within the eight models for stem hypotheses, the results using correlated components are better than those using orthogonal components. Together with the large content component under style model (3) (.207, which is about 3½ times the size of the agreement or endorsement components), this suggests that the Play construct is well represented by all three stem groups.

When we examine the results for Succorance and Understanding in Tables 3.9 and 3.10, we find that the stem hypotheses yield consistently better fits in models with correlated components, while the style hypotheses yield better fits in models with uncorrelated components. For both content areas, however, the only model resulting in acceptable fit is the most restricted stem model (8) ($\chi^2(14) = 17.28$, $\underline{p} = .2418$ for Succorance, and $\chi^2(14) = 17.98$, $\underline{p} = .2079$ for Understanding). The model (8) solutions entail correlated components in C_2, heterogeneous error in E_2, and scaling factors in A_2. For Succorance, the parameter estimates are:

$$C = \begin{pmatrix} .538 & & & \\ .132 & .513 & & \\ .237 & .186 & .482 & \\ .383 & .180 & .262 & .464 \end{pmatrix}$$

$$D = \text{diag}(.449, .815, .572, .593)$$

$$E^{\frac{1}{2}} = \text{diag}(.467, .483, .469, .732, .532, .621, .774, .650) .$$

For Understanding, the parameter estimates are:

$$C = \begin{pmatrix} .567 & & & \\ .229 & .544 & & \\ .235 & .187 & .668 & \\ .242 & .299 & .068 & .353 \end{pmatrix}$$

$$D = \text{diag}(.627, .853, .343, .806)$$

$$E^{\frac{1}{2}} = \text{diag}(.551, .489, .470, .620, .786, .540, .803, .453) .$$

(Recall that D symbolizes the lower right submatrix of A_2, defined on page 3-41). All of the covariances in C exceed twice their standard errors, for both Succorance and Understanding, except for the $\underline{c}_{43}$ element for Understanding (where $\underline{\text{s.e.}}(\underline{c}_{43}) = .049$). Since the covariances tend to be large relative to the variances, it suggests that, despite the differences between the stem groups, the Succorance and Understanding constructs are measured well by each of the stem groups, and would be well-represented by the principal component or by an equally-weighted composite of all eight scales in each case. The scaling factors in D imply that the attitude scales have somewhat lower content saturations than the self-descriptive scales, for both content areas. There is little evidence for a form component, which would imply that the self-descriptive and attitude scales are measuring somewhat different things, so the attitude scales seem to measuring the same thing as the self-descriptive scales, but less well.

When we examine the solutions for the best-fitting style model (8) for Succorance and Understanding, there is further evidence that the style hypotheses are weak. For Succorance, the estimate of C converges on an improper solution:

$$C = \begin{pmatrix} .210 & & & \\ .004 & .030 & & \\ .028 & .020 & .059 & \\ .026 & .003 & .013 & -.007 \end{pmatrix},$$

with standard errors

$$\underline{s.e.}(C) = \begin{pmatrix} .048 & & & \\ .015 & .012 & & \\ .018 & .010 & .017 & \\ .017 & .007 & .007 & .010 \end{pmatrix}.$$

The fourth variance component (for the form facet) is negative, which is usually a sign that this component should be dropped from the model. A series of 2-facet models for agreement and endorsement were attempted, but none reached an acceptable degree of fit. These analyses did, however, tend to confirm that consistent (nonnegative) variance components for agreement and endorsement can be obtained. For Understanding, the estimate of C is a proper one:

$$C = \begin{pmatrix} .166 & & & \\ -.005 & .067 & & \\ -.001 & -.009 & .038 & \\ .036 & .009 & .009 & .020 \end{pmatrix},$$

with standard errors

$$\underline{s.e.}(C) = \begin{pmatrix} .035 & & & \\ .017 & .019 & & \\ .013 & .008 & .011 & \\ .002 & .007 & .006 & .008 \end{pmatrix}.$$

Here the form component is small, but individually significant ($2\frac{1}{2}$ times its standard error). The form component is also highly correlated with content, since we may estimate $\underline{r}_{41}$ as $.036/\sqrt{.166 \times .020} = .625$, implying that subjects with a large self-descriptive - attitude difference tend to have higher scores on the common Understanding content of the scales.

It may be noted that estimates of agreement and endorsement obtained by modeling each PRF content area separately are intrinsically weaker than the estimates obtained by Morf, who effectively used his entire body of data to estimate them. The problem we have here is that the evidence of the UMLFA solutions reported in Table 3.3 shows the dimensionality of the Succorance and Understanding scales to be four (or at least three, if the 4-factor solutions are considered to be overfitted), but none of the models which restrict the covariance structure to resemble a structure consistent with a response style hypothesis comes close to providing an acceptable degree of fit. If present, the response style variance of the scales is swamped by the variance attributable to the different stem groups.

Returning now to Table 3.7, for Exhibition, it may be useful to explore ways of relaxing or modifying the restrictions implied by the series of models in Figure 3.1, in order to provide better 3-factor fits to stem and style hypotheses. The UMLFA results in Table 3.3 seem to indicate pretty clearly that the dimensionality of the Exhibition scales is three, and I have argued from the results using procrustes rotations in Tables 3.4 and 3.5 that a stem hypothesis fits the data better than a style hypothesis; yet none of the covariance structure analyses reported in Table 3.7 yields an acceptable degree of fit. Comparing the best-fitting stem model (8) with the best-fitting style model (8), the stem model is numerically better and, evaluating the chi square difference (21.89) on 1 df, it may be regarded as significantly better; but the stem model does not fit well, since $\chi^2(18) = 41.13$, $p = .0015$.

The pattern matrices B_1 and B_2 used for the ACOVSF solutions are overrestricted, since more restrictions are used than are needed for a unique solution for A, C and E .

Without lessening the degree of restriction, slightly different patterns for B_1 and B_2 could be rationalized. For example, Exhibition scale 6 (EAPF) has weights of 1 on stem Factors II and III in the B_1 pattern matrix. Ignoring the effect of the scaling factors in A_2 , this has the effect of giving scale 6 the same weight as scales 2, 3 and 7 on Factor II, and the same weight as scales 4 and 8 on Factor III, thus tending to force scale 6 to have more reliable variance than it has actually got, relative to these other scales. In retrospect, better fits to stem hypotheses might have been obtained by reducing the weights for scale 6 from 1. to $\frac{1}{2}\sqrt{2}$, say. Another way of adjusting the B_1 and B_2 patterns would be to "condition" them to the communalities of the UMLFA solutions, in the manner described for the procrustes rotations, earlier in the chapter. The purpose served by conditioning in the case of the procrustes rotations was to make the stem and style patterns more comparable under orthogonal procrustes rotation, where there is no C matrix to compensate for certain differences in the rational targets; conditioning the targets to the communalities might improve the fit to both stem and style patterns in the ACOVSF solutions as well. Note that any adjustment of the pattern matrices which amounts to multiplying each row by a constant may be carried out under the general ACOVSF model [3.12] by declaring A to be a diagonal matrix of fixed constants. For purposes of interpretation, it seems desirable to retain the pattern matrices B_1 and B_2 in a simple, fixed, rational form, however; and for this reason, it does not seem desirable to eliminate the restrictions on the patterns in order to achieve better fits to the hypotheses.

On examining the procrustes rotations to stem hypotheses in Table 3.5, it looks like better fits to stem hypotheses could be achieved by using some value other than zero in B_1, such as a low positive loading of .10; but this violates the principle that we want to retain a simple rational form for B_1. It is worth noting, however, that B_1 was established by using an all-or-none rule based on stem overlap, with loadings of 1 assigned to define factors based on sets of tests sharing three or more stems; thus B_1 does not properly reflect differences in the degree of stem overlap. An interesting analysis which would reflect the degree of stem overlap would be to construct a "confusion" matrix for stem overlap, and test the goodness of fit of the obtained covariances to the confusion matrix. This could be done for the Exhibition scales, say, using the entries in Table 3.1. Replacing the A's and B's in Table 3.1 by 1's, and the blanks by 0's, will yield an 18×8 matrix (for the content area of Exhibition), which we may call X; then $\Phi = X'X$, say, is an 8×8 confusion or overlap matrix for the stems. A reasonable way of testing whether stem confusion is responsible for the covariances among the measures would be to fit models of the kind:

$$\hat{\Sigma} = D \Phi D + E ,$$

where D is an 8 x 8 scaling matrix, and E is an 8 x 8 diagonal matrix of errors. For such an analysis, we would probably want to either rescale Φ by its diagonal elements, or else eliminate the adjustments for scale lengths used for the obtained variance-covariance matrices.

There are ways of achieving better fits to stem and style hypotheses (essentially by modifying the hypotheses!), but it is probably not worthwhile to spend further time and effort to try out some of these ways on the Morf data, since the data base is flawed. Owing to the lack of counterbalancing in the data, the stem and style hypotheses are inevitably confounded: a fit to style hypotheses cannot be interpreted unequivocally because variance attributable to stem groups may be present in the data, and a fit to stem hypotheses cannot be interpreted unequivocally because variance attributable to response styles may be present in the data. A prima facie case for confounding has been made by displaying the patterns of stem overlap in Table 3.1, and the reanalyses of the data have served to show empirically that stem hypotheses provide a rival explanation for the data, at least for the content areas of Succorance and Understanding, and at least as far as the reanalysis has proceeded.

Reanalysis of Morf Data--F-scale

Morf also used the F-scale reversals of Clayton and Jackson (1961) in his test instrument, and a reanalysis of the portion of his data involving the F-scale may be briefly reported. As discussed in Chapter 1, this version of the F-scale has two facets, and varies true and false keyed responses (agreement) and absolute versus relativistic phrasing (overgeneralization). These reversals were developed prior to the distinction which has been made between agreement and endorsement acquiescence, but most of the reversals are of the agreement kind (polar opposite, rather than negation reversals). On examination of the items making up the subscales, it can be seen that the content stems are only partly counterbalanced for the design facets. There were 26 items written initially, with 12 true keyed items and 14 false keyed items based on different stems. These were each rewritten to provide two variants of each stem, one in absolutely-worded, sweeping form (such as "All X are always Y"), and the other in a more relative, probabilistically-worded form (such as, "At least a few X are Y, once in a while"). Thus, the absolutely and relativistically phrased items are based on the same 26 stems, and may be considered counterbalanced; but the true and false keyed items are based on different sets of stems, so the agreement facet is not counterbalanced.

Table 3.11 presents the variance-covariance matrix for the four variants of the F-scale. The scales have been adjusted for the differing numbers of items per scale, and then rescaled so that the most variable scale (FRT) has a standard deviation of one, in the manner described in detail for the PRF scales. A 2-facet solution for content, agreement and overgeneralization fits the data ($\chi^2(3) = 3.51$, $p = .3190$), but yields the following estimates for C and its standard errors:

$$C = \begin{pmatrix} .203 & & \\ .099 & .216 & \\ .001 & .005 & .004 \end{pmatrix}, \text{ and}$$

$$\underline{\text{s.e.}}(C) = \begin{pmatrix} .031 & & \\ .023 & .032 & \\ .012 & .012 & .013 \end{pmatrix}.$$

The overgeneralization component ($\underline{c}_{33}$) is very small, and neither it nor its covariances with content and agreement exceed their standard errors. Accordingly, the best-fitting 1-facet solution was accepted, and is reported in the middle panel of Table 3.11. (The series of analyses varied the form of C, which could be either diagonal or general, and the form of E, which could be either homogeneous or heterogeneous.)

Stem and style hypotheses for the F-scale cannot be distinguished statistically in Morf's data, since the stem and style pattern matrices differ by no more than a nonsingular

Table 3.11

Reanalysis of Morf's F-scale Data

Variance-Covariance Matrix

	FRT	FRF	FAT	FAF
FRT	1.0000			
FRF	-.0019	.6168		
FAT	.6035	-.0293	.9674	
FAF	.0113	.2280	-.0307	.5113

2-facet (Content and Agreement) Style Pattern Solution

$$B_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad C_1 = \begin{pmatrix} .202 & .099 \\ .099 & .214 \end{pmatrix} \qquad E = .598\,I$$

Stem Pattern Solution

$$B_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad C_2 = \begin{pmatrix} .1535 & -.0030 \\ -.0030 & .0545 \end{pmatrix} \qquad E = .598\,I$$

transformation. In detail, the stem and style pattern matrices are:

$$B_1' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \text{ and } B_2' = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Since $B_1 = B_2 T$, where $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, a matrix C_1 corresponding to the style pattern can be transformed to the matrix C_2 corresponding to a stem pattern by means of the equation $C_2 = (T^{-1}) C_1 (T^{-1})'$. The solution corresponding to the stem pattern appears in the lower panel of Table 3.11, and both solutions fit equally well ($\chi^2(6) = 3.78$, $p = .7064$).

On examination of the covariances among the four variants of the F-scale, it is apparent that the versions which share the same stems (FRT and FAT, sharing set 1, and FRF and FAF, sharing set 2) are the only variants having substantial correlations. On the evidence presented in Table 3.11, a plausible interpretation of the data is that the set of true keyed items measure one thing, while the set of false keyed items measure something entirely different. This is no more proved, however, than the interpretation that each scale measures <u>two</u> things, namely, F-scale content and agreement response style. In order to rule out a stem interpretation of the data, the agreement facet would have to be counterbalanced, as discussed in the last section of the chapter.

Reanalysis of Heterogeneous MMPI and PRF Scales

This section deals with a part of the Morf data where counterbalancing was used, and which appears to yield significant agreement and endorsement components. Morf constructed four MMPI scales and four PRF scales, which were heterogeneous and balanced for content. The construction of the PRF scales is a bit more systematic, and will be discussed first.

PRF scales. Morf began with two true and two false keyed items from each of 18 separate PRF content scales, yielding a total of 72 items. Negatively phrased reversals were written for each item (or positively phrased reversals, for items which were negatively phrased to begin with), to yield a total of 144 items based on 72 stems. Thus, for each of the original content scales, there were eight items:

Item	Stem	Content Keying	Final Keying
1	1	pt	PT
2	2	pf	PF
3	3	nt	NT
4	4	nf	NF
5	1	nf	NT
6	2	nt	NF
7	3	pf	PT
8	4	pt	PF

From each such set of eight items, items 1-4 were randomly selected, with the restrictions that they had to be based on four different stems, and had to be content-keyed with the pattern (pt, pf, nt, nf), as shown in the column labeled "Content Keying." Then the reversals of items 1-4 became items 5-8, and had to be content-keyed with the pattern (nf, nt, pf, pt), since the negation reversal of a pt item is an nf item, and so on. Items 1-4 were then assigned to "correct" scales with the pattern (PT, PF, NT, NF), while items 5-8 were assigned to "incorrect" scales, as shown in the column labeled "Final Keying." This procedure was followed for each of the 18 original content areas, and resulted in the four final 36-item scales, each balanced for content and also extremely heterogeneous for content.

If we trace the fate of the item stems for each content area, we find the following pattern:

Stems	Final Scales			
	PT	PF	NT	NF
1	pt	-	nf	-
2	-	pf	-	nt
3	pf	-	nt	-
4	-	pt	-	nf

After the arbitrary scoring of the final scales, the pattern matrix for the style hypotheses is:

$$[3.16] \qquad B' = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{matrix} \text{(agreement)} \\ \text{(endorsement)} \end{matrix} .$$

By using the vectors of [3.16] to combine the four heterogeneous scales, an agreement and endorsement composite are created, and were intended for use as acquiescence markers. Note in passing that while the agreement composite is very heterogeneous for content, it is not balanced for stems since it is essentially a contrast between those items falling in stem groups 1 and 3, and those items falling in stem groups 2 and 4. The endorsement composite based on the scales <u>is</u> balanced for stems, however, since it is based on a contrast of items in all four stem groups with negation reversals of the same items. (An item design in which the agreement and endorsement facets are both counterbalanced for stems, using polar opposite as well as negation reversals, will be discussed in the last section of the chapter.)

The variance-covariance matrix for the heterogeneous PRF scales is shown in Table 3.12. A series of maximum likelihood analyses were undertaken, which indicate that the dimensionality of the variance-covariance matrix is two. When one or more of the vectors in

$$[3.17] \qquad B^{*\prime} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

(where B^* is the orthogonal complement of B from [3.16]), negligible or negative variance components result. Attempts to fit a rank 1 solution using either the agreement or endorsement pattern from [3.16] also yielded poor fits to the data.

Table 3.12

Reanalysis of Heterogeneous PRF Scales

Variance-covariance Matrix

	HPPT	HPPF	HPNT	HPNF
HPPT	1.0000			
HPPF	-.2387	.7552		
HPNT	-.1856	.0897	.7007	
HPNF	.0980	-.1116	-.2332	.6857

Results for 2-facet Solutions

Model	C	E	χ^2	df	p
1	diag.	homog.	13.31	7	.0650
2	genl.	"	11.53	6	.0733
3	diag.	heterog.	2.73	4	.6048
4	genl.	"	2.72	3	.4363

Parameter Estimates for Model 1

$$C = \begin{pmatrix} .057 & \\ & .179 \end{pmatrix} \qquad E^{\frac{1}{2}} = .741\, I$$

Parameter Estimates for Model 3

$$C = \begin{pmatrix} .059 & \\ & .176 \end{pmatrix}$$

$$E^{\frac{1}{2}} = \text{diag}(.867,\ .727,\ .680,\ .676)$$

The results from four 2-facet solutions which varied the form of C (diagonal or general) and the form of E (homogeneous or heterogeneous error) are shown in the second panel of Table 3.12. Solutions which vary the form of C but are otherwise identical (model 1 vs. 2, and model 3 vs. 4) are not significantly different, so we can conclude that models with uncorrelated components are adequate. This narrows the choice to models 1 and 3, for which parameter estimates are given in the bottom panels of Table 3.12. Model 1 fits at the .05 level, but there is significant improvement in fit when we allow heterogeneous error in model 3 ($\chi^2(3) = 10.58$, $\underline{p} = .0142$). Thus, we may adopt model 3 (which does not yield greatly different variance components than model 1).

It will be useful to see what the parameter estimates for model 3 imply about our ability to measure agreement and endorsement. The variances of balanced <u>36-item estimates</u> of α and η, the true agreement and endorsement components, may be found as:

[3.18] $$\hat{\sigma}_{\alpha}^2 = \sigma_{\alpha}^2 + \sigma^2_{\text{pooled}} = \underline{c}_{11} + \tfrac{1}{4}\sum_{1} \underline{e}_{ii}$$

$$= .059 + .550 = .609 \text{, and}$$

[3.19] $$\hat{\sigma}_{\eta}^2 = \sigma_{\eta}^2 + \sigma^2_{\text{pooled}} = \underline{c}_{22} + \tfrac{1}{4}\sum_{1} \underline{e}_{ii}$$

$$= .176 + .550 = .726 .$$

This implies that the parallel-form reliabilities of such 36-item estimates would be:

[3.20] $$\hat{\rho}_{\alpha\alpha} = \sigma_{\alpha}^2 / \hat{\sigma}_{\alpha}^2$$

$$= .059/.609 = .097 \text{ , and}$$

[3.21] $$\hat{\rho}_{\eta\eta} = \sigma_{\eta}^2 / \hat{\sigma}_{\eta}^2$$

$$= .176/.726 = .242 \text{ .}$$

Since the scales were actually administered partly on one test occasion and partly on a second test occasion, these estimates are perhaps intermediate between the reliabilities based on tests given at the same and at different test occasions. Note that because the variance components are based on 36-item scales, equations [3.18]-[3.21] apply to balanced 36-item scales of agreement and endorsement. The actual balanced estimates of agreement and endorsement in the data are based on four times as many items, so the "obtained" reliabilities (distinguished from the estimates of [3.20] and [3.21] by omitting the caret) may be estimated by applying the Spearman-Brown formula to the results in [3.20] and [3.21], which yields:

[3.22] $$\rho_{\alpha\alpha} = 4\hat{\rho}_{\alpha\alpha}/(1 + 3\hat{\rho}_{\alpha\alpha}) = .300 \text{ , and}$$

[3.23] $$\rho_{\eta\eta} = 4\hat{\rho}_{\eta\eta}/(1 + 3\hat{\rho}_{\eta\eta}) = .561 \text{ .}$$

Balanced estimates of agreement and endorsement can be obtained from the heterogeneous PRF scales, using

[3.24] $\hat{\alpha} = \frac{1}{4}(HPPT - HPPF + HPNT - HPNF)$ and

[3.25] $\hat{\eta} = \frac{1}{4}(HPPT - HPPF - HPNT + HPNF)$,

or we can obtain the same estimates by using an item analysis program to score the items as 144-item scales. When we do the latter, the coefficient-α's for agreement and endorsement are .347 and .600, respectively, which are reasonably close to the reliabilities of .300 and .561 given by [3.22] and [3.23]. On theoretical grounds, we would expect coefficient-α to overestimate the component reliabilities of $\hat{\alpha}$ and $\hat{\eta}$, since the derivation of coefficient-α assumes essentially unifactorial items (Cronbach, 1951).

One point of interest in this analysis is that the endorsement component is markedly larger than the agreement component. Morf had predicted that self-descriptive scales (as the heterogeneous PRF scales are) would have larger endorsement than agreement components, but concluded that this hypothesis was not supported (Morf, 1968, p. 71). Since the heterogeneous PRF items provide better estimates of agreement and endorsement than any other group of scales in the data, Morf's conclusion is contradicted, and the hypothesis _is_ supported.

MMPI scales. Construction of heterogeneous MMPI scales was described as follows:

> Four heterogeneous acquiescence scales were constructed by selecting items from the MMPI with neutral desirability scale values (Messick and Jackson, 1961) and intermediate endorsement frequencies (Wiggins, 1964). Sixty per cent of these items were used in their original form. The remaining items were used in their reversed and negatively worded form. Where available, the reversals of Lichtenstein and Bryan (1965) were used; where not, new reversals were written. This procedure yielded a set of 60 positively worded, and a set of 60 negatively worded items. Each set was randomly divided into two 30-item scales, one keyed true, the other keyed false (Morf & Jackson, 1972, pp.337-338).

Each of the resulting scales-- PT, PF, NT and NF, as for the heterogeneous PRF scales--is balanced for content, but the set of scales is not balanced for stems, since the scales are based on a total of 120 different item stems. In order to conclude that the heterogeneous scales are really balanced for content, it would have to be shown that the items are measures of a single kind of content or that the scale construction procedure did not differentially distribute items measuring different kinds of content into the four heterogeneous scales.

The reanalysis of the MMPI scales is given in Table 3.13, in the same format used for the heterogeneous PRF scales. A series of analyses indicated 1-facet and 3-facet solutions were not tenable, but that 2-facet solutions using the pattern of B in [3.16] tended to fit the data. The second panel of Table 3.13 reports the results for the same four 2-facet models used with the heterogeneous PRF scales. Since models using

Table 3.13

Reanalysis of Heterogenous MMPI Scales

Variance-covariance Matrix

	HMPT	HMPF	HMNT	HMNF
HMPT	.7118			
HMPF	-.4958	1.0000		
HMNT	-.1277	.2324	.6129	
HMNF	.1730	-.2667	-.3448	.6679

Results for 2-facet Solutions

Model	C	E	χ^2	df	p
1	diag.	homog.	18.85	7	.0087
2	genl.	"	13.39	6	.0372
3	diag.	heterog.	9.25	4	.0551
4	genl.	"	6.23	3	.1010

Parameter Estimates for Model 3

$$C = \begin{pmatrix} .112 & \\ & .297 \end{pmatrix}$$

$$E^{\frac{1}{2}} = \text{diag}(.521,\ .691,\ .520,\ .550)$$

Parameter Estimates for Model 4

$$C = \begin{pmatrix} .115 & \\ .035 & .300 \end{pmatrix}$$

$$E^{\frac{1}{2}} = \text{diag}(.503,\ .684,\ .529,\ .558)$$

homogeneous error components (models 1 and 2) do not fit the data at the .05 level, the choice of models is narrowed to models 3 and 4. Parameter estimates for models 3 and 4 are given in the bottom panels of Table 3.13, and indicate that the error variance of the HMPF scale is at least 1.5 times that of the other scales (e.g., $.684^2 / .558^2 = 1.50$ for model 4). Since the less restricted model 4 does not result in significant improvement over model 3 ($\chi^2(1) = 3.02$, $p = .0821$), we may tentatively adopt model 3.

Model 3 was also accepted for the heterogeneous PRF scales, so the development in equations [3.18]-[3.23] may be used to obtain component reliabilities for balanced scales of agreement and endorsement based on the MMPI items, and we find: $\rho_{\alpha\alpha} = .576$ and $\rho_{\eta\eta} = .783$. These are considerably larger than the estimates of .300 and .561 obtained for the heterogeneous PRF scales, and are almost surely inflated by content imbalance in the MMPI scales. Note that although the agreement and endorsement components are larger in absolute terms than those found for the PRF, and are also larger relative to the estimates of error, the fit of the MMPI models is generally poorer.

Selection of MMPI items neutral in social desirability would probably bias the selection procedure in the direction of selecting items from the predominantly false-keyed R, Hy, Pa and Pd scales, because of the known imbalance of keying in the MMPI (Jackson & Messick, 1962). Bock et al. (1969) estimated the test-retest reliability of an acquiescence measure based on 120 Hy items (with true and false items included) to be .348. It is not clear whether Bock's estimate should be considered an estimate of reliability for agreement or endorsement acquiescence, but, in any event, his estimate is sharply lower than the estimates based on Morf's heterogeneous MMPI scales, and more nearly resembles the lower estimates based on the heterogeneous PRF scales.

One candidate for a contaminant of the MMPI scales is social desirability, and Morf and Jackson's (1972) analysis indicates that a strong association between the MMPI and the desirability measures is present. Figure 3.2 displays the Factor II (Endorsement) by III (Desirability) plot from the targeted rotation (Morf & Jackson, 1972, p. 342), showing only those measures comprised of heterogeneous MMPI or PRF items or targeted for the desirability factor. With the measures HMPF, HMNT, HPPF and HPNT reflected in the endorsement direction, and AU (endorsement of undesirable adjectives) reflected in the desirability direction, we can run vectors through the centroids of the MMPI, PRF and desirability clusters in this plane. The MMPI, PRF and desirability clusters are quite distinct, and the MMPI centroid correlates more highly with the desirability centroid (r = −.66) than with the PRF centroid (r = .39). The

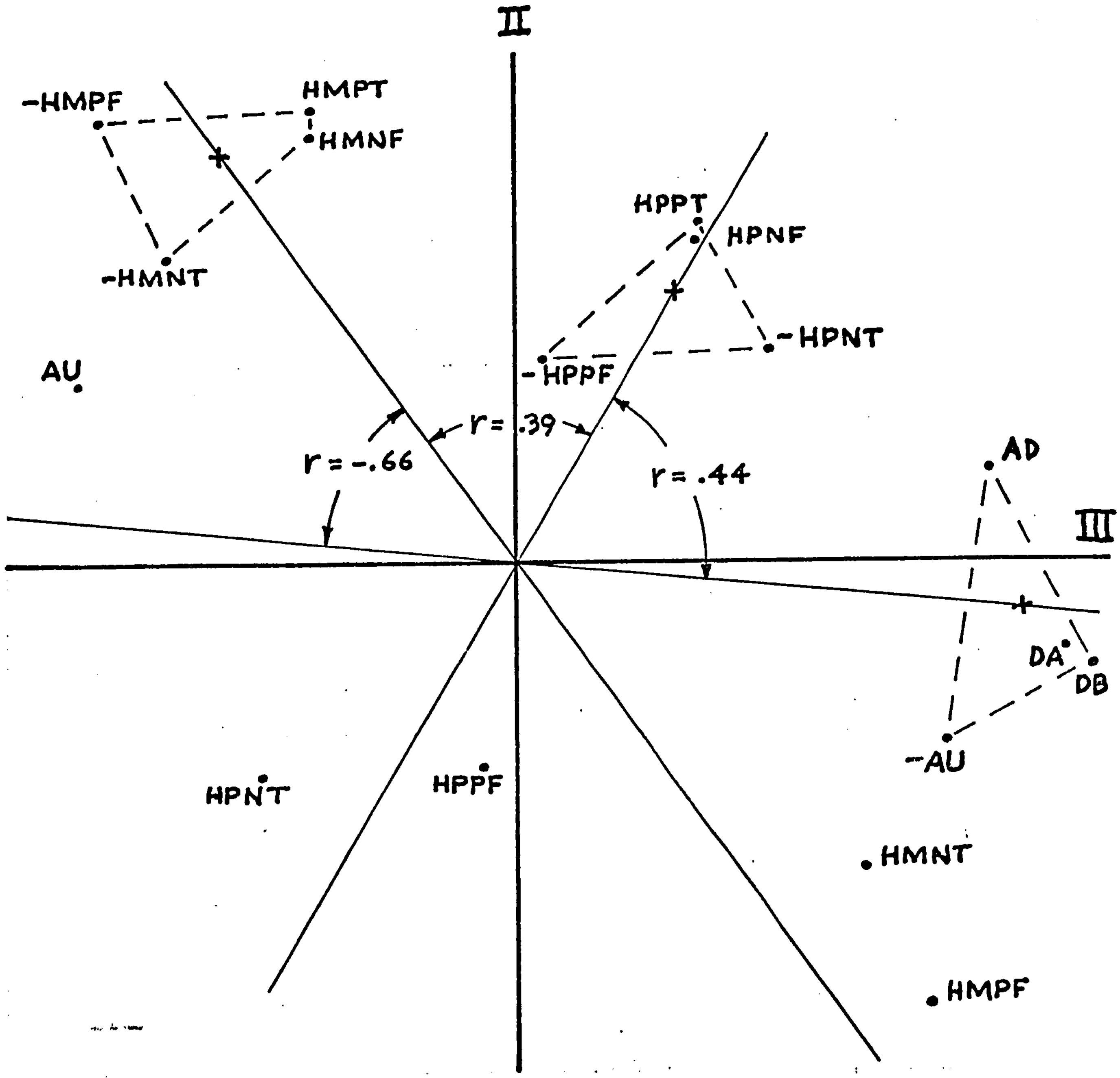


Figure 3.2. Factor II (Endorsement) by Factor III (Desirability) plot, using marker measures from Morf and Jackson (1972). Reflected measures are preceded by a minus sign, and the MMPI, PRF and desirability centroids are indicated by crosses.

MMPI-desirability relationship must be a strong one, since it shows up despite the targeting of the MMPI measures for factors other than desirability. Endorsement estimated from the PRF scales and endorsement estimated from the MMPI scales would appear to be measuring rather different things. As the correlations with the desirability centroid indicate, MMPI endorsement is rather highly and negatively correlated with desirability (*r* = -.66), while PRF endorsement is more moderately and positively correlated with desirability (*r* = .44).

Morf and Jackson noted the appearance of the MMPI scales on the desirability factor, and suggested that it was

> probably explainable in terms of the fact that although these scales were selected from the middle range of rated desirability on the MMPI, they tend to be more in the undesirable than in the desirable direction because of the predominantly undesirable content of the majority of MMPI items (Morf & Jackson, 1972, p. 346).

If the item selection and rewriting process for the MMPI scales did not control for desirability, however, it would seem difficult to claim that the scales are balanced for "content." Recall that this process resulted in positively and negatively phrased groups of MMPI items, each having equal numbers of items which were *true* and *false* keyed from an MMPI content point of view; these were then *randomly* *true* and *false* keyed to control for content. The same process should also control for desirability, however, unless the positively and negatively phrased groups of items were dissimilar in ways other than the direction of phrasing. Note that a response style theory of the data could conceivably accomodate a positive correlation between desirability and endorsement, such as we see with desirability and PRF endorsement in Figure 3.2. What it cannot accomodate is the large difference between MMPI endorsement and PRF endorsement that we also see. I would conclude that the item selection and rewriting process failed in the case of the heterogeneous MMPI scales.

The heterogeneous MMPI and PRF scales measure agreement more consistently than they measure endorsement. Figure 3.3 displays the Factor I (Agreement) by III (Desirability) plot from the targeted rotation. In Figure 3.3, the HMPF, HMNF, HPPF and HPNF scales are reflected in the agreement direction, and AU is again reflected in the desirability direction. The resulting MMPI and PRF centroids lie quite close to each other and to the Factor I axis. The manner in which the MMPI and PRF scales (as reflected) are lined up parallel to the desirability axis is consistent with the negative relationship between desirability and MMPI endorsement, and the positive relationship between desirability and PRF endorsement, which we saw in Figure 3.2.

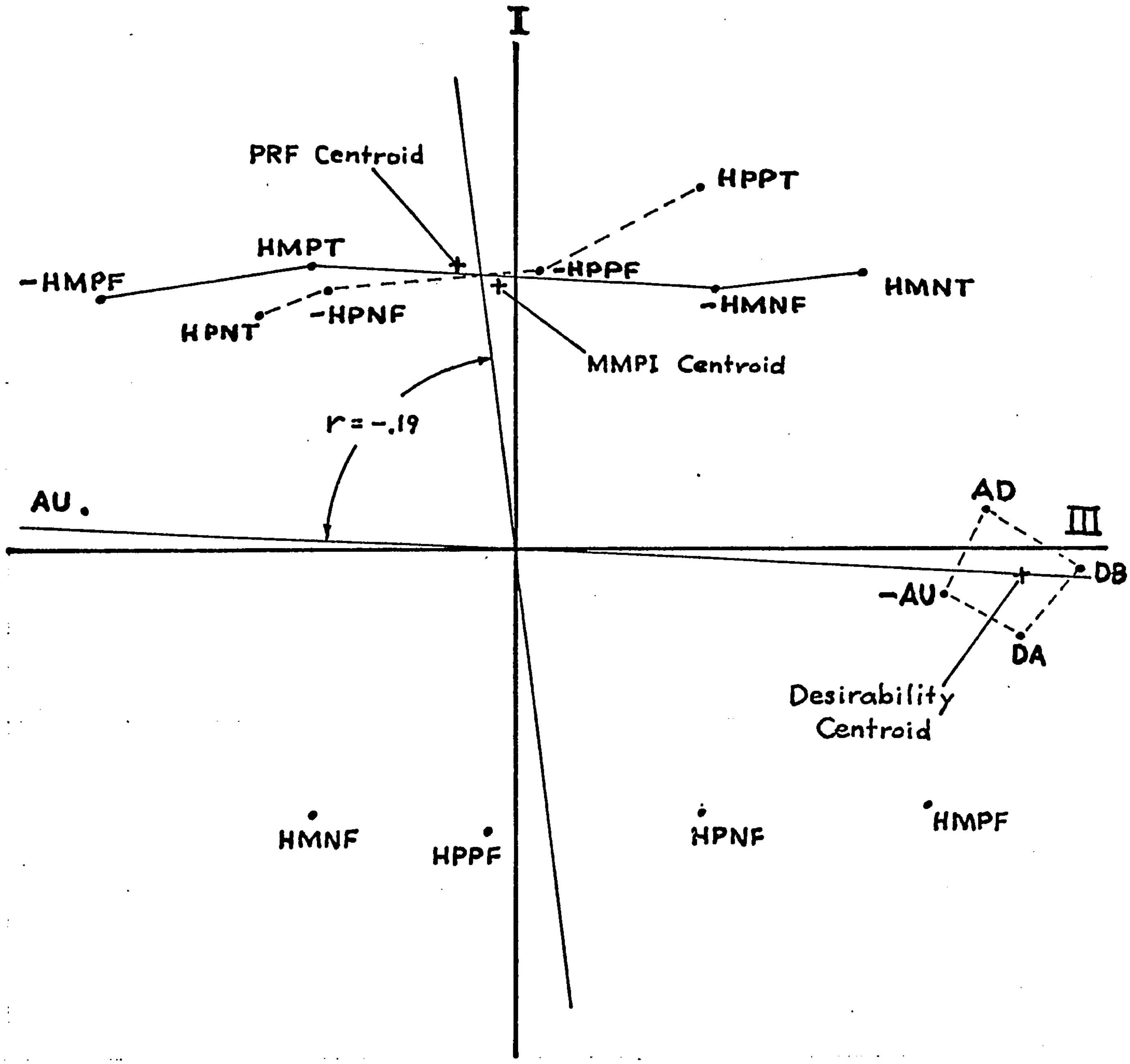


Figure 3.3. Factor I (Agreement) by Factor III (Desirability plot, using marker measures from Morf and Jackson (1972). Reflected measures are preceded by a minus sign, and the MMPI, PRF and desirability centroids are indicated by crosses.

Conclusions from the Reanalysis

My approach to the Morf data has been to use a series of subanalyses of small parts of the data base, each having a hypothesized ANOVA-like covariance structure. This approach contrasts sharply with Morf's approach, which was to fit the entire matrix of correlations to an overall target. There is merit in both approaches, and they need not be mutually exclusive.

The analysis by parts has tended to show that some of the data (Succorance, Understanding) showed a better fit to stem overlap patterns than to response style patterns; other parts (Play) showed a better fit to style patterns; still other parts (F-scale, MMPI) were equivocal, and could be interpreted in terms of either stem or style hypotheses; and finally, one part (heterogeneous PRF scales) could reasonably be interpreted in terms of style hypotheses, and yielded modestly reliable but usable estimates of agreement and endorsement acquiescence.

One could argue that the overall analysis controls, in some sense, for minor imperfections in the individual pieces of the data set. In the case of the MMPI measures, for example, we have seen that the presence of desirability in the MMPI scales could be inferred from the loadings on Factor III, and this would not be detected if the MMPI measures were not analyzed simultaneously with the desirability measures. It would perhaps be useful to begin with an overall factor analysis, then attempt to model individual parts of the data using maximum likelihood methods, and use the knowledge built up in the analysis of subsets of the data to design an improved overall analysis. In the present case, the reanalysis by parts has raised questions about the appropriateness of an overall analysis--or at least about an analysis targeted to acquiescence factors--both by a showing that most parts of the data admit plausible rival interpretations, and by an examination of the scale construction process and the actual composition of the final scales.

The most serious problem with the Morf data is that, with the exception of the heterogeneous PRF scales (and then only for the facet of endorsement), the positively-weighted and negatively-weighted scales defining the agreement and endorsement facets are systematically based on different groups of stems, even though stems are used at least twice in most of the data (specifically, in the four PRF content areas, the F-scale, and the heterogeneous PRF scales). What makes the lack of counterbalancing bad is that the items for a particular scale are rarely unifactorial, even from a pure content point of view. Earlier, I cited the example of a group of Exhibition stems which were concerned with artistic or athletic performance, and seemed to be rather different from the other Exhibition stems. The lack of counterbalancing maximizes the chance that such content-based differences in the exemplars of a trait will be confounded with the estimation of response styles. Note that where stems were counterbalanced for design facets in the Morf data, as in the case of the experimental form facet of the PRF content areas and the overgeneralization facet of the F-scale, the resulting variance components have tended to be small or nonexistent.

Counterbalancing. Recall the discussion of the heterogeneous PRF scales, where it was shown that the stems were counterbalanced for endorsement but not for agreement. A balanced incomplete block (BIB) design may be used to counterbalance the design facets across stems. If we use blocks of size two (i.e., two item variants for each stem), the appropriate design for a 2 x 2 layout is Plan I in Table 3.14. This design requires six stems, and it pairs items of each type with every other type. Table 3.14 shows the design as it would apply to the agreement and overgeneralization facets of the F-scale. The basic design can be repeated any number of times, to yield multiples of 12 items (total) based on multiples of 6 stems. For example, four repetitions of the 6-block pattern will yield a total of 48 items, based on 24 stems.

Table 3.14

Balanced Incomplete Block Plan I[a],
for a 2 x 2 Design on Test Items

Replication	(Stem)	Item Type			
		AT	AF	RT	RF
I.	1	X	X		
	2			X	X
II.	3	X		X	
	4		X		X
III.	5	X			X
	6		X	X	

[a]After Cochran and Cox (1957) Plan 11.1, p. 471. Facets represent Overgeneralization (A vs. R) and Agreement (T vs. F).

In the Morf data, the F-scale resembles replication group II in Table 3.14, where the estimate of the agreement facet is confounded with blocks. In replication group I, on the other hand, the overgeneralization facet is confounded with blocks. By having replication groups of all three kinds, the estimates of the style variables are not systematically confounded with block or stem variation. A somewhat more complex plan must be used for the four PRF content areas, where a 2 x 2 x 2 facet design is needed. To match each of the eight item types with every other item type, in blocks of size two, a minimum of 28 blocks or stems is required. Plan II in Table 3.15 gives the appropriate BIB design.

Repetition of stems. In the above discussion, it was implicit that the same stem should not be repeated too many times; otherwise subjects may have specific recall of the way in which similar items were answered, or may feel that the test is repetitious and that they are being made to respond to more items than are really necessary. For ease of analysis, it would be convenient to have each subject respond to complete blocks based on four or eight variants of a stem, but, intuitively, that seems excessive. Having two variants for each stem entails the complexities of a BIB design, but it is administratively feasible.

To minimize specific recall, items based on the same stem must be given at widely spaced intervals within a questionnaire,

Table 3.15

Balanced Incomplete Block Plan II,[a] for a 2 x 2 x 2 Design on Test Items

Repli- cation	Block (Stem)	Item Type							
		SPT	SPF	SNT	SNF	APT	APF	ANT	ANF
I.	1	X	X						
	2			X	X				
	3					X	X		
	4							X	X
II.	5	X		X					
	6		X						X
	7				X	X			
	8						X	X	
III.	9	X			X				
	10		X					X	
	11			X			X		
	12					X			X
IV.	13	X				X			
	14		X	X					
	15				X			X	
	16						X		X
V.	17	X					X		
	18		X		X				
	19			X					X
	20					X		X	
VI.	21	X						X	
	22		X				X		
	23			X		X			
	24				X				X
VII.	25	X							X
	26		X			X			
	27			X				X	
	28				X		X		

[a]After Cochran and Cox (1957) Plan 11.9, p. 473. Facets represent Form (S vs. A), Endorsement (P vs. N), and Agreement (T vs. F).

and the maximum interval is approximately half the length of the test. Morf's general practice was to administer items based on the same stem at two different test sessions, but this has the disadvantage that trait instability, if present, gets analyzed as part of the treatment. A subject may feel more or less "exhibitionistic" at different test sessions, for example, and his responses to a measure based on exhibition content will vary to the extent that the measure is sensitive to temporarl variation in the trait. If items within each block of a BIB design were to be administered at separate test sessions, however, trait instability would be confounded with the within-block variation and wrongly attributed to the design facets.

The approach I am adopting is that of administering both variants of a stem at a single test session, which is the approach adopted by Bock et al. (1969) in their analysis of the MMPI _Pt_ and _Hy_ scales, and the resulting data may be analyzed using extensions of their basic design.

Chapter 4

Procedure and Hypotheses

Procedure

Subjects. Data were collected from a total of 279 subjects, at 13 test sessions held during regular class time and lasting approximately one hour. All the subjects were undergraduates at New York City area colleges. Subjects who were unable to complete the questionnaire in the allotted time and those with very low vocabulary scores or large numbers of responses to the infrequency scale were eliminated from the analysis, resulting in a total of 199 subjects with usable data. The selection criteria will be discussed in detail in the next chapter.

Measures. The measures used are listed in Table 4.1. Each subject was administered a Research Questionnaire covering seven content areas, and an 18-item Vocabulary Test. The Research Questionnaire contained 20 "experimental scales" (160 items) for the content areas of Play, Understanding, and Authoritarianism (F-scale). Additionally, it contained four "marker scales" (64 items) for the content areas of Cognitive Structure, Desirability, Impulsivity and Infrequency.

The experimental scales for Play and Understanding were each based on 56 items, forming eight 7-item subscales. As proposed in Chapter 3, the items for both content areas were

Table 4.1

List of Measures

Label	Content	Format	Wording	Keying	Items
		Experimental Scales			
PSPT	Play	self-descriptive	positive	true	7
PSPF	"	"	"	false	"
PSNT	"	"	negative	true	"
PSNF	"	"	"	false	"
PAPT	"	attitude	positive	true	"
PAPF	"	"	"	false	"
PANT	"	"	negative	true	"
PANF	"	"	"	false	"
USPT	Understanding	self-descriptive	positive	true	7
USPF	"	"	"	false	"
USNT	"	"	negative	true	"
USNF	"	"	"	false	"
UAPT	"	attitude	positive	true	"
UAPF	"	"	"	false	"
UANT	"	"	negative	true	"
UANF	"	"	"	false	"
FSRT	F-scale	attitude	relative	true	12
FSRF	"	"	"	false	"
FSAT	"	"	absolute	true	"
FSAF	"	"	"	false	"

Questionnaire Marker Scales

Label	Content	Source	Items
DES	Desirability	PRF Form AA	20
IMP	Impulsivity	PRF Form E	16
INF	Infrequency	PRF Form AA	12
COG	Cognitive Structure	PRF Form E	16

Other Measures

Label	Description
VOC	Vocabulary (18 items)
SPD	Test-taking speed (items/minute)

based on 28 stems, with each stem used twice in the Research Questionnaire. The BIB design given schematically in Table 3.15 was imposed on the Play and Understanding items, in order to measure the response style facets of agreement (T vs. F), endorsement (P vs. N), and form (S vs A).

The Research Questionnaire also contained a 48-item experimental F-scale. It resembles the F-scale of Clayton and Jackson (1961), except that the design facets are counterbalanced across stems. There were 24 stems, with each stem used twice. The BIB design given schematically in Table 3.14, replicated four times, was imposed on the items, in order to measure the response style facets of agreement (T vs. F) and overgeneralization (A vs. R).

In addition to these experimental scales, four intact Personality Research Form scales were embedded in the test instrument: a 16-item Impulsivity scale (from PRF Form E), a 16-item Cognitive Structure scale (Form E), a 20-item Desirability scale (Form AA), and a 12-item Infrequency scale (Form A). Impulsivity, Cognitive Structure and Desirability are relevant to the response style hypotheses, and the Infrequency scale serves as a validity check. The marker variables are discussed further in the section on hypotheses below.

Together, the experimental and marker scales total 224 items. The actual Research Questionnaire was 464 items long, with only the first 224 items being scored. The 240 additional items were heterogeneous PRF and MMPI items for content areas other than the ones of primary interest. The additional items served several purposes. They kept faster students busy while the slower ones were still working, discouraged leafing back and forth through the questionnaire, and encouraged a rapid test-taking place. Since subjects were not expected to finish the Research Questionnaire (test taking was halted after 40-45 minutes, or after it became apparent that almost all subjects had completed the the first 224 items), the last item completed served as an incidental measure of test-taking speed.

The last formal measure was the Vocabulary Test. The vocabulary test was adapted from test V-1 of the French, Ekstrom and Price (1963) _Kit of Reference Tests for Cognitive Factors_. It is an 18-item multiple-choice test, timed at 4 minutes, and considered suitable for students in grades 7-12. Initially, I had planned to use the "wide-range" test V-3 from the _Kit_ (considered suitable for students in grades 7-16). In a pilot study, this test proved too difficult for most of the subjects--the best score was 15 correct out of a possible 24 items--and the easier test was substituted.

Administration. The Research Questionnaire was administered first. The face sheet is included as Appendix A, and the experimental and marker scale items are listed (in scale order) in Appendix B. In the actual Research Questionnaire, the items appeared in a different order designed to maximize separation of items based on the same stem or measuring the same content. Items based on the same stem were separated by an interval of 112 items, while items measuring the same content were separated by an interval of at least four. In Appendix B, the first three columns represent the item number of the item in the final Research Questionnaire.

After the Research Questionnaires were administered, the Vocabulary Test was given. Thus, the entire testing procedure, together with an introduction to the study and instructions for the tests, was intended to fit within a one-hour test session.

Design of the experimental scales and results of a pilot study. Appendix B items written for all four PRF content areas used in Morf's dissertation, but after piloting, the items for the content areas of Exhibition and Succorance were dropped in order to shorten the questionnaire.

Writing of the experimental scales began with the items from Morf's study, which had been sorted by scale and matched for content stems (as tabulated in Table 3.1). From this initial item pool, I prepared pairs of items based on the same stems in order to fill out the BIB design. Some of the item pairs (e.g., the SPT-APT pair for the PRF content areas) were available in the initial item pool, but the items had to be extensively rewritten to provide all possible combinations of items. Some of the stems from the initial item pool proved to be unsuitable. A fairly large group of PRF items were rejected because they have an implicit forced-choice formal ("I would rather X than Y) which is difficult to reverse simultaneously with respect to the agreement and endorsement facets. An item of this type from Morf's study was:

> When I go to a doctor, I like to be businesslike and not ask unnecessary personal questions. (SSPF)

Morf had keyed this item for the positively-phrased, false, self-descriptive Succorance subscale, but it also contains a negation which makes it inappropriate for a positive scale. This item was rewritten as:

> When I go to a doctor, I never ask unnecessary questions. (SSNF)

In some cases, the item had to be dropped. To provide the required number of stems for the BIB design, stems from the published forms of the PRF were taken and adapted as needed.

Morf had made a point of taking items which had not been included in the final form(s) of the PRF because of their low content saturation. I would not make a similar claim for the experimental scales in the present study, because of the use of items from the published PRF and the extensive rewriting that was required.

After an initial set of items with all the required pairs was obtained, the items were listed on separate slips of paper. A colleague was asked to sort them according to the design facets, as a check on the success of the rewriting process. The instructions for the Exhibition items begain with trait descriptions from the PRF Manual (Jackson, 1974):

> The items on the attached slips were intended to form a scale to measure the Need for Exhibition, defined as follows:
>
> "Description of high scorer:

"Wants to be the center of attention; enjoys having an audience; engages in behavior which wins the notice of others; may enjoy being dramatic or witty.

"Defining trait adjectives:

"colorful, entertaining, unusual, spellbinding, exhibitionistic, conspicuous, noticeable, expressive, ostentatious, immodest, demonstrative, flashy, dramatic, pretentious, showy."

1. Please sort the items into two separate piles, containing:

   a. true-keyed items (high scorers would respond true)

   b. false-keyed items (high scorers would respond false)

[At each step, the slips were collected, the results were recorded, and the slips were re-shuffled.]

2. Please sort the items into two separate piles, containing:

   a. positively-phrased items (without negations)

   b. negatively-phrased items (with negations)

   The verb form is the main basis for this distinction. "I do X" or "People should do X" are positively phrased, while "I do not X" or "People shouldn't X" are negatively phrased. The presence of the term "not" or the contraction "n't" is the best indicator of negation, but negation can also be indicated by the terms "seldom," "never," "rarely," "hardly ever," and the like.

3. Please sort the items into two separate piles, containing:

   a. self-descriptive items (e.g., "I do X" or "I do not X")

   b. attitude or opinion items (e.g., "Most people do X" or "People should do X")

   The distinction here is that for self-descriptive items, a true or false response implies that the respondent does or does not have a particular characteristic, while for attitude items, a true or false response implies that the characteristic is generally true or desirable for people (or generally not true or not desirable), whether or not the respondent has the characteristic.

After each scale was sorted three times (twice in the case of the F-scale), discrepancies between the intended and the rated classifications of the items were discussed and resolved by mutual agreement. The first sorting (by true and false keying) was the most important one, since it indicated wehether the items could be correctly identified from a content standpoint. The second and third sortings were straightforward, since they depended on grammatical markers in the phrasing of the items.

After this procedure was completed, several items were rejected as simply poor measures of the content, and revised items were substituted. This done, the items were combined with the marker scales and printed as a 576-item test booklet --the first form of the Research Questionnaire--consisting of 336 items intended for scoring, followed by 240 filler items. This was administered to an intact class of undergraduate introductory psychology students. It soon became apparent that the majority of the students would not be able to complete the 336 critical items of this first version of the Research Questionnaire in the time allotted, and test-taking was halted after 35 minutes. The number of items completed was tabulated, with the following results: 17 of 25 students completed between 180 and 280 items (corresponding to test-taking speeds of 5.2 to 8 items/minute), with a modal value at about 240. Three students completed more than 300 items; judging from their vocabulary scores, they were not among the brighter students in the class. Finally, five students completed between 80 and 150 items; several of these appeared to have a language handicap (one student was observed referring to a foreign language-English dictionary, for almost every question).

As a result of the pilot study, the Research Questionnaire was shortened by reducing the number of critical items from 336 to 224 (by eliminating 112 items written for the Exhibition and Succorance scales). The Play scale was retained because it was the one scale which appeared to show strong response style effects in the reanalysis of the Morf data (though a response style interpretation was vitiated by the lack of counterbalancing). The Understanding scale was retained because it serves as an indirect measure of ability (persons with high ability are likely to have a high need for understanding). Accordingly, the Exhibition and Succorance items were dropped. My aim was to design a Research Questionnaire which could be administered, together with the Vocabulary Test, in a one-hour test session. This was an important practical consideration, making it easier to obtain the large number of subjects required. With the majority of subjects having test-taking speeds of 5.2 to 8 items/minute, I estimated that 28-43 minutes would be enough time to complete the 224 critical items of the shortened Research Questionnaire.

Hypotheses

Style components. Nonzero agreement, endorsement and form components are expected for the Play and Understanding scales. Similarly, nonzero agreement and overgeneralization components are expected for the F-scale. In a sense, the overall null hypothesis for the dissertation is that none of these components will be found in suitably counterbalanced data. The analysis will use maximum likelihood methods to estimate the size of the components and test them for significance. It also provides estimates of the correlations between the style components and content for each of the content areas.

Self-descriptive and attitude items. It is expected that self-descriptive items will elicit more endorsement than agreement responding, and that attitude items will elicit more agreement than endorsement responding. As discussed in Chapter 3, these were hypotheses advanced by Morf, and limited support for both hypotheses was found in the reanalysis of his data.

Construct validity of agreement and endorsement. As discussed in Chapter 1, agreement is thought to be related to verbal interpretive skill, and endorsement to impulsivity, tempo and speed. There are four principal measures included for purposes of construct validation of agreement and endorsement--vocabulary, impulsivity, cognitive structure, and the measure of test-taking speed derived from the surplus items on the Research Questionnaire. Vocabulary serves as a reasonable proxy for verbal interpretive skill, and impulsivity is measured directly by the Impulsivity subscale of the PRF. The speed and cognitive structure measures require some discussion: Speed is partly a function of verbal ability--other things being equal, brighter subjects can be expected to complete more items in the allotted time than duller subjects. The Cognitive Structure scale is included somewhat speculatively. Persons scoring high on the scale are described as follows:

> Does not like ambiguity or uncertainty in information; wants all questions answered completely; desires to make decisions based upon definite knowledge, rather than upon guesses and probabilities (Jackson, 1964, p. 6).

Cognitive Structure has high negative correlations with Impulsivity (-.53 to -.68, Jackson, 1974), and appears to be primarily a reflected measure of impulsivity. The correlations with Impulsivity do not exhaust the reliable variance of the scale, however, and an examination of the items indicates that the Cognitive Structure scale may be tapping a component of intellectual orderliness and desire for consistency which is not otherwise measured.

The conceptual model for the hypotheses concerning the four validity measures is that most of their reliable variance lies essentially in a plane, as shown in Figure 4.1. The dashed vectors in Figure 4.1 indicate the hypothesized

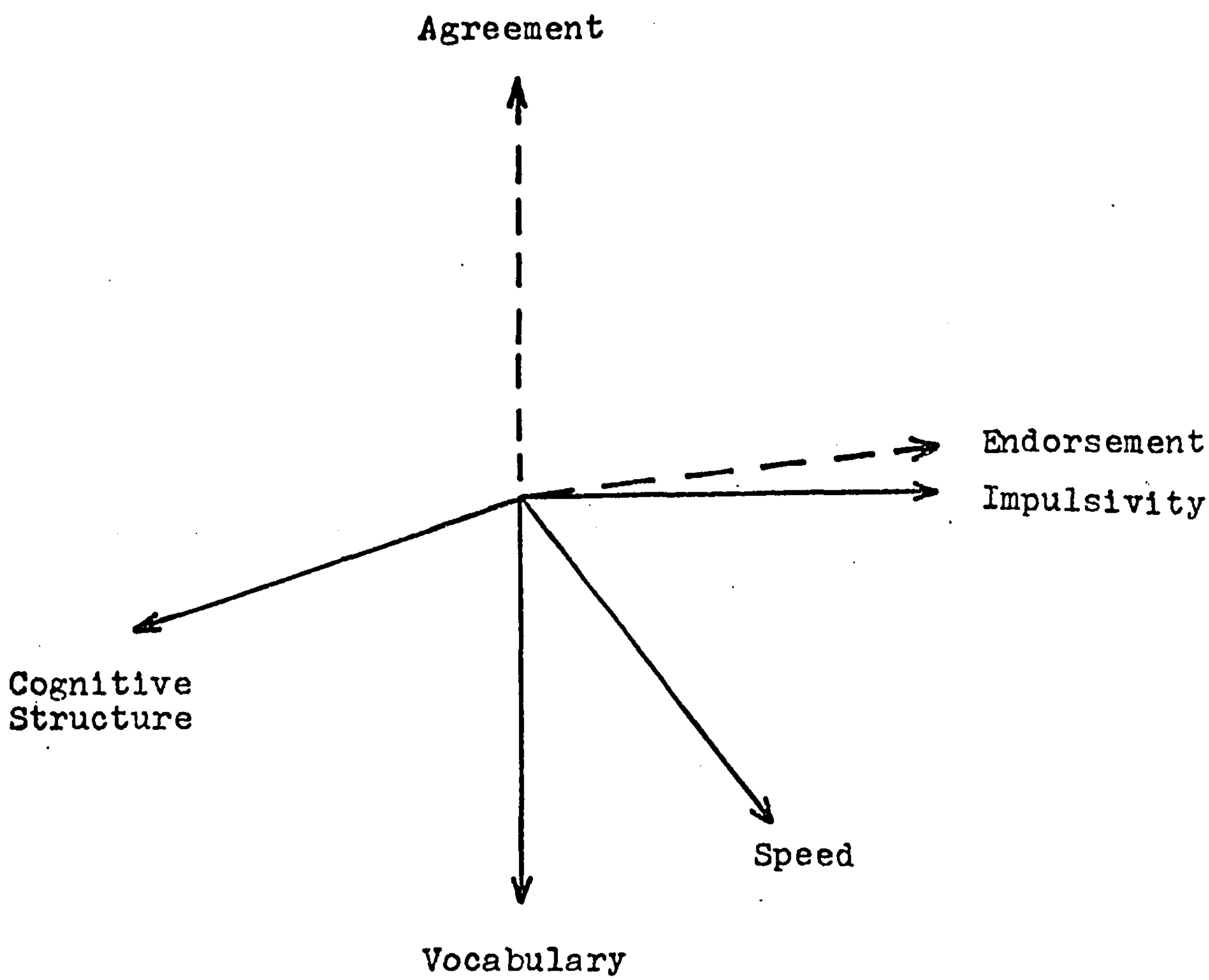


Figure 4.1. Hypothesized relationships of agreement and endorsement with the construct validity measures--Vocabulary, Impulsivity, Cognitive Structure and Speed.

projections of agreement and endorsement measures (linear combinations of the subscales for the Play, Understanding and F-scale content areas) onto the plane. Vocabulary and impulsivity are expected to be orthogonal, with speed and cognitive structure correlated with both. Multivariate multiple regression may be used to determine whether vocabulary, speed and reflected cognitive structure have the expected negative relationship with agreement (after impulsivity is partialled out), and whether impulsivity, speed and reflected cognitive structure have the expected positive relationship with endorsement (after vocabulary is partialled out).

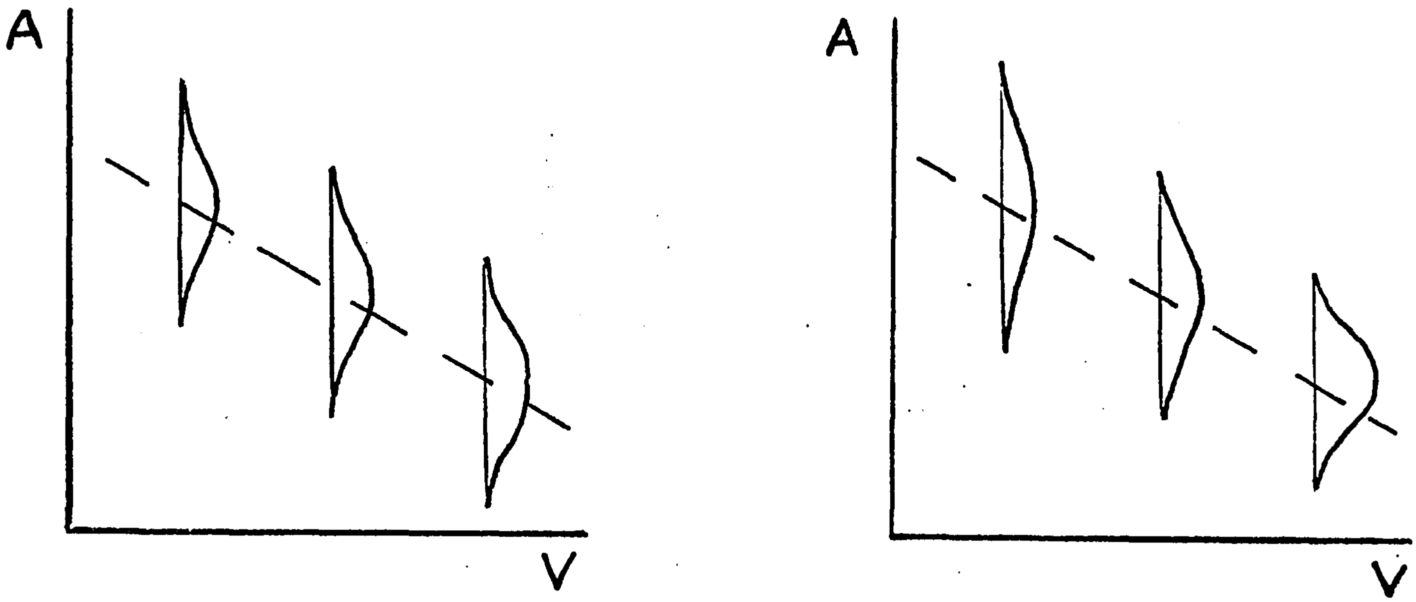
<u>The two-factor theory of acquiescence</u>. In the last section, agreement and endorsement were presumed to be orthogonal, but that is a matter for empirical investigation. By treating the Play and Understanding content areas separately, three measures of agreement and two measures of endorsement may be obtained. A measure of acquiescence (presumably agreement) is also available from the F-scale content area. These five measures should show discriminant and convergent validity, with the agreement measures correlating more highly among themselves than with the endorsement measures, and vice versa. The sub-matrix of S scores is expected to have a two-factor simple structure, which may be fitted and tested by maximum likelihood methods. Past attempts to find correlations

between acquiescence measures based on different instruments or content areas have been disappointing, but these attempts have not used data in which the two kinds of acquiescence could be distinguished.

Linear and nonlinear relationships with acquiescence. It will also be of interest to explore and display the relationships between other variables in the study and the network of measures shown in Figure 4.1, but no specific hypotheses are advanced. The data base will also be large enough to permit an exploration of interactions between the response style and validity measures. A plausible hypothesis would be that subjects with low vocabulary scores have larger response style components than subjects with high vocabulary scores, for example, because the former are less able to understand the items and therefore less able to respond consistently in terms of content. Interactions with response style measures would imply nonlinear relationships between variables in the study.

As discussed in Chapter 1, scores on a test and acquiescence measured from the same test tend to be nonlinearly related (Cronbach, 1950; Messick, 1967). On a true-false vocabulary test, for example, a subject who knows all the answers will have no opportunity to display his acquiescence or lack of acquiescence; less obviously, a subject who gets every answer wrong, as a result of very poor guessing, will also be treated as neutral in acquiescence. Similarly, subjects who very high or low on a personality measure cannot have very high acquiescence scores derived from the same measure.

A somewhat different issue is posed by the possibility of nonlinear relationships between acquiescence and measures _other_ than the ones used to estimate acquiescence. For example, Figure 4.1 depicts a negative relationship between vocabulary and agreement. We may also want to distinguish between the situations depicted in Figures 4.2a and 4.2b: Figure 4.2a shows a linear negative relationship between verbal ability and agreement; Figure 4.2b also shows a negative relationship, but with marked nonlinearity (heteroscedasticity) such that subjects low in verbal ability are more variable in agreement tendencies than subjects high in verbal ability. Since agreement can be measured from personality tests, and verbal ability from an independent vocabulary test, there will not be an artifactual nonlinear relationship between the two variables. Examination of residual plots and specific tests for heteroscedasticity can help determine whether the vocabulary-agreement relationship resembles Figure 4.2a or 4.2b (Draper & Smith, 1966, pp. 86-97; Johnston, 1972, pp. 214-221). If the situation depicted in Figure 4.2b is found, it would imply that agreement is more of a measurement problem in some populations than in others. Similarly, it would be of interest to determine whether subjects low on the cognitive structure measure or high on the impulsivity measure tend to be more variable with respect to either agreement or endorsement.



(a)

Linear relationship

(b)

Nonlinear relationship
(heteroscedasticity)

Figure 4.2. Alternate forms of regression of agreement (A) on verbal ability (V) consistent with a negative relationship overall.

Chapter 5

Results

The bad news is that when data collection was completed, I discovered that I had two distinct subsamples in the data, differing in means and regression slopes on a number of key variables. At the outset of data collection, there was no intention of setting up a within-groups design, but it became apparent that the two groups of subjects were very different and, reluctantly, I concluded that some of the major analyses would have to be done separately within groups. The good news is that the two groups differ in ways which have a bearing on the research issues of the dissertation, and particularly on the hypotheses about nonlinearity proposed at the end of Chapter 4.

The plan of organization for this chapter is that the first section will be concerned with sample characteristics and the differences between the two subsamples. The next section reports analyses of covariance structure, separately by group. In these first two sections, we see that, of the four response styles proposed, only agreement appears to be strongly represented in the data and to have meaningful relationships with other variables. The final sections of the chapter attempt a more integrated analysis of the data base as a whole, using multivariate multiple regression and canonical analysis to identify the relationships between the experimental and the marker variables.

Sample Characteristics

Subjects

Data were collected from a total of 279 subjects, at 13 test sessions during regular class time. 173 subjects were from the Manhattan branch of a four-year "technical" college, and 106 subjects were from several "liberal arts" colleges in the New York City area. The liberal arts colleges are more academically oriented than the technical college, and, for short, I will refer to the two samples as the Technical and Academic groups.

The student body at the technical college is most unusual. A recent student newsletter reported that 84% of the students were male, and that 52% were from foreign countries. A majority of the American students seem to be black or hispanic, but detailed figures were not reported. During and after testing, it became apparent that the students at the technical college are markedly deficient in English language skills. Even after eliminating subjects failing to complete enough items to yield a scorable Research Questionnaire and subjects with the lowest Vocabulary scores, the Technical group averages about 3.6 fewer correct Vocabulary items than the Academic group.

The subjects at the academic colleges are from three sites. All of the academic colleges are coeducational, but two of the colleges were formerly girls' schools, and all three colleges are predominantly female (79% in my data). One consequence of the differences between samples is an essentially spurious correlation between Sex and Vocabulary. In a College x Sex x

Vocabulary crosstabulation of the subjects (with subjects dichotomized on Vocabulary), the piling up of subjects in the Technical-Male-Low Vocabulary and the Academic-Female-High Vocabulary cells results in a significant overall association between Sex and Vocabulary ($\phi = .286$, $\chi^2(1) = 16.3$, $p < .001$, as shown in Table 5.1). The within-group association between Sex and Vocabulary is not significant, however, whether one looks at the within-college ϕ-coefficients, within-group regressions of Vocabulary on Sex, or a log-linear decomposition of the 3-way table (see Table 5.1 for details).

The Technical group consists of a large number of students who are present in smaller numbers in most undergraduate populations. Most colleges try to accomodate limited numbers of students from foreign countries or educationally disadvantaged backgrounds, through open-admission policies, formal or informal quota systems, and the like, but their numbers tend to be small as a proportion of the total student body. Such students would normally be at high risk for academic failure and have fairly high drop-out rates. At the technical college, however, foreign and educationally disadvantaged subjects seem to predominate. It seemed likely to me that students with low English language skills would show more acquiescence and other stylistic responding and since I wanted to study the relationship between Vocabulary and stylistic responding, I was not concerned about the fact that the Technical group would be

Table 5.1

Vocabulary (Low and High) x Group x Sex

Crosstabulation of Subjects

Vocabulary	Technical Group		Academic Group	
	Male	Female	Male	Female
Low	53	11	9	24
High	22	11	14	55

$\phi_{SEX,VOC} = .183$, n.s. $\phi_{SEX,VOC} = .078$, n.s.

Note: Subjects dichotomized at median on Vocabulary. With data collapsed across group, $\phi_{SEX,VOC} = .286$, $\chi^2(1) = 16.3$, $p < .001$.

lower in Vocabulary than my other subjects. The group differences in Vocabulary and the other variables were so strong, however, that it became necessary to include group as a variable in the analysis.

Exclusion criteria. One indication of the differences between groups is that the Technical group yielded a large number of unscorable protocols, owing to failure to complete the Research Questionnaire in the time allotted. Subjects were retained for analysis if: (a) they completed at least 148 items from the 160-item experimental series; (b) they had no more than three scored responses on the Infrequency scale; and (c) they had vocabulary scores of at least 6 (with the test scored by the formula $3R - W$, to correct for guessing; thus, a score of 6 represents a minimum of 2 correct responses). Application of these three criteria led to the rejection of 25 Technical and no Academic students by criterion (a), 3 Technical and 3 Academic students by criterion (b), and 48 Technical and 1 Academic students by criterion (c). The final sample contains 97 Technical and 102 Academic students.

Marker variables

It will be useful to begin by re-examining the marker variables in the light of the data. Table 5.2 displays their means and standard deviations, by group. All variables in Table 5.2 are in forced-normal form, with means of 0 and standard deviations of 1 for both groups combined. We find that the Academic and

Table 5.2

Means and Standard Deviations for Marker Variables, by Group

	Means			Std. Devs.			
	Tech.	Acad.	*t*	Tech.	Acad.	F_{max}	α
AGE	.16	-.15	2.24*	1.03	.95	1.19	-
SPD	-.48	.47	-7.85***	.89	.87	1.05	-
VOC	-.42	.40	-6.35***	1.00	.82	1.47	.74
COG	.31	-.30	4.51***	.88	1.02	1.35	.63
DES	-.26	.25	-3.70***	.98	.96	1.04	.47
IMP	-.09	.09	-1.21	.96	1.03	1.15	.71
INF	.28	-.26	3.93***	1.01	.92	1.22	.12

Note: All variables are in forced-normal form and have means of 0 and standard deviations of 1 for the entire sample.

N = 97 for Technical group and N = 102 for Academic group; df = 197 for t and approximately (100,100) for F_{max}.

*$p < .05$

**$p < .01$

***$p < .001$

Technical groups have significantly different means on all variables in Table 5.2 except Impulsivity. Subjects in the Technical group are slightly older, and have much slower rates of responding to the Research Questionnaire; they have lower Vocabulary and Desirability scores, and higher Cognitive Structure and Infrequency scores. Table 5.3 presents the results of a series of stepwise regression analyses, in which each of the marker variables was regressed on a basic predictor set consisting of Group, Speed, Vocabulary, the Group x Speed and Group x Vocabulary crossproducts, Age and Sex. For convenience, I'll refer to these by the 3-character mnemonics--GRP, SPD, VOC, etc.--appearing in the table. GRP, SPD and VOC were forced to enter whether significant or not, while the remaining variables only entered if they had significant partial regression coefficients ("F-to-enter" > 3.89, $p < .05$) in subsequent steps. (In models for predicting criterion variables which were also in the basic predictor set, of course, the variable predicted was omitted from the predictor set.)

All of the multiple correlations in Table 5.3 are significant except that, if we adopt $\alpha = .05/8$ for the set of eight criteria, Impulsivity fails to reach significance at the .006 level ($R = .22$, $p = .014$) Now we examine the marker variables in more detail.

Table 5.3

Regression Coefficients for Predicting Marker Variables
from Group, Speed, Vocabulary, Crossproducts with Group, Sex and Age

Cri-terion	Std. Dev.	Multi-ple R	Inter-cept	Forced Predictors			Stepwise Predictors			
				GRP	SPD	VOC	GXS	GXV	SEX	AG
SEX	50	62***	93	39***	17***	00	-	-	NA	-
AGE	100	31***	05	-09	-79**	02	36*	-	-	NA
SPD	100	64***	-152	37**	NA	25***	NA	-	64***	-15*
VOC	100	51***	-70	52***	77***	NA	-30*	NA	-	-
DES	100	36***	-51	39*	-06	71**	-	-33*	-	-
COG	100	37***	67	-44**	-13	00	-	-	-	16*
IMP	100	22*	14	-09	18*	10	-	-	-	-
INF	100	40***	51	-38*	-80**	-17*	52***	-	-	-
Std. Dev.	-	-	-	50	100	100	156	150	50	100

Note: All figures are reported to two decimal places. SEX is scaled 1 for Male and 2 for Female, and GRP is scaled 1 for the Technical group and 2 for the Academic group. All other variables (except crossproducts) are in forced-normal form.

Regression coefficients and multiple *R*s are from the final model in a stepwise regression analysis in which GRP, SPD and VOC were forced to enter, whether significant or not, while SEX, AGE and the crossproducts were only allowed to enter if they had significant partial regression coefficients at the applicable step ("*F*-to-enter" > 3.89, $p < .05$).

$N = 199$. *$p < .05$; **$p < .01$; ***$p < .001$.

Sex and Age. SEX is correlated with GRP, as indicated earlier, and also correlated with SPD after adjusting for the group differences (females are faster than males, a common finding for tasks involving clerical or perceptual-motor speed--e.g., Broverman & Klaiber, 1969). The sex difference in Vocabulary observed for the total sample (Table 5.1) is eliminated in the regression analysis, and the final model is: $\widehat{SEX} = .93 + .39\ GRP + .17\ SPD + .00\ VOC$. I could have rerun the analysis dropping VOC, and the regression coefficients for GRP and SPD might have changed slightly, but I chose not to, since the conclusions reached would not be affected.

For predicting AGE, there is a Group x Speed interaction, which may be interpreted by writing the separate equations by group:

Tech: $\widehat{AGE} = -.04 - .43\ SPD + .02\ VOC$

Acad: $\widehat{AGE} = -.13 - .07\ SPD + .02\ VOC$

This suggests that the older subjects in the Technical group are slower than the younger subjects, but there is essentially no difference between older and younger subjects in the Academic group.

Since there are quite a few interactions with GRP in the data, it may be useful to sketch the manner in which I will interpret them. The full model for AGE in Table 5.3 may be written:

$$\hat{AGE} = .05 - .09\ GRP - .79\ SPD + .36\ GXS + .02\ VOC$$
$$= (.05 - .09x1) - (.79 - .36x1)\ SPD + .02\ VOC$$
(for the Technical group)
$$= (.05 - .09x2) - (.79 - .36x2)\ SPD + .02\ VOC$$
(for the Academic group) ,

which yields the separate equations by group mentioned in the last paragraph. The saddle point for the regressions is at the point $SPD = -\underline{b}_{GRP} / \underline{b}_{GXS} = -(-.09)/.36 = .25$, which implies that (ignoring Vocabulary) in the region where the Speed scores equal .25, there is no difference between the groups on Age. Using the Johnson-Neyman technique (e.g., Kerlinger & Pedhazur, 1973, pp. 256-257), we can set confidence bounds on this region. The .95 confidence bounds are at $SPD = -.93$ and $SPD = 2.78$, which means that the groups are not significantly different in AGE in the region $-.93 \leq SPD \leq 2.78$. Below this region, the Technical students are significantly older than the Academic students. The Johnson-Neyman region is, in principle, two-sided, and the model asserts that subjects above 2.78σ on SPD would be older in the Academic group than in the Technical group. 2.78σ lies outside the range of the SPD variable, however, so can only make a firm statement about the subjects below $-.93\sigma$.

Speed and Vocabulary. In crude form, the groups show differences in both means and variances on these variables. (The Technical group has a lower mean and is more variable on Vocabulary, and has a lower mean and is less variable on Speed.)

Both measures were given a forced-normal transformation, and will be used in that form throughout. In forced-normal form, the mean differences remain, but the variability differences are eliminated for Speed and markedly reduced for Vocabulary ($\underline{F}_{max}(96,101) = 1.47$, $\underline{p} = .059$ for Vocabulary). For predicting SPD, we find a positive association with VOC and SEX, and a negative association with AGE. For predicting VOC, there is a Group x Speed interaction which yields the separate equations

$$\text{Tech:} \quad \widehat{VOC} = -.18 + .47 \text{ SPD}$$

$$\text{Acad:} \quad \widehat{VOC} = .34 + .17 \text{ SPD}$$

with a saddle point at SPD = 1.73 . Thus, the Vocabulary-on-Speed regression is stronger in the Technical group than in the Academic group. The Johnson-Neyman region is $.63 \leq \text{SPD} \leq \underline{\text{max}}$. , indicating that among the slower and average students, the Technical group has lower Vocabulary scores than the Academic students, but among the faster students there is no difference in Vocabulary scores.

The Vocabulary measure suffers from restriction of range at the upper end (6 subjects in the Academic group and 1 subject in the Technical group achieved the maximum score), which makes me regret the decision to use the narrow-range test after pretesting. The effect of the forced-normal transformation on Vocabulary is to stretch out the upper tail and bring in the lower tail on the measure. The interaction with Speed occurs just as strongly when the crude Vocabulary measure is substituted in the regression equation just discussed, however, and is not an artifact of the forced-normal transformation. It may be partly a function of the ceiling on the Vocabulary measure, but evidence presented later on will suggest that Speed serves as an auxiliary measure of ability in the Technical group, and is more like a personality variable measuring personal tempo or simply clerical speed in the Academic group. At a number of points in the analysis, we will find a Group x Speed interaction entering the regression equations.

Desirability. Here is an instance of a variable with a Group x Vocabulary interaction. The equations by group are

$$\text{Tech:} \quad \widehat{DES} = -.12 - .06\ SPD + .38\ VOC$$

$$\text{Acad:} \quad \widehat{DES} = .27 - .06\ SPD + .07\ VOC$$

with a saddle point at $VOC = 1.18$ and a Johnson-Neyman region of $.28 \leq VOC \leq \underline{max}$. Here we find a fairly strong relationship between Desirability and Vocabulary in the Technical group (subjects with higher verbal ability give more desirable responses), but essentially no relation between Desirability and Vocabulary in the Academic group. The Academic group has higher Desirability scores overall, but for subjects in the upper third of the VOC distribution (above $VOC = .28$) there is no difference between the groups.

The Desirability measure is odd in several respects. The true subscale has an alpha of only .15, while the false subscale has an alpha of .57 (the alpha for the combined scale is .47), and the true and false parts correlate only .13. An examination of the items suggests that the true and false subscales are specialized in meaning. The items which correlate most highly with the _false_-keyed subscale are:

> I often question whether life is worthwhile. (45% _false_)
>
> I often have the feeling that I am doing something evil. (71% _false_)
>
> I almost always feel sleepy and lazy. (83% _false_)
>
> I find it very difficult to concentrate. (78% _false_)

The items which correlate most highly with the _true_ subscale are:

> I often take some responsibility for looking out for newcomers in a group. (71% _true_)
>
> My memory is as good as other people's. (70% _true_)
>
> I am always prepared to do what is expected of me. (51% _true_)
>
> I am able to make correct decisions on difficult questions. (70% _true_)

The _false_ items resemble psychopathology items such as the MMPI _A_ scale items, while the _true_ items seem more like impression-management or plus-getting items. Questioning whether life is worthwhile and feeling evil, sleepy, lazy and unable to concentrate do not have the same implications for personality as failing to have a good memory, do "what is expected," make correct decisions, or look out for newcomers. I considered a separation of the _true_ and _false_ Desirability subscales for a while, but preliminary analyses did not look promising. Finally I decided to leave the Desirability scale intact, on the theory that the presence of the _true_-keyed items would serve to help suppress any agreement variance which was present in the _false_ subscale.

Impulsivity and Cognitive Structure. Impulsivity is not well-predicted in Table 5.3. It shows the expected positive relationship with SPD, and while the relationship is not a strong one, later analyses will show that SPD is consistently associated with an Impulsivity-Cognitive Structure factor. Impulsivity is a well-behaved scale. The true and false parts have reasonable alphas (.60 and .58), and correlate .44. The combined scale has an alpha of .71.

Cognitive Structure was intended as a second (reflected) measure of Impulsivity, and the two scales have a correlation of -.56, which is very reasonable. The true and false parts of the Cognitive Structure scale have alphas of .63 and .35, and correlate .27. The combined scale has an alpha of .61. The regression coefficients in Table 5.3 indicate a mean difference between the groups and a positive relationship with Age (older subjects and subjects in the Technical group have a higher need for Cognitive Structure).

Infrequency. Subjects with scores higher than 3 on the Infrequency scale were eliminated from the analysis, so the resulting variable has a range of four. In Table 5.3, we find a mean difference between the groups, a slight negative relationship with VOC, and a Group x Speed interaction. Separate equations yield

$$\text{Tech:} \quad \widehat{INF} = .13 - .28\ SPD - .17\ VOC$$

$$\text{Acad:} \quad \widehat{INF} = -.25 + .24\ SPD - .17\ VOC$$

with a saddle point at $SPD = .73$. In the Academic group, the faster subjects are making more infrequent responses, but in the Technical group, it is the slower students who are making more infrequent responses. This suggests that the faster subjects in the Academic group are making more infrequent responses through carelessness, but in the Technical group, where there is a high SPD-VOC correlation ($\underline{r}$ = .41), it is the slower and less able students who are making more infrequent responses.

Infrequency has a low internal consistency (alpha = .12), which is about the best we can expect for a scale having such low endorsement frequencies (3-11% endorsement in the keyed direction for the items of the scale). The effect of the forced-normal transformation on this measure is to substitute scale values of -.77, .27, 1.05 and 1.82 for the numbers 0, 1, 2 and 3, which is tending to treat the 1's, 2's and 3's as more alike one another, and more dissimilar from the 0's. This is a reasonable metric for the variable, which is being treated as an index of something having a more-or-less normal distribution, and prevents the 3's from having an inordinate effect on regressions involving the Infrequency measure.

Experimental Variables

Table 5.4 presents the summary statistics for the 20 experimental subscales, separately by group. The subscales were originally scored by counting 2 for items answered in the +content direction and 0 for items answered in the -content direction. When one member of a pair of matched experimental items was not answered, it was scored like the other member of the pair (2 for +content and 0 for -content). In the few instances where both members of a pair were omitted by a respondent, the items were both scored 1. (This method for handling missing data prevents omitted items from making a spurious contribution to the estimates of response style scores.) Then the subscales were "semi-standardized" by scaling the Play and Understanding scores by $\sigma_{PSPF} = 3.515$ and the F-scale scores by $\sigma_{FSAT} = 5.503$. These scaling factors preserve all of the variance ratios for subscales in each content area, and give the measures a maximum variance of about 1.00.

Group differences. The most interesting feature of Table 5.4 can be found by running down the column of *t*-tests for mean differences between the groups. The signs of the *t* values indicate that, without exception, the Technical group has higher scores on every true subscale and lower scores on every false subscale. Since the items have been scored in the content direction, this means that the subjects in the Technical group are responding true more often than the subjects in the Academic group. It is the first indication of the strength of the group

Table 5.4

Summary Statistics for Experimental Scales, by Group

	Means			Std. Devs.			α
	Tech.	Acad.	t	Tech.	Acad.	F_{max}	
PSPT	2.02	1.91	1.00	.82	.82	1.00	.36
PSPF	1.94	2.60	-4.90***	.97	.93	1.09	.57
PSNT	1.94	1.53	3.26***	.87	.88	1.00	.48
PSNF	2.01	2.49	-3.44***	1.06	.87	1.48	.55
PAPT	2.48	2.21	1.44	.76	.91	1.39	.38
PAPF	2.61	3.29	-6.09***	.92	.62	2.18***	.54
PANT	2.49	2.34	1.30	.84	.84	1.01	.30
PANF	1.99	2.68	-5.97***	.87	.74	1.36	.48
USPT	1.94	1.93	0.11	.99	.90	1.22	.53
USPF	1.97	2.39	-4.06***	.77	.72	1.12	.31
USNT	2.34	2.04	2.94**	.70	.72	1.04	.08
USNF	2.58	3.19	-5.50***	.91	.63	2.05***	.44
UAPT	3.16	2.96	2.20*	.71	.62	1.33	.22
UAPF	2.87	3.18	-3.98***	.57	.53	1.19	.17
UANT	2.73	2.52	1.73	.83	.83	1.02	.44
UANF	2.21	2.77	-5.27***	.85	.62	1.88**	.35
FSRT	2.87	2.15	6.49***	.79	.77	1.06	.59
FSRF	2.10	2.16	-0.56	.78	.66	1.40	.41
FSAT	2.48	1.49	8.06***	.97	.76	1.59*	.72
FSAF	2.39	3.06	-6.53***	.83	.61	1.89**	.53

Note: Subscales were originally scored by counting 2 for items answered in the content direction, 0 for items answered in the negative content direction, and 1 for omitted responses. Then the scales were "semistandardized" by scaling the Play and Understanding items by $\sigma_{PSPF} = 3.515$ and the F-scale items by $\sigma_{FSAT} = 5.503$.

See note to Table 5.2 for sample sizes, degrees of freedom and significance levels.

difference in agreement which runs through the data. There is also a noticeable tendency for the subjects in the Technical group to be more variable than the subjects in the Academic group; 15 of 20 variances are larger in the Technical group (p = .042 by a sign test).

The alpha coefficients of the subscales are low but reasonable for 7-item and 12-item scales, and average .42. Morf and Jackson (1972) reported an average alpha of .32 for the same group of subscales. My subjects are probably much more heterogeneous than Morf and Jackson's subjects, however, which is probably the main reason why my alphas tend to be larger than Morf and Jackson's.

Correlations among the Experimental Measures. We may further evaluate the group differences on the experimental variables by using the same basic predictor set of marker variables used in Table 5.3. Instead of applying the model to each of the 20 subscales, however, it will be helpful to begin by computing the *a priori* estimates of the content and style variables which are embedded in the 20 subscales. This is accomplished by the matrix multiplication $X^* = XT$ represented symbolically in Table 5.5. Let X be the $N \times 20$ data matrix containing the experimental subscales, and T be the 20×11 block-diagonal matrix of transformation given in Table 5.5; X^*, then, contains the estimates of Play content, agreement, endorsement and form (PC, PA, PE, and PF), Understanding content, agreement, endorsement and form (UC, UA, UE and UF), and F-scale content, agreement and overgeneralization (FC, FA and FO). These two-character

Table 5.5

Computation of Experimental Content and Style Measures

Experimental Measures		Experimental Subscales	Matrix of Transformation

$$
\begin{pmatrix} PC \\ PA \\ PE \\ PF \\ UC \\ UA \\ UE \\ UF \\ FC \\ FA \\ FO \end{pmatrix}^{tr}
=
\begin{pmatrix} PSPT \\ PSPF \\ PSNT \\ PSNF \\ PAPT \\ PAPF \\ PANT \\ PANF \\ USPT \\ USPF \\ USNT \\ USNF \\ UAPT \\ UAPF \\ UANT \\ UANF \\ FSAT \\ FSAF \\ FSRT \\ FSRF \end{pmatrix}^{tr}
\begin{pmatrix}
1 & 1 & 1 & 1 & & & & & & & \\
1 & -1 & -1 & 1 & & & & & & & \\
1 & 1 & -1 & 1 & & & & & & & \\
1 & -1 & 1 & 1 & & & & & & & \\
1 & 1 & 1 & -1 & & & & & & & \\
1 & -1 & -1 & -1 & & & & & & & \\
1 & 1 & -1 & -1 & & & & & & & \\
1 & -1 & 1 & -1 & & & & & & & \\
 & & & & 1 & 1 & 1 & 1 & & & \\
 & & & & 1 & -1 & -1 & 1 & & & \\
 & & & & 1 & 1 & -1 & 1 & & & \\
 & & & & 1 & -1 & 1 & 1 & & & \\
 & & & & 1 & 1 & 1 & -1 & & & \\
 & & & & 1 & -1 & -1 & -1 & & & \\
 & & & & 1 & 1 & -1 & -1 & & & \\
 & & & & 1 & -1 & 1 & -1 & & & \\
 & & & & & & & & 1 & 1 & 1 \\
 & & & & & & & & 1 & -1 & -1 \\
 & & & & & & & & 1 & 1 & -1 \\
 & & & & & & & & 1 & -1 & 1
\end{pmatrix}
$$

$$
\underset{\underline{N},11}{X^{*}} = \underset{\underline{N},20}{X} \cdot \underset{20,11}{T}
$$

mnemonics will be used throughout the chapter to refer to the *a priori* estimates of the content and style variables which will be called, collectively, the 11 "experimental measures," to distinguish them from the 20 "experimental subscales."

Before examining the multiple regressions for predicting the experimental measures, it will be useful to look at Table 5.6, which displays their intercorrelations. Little subtables from Table 5.6 are of interest. Notice that the content measures are basically uncorrelated:

UC	.125	
FC	.003	-.145*
	PC	UC

The only significant correlation here indicates a slight tendency for subjects with a higher need for Understanding to have lower F-scale scores. Now look at the correlations among the agreement measures:

UA	.704***	
FA	.746***	.663***
	PA	UA

These appear to be the highest correlations ever reported for independent measures of agreement. The content-agreement correlations also show a consistent pattern:

PC	-.210***	-.296***	-.298***
UC	-.183*	-.033	-.223***
FC	.260***	.173*	.308***
	PA	UA	FA

Table 5.6

Correlations among the Experimental Measures

	PC	PA	PE	PF	UC	UA	UE	UF	FC	FA
PC	-									
PA	-210**	-								
PE	120	-145*	-							
PF	136	042	029	-						
UC	125	-183**	175*	-014	-					
UA	-296***	704***	-057	008	-033	-				
UE	117	-156*	279***	085	268***	-108	-			
UF	-054	070	118	088	269***	047	152	-		
FC	003	260***	047	112	-145*	173*	034	-120	-	
FA	-298***	746***	036	163*	-223**	663***	-071	084	308***	-
FO	-036	352***	020	153*	-202**	200**	-116	061	263***	39[illegible]

Note: N = 199

*p < .05

**p < .01

***p < .001

Subjects with higher agreement scores tend to have lower scores for Play and, to some extent, for Understanding content, and higher scores on the F-scale. Thus, the agreement measures are showing convergent and discriminant validity within the matrix presented in Table 5.6. Given the low correlations among the content measures, the high correlations among the agreement measures cannot be explained in terms of confounding with Play, Understanding, or F-scale content.

The other style measures present a mixed picture. The only form-form correlation ($\underline{r}_{PF,UF} = .088$, n.s.) is not significant, so the form facet does not have convergent validity. The only endorsement-endorsement correlation ($\underline{r}_{PE,UE} = .279$***) is significant, but it is not nearly as strong as the agreement-agreement correlations, and is rivaled in size by other correlations involving the endorsement measures (in particular, $\underline{r}_{PE,UC} = .175$* and $\underline{r}_{UE,UC} = .268$***). As we shall see later, analyses of covariance structure will provide little evidence for a separate endorsement component, and the endorsement measures show little relationship with the marker variables.

The F-scale measures have uniform moderate correlations with one another:

FA	.308***	
FO	.263***	.395***
	FC	FA

The overgeneralization measure behaves more like F-scale content than like F-scale agreement (examine the bottom three rows in Table 5.6). My best guess about the overgeneralization facet is that it is a content facet with scrambled scale saliences.

<u>Multiple regressions with experimental measures</u>. Table 5.7 presents the results of a regression analysis of the experimental measures. As before, GRP, SPD and VOC were forced to enter, but AGE, SEX and the crossproducts with GRP only entered if they made a significant contribution. Under these conditions, SEX did not enter any of the equations, and is omitted from the table. For these analyses, the Play and Understanding measures have been scaled by a factor of $\sigma_{PC} = 13.62$, and the F-scale measures by $\sigma_{FC} = 9.74$, so variance ratios within a group of measures based on the same content area are meaningful. From the column of standard deviations, we can see that the content and agreement measures are roughly equal in variability for Play and Understanding, but that F-scale agreement is more variable than F-scale content ($F_{max} = 1.85$, $p < .01$). Comparing each of the other style measures with the corresponding agreement measures, we find that the agreement measures are much more variable (the smallest F_{max} is 2.04, $p < .001$, for UA and UF). These differences in variability among the <u>a priori</u> measures have implications for the size of the content and style variance components, which can be examined in a more detailed way when we look at the analyses of covariance structure in a later section.

All of the multiple correlations for the content, agreement and overgeneralization measures are significant at the .001 level in Table 5.7. Three of the four multiple correlations for the endorsement and form measures are not significant at the .05 level. The result for the UE measure ($R = .21$, $p = .038$) is marginal at best, and if we adopt $\alpha = .05/11$ for the set of 11 analyses,

Table 5.7

Regression Coefficients for Predicting Experimental Measures from Group, Speed, Vocabulary, Crossproducts with Group, and Age

Criterion	Std. Dev.	Multiple *R*	Intercept	Forced Predictors GRP	SPD	VOC	Stepwise Predictors GXS	GXV	AGE
PC	100	38***	-17	11	18*	08	-	-	-21**
UC	79	35***	-21	18	43*	16**	-26*	-	15**
FC	100	33***	71	-42**	-03	39	-	-34*	-
PA	97	56***	58	-69***	-51*	-18**	37**	-	21***
UA	80	55***	52	-53***	-53**	-04	32**	-	21***
FA	136	62***	107	-91***	-73*	-34***	39*	-	18*
PE	52	15	-17	02	07	-08	-	-	-
UE	50	21*	13	15	-01	05	-	-	-
PF	48	16	-34	-08	03	-06	-	-	-
UF	56	08	-67	10	-01	-03	-	-	-
FO	68	42***	-02	-41***	03	-15**	-	-	-
Std. Dev.	-	-	-	50	100	100	156	150	100

Note: The Play and Understanding criteria were scaled by $\sigma_{PC} = 13.62$, and the F-scale criteria by $\sigma_{FC} = 9.74$. SEX was included in the predictor set, but never entered any of the models, so it is omitted from this table.

All notes for Table 5.3 also apply to this table.

the result for UE should also be regarded as not significant.

The three content measures have modest, reasonable relationships with the variables in the predictor set (R = .33-.38, $p < .001$). We find that subjects with a higher need for Play tend to be younger and faster. Subjects with a higher need for Understanding tend to be older. There is a Group x Speed interaction for predicting UC, which yields the equations

$$\text{Tech:} \quad \hat{UC} = -.03 + .17\ SPD + .16\ VOC + .15\ AGE$$

$$\text{Acad:} \quad \hat{UC} = .15 - .09\ SPD + .16\ VOC + .15\ AGE$$

with a saddle point at SPD = .69 and a Johnson-Neyman region of $-.28 \geq SPD \geq \underline{max}$. For subjects in the lower third of the SPD distribution, the Academic subjects have higher UC scores than the Technical subjects. For F-scale content, there is a Group x Vocabulary interaction which yields the equations

$$\text{Tech:} \quad \hat{FC} = .29 + .05\ VOC - .03\ SPD$$

$$\text{Acad:} \quad \hat{FC} = -.15 - .29\ VOC - .03\ SPD$$

with a saddle point at VOC = -1.24 and a Johnson-Neyman region of $\underline{min.} \geq VOC \geq -.32$. The Technical group has higher scores on F-scale content, but only for subjects in the upper tw -thirds of the VOC distribution. F-scale content is commonly found to have a negative relationship with ability measures (e.g., Byrne, 1966, p. 267), which is what $b_{VOC} = -.29$ for the Academic group implies. For the Technical group, however, $b_{VOC} = .05$, and there is essentially no relationship between FC and VOC.

Looking now at the results for the three agreement measures, we find that they have stronger relationships with the predictor set ($\underline{R}$ = .55-.62, $\underline{p}$ < .001) than the content measures had. Moreover, they have impressively uniform relationships with the predictor set. What is important here is the sign pattern rather than the absolute magnitude of the regression coefficients. We find that older subjects and subjects with lower VOC scores tend to have higher agreement scores, regardless of group. In each case there is a Group x Speed interaction, and we may write separate equations by group as

Tech: PA = -.11 - .14 SPD - .18 VOC + .21 AGE

UA = -.01 - .21 SPD - .04 VOC + .21 AGE

FA = .16 - .34 SPD - .34 VOC + .18 AGE

Acad: PA = -.80 + .23 SPD - .18 VOC + .21 AGE

UA = -.54 + .11 SPD - .04 VOC + .21 AGE

FA = -.75 + .05 SPD - .34 VOC + .18 AGE

with saddle points at about 2σ on SPD (SPD = 1.72-2.08). Except for subjects in the upper range of the SPD variable (above SPD = .77-1.01), the Technical group has higher agreement scores than the Academic group. Looked at from the side of the SPD variable, the Group x Speed interaction implies that the faster subjects tend to have lower agreement scores than slower subjects in the Technical group, but higher agreement scores in the Academic group.

In Chapter 4, it was hypothesized that agreement acquiescence would be associated with verbal ability and endorsement acquiescence with speed and impulsivity. We have not looked at the relationship between impulsivity and the response style measures yet, but so far, the two-factor theory of acquiescence is disconfirmed. The endorsement measures in Table 5.7 have no relationship to either VOC or SPD, and as we saw earlier, the endorsement measures are not strongly correlated with one another. There is good evidence for agreement acquiescence in the data, and agreement shows the predicted negative relationship with VOC. The SPD variable is related to agreement rather than to endorsement; Moreover, quite unexpectedly, the agreement-on-SPD regressions are different in sign for the two groups. The best clue we have about the basis of this Group x Speed interaction for predicting agreement is the Group x Speed interaction for predicting VOC. Recall that SPD is positively associated with VOC in the Technical group ($r_{SPD,VOC} = .41$, $p < .001$), but has a low positive ($r_{SPD,VOC} = .17$, n.s.) association with VOC in the Academic group. This suggests that SPD functions like an ability measure in the Technical group, but is relatively independent of ability in the Academic group. If agreement is primarily a function of verbal ability--duller subjects having more acquiescence owing to inability to understand the items well enough to respond sensibly to them--we may be able to find a positive relationship with SPD when SPD is also a function of ability. Where SPD is independent of ability, as in the Academic group, we may be able to find that the faster subjects are not considering the items as carefully as the slower subjects, and are making more of the content-inconsistent "errors" which get scored as agreement acquiescence.

Conclusions--sample characteristics. We find good evidence for the presence of agreement in the data, but little evidence for the other response styles. There is a strong group difference in agreement response tendency. In Table 5.7, 30-38% of the variance of the agreement measures was accounted for by the variables in the predictor set. Of this total, 18-26% of the agreement variance was accounted for by the group difference alone, leaving 12-13% additional variance to be accounted for by the other variables in the predictor set. It is instructive but also a nuisance for the analysis to find so many variables having different regression slopes in the two groups. Of the 19 regression models in Tables 5.3 and 5.7, seven show a Group x Speed interaction and two show a Group x Vocabulary interaction. Based on the many indications of differences in variances and covariances for the groups, the analyses of covariance structure will be done separately by group. We turn now to those analyses.

Analyses of Covariance Structure--Experimental Scales

In this section, we consider ACOVS solutions for the experimental Play, Understanding, and F-scale subscales. Each content area will be modeled separately by group, yielding six analyses in all. The plan for analysis is that we will make a preliminary determination of the dimensionality or rank r for a group of p subscales, and then fit a range of "simply-patterned" ACOVS models to the data, following the principles for analysis outlined at the end of Chapter 2.

Play

Tables 5.8 and 5.9 provide summary information on the covariance structure of the eight experimental Play subscales, separately by group. In each table, the top panel reports the observed covariances S (scaled by the standard deviation of the PSNT scale in the combined sample), followed by the eigenvalues of S and R . We note that two of the roots of R are greater than one in each case, providing our first indicant of the dimensionality of the data.

The second panel reports the transform $V = P' S P$, where

$$P = 1/\sqrt{8} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \begin{matrix} \zeta & \text{(content)} \\ \alpha & \text{(agreement)} \\ \alpha\eta & \\ \eta & \text{(endorsement)} \\ \beta & \text{(form)} \\ \alpha\beta & \\ \alpha\beta\eta & \\ \beta\eta & \end{matrix}$$

P is a symmetric, orthonormal variant of the Hadamard solution

Table 5.8

Summary Information on Play Subscales (Technical Group, N = 97)

Covariance Matrix S

1	PSPT	6653							
2	PSPF	0108	9349						
3	PSNT	3322	0004	7651					
4	PSNF	0104	5162	-0867	11297				
5	PAPT	2564	0363	2357	1650	5845			
6	PAPF	0339	4176	-1009	3727	0365	8433		
7	PANT	1960	-0609	2399	-0781	1712	-1425	7063	
8	PANF	0481	4562	-0717	4377	1037	4208	-0337	7556
Roots(S)		22857	14172	6460	5384	5040	3572	3359	3004
Roots(R)		29036	17768	8499	6939	5962	4755	4265	3531

V = P' S P

1	PC	17789							
2	PA	-4152	18436						
3	PAE	0549	-1629	4167					
4	PE	1526	-1389	0002	5047				
5	PF	1325	0139	-0175	0120	4869			
6	PAF	-0626	1470	1257	-0696	-0371	5233		
7	PAEF	-0340	0685	-0205	-0974	-0533	0062	4367	
8	PEF	-0540	-0092	-0169	0436	-0531	-0271	-0189	3938

Table 5.8 (continued)

Likelihood-Ratio Tests on Dimensionality

r Roots or Variances	Residual Roots of R			Homogeneity of Variances			Sphericity Tests on V		
	χ^2/df	df	p	χ^2/df	df	p	χ^2/df	df	p
8	6.39	28	<.001	23.15	7	<.001	5.55	35	<.001
7	3.89	27	<.001	18.24	6	<.001	4.97	27	<.001
6	.89	20	.565	.62	5	.686	1.05	20	.392
5	.87	14	.548	.52	4	.720	.76	14	.715
4	.70	9	.676	.39	3	.759	.54	9	.850
3	.30	5	.902	.13	2	.880	.17	5	.973
2	.24	2	.775	.08	1	.782	.12	2	.886

UMLFA Results

r Factors	χ^2_{fit}	df	p	χ^2_{diff}	df	p	ρ_T	% Var.
1	71.67	20	<.001	-	-	-	.52	28
2	11.50	13	.569	60.17	7	<.001	1.02	44
3	6.57	7	.475	4.92	6	.554	1.01	55
4	.80	2	.672	5.78	5	.329	1.11	60

Notes: The coefficients in S and V are reported to 4 decimal places (decimal omitted). Indicants of dimensionality for the data, using various criteria, have been circled (see text).

Table 5.9

Summary Information on Play Subscales (Academic Group, N = 102)

		Covariance Structure S							
1	PSPT	6669							
2	PSPF	1700	8592						
3	PSNT	2132	0939	7667					
4	PSNF	2244	3604	0099	7649				
5	PAPT	2475	0896	3673	1279	8121			
6	PAPF	1117	2235	-0263	2158	1185	3875		
7	PANT	1818	1646	2483	1073	1111	1265	7115	
8	PANF	1698	3189	-0714	2816	0382	2370	0665	5539
Roots(S)		18771	11855	6483	5157	4251	4013	2805	1893
Roots(R)		26841	18285	8835	8332	5800	3776	3351	2409
		V = P' S P							
1	PC	18222							
2	PA	-0180	10617						
3	PAE	0707	-0270	3600					
4	PE	0738	-1180	1113	5184				
5	PF	1676	-0255	-0030	0190	4433			
6	PAF	-1783	0933	0100	0593	-0206	3559		
7	PAEF	0722	-2347	-0220	-0894	0175	0241	5008	
8	PEF	-0249	-0983	-0932	0574	-0345	-0059	1232	4605

Table 5.9 (continued)

Likelihood-Ratio Tests on Dimensionality

r Roots or Variances	Residual Roots of R χ^2/df	df	p	Homogeneity of Variances χ^2/df	df	p	Sphericity Tests on V χ^2/df	df	p
8	6.35	28	<.001	19.66	7	<.001	5.71	35	<.001
7	2.93	27	<.001	7.89	6	<.001	3.70	27	<.001
6	1.31	20	.137	1.29	5	.267	1.94	20	.007
5	1.17	14	.257	1.17	4	.324	1.32	14	.187
4	1.31	9	.192	.92	3	.430	1.01	9	.429
3	.94	5	.419	.78	2	.456	.38	5	.862
2	.18	2	.822	.00	1	.999	.04	2	.960

UMLFA Results

r Factors	χ^2_{fit}	df	p	χ^2_{diff}	df	p	ρ_T	% Var.
1	65.63	20	<.001	-	-	-	.57	24
2	14.44	13	.344	51.20	7	<.001	.98	43
3	7.37	7	.392	7.07	6	.315	.99	54
4	3.44	2	.179	3.93	5	.560	.87	56

Notes: The coefficients in S and V are reported to 4 decimal places (decimal omitted). Indicants of dimensionality for the data, using various criteria, have been circled (see text).

matrix introduced in Chapter 2, equation [2.7c]. The components implied by each column (or row) of P are labeled at the right of the matrix. The effects of most interest--in a notation which will be carried through the discussion of ACOVS models--are for content (ζ), agreement (α), endorsement (η), and form (β).

On examining V , we note immediately that the v_{11} and v_{22} elements, corresponding to the content and agreement components, are considerably larger than the remaining diagonal elements. This implies that we will almost certainly be able to identify content and agreement components in the data.

The third panel of the tables provides several likelihood-ratio tests on the dimensionality of V , and the fourth panel presents the results of a UMLFA analysis. There are six rough tests for dimensionality of the data in Tables 5.9 and 5.10:

1. Roots of R greater than one, in the first panel. This is the familiar Guttman-Kaiser rule for factor analysis (Kaiser, 1961).

2. Residual roots of R , in the third panel. The entry for $s = 8$ is Wilks's (1932) test that the measures are uncorrelated, and the remaining tests in that column are Bartlett's (1950, 1951a, 1951b) tests for significance of the leading principal components.

3. Homogeneity of variances, in the third panel (middle section). This is Bartlett's (1937) test for the homogeneity

or equality of independent variances, applied to the diagonal elements of V . The elements of V are not experimentally independent, but it is feasible to perform tests of independence (item 2) and equality (item 3) sequentially (e.g., Bock, 1975, pp. 459-460). The entry for $\underline{s} = 8$ is a test of equality for all eight diagonal elements, while the entries for $\underline{s} \leq 7$ are tests for equality of the $\underline{s}$ smallest elements among the last seven.

4. <u>Sphericity tests on</u> V , in the third panel (last section). The entry for $\underline{s} = 8$ is Mauchly's (1940) test of sphericity for Σ , and the entry for $\underline{s} = 7$ is Huynh and Feldt's (1970) test that Σ has the "Type H" pattern. The remaining entries in this column are the extensions of these two tests which were proposed in Chapter 2. They are tests of sphericity on the lower-right $\underline{s} \times \underline{s}$ submatrix of V , after reordering V so that the diagonal elements are in descending order by size. The test for dimensionality is conducted in the same way for each of items 2-4: If the last value of $\underline{s}$ for which the test shows significant lack of fit, or significant differences among the roots or variances tested for equality, is $\underline{t}$, then the implied dimensionality for the data is $\underline{r} = \underline{p} - \underline{t}$.

5. <u>UMLFA tests for goodness-of-fit</u>, in the fourth panel. The implied dimensionality of the data is the first value of $\underline{r}$ showing adequate fit.

6. <u>UMLFA tests for difference in fit</u>, in the fourth panel. Here the implied dimensionality is the last value of $\underline{r}$ showing

a significant difference between the solutions for $\underline{r}$ and $\underline{r}$-1 factors.

For the Technical group, all six tests indicate a dimensionality of two for the data, and for the Academic group, five of the six tests also agree on a dimensionality of two. An examination of the elements of V indicates that the probable components are content and agreement, and we can make rough estimates of the content and agreement components using the component analysis methods of the early part of Chapter 2. Assuming that V has the form of equation [2.9b], with two components and homogeneous error, the last six diagonal elements of V estimate the pooled error component, so we have:

$$\overline{\hat{\sigma}^2} = \sum_{j=3}^{8} \underline{v}_{jj} / 6 .$$

Since $\underline{E}(\underline{v}_{11}) = \underline{p}\,\sigma_\xi^2 + \sigma^2$ and $\underline{E}(\underline{v}_{22}) = \underline{p}\,\sigma_\alpha^2 + \sigma^2$,

$$\hat{\sigma}_\xi^2 = (\underline{v}_{11} - \overline{\hat{\sigma}^2})/8 \quad \text{and}$$

$$\hat{\sigma}_\alpha^2 = (\underline{v}_{22} - \overline{\hat{\sigma}^2})/8 .$$

We find $\overline{\hat{\sigma}^2} = .460$ for the Technical group, and $\overline{\hat{\sigma}^2} = .440$ for the Academic group. These are not significantly different ($\underline{F}_{max}(100,100) = 1.05$, n.s.), so we can pool the estimates further to obtain $\overline{\hat{\sigma}^2} = .450$ for both groups. Thus, we obtain the rough estimates:

Technical group: $\hat{\sigma}_\xi^2 = .166$, $\hat{\sigma}_\alpha^2 = .174$;

Academic group: $\hat{\sigma}_\xi^2 = .171$, $\hat{\sigma}_\alpha^2 = .076$.

As we shall see, these are fairly close to the estimates obtained by more sophisticated ACOVS methods. The agreement component is actually larger than the content component for the Technical group, but less than half the size of the content component in the Academic group. By any reasonable statistical test, the first difference is not significant, while the second difference is. Perhaps the most defensible test would be a test for correlated differences between variances:

$$F_{1,N-2} = (N-2)(v_{11}-v_{22})^2 / 4v_{11}v_{22}(1-v_{12}^2/v_{11}v_{22})$$

(after Formula 15-11, Walker & Lev, 1969, p. 264). This formula is for testing whether the variances of the a priori estimates of content and agreement are significantly different, within group, and we obtain $F(1,95) = .03$, n.s., for the Technical group, and $F(1,100) = 7.47$, $p < .01$, for the Academic group. Similarly, the groups do not differ in variance on the content estimates ($F_{max}(100,100) = 1.02$, n.s.), but do differ in variance on the agreement estimates ($F_{max}(100,100) = 1.74$, $p < .001$).

Two other features of Tables 5.8 and 5.9 which may be noted are the Tucker and Lewis (1973) factor reliability coefficients ρ_T and the precent variance figures in the fourth panel. ρ_T should be close to one for a good UMLFA solution (it is). The percent variance figures are obtained as tr(BB')/tr(S) for the UMLFA solution, and we find 44% of the variance accounted for in the Technical group and 43% in the Academic group. Since the average coefficient-α for the Play subscales was .46, we cannot expect to account for more variance than this.

Table 5.10 presents the results for a range of 2-component ACOVS solutions for the Play subscales. All of the models fit for the Technical group, including the most restricted model 1, while none of the models fit for the Academic group, including the least restricted model 8. In the second panel of Table 5.10 are averaged χ^2 values for goodness of fit and averaged differences in χ^2 values for models differing in the restrictions placed on A, on C, and on E. (It is possible to take the vector of eight χ^2 values-- x, say--and the vector of eight df values d, and premultiply each by the Hadamard solution matrix in order to decompose the χ^2 into "effects" attributable to the 2 x 2 x 2 design on the range of solutions. This can be defended heuristically if not statistically, and gives a nice overview of the effects of various kinds of restrictions.) We note that for the Technical group, none of the design effects improves the fit to the data, on average. For the Academic group, homogeneous error variances improve the fit, on average ($\chi^2(7) = 26.53$, $p < .001$), but the other design variations do not result in much improvement.

Selected ACOVS solutions for the Technical group are given in Table 5.11 and plotted in Figure 5.1. The selected solutions are the 2-factor UMLFA solution, the most restricted model 1 from Table 5.10, and a third model 1A with the same number of restrictions as model 1 but with the restrictions distributed in a slightly different way. In model 1, C is constrained to be diagonal, but in the other ACOVS model, C is constrained to be proportional to a correlation matrix. Both ACOVS models yield a reasonable

Table 5.10

Results for 2-Component ACOVS Solutions (Play Subscales)

Model	Parameter Type A	B	C	E	df	Tech. Gp. χ^2	p	Acad. Gp. χ^2	p
1	1	2	1	1	33	40.57	.171	73.89	<.001
2	"	"	"	2	26	28.17	.350	44.29	.014
3	"	"	2	1	32	35.38	.311	73.88	<.001
4	"	"	"	2	25	23.63	.541	44.20	.010
5	2	"	1	1	27	32.29	.222	63.44	<.001
6	"	"	"	2	20	20.50	.427	39.21	.006
7	"	"	2	1	26	30.07	.265	60.66	<.001
8	"	"	"	2	19	19.44	.429	38.07	.006

Effect Summary

	df	Tech. Gp. χ^2	p	Acad. Gp. χ^2	p
Mean χ^2	26	28.76	.322	54.71	<.001
A_1 vs. A_2	6	6.36	.384	8.72	.190
C_1 vs. C_2	1	3.25	.071	1.01	.316
E_1 vs. E_2	7	11.64	.113	26.53	<.001

Note: For parameter types, see Table 2.8. Models with circled probabilities are discussed in detail elsewhere.

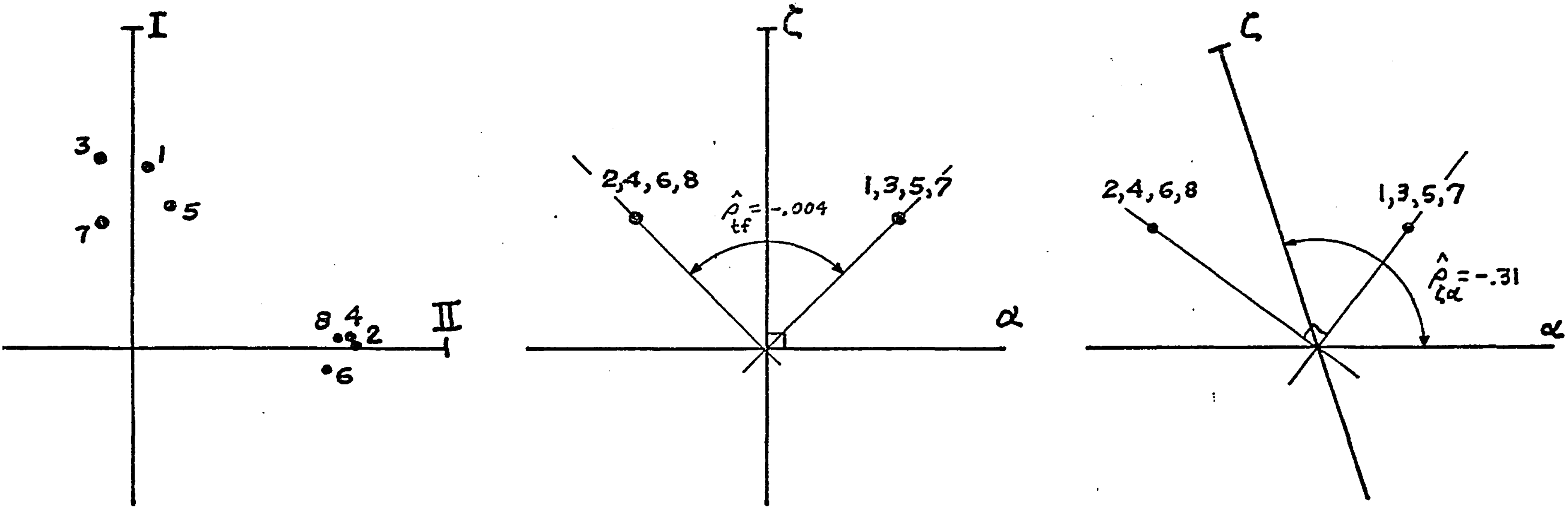
Table 5.11

Selected 2-Component Solutions for Experimental Play Subscales (Technical Group)

	UMLFA (Varimax)			ACOVS Model 1			ACOVS Model 1A		
	I	II	e_j	ζ	α	e_j	ζ	α	e_j
PSPT	558	048	352	1.*	1.*	461	1.*	1.*	461
PSPF	001	689	460	1.*	-1.*	461*	1.*	-1.*	461*
PSNT	587	-091	412	1.*	1.*	461*	1.*	1.*	461*
PSNF	009	684	661	1.*	-1.*	461*	1.*	-1.*	461*
PAPT	442	128	372	1.*	1.*	461*	1.*	1.*	461*
PAPF	-064	609	468	1.*	-1.*	461*	1.*	-1.*	461*
PANT	390	-105	543	1.*	1.*	461*	1.*	1.*	461*
PANF	022	667	310	1.*	-1.*	461*	1.*	-1.*	461*
C	1.*			165			169		
	0.*	0.*		0.*	173		-052	169*	

UMLFA (Varimax)	ACOVS Model 1	ACOVS Model 1A
$\chi^2(13) = 11.30$	$\chi^2(33) = 40.57$	$\chi^2(33) = 35.42$
$p = .569$	$p = .171$	$p = .355$

Note: Parameter estimates are reported to 3 decimal places except where a decimal is explicitly provided. Starred parameters were fixed or contrained to be equal to other parameters.



(a) UMLFA (Varimax)

$\chi^2(13) = 11.50$

$p = .569$

(b) ACOVS Model 1

$\chi^2(33) = 40.57$

$p = .171$

(c) ACOVS Model 1A

$\chi^2(33) = 35.42$

$p = .355$

Figure 5.1. Selected solutions for Play subscales (Technical group).

approximation to the UMLFA solution, which shows a clear separation between the true (odd-numbered) and false (even-numbered) tests. In order to plot the two ACOVS models on the same scale as the UMLFA model in Figure 5.1, the solutions in Table 5.11 were transformed by getting $B^* = B\,\mathrm{diag}(C)^{\frac{1}{2}}$ and $C^* = \mathrm{diag}(C)^{-\frac{1}{2}}\, C\, \mathrm{diag}(C)^{-\frac{1}{2}}$; the transformation puts 1's in the diagonal of C^* , but does not affect the fit of the solution.

Models 1 and 1A both fit the data even though the true tests are constrained to correlate perfectly in component-space, and so are the false tests. The two restricted solutions do not differ significantly from the UMLFA solution ($\chi^2(20) = 29.27$, n.s., for the UMLFA-model 1 difference, and $\chi^2(20) = 24.12$, n.s., for the UMLFA-model 1A difference). Models 1 and 1A have the same degrees of freedom, and cannot be formally tested for difference in fit, but the χ^2 difference can be evaluated heuristically using 1 df: This yields $\chi^2(1) = 5.15$, $p < .01$, and we can conclude that model 1A fits marginally better. Note that model 1 has the restrictions $\hat{\sigma}_t^2 = \hat{\sigma}_f^2$ and $\hat{\rho}_{\zeta\alpha} = 0$, while $\hat{\sigma}_\zeta^2$, $\hat{\sigma}_\alpha^2$ and $\hat{\rho}_{tf}$ are unrestricted. Model 1A has the restrictions $\hat{\sigma}_\zeta^2 = \hat{\sigma}_\alpha^2$ and $\hat{\rho}_{tf} = 0$, while $\hat{\sigma}_t^2$, $\hat{\sigma}_f^2$ and $\hat{\rho}_{\zeta\alpha}$ are unrestricted. A slightly better fit is achieved with the second set of restrictions.

Under either restricted model, the true and false tests correlate approximately zero, a result which is also evident when we examine the observed covariances in Table 5.8. Can we have blundered somehow? The subscales had low reliabilities (average coefficient-α about .46): Can the result we have obtained be due to random responding? No: If it were due to random responding, the correlations within the true and false sets of tests would not be as high as they are. The components in Table 5.11 are variance components for seven-item subscales. under either either model 1 or 1A, the reliabilities of 56-item estimates of content or agreement would be about

$$\hat{\rho}_{\xi\xi} = \hat{\rho}_{\alpha\alpha} = (8 \times .17)/(8 \times .17 + .46) = .75 ,$$

which is fairly respectable. Are we inadvertently measuring different kinds of content with the true and false subscales? Well, perhaps: But then we would not expect the a priori estimates of agreement obtained from the three content areas to correlate as highly as they do (.66-.75 overall, as indicated earlier in the chapter).

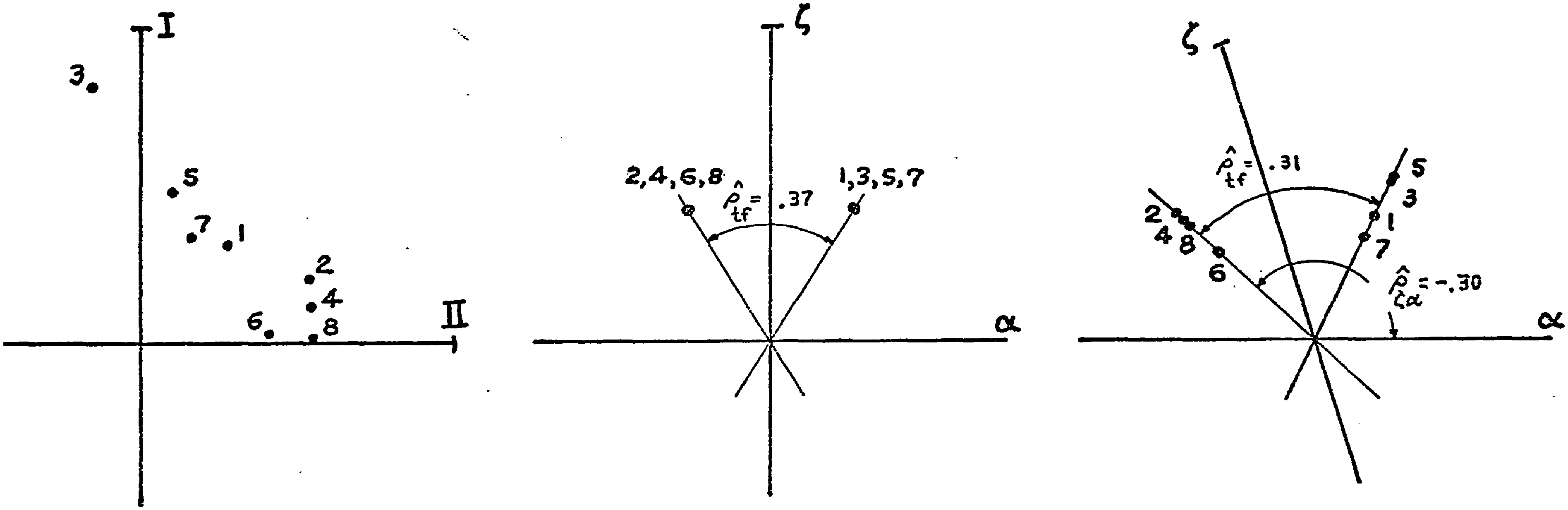
Fitting ACOVS models to the Play subscales for the Technical group was very easy--perhaps deceptively so. When we turn to the data for the Academic group, however, none of the simply-patterned 2-component models provides an acceptable fit, as shown in Table 5.10. Heterogeneous error variances improve the fit, and models 2, 4, 6, and 8 using $E = E_2$ do not differ significantly from one another. Selected solutions for the Academic group are given in Table 5.12 and Figure 5.2, including the UMLFA solution and the model 2 and 8 ACOVS solutions. While there is a clear separation of the true and false scales, the

Table 5.12

Selected 2-Component Solutions for Experimental Play Subscales (Academic Group)

	UMLFA (Varimax)			ACOVS Model 2			ACOVS Model 8			
	I	II	$\underline{e}_j$	ζ	α	$\underline{e}_j$	$\underline{a}_j$	ζ	α	$\underline{e}_j$
PSPT	320	284	484	1.*	1.*	436	1.*	1.*	1.*	462
PSPF	204	548	518	1.*	-1.*	534	1.*	1.*	-1.*	497
PSNT	814	-147	082	1.*	1.*	510	1240	1.*	1.*	453
PSNF	114	541	459	1.*	-1.*	489	918	1.*	-1.*	203
PAPT	471	112	578	1.*	1.*	558	1249	1.*	1.*	494
PAPF	048	416	212	1.*	-1.*	187	681	1.*	-1.*	219
PANT	332	167	574	1.*	1.*	537	795	1.*	1.*	582
PANF	008	550	251	1.*	-1.*	286	884	1.*	-1.*	270
C	1.*			164				184		
	0.*	1.*		0.*	075			-040	099	
	$\chi^2(13) = 14.44$			$\chi^2(26) = 44.29$				$\chi^2(19) = 38.07$		
	$\underline{p} = .344$			$\underline{p} = .014$				$\underline{p} = .006$		

Note: Parameter estimates are reported to 3 decimal places except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.



(a) UMLFA (Varimax)

$\chi^2(13) = 14.44$

$p = .344$

(b) ACOVS Model 2

$\chi^2(26) = 44.29$

$p = .014$

(c) ACOVS Model 8

$\chi^2(19) = 38.07$

$p = .006$

Figure 5.2 Selected solutions for Play subscales (Academic group).

UMLFA solution for the Academic group is much "noisier" than the UMLFA solution for the Technical group was.

Model 8 in Figure 5.2 illustrates the handling of the least-restricted simply-patterned model. It is plotted by finding $B^* = A\,B\,\mathrm{diag}(C)^{\frac{1}{2}}$ and $C^* = \mathrm{diag}(C)^{-\frac{1}{2}}\,C\,\mathrm{diag}(C)^{-\frac{1}{2}}$. Note that all of the projections of the tests lie on one of two vectors in model 8, so the correlation between any pair of true and false tests is constant in component-space; simply-patterned ACOVS models will always have this appearance under the least-restricted model. Model 8 fits significantly poorer than the UMLFA solution ($\chi^2(6) = 23.63$, $p < .001$), and does not fit overall, so we cannot use it.

Three-component ζ,α,η and ζ,α,β solutions were obtained, along with the 4-component ζ,κ,η,β solution, and none achieved an acceptable degree of fit. Estimates of the endorsement and form components were close to zero in each of these solutions, and I finally concluded that it was not possible to fit a simply-patterned ACOVS model for the Academic group.

Judging from the UMLFA solution, the principal impediment to fitting a simply-patterned model is that test 3 (PSNT) behaves differently than the other true measures. Sörbom (1974) has suggested an analytic procedure for improving the fit of overrestricted models, which also calls attention to test 3. Sörbom's procedure is to examine the first derivatives obtained at the point of solution, free the element having the largest partial first derivative, and then reanalyze the data. If the fit is still not adequate, the derivatives of the new solution may be obtained, and again the element having the largest partial first derivative is freed. From a given starting point, this process yields a unique solution in a manner analogous to the forward selection procedure used in stepwise multiple regression (e.g., Draper & Smith, 1966, pp. 169 ff.) Since none of the simply-patterned models fit the data, Sörbom's procedure may be adopted here. Suppose we confine our attention to the partials of B , and begin with model 2 from the 2-component solution given in Table 5.12. The partials are:

$$\partial \underline{F}/\partial B' = \begin{pmatrix} -103 & -114 & 125 & -081 & -024 & 147 & 031 & 020 \\ 111 & 000 & \boxed{-169} & -014 & -010 & -079 & 081 & 106 \end{pmatrix}$$

The $\partial \underline{F}/\partial \underline{b}_{32}$ element, corresponding to the "agreement" loading for test 3, is the largest. When this element is freed, we obtain the new solution displayed in Table 5.13 and Figure 5.3. The fit for the new solution is adequate ($\chi^2(25) = 28.49$, $\underline{p} = .286$), and the new solution fits significantly better than model 2 ($\chi^2(1) = 15.80$, $\underline{p} < .001$). The new solution in Figure 5.3 resembles a tidy version of the UMLFA solution in Figure 5.2 (after reflection and rotation). The new solution puts one vector through all of the false tests and another correlated ($\hat{\rho} = .54$) vector through three of the true tests. Test 3 is off by itself, approximately orthogonal to the false tests.

The solution in Table 5.13 is a good "simple structure" solution, but it is not very satisfying from the standpoint of a component model for the data. It is not clear where we should

Table 5.13

Final Solution for Play Subscales (Academic Group)--
Model 2 with $\underline{b}_{32}$ free

	ζ	α	$\underline{e}_j$
PSPT	1.*	1.*	491
PSPF	1.*	-1.*	575
PSNT	1.*	3194	145
PSNF	1.*	-1.*	491
PAPT	1.*	1.*	575
PAPF	1.*	-1.*	180
PANT	1.*	1.*	558
PANF	1.*	-1.*	287

C	168	
	0.*	050

$$\chi^2(25) = 28.49$$
$$\underline{p} = .286$$

Note: Parameter estimates are reported to three decimal places except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.

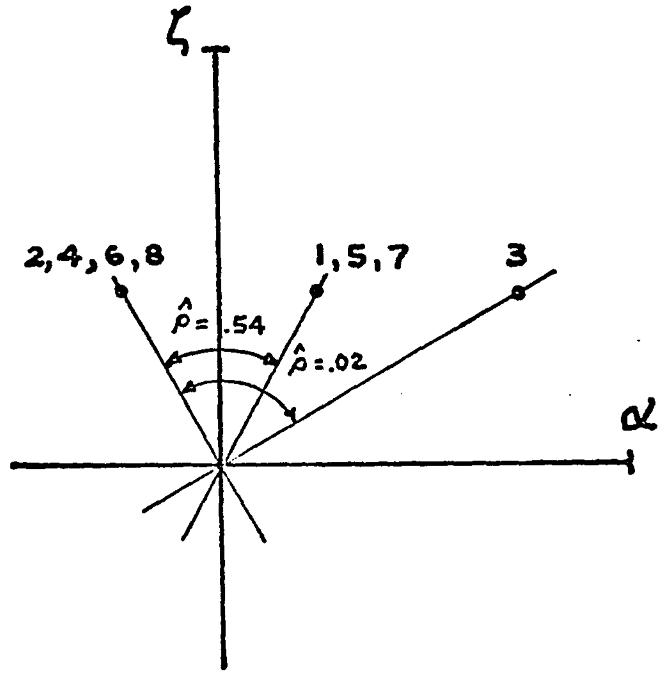


Figure 5.3. Final solution for Play subscales (Academic group). Model 2 with $\underline{b}_{32}$ element unconstrained.

locate "agreement" in Figure 5.3, for example, nor why the solution for the Academic group is not as tidy as the solutions obtained earlier for the Technical group. This failure to find a simply-patterned solution for the Academic group does not rule out content and agreement as sources of the observed covariances, but the data do not sustain a simple view of the tests as providing parallel or tau-equivalent measures of the two components. The PSNT test elicits more apparent agreement than the other true measures.

Understanding

Tables 5.14 and 5.15 present summary information on the covariance structure of the Understanding subscales, in the same format as Tables 5.8 and 5.9 for Play. When we examine the elements of V for Understanding, we find that the largest elements for both datasets are $\underline{v}_{11}$ and $\underline{v}_{22}$, corresponding to the content and agreement components, followed by $\underline{v}_{55}$, corresponding to the form component. Taking the last six diagonal elements of V to estimate the mean error variance $\overline{\hat{\sigma}^2}$, provisional estimates of the content and agreement components are:

$$\text{Tech.: } \hat{\sigma}_{\zeta}^2 = .101, \quad \hat{\sigma}_{\alpha}^2 = .113, \quad \text{based on } \overline{\hat{\sigma}^2} = .429 \,;$$

$$\text{Acad.: } \hat{\sigma}_{\zeta}^2 = .086, \quad \hat{\sigma}_{\alpha}^2 = .037, \quad \text{based on } \overline{\hat{\sigma}^2} = .374 \,.$$

Again we notice that the content and agreement components are about the same size for the Technical group, but that the agreement component for the Academic group is considerably smaller than

Table 5.14

Summary Information on Understanding Subscales (Technical Group, $\underline{N}$ = 97)

Covariance Matrix S

1	USPT	9878							
2	USPF	0510	5854						
3	USNT	2508	-0144	4946					
4	USNF	2445	3091	0428	8227				
5	UAPT	2429	-0463	0939	-0310	5047			
6	UAPF	-0267	1845	-0470	2062	0042	3300		
7	UANT	1580	-1212	1152	-0063	1880	0298	6961	
8	UANF	-0361	2706	-1307	3242	-0440	2259	-0535	7249
Roots(S)		14909	13382	6704	4522	3847	3186	3017	1895
Roots(R)		23163	18669	8984	7539	6262	4557	3761	2377

V = P' S P

1	UC	12394							
2	UA	-0904	13318						
3	UAE	-0118	1120	3978					
4	UE	2551	-0768	0701	4709				
5	UF	2127	-0051	0806	1131	6643			
6	UAF	-0230	0190	0188	0657	0973	3114		
7	UAEF	1052	-0534	0175	0396	-0378	0289	3161	
8	UEF	1850	1696	-0473	0681	0709	-0388	0330	4145

Table 5.14 (continued)

Likelihood-Ratio Tests on Dimensionality

r Roots or Variances	Residual Roots of R			Homogeneity of Variances			Sphericity Tests on V		
	χ^2/df	df	p	χ^2/df	df	p	χ^2/df	df	p
8	5.09	28	<.001	17.11	7	<.001	5.14	35	<.001
7	3.16	27	<.001	14.57	6	<.001	4.67	27	<.001
6	1.45	20	.089	3.97	5	.001	2.32	20	<.001
5	.79	14	.676	1.53	4	.190	1.42	14	.133
4	.56	9	.834	1.07	3	.359	.87	9	.551
3	.20	5	.962	.92	2	.398	.62	5	.685
2	.22	2	.803	.01	1	.942	.41	2	.665

UMLFA Results

r Factors	χ^2_{fit}	df	p	$\chi^2_{diff.}$	df	p	ρ_T	% Var.
1	61.53	20	<.001	-	-	-	.49	21
2	18.69	13	.133	42.08	7	<.001	.89	40
3	5.60	7	.587	13.08	6	.046	1.05	48
4	1.79	2	.408	3.81	5	.577	1.03	61

Notes: The coefficients in S and V are reported to 4 decimal places (decimal omitted). Indicants of dimensionality for the data, using various criteria, have been circled (see text).

Table 5.15

Summary Information on Understanding Subscales (Academic Group, N = 102)

Covariance Matrix S									
1	USPT	8094							
2	USPF	0859	5218						
3	USNT	2742	1208	5123					
4	USNF	1687	1714	0645	4010				
5	UAPT	2651	0526	0866	0851	3804			
6	UAPF	-0225	0502	-0675	0549	0417	2781		
7	UANT	1156	0636	1486	0742	0458	-0166	6811	
8	UANF	-0048	0798	0008	1038	0682	0729	0629	3852
Roots(S)		12666	6564	6120	4476	3087	2617	2133	2031
Roots(R)		23610	13931	12828	7910	6204	5231	4147	2252
V = P' S P									
1	UC	10578							
2	UA	2004	6690						
3	UAE	0058	-0139	3984					
4	UE	0946	0282	1288	4271				
5	UF	2176	1249	0901	0364	5147			
6	UAF	-0598	0758	0540	0894	0368	3074		
7	UAEF	1376	0569	-0648	-0505	-0032	-0568	2687	
8	UEF	0376	0957	-0143	0313	-0326	-0257	0327	3261

Table 5.15 (continued)

Likelihood-Ratio Tests on Dimensionality

r Roots or Variances	Residual Roots of R χ^2/df	df	p	Homogeneity of Variances χ^2/df	df	p	Sphericity Tests on V χ^2/df	df	p
8	3.85	28	<.001	11.33	7	<.001	4.28	35	<.001
7	2.00	27	(<.001)	5.21	6	<.001	2.92	27	<.001
6	1.46	20	.085	2.90	5	(.013)	2.34	20	<.001
5	1.35	14	.169	1.81	4	.123	2.48	14	(.002)
4	1.19	9	.299	1.37	3	.251	1.67	9	.090
3	1.39	5	.225	.49	2	.613	1.31	5	.257
2	1.55	2	.212	.46	1	.500	2.22	2	.109

UMLFA Results

r Factors	χ^2_{fit}	df	p	χ^2_{diff}	df	p	ρ_T	% Var.
1	44.54	20	<.001	-	-	-	.57	22
2	17.11	13	(.194)	27.43	7	(<.001)	.89	37
3	4.79	7	.686	12.32	6	.055	1.11	45
4	.76	2	.683	4.03	5	.546	1.22	54

Notes: The coefficients in S and V are reported to 4 decimal places (decimal omitted). Indicants of dimensionality for the data, using various criteria, have been circled (see text).

either the content component for the Academic group or the agreement component for the Technical group. This is similar to the result observed for the Play scales, and once again illustrates the greater influence of agreement with the Technical group.

For both groups, three of the rough tests on dimensionality indicate $\underline{r} = 2$, while the other three tests indicate $\underline{r} = 3$ or (in one case) $\underline{r} = 4$. On fitting a range of 2-component models to the datasets, it is difficult to fit a simply-patterned model without relaxing one of the fixed parameters in B . By fitting 3-component models which include a form component, however, reasonable simply-patterned models can be obtained.

Table 5.16 reports the results for the same range of 2-component models used with the Play subscales. Only one of the 16 solutions provides an acceptable fit for the data--the least-restricted model 8 for the Academic group. For comparison purposes, model 8 for the Technical group can also be examined. The UMLFA solutions for the two groups, and the results using model 8 are given in Tables 5.17 and 5.18, and plotted in Figures 5.4 and 5.5. Solutions obtained from model 8 by relaxing the $\underline{b}_{42}$ element are also reported.

There is a clear separation between the _true_ and _false_ subscales for both groups, and especially for the Technical group. One noticeable feature of the Understanding data is that test 1 (USPT) has larger variances than the other tests in Tables

Table 5.16

Results for 2-Component ACOVS Solutions (Understanding Subscales)

Model	Parameter Type A	B	C	E	df	Tech. Gp. χ^2	p	Acad. Gp. χ^2	p
1	1	2	1	1	33	85.34	<.001	89.28	<.001
2	"	"	"	2	26	41.99	.025	52.89	.001
3	"	"	2	1	32	84.86	<.001	83.39	<.001
4	"	"	"	2	25	41.00	.023	50.34	.002
5	2	"	1	1	27	63.22	<.001	63.23	<.001
6	"	"	"	2	20	32.87	.035	33.81	.027
7	"	"	2	1	26	58.89	<.001	58.85	<.001
8	"	"	"	2	19	32.38	.028	26.13	.127

Effect Summary

	df	Tech. Gp. χ^2	p	Acad. Gp. χ^2	p
Mean χ^2	26	55.07	<.001	57.24	<.001
A_1 vs. A_2	6	16.46	.012	23.47	<.001
C_1 vs. C_2	1	1.57	.210	5.13	.024
E_1 vs. E_2	7	36.02	<.001	32.90	<.001

Note: For parameter types, see Table 2.8. Models with circled probabilities are reported in detail elsewhere.

Table 5.17

Selected 2-Component Solutions for Experimental Understanding Subscales (Technical Group)

	UMLFA (Varimax)			ACOVS Model 8				ACOVS Model 8A			
	I	II	e_j	a_j	ζ	α	e_j	a_j	ζ	α	e_j
USPT	765	155	379	1.*	1.*	1.*	608	1.*	1.*	1.*	424
USPF	-042	495	338	1.*	1.*	-1.*	336	1.*	1.*	-1.*	345
USNT	351	-055	368	551	1.*	1.*	379	456	1.*	1.*	377
USNF	172	633	392	1153	1.*	-1.*	491	1635	1.*	-525	405
UAPT	309	-040	407	634	1.*	1.*	352	417	1.*	1.*	407
UAPF	-074	354	199	753	1.*	-1.*	188	762	1.*	-1.*	190
UANT	262	-061	624	564	1.*	1.*	575	361	1.*	1.*	623
UANF	-168	563	380	1145	1.*	-1.*	398	1185	1.*	-1.*	387
C	1.*				148				180		
	0.*	1.*			033	166			081	222	

UMLFA (Varimax): $\chi^2(13) = 18.69$, $p = .133$

ACOVS Model 8: $\chi^2(19) = 32.38$, $p = .028$

ACOVS Model 8A: $\chi^2(18) = 24.98$, $p = .126$

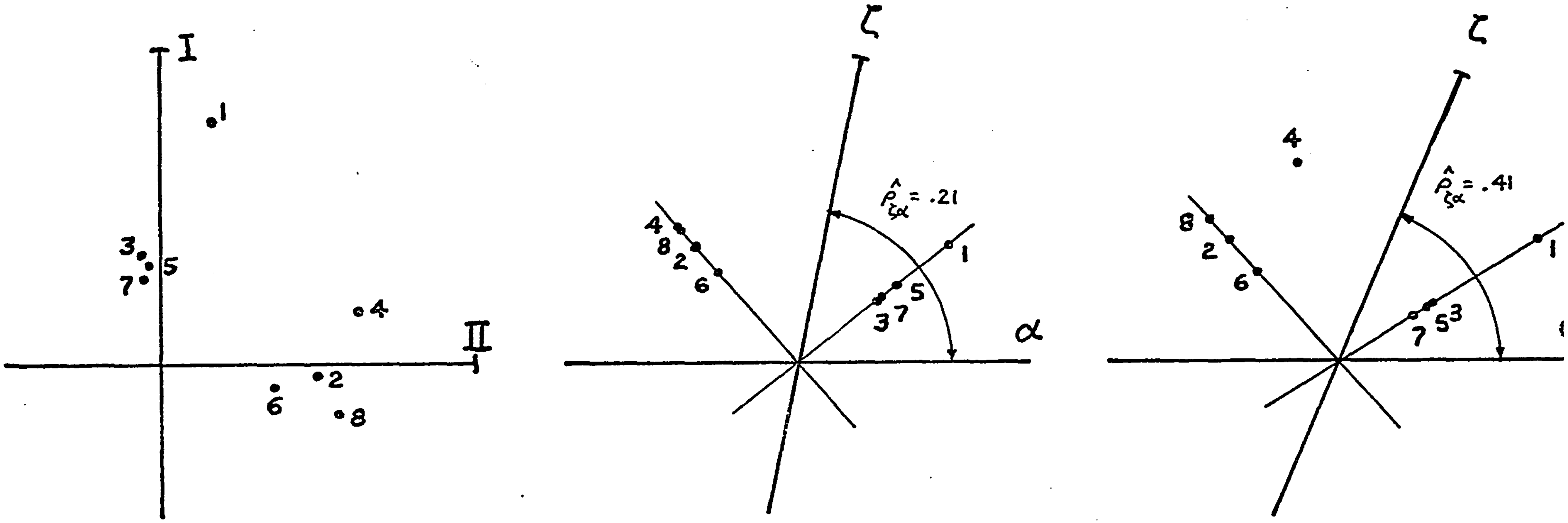
Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.

Table 5.18

Selected 2-Component Solutions for Experimental Understanding Subscales (Academic Group)

	UMLFA (Varimax)			ACOVS Model 8				ACOVA Model 8A			
	I	II	e_j	a_j	ζ	α	e_j	a_j	ζ	α	e_j
USPT	895	-053	005	1.*	1.*	1.*	222	1.*	1.*	1.*	167
USPF	117	352	384	1.*	1.*	-1.*	406	1.*	1.*	-1.*	381
USNT	309	037	415	457	1.*	1.*	389	423	1.*	1.*	397
USNF	210	360	228	1514	1.*	-1.*	136	1244	1.*	-720	198
UAPT	302	109	277	442	1.*	1.*	266	408	1.*	1.*	274
UAPF	-015	174	247	335	1.*	-1.*	265	410	1.*	-1.*	254
UANT	137	134	644	237	1.*	1.*	648	209	1.*	1.*	653
UANF	012	296	297	607	1.*	-1.*	342	669	1.*	-1.*	321
C	1.*				228				224		
	0.*	1.*			118	123			125	167	
	$\chi^2(13) = 17.11$				$\chi^2(19) = 26.13$				$\chi^2(18) = 24.26$		
	$p = .194$				$p = .127$				$p = .147$		

Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.



(a) UMLFA (Varimax)

$\chi^2(13) = 18.69$

$\underline{p} = .133$

(b) ACOVS Model 8

$\chi^2(19) = 32.38$

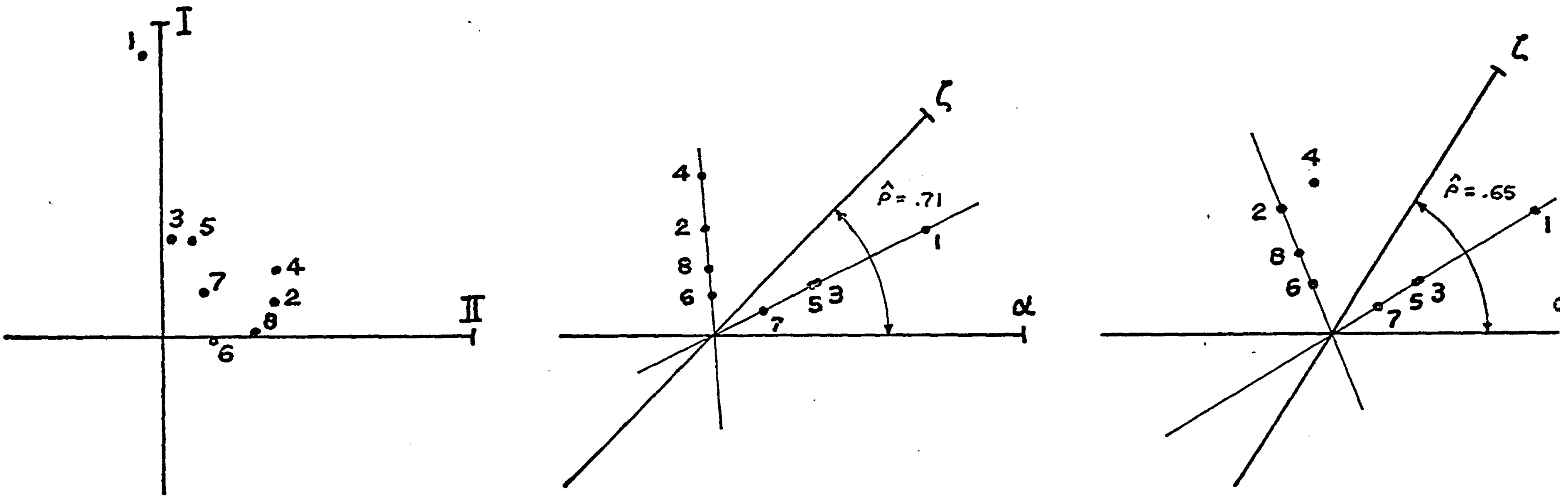
$\underline{p} = .028$

(c) ACOVS Model 8A

$\chi^2(18) = 24.98$

$\underline{p} = .126$

Figure 5.4. Selected solutions for Understanding subscales (Technical group).



(a) UMLFA (Varimax)

$\chi^2(13) = 17.11$

$p = .194$

(b) ACOVS Model 8

$\chi^2(19) = 26.13$

$p = .127$

(c) ACOVS Model 8A

$\chi^2(18) = 24.26$

$p = .147$

Figure 5.5. Selected solutions for Understanding subscales (Academic group).

5.14 and 5.15, and also has larger projections in component-space in Figures 5.4 and 5.5. It is tempting to speculate that there is something unusual about the USPT scale. A review of the Understanding items in Appendix B does not reveal any qualitative difference between the USPT items and the other items, that I can see. The USPT subscale has a higher coefficient-α than the other Understanding subscales (cf. Table 5.4), and may just contain better items. The USPT subscale is reasonably collinear with the other _true_ measures, and an examination of the partial first derivatives obtained with model 8, for both groups, indicates that the largest partial is for test 4 (USNF). Accordingly, solutions were obtained with the $\underline{b}_{42}$ element freed, which resulted in an acceptable fit for the Technical group and a slightly improved fit for the Academic group.

When 3-component or 4-component solutions were attempted for the Play subscales, the estimates of components other than content and agreement were invariably less than twice their standard errors. When solutions involving null components are attempted, technical difficulties often occur with the ACOVSF program, such as excessive iterations or improper solutions (zero or near-zero scaling factors or error variances, singular or non-Gramian C matrices, and the like), and these were also found in trying to fit higher-order ACOVS solutions to the Play subscales. With the Understanding subscales, however, reasonable 3-component solutions using the form component can be obtained for both groups.

Table 5.19 presents the results for a range of 3-component ζ, α, β solutions. Models 2, 4, 6 and 8 fit for the Technical group, and models 4, 6 and 8 fit for the Academic group. Tables 5.20 and 5.21 report the UMLFA 3-factor and model 4 solutions for the two groups. It should be noted that model 4 does not differ significantly from model 2 for the Technical group ($\chi^2(3) = 2.98$, $p = .395$), and model 4 does differ significantly from model 8 for the Academic group ($\chi^2(7) = 14.91$, $p = .037$), so we could have accepted model 2 for the Technical group and model 8 for the Academic group. Model 4 provides better comparability between groups, however, and will be interpreted.

The UMLFA solution for the Technical group shows the form effect (in Table 5.20). Factor I is marked by the false scales, Factor II is marked by the true scales, and Factor III by a self-descriptive vs. attitude contrast. Model 4 does not differ significantly from the UMLFA solution ($X^2(15) = 19.97$, n.s.), and shows a small variance component for form. Based on a pooled error estimate of $\overline{\hat{\sigma}^2} = .392$ (obtained by averaging the e_j column for model 4), the estimated reliability of a 56-item measure of the form effect would be

$$\hat{\rho}_{\beta\beta} = (8 \times .033)/(8 \times .033 + .392) = .40 ,$$

which is not very high. (By contrast, $\hat{\rho}_{\zeta\zeta} = .66$ and $\hat{\rho}_{\alpha\alpha} = .69$).

The UMLFA solution for the Academic group is not as clear as the one for the Technical group. Also, test 1 (USPT) goes Heywood

Table 5.19

Results for 3-Component ACOVS Solutions (Understanding Subscales)

Model	Parameter Type A	B	C	E	df	Tech. Gp. χ^2	p	Acad. Gp. χ^2	p
1	1	3	1	1	32	71.57	<.001	82.00	<.001
2	"	"	"	2	25	28.55	.283	44.26	.010
3	"	"	2	1	29	65.66	<.001	64.50	<.001
4	"	"	"	2	22	25.57	.270	29.41	.134
5	3	"	1	1	25	39.72	.031	50.97	.002
6	"	"	"	2	18	19.53	.360	20.13	.326
7	"	"	2	1	22	31.00	.096	46.70	.002
8	"	"	"	2	15	14.53	.485	14.51	.488

Effect Summary

	df	Tech. Gp. χ^2	p	Acad. Gp. χ^2	p
Mean χ^2	23.5	37.02	.038	44.06	.006
A_1 vs. A_2	7	21.64	.003	21.97	.003
C_1 vs. C_2	3	5.65	.130	10.56	.014
E_1 vs. E_2	7	29.94	<.001	33.96	<.001

Note: For parameter types, see Table 2.8. Components are content, agreement, and form. Models with circles probabilities are reported in detail elsewhere.

Table 5.20

Selected 3-Component Solutions for Experimental Understanding Subscales (Technical Group)

	UMLFA (Unrotated) I	II	III	e_j	ACOVS Model 4 ζ	α	β	e_j
USPT	277	769	153	296	1	1	1	672
USPF	490	-126	125	314	1	-1	1	289
USNT	-003	336	020	381	1	1	1	303
USNF	636	068	088	405	1	-1	1	473
UAPT	015	328	-199	357	1	1	-1	324
UAPF	366	-130	-168	151	1	-1	-1	125
UANT	-015	314	-461	385	1	1	-1	503
UANF	524	-234	-059	392	1	-1	-1	448

C	UMLFA I	II	III	ACOVS ζ	α	β
	1.*			094		
	0.*	1.*		-016	110	
	0.*	0.*	1.*	015	-004	033

UMLFA: $\chi^2(7) = 5.60$, $p = .587$

ACOVS Model 4: $\chi^2(22) = 25.57$, $p = .270$

Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.

Table 5.21

Selected 3-Component Solutions for Experimental Understanding Subscales (Academic Group)

	UMLFA (Unrotated) I	II	III	e_j	ACOVS Model 4 ζ	α	β	e_j
USPT	897	-004	-003	005	1.*	1.*	1.*	469
USPF	098	394	039	356	1.*	-1.*	1.*	382
USNT	308	180	414	214	1.*	1.*	1.*	289
USNF	189	339	-105	239	1.*	-1.*	1.*	257
UAPT	296	113	-085	273	1.*	1.*	-1.*	275
UAPF	-025	144	-211	212	1.*	-1.*	-1.*	193
UANT	130	190	165	601	1.*	1.*	-1.*	602
UANF	-004	276	-125	293	1.*	-1.*	-1.*	288

	UMLFA			ACOVS Model 4		
C	1.*			086		
	0.*	1.*		017	039	
	0.*	0.*	0.*	025	019	023

UMLFA: $\chi^2(7) = 4.79$, $p = .686$

ACOVS Model 4: $\chi^2(22) = 29.41$, $p = .134$

Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.

(indicated by e_1 = .005 in Table 5.21), so the UMLFA solution is an improper one. Model 4 fits, however, and based on 56-item scales and $\hat{\sigma}^2 = .344$, estimated reliabilities under the model would be: $\hat{\rho}_{\zeta\zeta} = .67$, $\hat{\rho}_{\alpha\alpha} = .48$, and $\hat{\rho}_{\beta\beta} = .35$. The form effect is even smaller for the Academic group than it was for the Technical group. It is also rather highly correlated with the content and agreement components ($\hat{\rho}_{\zeta\beta} = .56$ and $\hat{\rho}_{\alpha\beta} = .63$). High scores on the form component indicate a tendency to endorse self-descriptive rather than attitude items, and these correlations for the Academic group indicate that subjects with high scores on the form component are also more likely to have high scores on agreement and Understanding content.

The form component should probably be regarded as a content rather than a "style" component, since the form component for Understanding is not correlated with the form component for Play, as indicated earlier. In any event, the form component appears to be too small to affect the behavior of the scales. For all practical purposes, the important components of the Understanding measures are content and agreement, and agreement is more important in the Technical than in the Academic group.

F-Scale

The F-scale only has four subscales, and can be reported more compactly than the data for the Play and Understanding content areas. Summary information on the covariance structure is given in Table 5.22 for both groups. The transformation matrix P used to obtain V is

$$P = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad \begin{matrix} \zeta & \text{FC} & \text{(content)} \\ \alpha & \text{FA} & \text{(agreement)} \\ \alpha\omega & \text{FAO} & \\ \omega & \text{FO} & \text{(overgeneralization)} \end{matrix}$$

The components corresponding to the rows (or columns) of P are given at the right of the matrix. Tests on dimensionality are not reported in Table 5.22, but the roots-greater-than-one test and the three likelihood-ratio tests reported for the Play and Understanding subscales all indicate a dimensionality of two, for both groups. With only four subscales, UMLFA solutions cannot be obtained for more than one factor, but the 1-factor UMLFA solutions are clearly rejected ($\chi^2(2) = 18.45$, $p < .001$, for the Technical group, and $\chi^2(2) = 13.13$, $p = .001$, for the Academic group). For both groups, v_{22} corresponding to agreement is the largest element of V, and v_{11} corresponding to content is the next largest.

The results for a range of ACOVS solutions are given in Table 5.23. Three-component ζ,α,ω solutions were obtained for both groups but were not satisfactory (models 1-4). Two-component ζ,α solutions were obtained, and yielded a satisfactory fit for the Academic group (models 5-8). Finally, based on the results

Table 5.22

Summary Information on F-Scale

		Covariance Matrix S				V = P' S P					Rt(S)	Rt(R)
Technical Group												
1	FSAT	9262				FC	7222				15285	19511
2	FSAF	-3574	6908			FA	1500	14225			7176	10500
3	FSRT	4451	-2460	6283		FAO	-0781	-0262	3363		3214	3950
4	FSRF	-0867	2735	-0102	6068	FO	1331	2691	-0216	3712	2846	3400
Academic Group												
1	FSAT	5820				FC	7206				9461	20892
2	FSAF	-0656	3650			FA	2017	7487			5405	11508
3	FSRT	3485	-0285	5939		FAO	-0761	-0099	2473		2568	5734
4	FSRF	0198	1335	0462	4339	FO	0384	0357	-0133	2580	2314	3663

Note: All coefficients are reported to 4 decimal places (decimal omitted).

Table 5.23

Results for ACOVS Solutions (F-scale)

Model	Parameter Type B	Parameter Type C	Parameter Type E	df	Tech. Gp. χ^2	Tech. Gp. p	Acad. Gp. χ^2	Acad. Gp. p
1	3	1	1	6	23.55	<.001	12.64	.049
2	"	"	2	3	19.47	<.001	11.45	.010
3	"	2	1	3	2.49	.477	3.57	.312
4	"	"	2	0	-	-	-	-
5	2	1	1	7	23.79	.001	12.69	.080
6	"	"	2	4	19.48	<.001	11.56	.021
7	"	2	1	6	21.66	.001	4.77	.574
8	"	"	2	3	17.49	<.001	3.53	.317
5A	2A	1	1	6	12.43	.053	9.30	.157
6A	"	"	2	3	3.72	.294	3.83	.280
7A	"	2	1	5	12.40	.030	3.76	.585
8A	"	"	2	2	2.94	.230	1.89	.642

Note: Models with circled probabilities are reported in detail elsewhere. Parameter types are:

$$B_3' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

$$B_2' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}, \qquad B_{2A}' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & \underline{b}_{42} \end{pmatrix}$$

C_1 = diagonal , C_2 = general symmetric

$E_1 = \sigma^2 I$, E_2 = diagonal

for the Technical group, four additional two-component models with a free $\underline{b}_{42}$ element were obtained, and yielded a satisfactory fit for the Technical group as well (models 5A-8A).

Details of selected ACOVS solutions are given in Table 5.24. According to Table 5.23, the 3-component model 3 yielded an adequate fit for both groups, but the obtained solutions in Table 5.24 are unsatisfactory. For the Technical group, the covariance components in C are non-Gramian (the correlations with the overgeneralization component are estimated as $\hat{\rho}_{\zeta\omega} = 1.12$ and $\hat{\rho}_{\alpha\omega} = 1.35$, which is a clear sign of trouble with the model), and $\hat{\sigma}^2_{\omega}$ is half the size of its standard error. For the Academic group, C is Gramian, but $\hat{\sigma}^2_{\omega}$ is one-fourth the size of its standard error.

The results for the Academic group (models 5 and 7) can be examined first in Table 5.24. The parameter estimates for model 5 imply (by the multiplication B C B' + E)

$$\hat{\Sigma} = \begin{bmatrix} 494 & & & \\ 0 & 494 & & \\ 241 & 0 & 494 & \\ 0 & 241 & 0 & 494 \end{bmatrix} .$$

Significant improvement is obtained using correlated components in model 7 ($\chi^2(1) = 7.92$, $\underline{p} = .005$), which yields the estimate

$$\hat{\Sigma} = \begin{bmatrix} 594 & & & \\ -007 & 394 & & \\ 341 & -007 & 594 & \\ -007 & 141 & -007 & 394 \end{bmatrix} .$$

If the two estimates of $\hat{\Sigma}$ are compared with the observed S

Table 5.24

Selected ACOVS Solutions for F-scale

	ζ	α	ω	e_j	ζ	α	e_j	ζ	α	e_j
Tech. Gp.	Model 3				Model 5A			Model 6A		
FSAT	1.*	1.*	1.*	336	1.*	1.*	334	1.*	1.*	438
FSAF	1.*	-1.*	-1.*	336*	1.*	-1.*	334*	1.*	-1.*	179
FSRT	1.*	1.*	-1.*	336*	1.*	1.*	334*	1.*	1.*	211
FSRF	1.*	-1.*	1.*	336*	1.*	-495	334*	1.*	-380	430
C	097				091			101		
	037	272								
	033	067	009		0.*	356		0.*	376	
	$\chi^2(3) = 2.49$				$\chi^2(6) = 12.43$			$\chi^2(3) = 3.72$		
	$p = .477$				$p = .053$			$p = .294$		
Acad. Gp.	Model 3				Model 5			Model 7		
FSAT	1.*	1.*	1.*	247	1.*	1.*	253	1.*	1.*	253
FSAF	1.*	-1.*	-1.*	247*	1.*	-1.*	253*	1.*	-1.*	253*
FSRF	1.*	1.*	-1.*	247*	1.*	1.*	253*	1.*	1.*	253*
FSRF	1.*	-1.*	1.*	247*	1.*	-1.*	253*	1.*	-1.*	253*
C	118				117			117		
	050	125								
	010	009	003		0.*	124		050	124	
	$\chi^2(3) = 3.57$				$\chi^2(7) = 12.69$			$\chi^2(6) = 4.77$		
	$p = .312$				$p = .080$			$p = .574$		

for the Academic group, in the lower panel of Table 5.22, it can be seen that correlated components maintain the near-zero *true-false* correlations, and improve the fit by allowing the *true* measures to have larger variances than the *false* measures. Comparing the observed S for the Academic group with the observed S for the Technical group in Table 5.23, it is possible to identify some of the features of the Technical group data which make it difficult to fit a simply-patterned model. The FSAT subscale has a noticeably larger variance than the other three. Also, two of the *true-false* covariances (those with the FSRF subscale) are near zero, while the other two are noticeably nonzero. An examination of the first derivatives obtained with models 5-8 for the Technical group indicated that the fit of the models was stressed most by the fixed b_{42} element. Accordingly, models 5A-8A were obtained with the b_{42} element unconstrained. Models 5A and 6A are reported in Table 5.24. The estimates of b_{42} preserve the sign pattern of B for models 5A and 6A, but improve the fit by allowing test 4 to be roughly orthogonal to the two *true* tests 1 and 3. Model 6A results in significant improvement over 5A ($\chi^2(3) = 8.71$, $p < .05$) by allowing heterogeneous error.

Fitting a hypothesized solution and then freeing one element calculated to improve the fit ought to leave one feeling uneasy. In Chapter 2, I emphasized the view that the proper aim of covariance structure analysis is not the finding of a single model which "fits best" in some sense, but rather the identification of features of the models which enable them to fit reasonably well. For the Academic group, we can fit a 2-component model with uncorrelated components; correlated components do and heterogeneous error variances do not improve the fit obtained with model 5. For the Technical group, it is clear that the major sources of variance of the subscales are content and agreement, but the data are too irregular to fit a simply-patterned model. Acceptable fits can be achieved only by relaxing one or more elements of B , particularly $\underline{b}_{42}$ or $\underline{b}_{12}$.

Models 5A and 6A for the Technical group have an oddly asymmetric appearance. When $\underline{b}_{42}$ is freed, the element with the largest first derivative is the $\underline{b}_{12}$ element, and a further series of models with the $\underline{b}_{12}$ and $\underline{b}_{42}$ elements both free were obtained, and lend themselves to a simple interpretation. The results for model 5B for both groups are reported in Table 5.25 and Figure 5.6. We find the signs of the simply-patterned B matrix preserved, but for both groups the subscales are lined up along the agreement axis in the order (from high positive to high negative): FSAT, FSRT, FSRF, FSAF. Although there was no evidence for a separate overgeneralization component, the absolute-relative distinction can be seen in the size of the loadings of the scales on the agreement axis. The absolutely-worded tests provide a stronger measure of agreement than the relatively-worded tests, if model 5B is accepted for the data. The finding of

Table 5.25

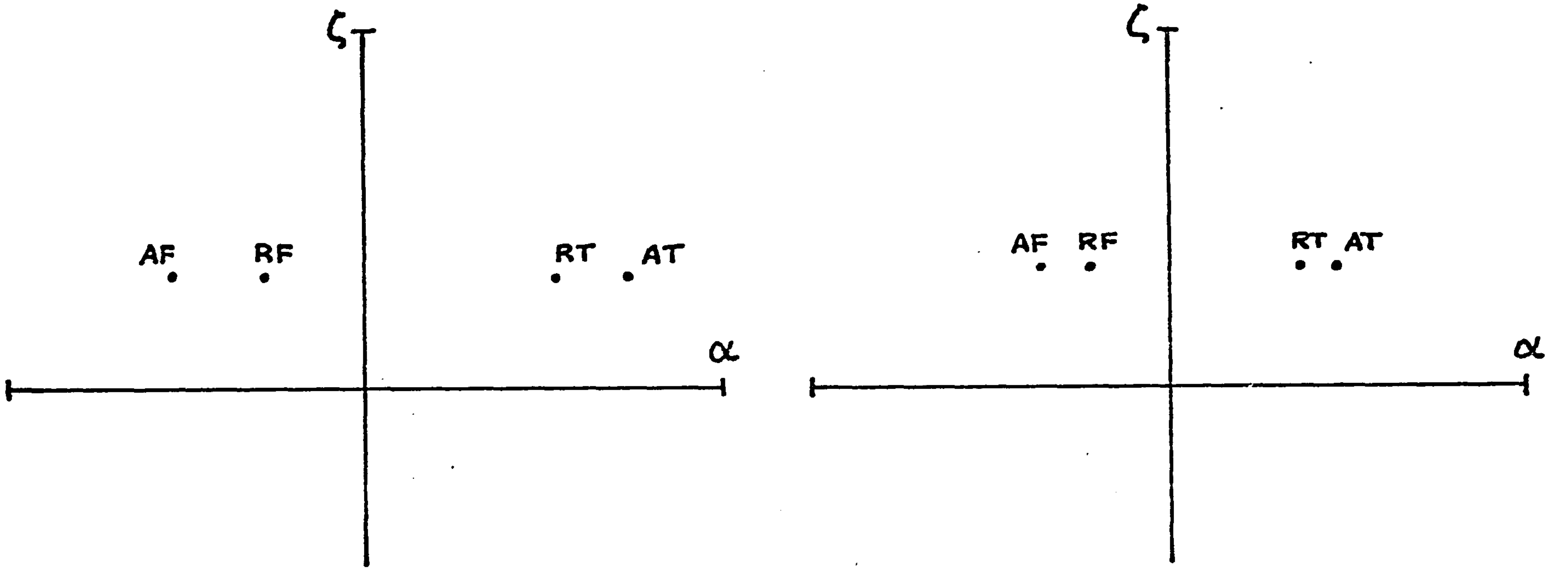
Final ACOVS Solutions (Model 5B) for F-Scale

	Technical Group ζ	Technical Group α	Technical Group $\underline{e}_j$	Academic Group ζ	Academic Group α	Academic Group $\underline{e}_j$
FSAT	1.*	1363	323	1.*	1262	251
FSAF	1.*	-1.*	323*	1.*	-1.*	251*
FSRT	1.*	1.*	323*	1.*	1.*	251*
FSRF	1.*	-508	323*	1.*	-613	251*
C	093			107		
	0.*	293		0.*	132	

Technical Group: $\chi^2(5) = 7.16$, $\underline{p} = .209$

Academic Group: $\chi^2(5) = 7.82$, $\underline{p} = .166$

Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.



(a) ACOVS Model 5B, Technical Group

$\chi^2(5) = 7.16$, $\underline{p} = .209$

(b) ACOVS Model 5B, Academic Group

$\chi^2(5) = 7.82$, $\underline{p} = .166$

Figure 5.6. Final F-Scale Solutions for both groups.

larger agreement components for the absolutely-worded scales is plausible: If a subject responds true both to absolute-true items and absolute-false reversals, it is more of a contradiction than if he responds true both to relative-true and relative-false items.

Using the estimates in Table 5.25, we find larger agreement than content components for both groups. $\hat{\sigma}_{\zeta}^2$ is about the same size in both groups, but $\hat{\sigma}_{\alpha}^2$ is larger in the Technical than in the Academic group. Estimated reliabilities for 48-item scales would be:

Tech.: $\hat{\rho}_{\zeta\zeta} = .54$, $\hat{\rho}_{\alpha\alpha} = .78$, with $\hat{\sigma}^2 = .32$;

Acad.: $\hat{\rho}_{\zeta\zeta} = .63$, $\hat{\rho}_{\alpha\alpha} = .68$, with $\hat{\sigma}^2 = .25$.

Thus, with both groups, this set of F-scale reversals provides a more reliable measure of agreement than of content. These reliabilities are based on the formula we have used all along (e.g., $\hat{\rho}_{\zeta\zeta} = 4\hat{\sigma}_{\zeta}^2 / (4\hat{\sigma}_{\zeta}^2 + \hat{\sigma}^2))$, but in the present case they are, strictly speaking, estimated reliabilities based on 48-item scales of the FSAF and FSRT type only. Because model 5B deviates from the simply-patterned class of models, there is an ambiguity about the size of variance components and therefore about reliability estimates under the model. If we multiply the symbolic parameter matrices under model 5B, we find:

$$\hat{\Sigma} = \begin{pmatrix} 1 & \beta_1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -\beta_2 \end{pmatrix} \begin{pmatrix} \sigma_\zeta^2 & \\ 0 & \sigma_\alpha^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ \beta_1 & -1 & 1 & -\beta_2 \end{pmatrix} + \sigma^2 I$$

$$= \begin{pmatrix} \sigma_\zeta^2+\beta_1^2\sigma_\alpha^2 & & & \\ \sigma_\zeta^2-\beta_1\sigma_\alpha^2 & \sigma_\zeta^2+\sigma_\alpha^2 & & \\ \sigma_\zeta^2+\beta_1\sigma_\alpha^2 & \sigma_\zeta^2-\sigma_\alpha^2 & \sigma_\zeta^2+\sigma_\alpha^2 & \\ \sigma_\zeta^2+\beta_1\beta_2\sigma_\alpha^2 & \sigma_\zeta^2+\beta_2\sigma_\alpha^2 & \sigma_\zeta^2-\beta_2\sigma_\alpha^2 & \sigma_\zeta^2+\beta_2^2\sigma_\alpha^2 \end{pmatrix} + \sigma^2 I .$$

The free parameters β_1 and β_2 resemble scale factors, but apply only to the acquiescence components. (Models with scale factors applying to both the content and agreement components cannot be used with only four measures, because they result in an underidentified model.) As interpreted in Chapter 2, under the simply-patterned class of models, scale factors define a metric in which the component-space of the model evidences a kind of homogeneity. If we had another facet in the design on the F-scale measures, it might have been possible to fit a simply-patterned model. As the data stand, however, the β_1 and β_2 parameters are difficult to separate from the estimates of the components. There is no reason why, for example, we cannot regard the "true" agreement component as the component $\beta_1^2\sigma^2$ for the FSAT subscale, or $\beta_2^2\sigma^2$ for the FSRF subscale. For the Technical group, these alternate definitions yield $\hat{\sigma}_\alpha^2 = .293$, $\hat{\beta}_1^2\hat{\sigma}_\alpha^2 = .544$, and $\hat{\beta}_2^2\hat{\sigma}_\alpha^2 = .076$ --and these are not trivially different estimates.

A possible interpretation of the β_1 and β_2 parameters is that agreement is a bigger component of the FSAT subscale than of the FSAF and FSRT subscales and, in turn, the latter have larger agreement components than the FSRF subscale. This interpretation is difficult to justify in terms of the scale construction process, however: If the attempt to write item reversals has failed to yield homogeneous agreement components for the measures, then it may also have failed by reason of introducing extraneous content components.

Final remarks--ACOVS models. The finding that a simply-patterned ACOVS model fits the data provides partial evidence for the hypothesis that there are measurable variance components associated with the design on the measures, but it is necessarily incomplete evidence. Other sorts of evidence--convergent and discriminant validity for the design components, nonzero correlations with variables linked to the components by theory, and so on--are required for an interpretation of the components. On the other hand, failure to find a simply-patterned model for the data does not invalidate an interpretation of the components provided these other sorts of evidence are available. Failure to find a simply-patterned model does, however, make it difficult or impossible to derive further results from final solution, such as reliability estimates for the components. It is convenient to have a simply-patterned theory, but it may not always be possible to obtain simply-patterned data! Even though simply-patterned models do not fit well, however, it will usually be possible to use the component estimates to get a rough idea of the size of various effects in the data.

Canonical Analysis

So far, we have found evidence for the convergent and discriminant validity of the *a priori* agreement measures, and evidence for the presence of an agreement component for each of the three experimental content areas. There is no evidence for any of the other hypothesized response style components--endorsement, form or overgeneralization--except for a small form effect with the Understanding scales. There is also evidence for group differences in means, variances and regression slopes on a number of our variables. The Technical group has lower Vocabulary and Speed means, and higher agreement scores, and also tends to be more variable on the experimental measures. The increased variability on the experimental measures appears to be due, at least in part, to this group's larger variance components for agreement. There is also evidence that the regression of agreement on Speed is different for the two groups.

In this section, we will further examine the relationships between the experimental measures and outside variables. Although we could not always fit simply-patterned models to the experimental subscales, in the spirit of robust regression we will continue to use the equally-weighted *a priori* estimates of content and agreement introduced in the early part of the chapter. We will also continue to look at relationships for each group separately. Specifically, we will look at the regression of the PC, UC, FC, PA, UA and FA measures (set 1) on the SPD, VOC, DES, IMP, COG, INF and AGE measures (set 2). SEX and the measures of the other hypothesized response style components contribute little to the multivariate relationship between the sets, and are omitted from the analysis. The analysis of the relationships between the set 1 and 2 measures can take the form of a multivariate multiple regression or, equivalently, a canonical correlation analysis.

Table 5.26 reports the correlations among the 11 measures discussed in this section, separately by group. It is of interest to pull out the correlations among the three agreement measures, and for the Technical group they are:

UA	.725***	-
FA	.787***	.700***
	PA	UA

These are slightly higher than the correlations reported earlier for both groups combined. For the Academic group, the correlations are sharply diminished, but still decidedly nonzero:

UA	.478***	-
FA	.482***	.341***
	PA	UA

These correlations still exceed any of the other correlations involving the agreement measures for the Academic group. Using rough estimates of reliability proposed in the ACOVS section of this chapter, we can write the deattenuated correlations among the agreement measures as:

Table 5.26

Within-Group Correlations

	PA	UA	FA	PC	UC	FC	SPD	VOC	AGE	DES	IMP	COG	INF
PA	-	478***	482***	-013	-086	210*	208*	-048	064	-210	-039	153	023
UA	725***	-	341***	-207*	238*	032	133	077	047	018	012	080	-083
FA	787***	700***	-	-145	-046	275**	-012	-192	-015	-120	-089	188	055
PC	-229*	-256*	-291**	-	095	-113	115	079	-023	175	277**	-402**	-043
UC	-131	-070	-221*	080	-	-227*	-084	213*	144	273**	010	010	-012
FC	120	088	148	257*	029	-	-020	-257**	116	101	-094	283**	-082
SPD	-280**	-330***	-372***	354***	201*	-063	-	170	-064	-134	139	-133	203*
VOC	-326***	-225*	-412***	196	214*	039	411***	-	044	034	062	-053	-063
AGE	430***	538***	374***	-485***	143	-091	-360***	-158	-	-013	-005	145	-075
DES	-321**	-221*	-306**	189	407***	061	192	366***	-022	-	-235*	102	-235*
IMP	-142	-190	-218*	355***	-068	-056	242*	210*	-297**	-159	-	-513***	-004
COG	191	157	163	-202*	270**	196	167	-036	239*	155	-553***	-	-113
INF	333***	148	407***	-123	-386**	-075	-311**	302**	216*	-308**	014	-084	-

*$p < .05$

**$p < .01$

***$p < .001$

Note: Academic group (N = 102) above diagonal, and Technical group (N = 97) below diagonal.

PA	(75)			(57)		
UA	101	(69)		91	(48)	
FA	103	101	(78)	77	60	(68)
	PA	UA	FA	PA	UA	FA
	Tech. Group			Acad. Group		

(estimated component reliabilities are in the diagonal; decimals omitted). Thus the differences in the size of the agreement correlations for the two groups can be largely accounted for by the differences in component reliabilities. For the Technical group, the correlations are as high as they can be, given the estimated component reliabilities. For the Academic group, however, the deattenuated correlations are noticeably less than one, suggesting that the estimated component reliabilities are too high (especially for the F-scale) or that there are unmeasured sources of error which are attenuating the agreement correlations for this group.

Table 5.27 presents the results of a canonical analysis of the relationships between set 1 and set 2, for both groups. The upper panel shows the standard Bartlett (1948) tests for dimensionality, and we find a dimensionality of two for the Technical group and one for the Academic group. By Roy's (1939) greatest characteristic root criterion, the first two roots are significant for the Technical group and the first root is significant for the Academic group ($p < .05$). The lower panel of Table 5.27 presents the "canonical component" structure loadings, which can be characterized most simply as the correlations of the measures

Table 5.27

Canonical Analysis: Agreement and Content Predicted from Marker Variables

	Technical Group			Academic Group		
	I	II	$\underline{V}(z_j \mid \hat{X})$	I	II	$\underline{V}(z_j \mid \hat{X})$
Set 1: Agreement and Content Variables						
PA	63	-32	50	38	-16	35
UA	73	-04	54	02	05	00
FA	60	-47	58	41	05	17
PC	-72	05	53	-69	-17	50
UC	02	77	59	-36	48	36
FC	-05	10	01	30	51	35
Set 2: Marker Variables						
SPD	-52	27	34	-06	-31	10
VOC	-33	47	33	-32	-10	11
AGE	84	17	73	02	36	13
DES	-29	58	42	-47	59	57
IMP	-44	-06	20	-41	-21	21
COG	35	28	20	63	51	66
INF	19	72	56	13	-13	04
Variance Accounted For						
$\underline{V}(Z_1 \mid \hat{X})$	30	16	46	17	09	26
$\underline{V}(Z_2 \mid \hat{X})$	22	18	40	13	13	26
$\underline{V}(Z \mid \hat{X})$	26	17	42	15	11	26

Bartlett Decomposition

$\underline{df}$	$\underline{r}_c$	χ^2	$\underline{p}$	$\underline{r}_c$	χ^2	$\underline{p}$
42	74	152.1	<001	60	84.3	<001
30	66	80.6	<001	45	42.4	070
20	38	29.0	092	34	21.3	383
12	32	14.9	248	29	10.1	612
6	21	5.1	529	12	1.6	954
2	11	1.0	612	04	.2	978

Note: Leading decimals omitted. Upper panels contain canonical components (Jordan, 1975). Variance accounted for is reported using the generalized statistic $\underline{V}(A \mid B) = tr(A'B(B'B)^{-1}B'A)/tr(A'A)$, with A and B in standard-score form.

with the principal components of canonical variate space (Jordan, 1975). Using the notation of my 1975 paper, the canonical components may be obtained as

$$\hat{S} = \tfrac{1}{2} R \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} [\tfrac{1}{2}(I + P)]^{-\frac{1}{2}} ,$$

where R contains the correlations among the measures, C_1 and C_2 contain the set 1 and set 2 canonical weights, and P is a diagonal matrix containing the canonical correlations in descending order. The canonical components are orthogonal and have unit length, and as in factor analysis or principal component analysis, the sums of squares for rows of S are squared multiple correlations for predicting the measures from the canonical components, and the mean sums of squares for columns may be interpreted as the (average) variance of the measures accounted for by the components.

Although only one canonical component is significant for the Academic group, two components are reported in Table 5.27 for both groups, as an aid to plotting the canonical components in Figure 5.7. In Figure 5.7, the set 1 (experimental) variables are shown by x's , and the set 2 (marker) variables by •'s . A dashed line indicates the vector through the centroid of the three agreement measures. For the Technical group, we find that the first component is marked at the upper end by the three agreement measures and AGE, and at the lower end by IMP and PC. The

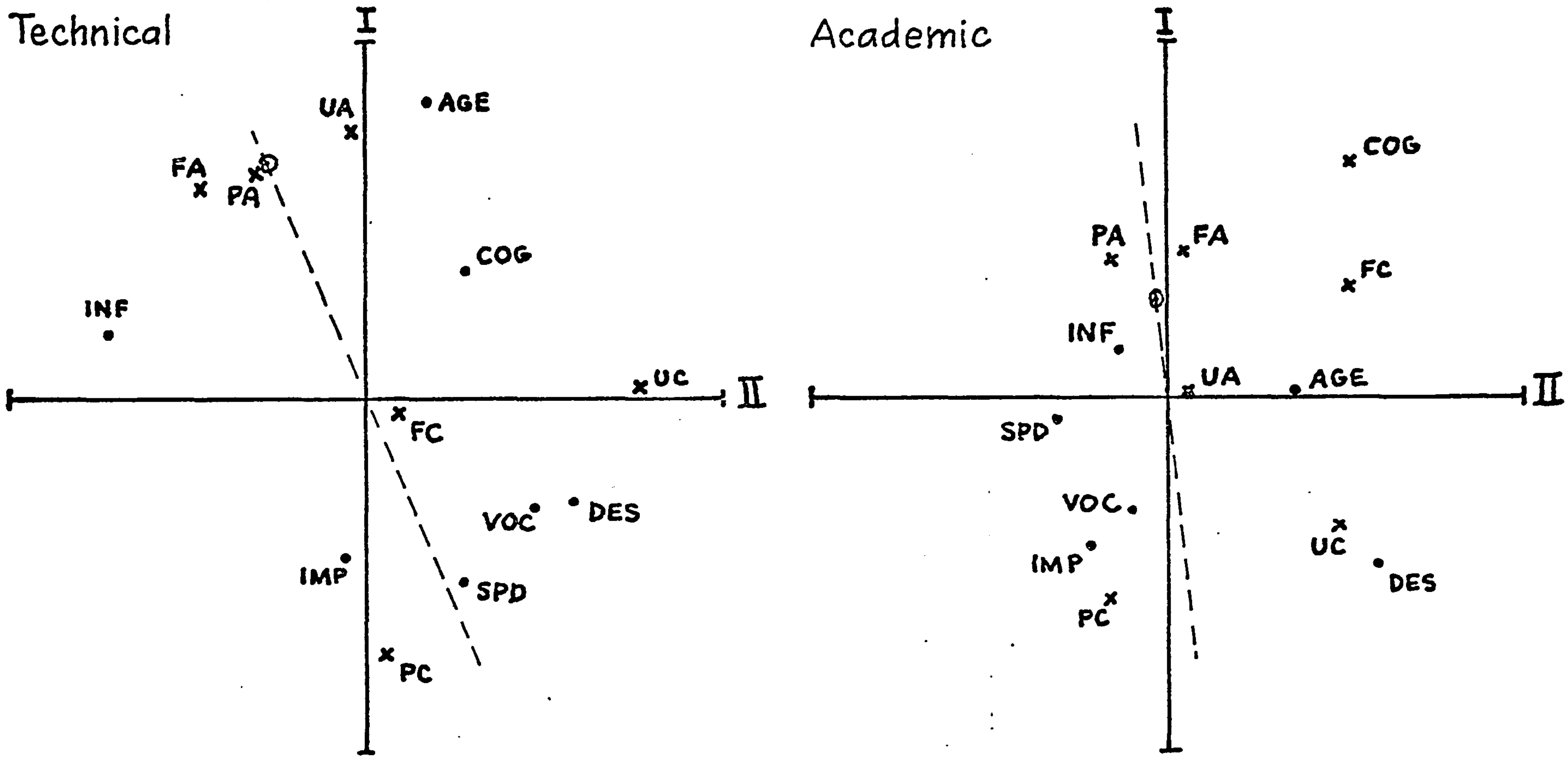


Figure 5.7. Canonical components: Agreement and content (×) versus marker variables (•). Centroid vector for agreement measures is indicated by a dashed line.

second component is marked by DES and UC at the positive end, and INF at the negative end. The SPD and VOC measures lie midway between the Impulsivity-Play-"Youth" and the Desirability-Understanding-"Low Infrequency" axes of canonical component-space. The relationship between agreement and the other measures can be broadly interpreted by noting the projections of the measures on the agreement centroid in Figure 5.7: In the Technical group, subjects with high agreement scores tend to be older and slower, have lower verbal ability, make more infrequent and fewer desirable responses, and have a lower need for Impulsivity, Play and Understanding.

When we turn to the results for the Academic group, given in the right halves of Table 5.27 and Figure 5.7, we find that the PA and FA measures lie on the single significant axis, but UA is not well represented in component space. Running down the column of loadings for the first component in Table 5.27, the measures with "large" (above .30) loadings are PC, COG, DES, FA, IMP, PA, UC and VOC (in descending order). If we compare the right and left halves of Figure 5.7 for similarity, we can see three clusters of measures having the same orientation for both groups: There is a PA-FA cluster, a PC-IMP-COG cluster (with COG reflected), and a UC-DES cluster. Variables other than the six I have mentioned do not replicate well across groups. Mentally rotate both solutions to the agreement centroid, however, and you can "see" a common factor defined by agreement at one end nearly uncorrelated PC-IMP-COG and UC-DES clusters at the other. Such mental rotation takes some liberties with the data, especially since the configuration I have described is one involving two dimensions, and we are only entitled to one in the case of the Academic group. The three clusters can be defended on a priori grounds. Certainly the agreement measures should form a cluster, and the main interpretive problem for this cluster would be explaining why UA misbehaves for the Academic group. The relationship between Impulsivity and Cognitive Structure was expected, and it also makes sense that these variables would be associated with Play content. (In the normative sample for the published form of these PRF scales, Play has its highest correlations with the Impulsivity scale; the latter has its highest correlations with the Cognitive Structure scale--cf. Jackson, 1974, p. 30). Also, an examination of the Understanding items will show that many of them have a fairly obvious desirability connotation. Correlations between Understanding and Desirability for the published PRF are not high (.29 for males and .25 for females--cf. Jackson, op. cit.), but the published scales were based on items systematically chosen to minimize the correlations of content scales with Desirability; perhaps the item-rewriting process used to construct the experimental scales for this study has allowed an association with Desirability to creep back into the Understanding subscales.

The main purpose of the canonical analysis is that of displaying the relationship of the agreement measures with outside variables. Since the experimental content measures have been included in set 1 along with the agreement measures, Play, Understanding and F-scale content are affecting the canonical relationship shown in Figure 5.7. The point of identifying the three clusters of variables in Figure 5.7 is not to defend an interpretation of their interrelationships, but to draw attention to role of content in determining the canonical structure.

A second canonical analysis is displayed in Table 5.28 and Figure 5.8, and puts the three agreement measures together as set 1, moving the experimental content measures into set 2. Again we find two dimensions for the Technical group and one for the Academic group. The leading canonical correlations in Table 5.28 are higher than the ones observed in Table 5.27, especially for the Academic group, indicating that there is some independent relationship between the experimental content and agreement measures which is now contributing to the canonical relationship between the sets. For the Technical group, in Figure 5.8, the new solution resembles the first one, except that the PC-IMP-COG and UC-DES clusters have been collapsed. The second dimension of the new solution is determined largely by an association between FA and INF, and between UA and AGE. For the Academic group, moving the experimental content measures to set 2 has radically altered the solution. The agreement centroid, happily marked by all three agreement measures, is unhappily aligned with the non-significant second component. The loadings on first component are positive for PA, FA, PC and FC, and negative for UA and UC. The DES and VOC measures also have negative loadings on the first component. One interpretation of this first component for the Academic group is that it is primarily

Table 5.28

Canonical Analysis: Agreement Predicted from Marker and Content Variables

	Technical Group			Academic Group		
	I	II	$\underline{V}(z_j \| \hat{X})$	I	II	$\underline{V}(z_j \| \hat{X})$
Set 1: Agreement Variables						
PA	83	05	70	46	62	60
UA	75	47	79	-40	72	68
FA	87	-19	80	33	58	45
Set 2: Marker Variables						
SPD	-50	-02	26	07	30	10
VOC	-53	30	37	-29	-08	09
AGE	63	49	64	01	09	01
DES	-44	10	21	-38	-20	19
IMP	-28	-01	08	-12	-08	02
COG	26	04	07	19	36	16
INF	51	-46	47	19	-06	04
PC	-40	-01	16	25	-40	22
UC	-25	27	14	-53	23	33
FC	19	-09	05	38	39	30
Variance Accounted For						
$\underline{V}(Z_1 \| \hat{X})$	67	09	76	16	42	58
$\underline{V}(Z_2 \| \hat{X})$	18	06	24	08	07	15
$\underline{V}(Z \| \hat{X})$	29	07	36	10	15	25
Bartlett Decomposition						
$\underline{df}$	$\underline{r}_c$	χ^2	$\underline{p}$	$\underline{r}_c$	χ^2	$\underline{p}$
30	659	85.36	< 001	522	53.38	007
18	522	34.71	011	377	23.42	178
8	263	6.37	607	303	9.03	341

Note: Leading decimals omitted. Upper panels contain canonical components (Jordan, 1975). Variance accounted for is reported using the generalized statistic $\underline{V}(A|B) = tr(A'B(B'B)^{-1}B'A)/tr(A'A)$, with A and B in standard-score form.

gcr ~ .23 A / MARKER + C

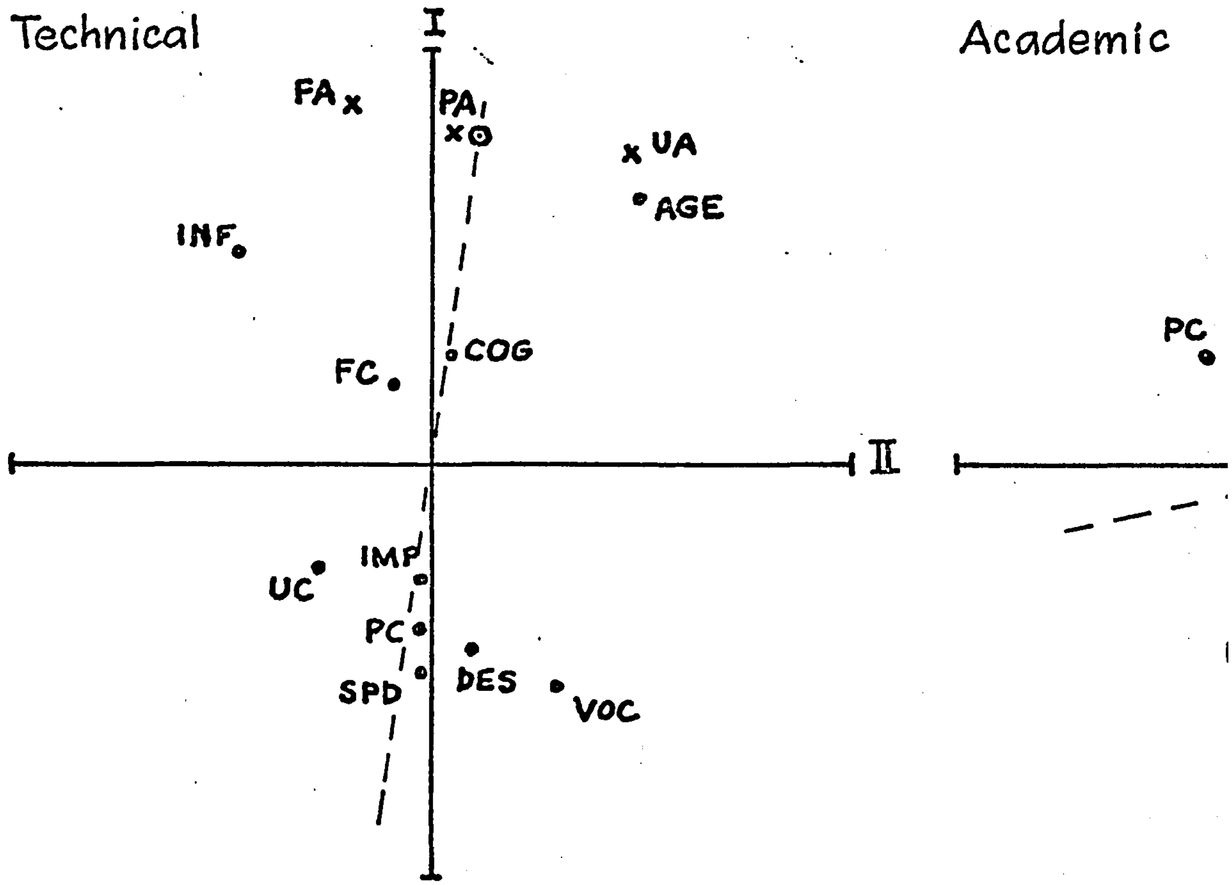
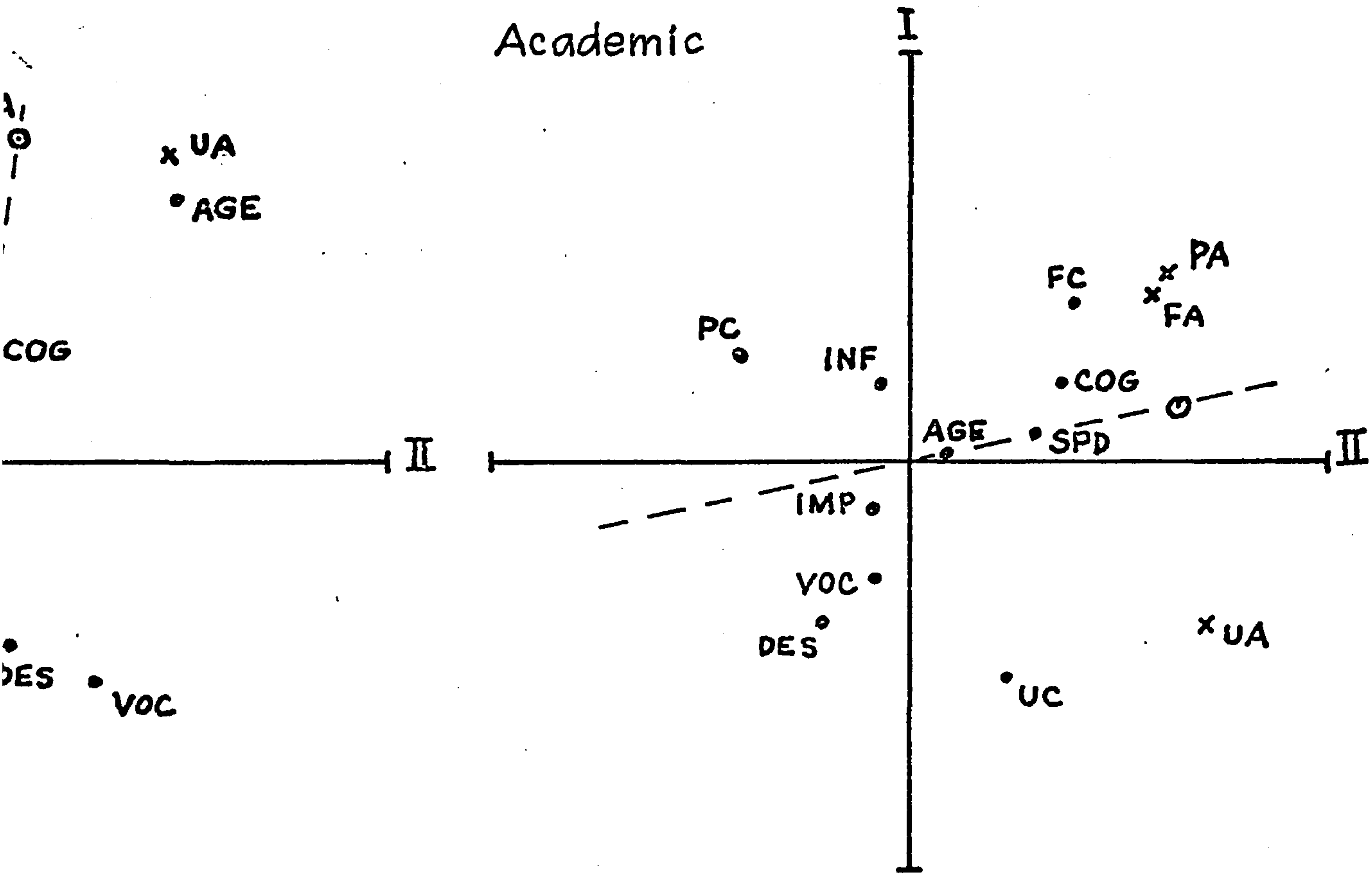


Figure 5.8. Canonical components: Agreement (x) v
iables (•). Centroid vector for agreement measures is i



anonical components: Agreement (x) versus marker and content var-
ld vector for agreement measures is indicated by dashed line.

attributable to overlap of the experimental content and experimental agreement measures--possibly mediated by Desirability.

At the very least, we would like to be able to say that a relationship between agreement and outside variables is statistically independent of content for the scales used to measure agreement. Accordingly, a third canonical analysis is reported in Table 5.29 and Figure 5.9. This is an analysis of the relationship between the three agreement measures and seven marker variables, with the three experimental content measures partialled out of the measures in both sets. Two dimensions are significant for the Technical group, but now there are no significant relationships between the two sets of measures for the Academic group. For both groups, the canonical correlations are lower than those observed for the first analysis reported in Table 5.7. For the Technical group, Figure 5.9 resembles the result in Figure 5.8: Subjects with high scores on the agreement measures tend to be older and slower, have lower verbal ability, and tend to give more infrequent and fewer desirable responses. For the Academic group, there are no significant relationships between the sets of variables. It is of interest to note the partial correlations among the agreement measures, which are:

UA	.699***		.527***	
FA	.761***	.670***	.457***	.336***
	PA	UA	PA	UA
	Tech. Group		Acad. Group	

These partial correlations are about the same order of magnitude

Table 5.29

Canonical Analysis: Agreement Predicted from Marker Variables, with Content Partialled Out

	Technical Group			Academic Group		
	I	II	$V(z_j \mid \hat{X})$	I	II	$V(z_j \mid \hat{X})$
			Set 1: Agreement Variables			
PA	82	-20	71	71	39	66
UA	81	30	75	04	61	37
FA	73	-39	69	43	-23	24
			Set 2: Marker Variables			
SPD	-36	02	13	25	62	44
VOC	-41	43	36	-12	39	17
AGE	75	29	66	09	15	03
DES	-38	15	17	-60	-03	36
IMP	-13	05	02	-19	22	08
COG	21	-11	05	38	-10	15
INF	39	-60	51	25	-29	14
			Variance Accounted For			
$V(Z_1 \mid \hat{X})$	62	09	72	23	19	42
$V(Z_2 \mid \hat{X})$	17	10	27	10	10	20
$V(Z \mid \hat{X})$	31	09	40	14	13	27
			Bartlett Decomposition			
df	r_c	χ^2	p	r_c	χ^2	p
21	582	64.73	<001	383	25.40	233
12	491	28.57	005	309	10.73	553
5	221	4.39	503	124	1.44	919

Note: Leading decimals omitted. Upper panels contain canonical components (Jordan, 1975). Variance accounted for is reported using generalized statistic $V(A \mid B) = tr(A'B(B'B)^{-1}B'A)/tr(A'A)$, with A and B in standard-score form.

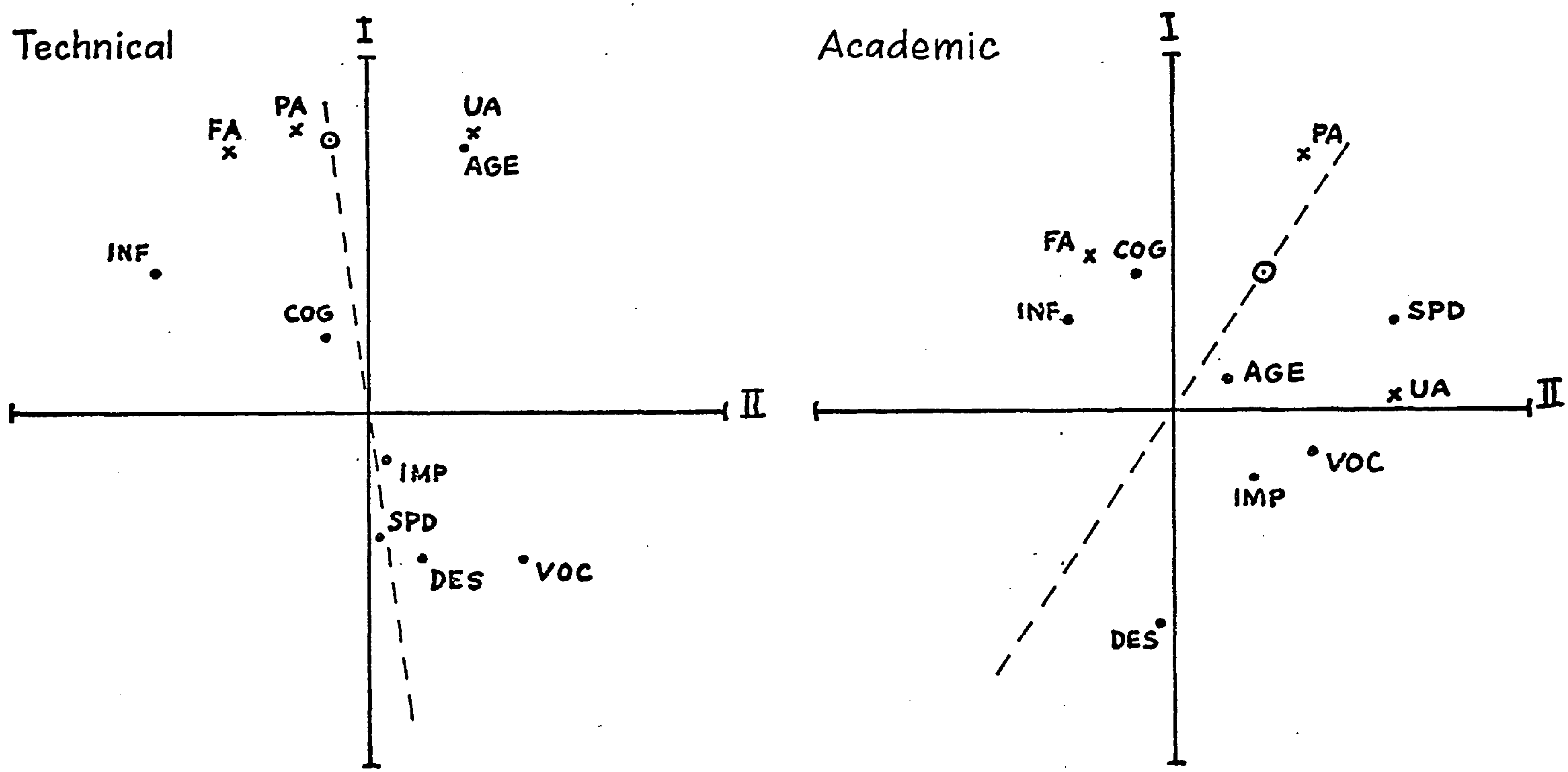


Figure 5.9. Canonical components: Agreement (×) versus marker variables (•), with content partialled out. Centroid vector for agreement measures is indicated by a dashed line.

as the zero-order correlations in Table 5.26, so it is clear that common content cannot explain the high correlations among the agreement measures.

Partialling the content measures out of both set 1 and set 2 evidently eliminates any trace of a relationship between the agreement and marker variables for the Academic group. In this case, a more appropriate analysis might be the one which Cooley and Lohnes (1971, pp. 202 ff.) call "multiple part correlation," in which the experimental content measures are partialled out of the agreement measures in set 1, but not out of the marker variables in set 2. This can be accomplished by a canonical analysis of

$$\tilde{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} - \begin{pmatrix} R_{13}R_{33}^{-1}R_{31} & R_{13}R_{33}^{-1}R_{32} \\ R_{23}R_{33}^{-1}R_{31} & 0 \end{pmatrix} ,$$

where the subscript 3 refers to the set of content variables. For computational convenience, $\tilde{R}$ may be rescaled to have unit diagonal elements, but that has no effect on the significance test of the relationship between set 1.3 and 2 (indicating the partialling by the familiar Yule dot-notation). This "part correlation" analysis was performed for the Academic group, and while it increased the leading canonical correlation slightly (from .383 to .400), the relationship between the agreement measures and the marker variables was still not significant ($\chi^2(21) = 26.91$, $p = .177$).

One further attempt was made to find a relationship between the agreement measures and the marker variables, independent of the experimental content measures, for the Academic group. An agreement composite (AGR) was created by summing the three agreement measures. Multivariate significance tests based on least-square principles tend to lose power and precision when highly correlated variables are included on one or the other side of the model (e.g., Cramer, 1975). Since the three agreement measures are highly correlated, the AGR composite should be more reliable than any of the individual measures, and more likely to show a significant relationship with the outside measures--if, in fact, there is a relationship present in the data. Predicting AGR from the three content measures yields the multiple correlation $\underline{R} = .283$ ($\underline{F}(3,98) = 2.85$, $\underline{p} = .041$). Adding the seven marker variables raises $\underline{R}$ to .388, but they do not make a significant contribution ($\underline{F}(7,91) = 1.07$, n.s.), and and the 10-variable model is not significant overall ($\underline{F}(10,91) = 1.61$, n.s.) Furthermore, none of the seven marker variables improves the prediction of AGR after the three content measures have entered. The next variable to enter would be DES, which does not quite make a significant contribution ($\underline{F}(1,97) = 3.59$, $\underline{p} = .061$). For the Academic group, then, there is apparently no detectable relationship between agreement and the marker variables, independent of the experimental content measures.

The extra analyses performed for the Academic group were also performed for the Technical group, but do not appreciably alter the conclusions already reached about the structure of relationships between agreement and the outside variables. In a stepwise regression analysis of AGR for the Technical group, with the three content measures forced to enter first, the final model is

$$\hat{AGR} = -04\ PC - 11\ UC + 21\ FC - 20\ VOC + 45\ AGE - 20\ DES\ ,$$

with the variables standardized in the subsample. All of the coefficients in the final model are significant at the .05 level or better, except PC and UC, and the multiple correlation is .642 ($F(6,96) = 10.51$, $p < .001$). Comparing this result with the canonical analysis in Figure 5.8, we find the sign pattern of the loadings on the agreement centroid preserved in the regression model. At the final step of the regression model, SPD, IMP, COG and INF also preserve the sign pattern of the canonical analysis, but do not have sufficient independent information left in them, individually or collectively, to make a significant contribution to the prediction of AGR. As a final statement for the Technical group, subjects with high agreement scores tend to be older, have lower verbal ability, and make more undesirable responses. At the zero-order level, these subjects also tend to be slower and more impulsive and to have a higher need for cognitive structure, but the measures of these traits do not have a relationship with agreement independent of experimental content, AGE, VOC and DES.

Agreement and Endorsement

One of the hypotheses advanced in Chapter 4 was that the three *a priori* measures of agreement and the two *a priori* measures of endorsement would have a 2-factor simple structure. This hypothesis is mooted, owing to our inability to find endorsement components for the Play and Understanding data. In this section, we will take a last look at the evidence for an endorsement component. Table 5.30 presents the covariances and correlations among the *a priori* measures of agreement and endorsement. For the Technical group, $r_{PE,UE} = .221$ ($p < .05$). $r_{PE,UE}$ can be considered an alternate-form reliability coefficient for the 56-item PE and UE measures of endorsement; even corrected for double-length, the observed correlation for the Technical group is not very promising. For the Academic group, $r_{PE,UE} = .337$ ($p < .001$), which is better than the value observed for the Technical group, but still not large.

It is fairly easy to fit 2-factor ACOVS models to the covariances in Table 5.30, and selected models for the Technical group are given in Table 5.31. The first model allows correlated factors and heterogeneous error variances, while the second is a more restricted model having uncorrelated factors and constraints on the error variances. (In the more restricted model, the PA and PE scales were constrained to have equal error variances, and the UA and UE scales were also constrained to have equal error variances. The rationale for this particular constraint is the *a priori* measures from the same content area can be reasonably

Table 5.30

Covariances and Correlations among Experimental Agreement and Endorsement Measures

Technical Group

	PA	UA	FA	PE	UE
PA	9820	6054	10545	-0740	-0343
UA	725***	7095	7967	-0094	-0409
FA	787***	700***	18281	0114	-0427
PE	-144	-022	016	2689	0573
UE	-069	-097	-063	221	2508

Academic Group

	PA	UA	FA	PE	UE
PA	5653	2145	3551	-0628	-0384
UA	478***	3563	1994	-0294	0150
FA	482***	341***	9618	0572	0701
PE	-159	-094	111	2760	0845
UE	-107	053	150	337***	2274

<u>Note</u>: Leading decimals omitted. Covariances above diagonals, and correlations below diagonals.

Table 5.31

2-Factor ACOVS Solutions for Technical Group

Model #7

	ζ	α	e_j
PA	898	0.*	176
UA	676	0.*	253
FA	1174	0.*	449
PE	0.*	232	215
UE	0.*	247	190
C	1.*		
	-170	1.*	

$\chi^2(4) = 7.08$

$p = .132$

Model #2

	ζ	α	e_j
PA	896	0.*	176
UA	681	0.*	234
FA	1174	0.*	449
PE	0.*	308	176*
UE	0.*	170	234*
C	1.*		
	0.*	1.*	

$\chi^2(7) = 8.37$

$p = .301$

Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.

expected to have the same error variances.) The more restricted model of Table 5.31 requires $\hat{\Sigma}$ to have the form

$$\hat{\Sigma} = \begin{pmatrix} x & & & & \\ x & x & & & \\ x & x & x & & \\ 0 & 0 & 0 & x & \\ 0 & 0 & 0 & x & x \end{pmatrix},$$

implying that the agreement-endorsement correlations are simultaneously zero. Both models fit adequately for the Technical group.

Selected models for the Academic group are given in Table 5.32. Models with 5 fixed zeros in B do not fit adequately for this group, and it is necessary to relax at least one additional element of B . Freeing the $\underline{b}_{32}$ element results in the greatest improvement in fit, and has the effect of allowing the FA measure to be more highly correlated with the endorsement measures than the PA and UA measures are. None of the agreement-endorsement correlations are significantly different from zero for the Academic group, but the ACOVS fitting function is sensitive to the differences among these correlations.

When the submatrix of <u>a priori</u> agreement and endorsement measures is examined, it is possible to find evidence that endorsement shows convergent validity. It is more difficult to find evidence that endorsement is correlated in a meaningful way with any of the other variables in the data. Table 5.33 displays the results of a canonical analysis between the marker and content measures (set 1) and the pooled agreement and endorsement measures AGR = PA + UA + FA and END = PE + UE . Table 5.33

Table 5.32

2-Factor ACOVS Solutions for Academic Group

Model #9

	ζ	α	$\underline{e}_1$
PA	648	0.*	146
UA	331	0.*	247
FA	647	372	523
PE	0.*	329	168
UE	0.*	257	161

C $\begin{pmatrix} 1.* & \\ -247 & 1.* \end{pmatrix}$

$\chi^2(3) = 2.42$

$\underline{p} = .490$

Model #4

	ζ	α	$\underline{e}_1$
PA	628	0.*	168
UA	357	0.*	203
FA	604	327	526
PE	0.*	334	168*
UE	0.*	223	203*

C $\begin{pmatrix} 1.* & \\ 0.* & 1.* \end{pmatrix}$

$\chi^2(6) = 7.43$

$\underline{p} = .283$

Note: Parameter estimates are reported to three decimal places, except where a decimal is explicitly provided. Starred parameters were fixed or constrained to be equal to other parameters.

Table 5.33

Canonical Analysis: Agreement and Endorsement Composites Predicted from Marker and Content Variables

	Technical Group			Academic Group		
	I	II	$\underline{V}(z_j \mid \hat{X})$	I	II	$\underline{V}(z_j \mid \hat{X})$
Set 1: Agreement and Endorsement Composites						
AGR	89	-19	83	70	-46	69
END	12	85	74	47	67	67
Set 2: Marker and Content Variables						
SPD	-46	29	30	24	-20	10
VOC	-49	09	25	-15	06	03
AGE	67	-08	46	13	06	02
DES	-45	-07	21	-15	40	18
IMP	-27	08	08	-01	23	05
COG	29	14	10	48	06	23
INF	42	-29	26	07	12	02
PC	-33	40	27	-17	37	17
UC	-12	63	41	26	38	21
FC	21	17	08	46	-18	25
Variance Accounted For						
$\underline{V}(Z_1 \mid \hat{X})$	40	38	78	35	33	68
$\underline{V}(Z_2 \mid \hat{X})$	16	08	24	07	06	13
$\underline{V}(Z \mid \hat{X})$	20	13	33	11	10	22
Bartlett Decomposition						
$\underline{df}$	$\underline{r}_c$	χ^2	$\underline{p}$	$\underline{r}_c$	χ^2	$\underline{p}$
20	645	74.30	<001	423	27.12	135
9	468	22.14	009	293	8.48	512

Note: Leading decimals omitted. Upper panels contain canonical components (Jordan, 1975). Variance accounted for is reported using generalized statistic $\underline{V}(A \mid B) = tr(A'B(B'B)^{-1}B'A)/tr(A'A)$, with A and B in standard-score form.

is in the format used earlier for reporting canonical analyses, and the canonical components are plotted in Figure 5.10. For the Technical group, there are two significant dimensions for the canonical relationship. The first canonical component pulls out the AGR composite, and the second pulls out the END composite. In the last section, we saw that the AGE, VOC and DES measures enter a stepwise multiple regression for predicting AGR, after content is partialled out. None of the marker variables has a significant zero-order relationship with the END composite, however, and only the UC measure enters the equation for predicting END. For the Academic group, there are no significant dimensions for the canonical relationship, and neither AGR nor END has a significant zero-order relationship with any of the marker variables. Only the UC measure comes close to having a significant relationship with END for the Academic group ($r_{UC,END} = .186$, $p = .062$). Given the low correlations between the alternate measures of endorsement, PE and UE, it is not surprising that we are unable to find relationships between the END composite and other variables.

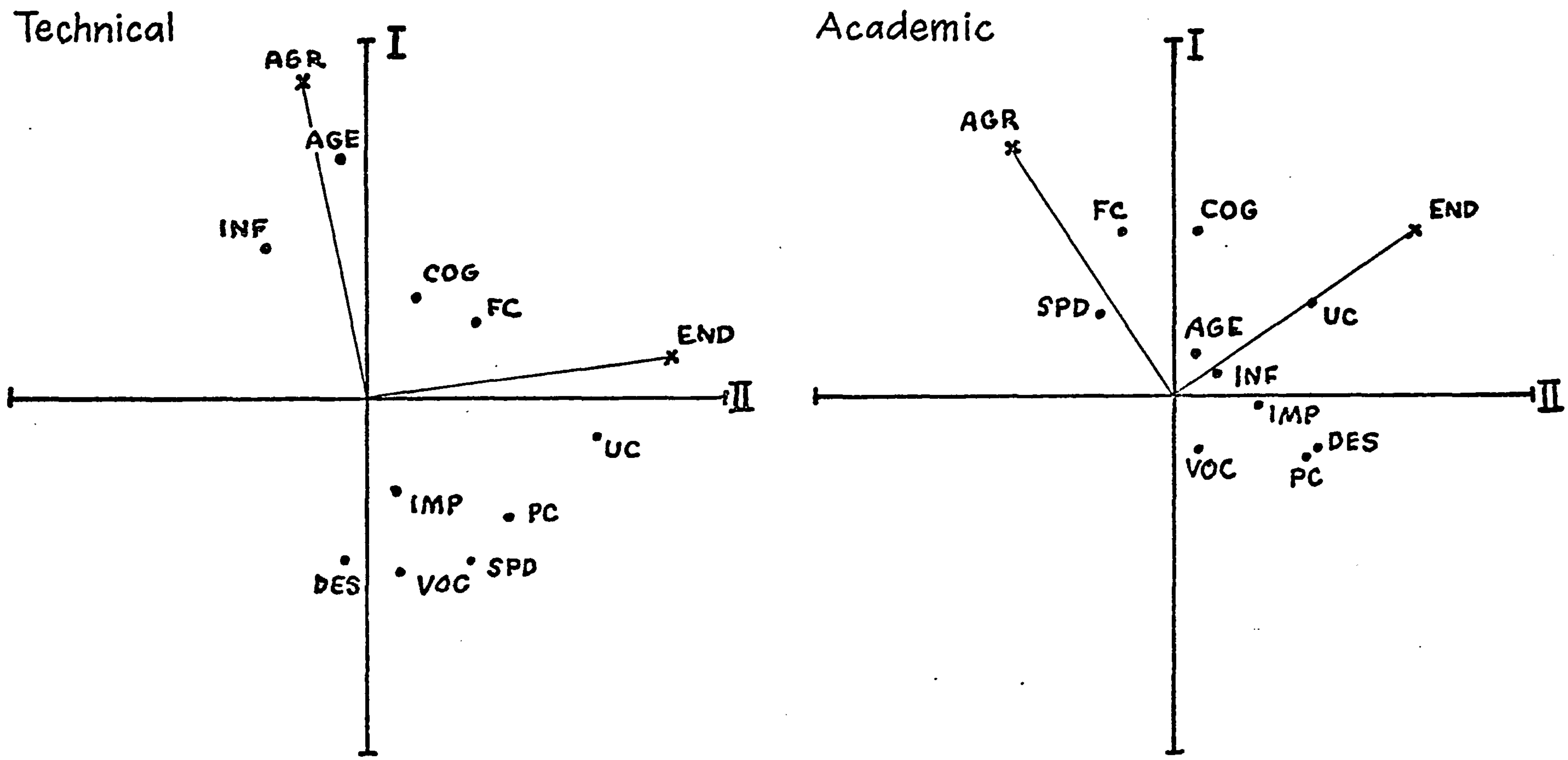


Figure 5.10. Canonical components: Agreement and endorsement composite variables (x) versus marker and experimental content variables (•).

Chapter 6

Summary and Conclusion

By way of summary, I will indicate what has been discovered for each set of hypotheses proposed in Chapter 4. The most interesting unanticipated finding--and the profoundest complication for the analysis--was the finding of strong group differences in the sample. The results for each of the hypotheses must be expressed differently for the two groups.

1. Style components. Of the hypothesized style components (agreement, endorsement, form and overgeneralization), only agreement, the original "acquiescence," is present in the data in any strength. For each of the experimental content areas, both groups have content components of similar size, but the Technical group has an agreement component on the order of twice the size of the agreement component for the Academic group.

Table 6.1 provides a summary for two kinds of component estimates available from the data. The upper panel reports the estimates obtained from the transformed matrix $V = P' S P$ under a simple 2-component content and agreement model which assumes homogeneous error and no scaling factors. Under these assumptions, reliability estimates for measures of the components based on all items in a particular content area can be obtained by a simple application of Spearman-Brown principles, and these reliability estimates are also reported. The lower panel reports component estimates obtained from a selected ACOVS model--either the "best"-fitting simply-patterned model or, for the Technical

Table 6.1

Summary of Component Estimates for Experimental Scales

Estimates	Play Tech.	Play Acad.	Understanding Tech.	Understanding Acad.	F-scale Tech.	F-scale Acad.
Component Estimates Using V						
Source Table	5.8	5.9	5.14	5.15	5.22	5.22
$\hat{\sigma}_{\xi}^2$	17	17	10	09	09	12
$\hat{\sigma}_{\alpha}^2$	17	08	11	04	27	12
$\hat{\sigma}_{\xi\alpha}$	05	00	01	03	04	05
$\overline{\hat{\delta}^2}$	45[a]	45[a]	43	37	35	25
$\hat{\rho}_{\xi\xi}$	75	75	65	65	51	65
$\hat{\rho}_{\alpha\alpha}$	76	57	68	44	75	66
Estimates Using Selected ACOVS Model						
Source Table & Model	5.11 #1	5.12 #8	5.17 #8	5.18 #8	5.25 #5B	5.24 #7
$\hat{\sigma}_{\xi}^2$	17	18[b]	15[b]	23[b]	09[c]	12
$\hat{\sigma}_{\alpha}^2$	17	10[b]	17[b]	12[b]	29[c]	12
$\hat{\sigma}_{\xi\alpha}$	00[d]	-04[b]	03[b]	12[b]	00[d]	05
$\overline{\hat{\delta}^2}$	46	40[e]	42[e]	33[e]	32	25
ACOVS fit	YES	NO[f]	NO[g]	YES[g]	YES	YES

[a]Error estimates pooled across groups for Play subscales.

[b]A matrix used, so estimates are in metric of SPT and SPF scal

[c]Heterogeneous agreement, in metric of FSAF and FSRT scales.

[d]Constrained to equal zero under model.

[e]Average of heterogeneous error estimates under model.

[f]Element $\underline{b}_{32}$ freed to obtain fitting model (cf. Table 5.13).

[g]Simple ACOVS models can be obtained by allowing small form components (cf. Tables 5.20 and 5.21).

group F-scale data, the model allowing heterogeneous agreement components. Footnotes indicate some of the special characteristics of the selected ACOVS models, and there are cross-references to the tables in Chapter 5 where the models are reported in more detail.

With certain qualifications, the simple estimates in the upper panel of Table 6.1 agree fairly well with the more sophisticated ACOVS estimates in the lower panel. One qualification is that the covariance components $\hat{\sigma}_{\xi\alpha}$ seem less stable than the variance components $\hat{\sigma}_{\xi}^2$ and $\hat{\sigma}_{\alpha}^2$, and I assume that the ACOVS fitting function is more sensitive to variance differences than to covariance differences among the components. For the Play and F-scale data, the variance components reported in the upper and lower panels agree very well, even though scaling factors were allowed for two of the four datasets in the lower panel. For the Understanding data, the components are noticeably larger in the lower panel, where they are in the metric of the <u>self-descriptive positive</u> subscales, than they are in the upper panel, where they are in the implicitly averaged metric of the simpler model. An explanation of the different estimates observed for the Understanding data would be that the <u>self-descriptive positive</u> subscales are "better" than the others, in the sense that they are more responsive to individual differences and therefore yield larger variance components. They are not "better" in the sense that they yield larger content and lower agreement components, however, since the ratio $\hat{\sigma}_{\xi}^2 / \hat{\sigma}_{\alpha}^2$ (about 1 for the Technical group and 2 for the Academic group) is the same in both panels.

For the Understanding data, we also found that 3-component models containing a small form component on the order of .02 or .03 provided an adequate fit to the data, and the content and agreement components obtained under the 3-component models are closely matched by the upper-panel estimates in Table 6.1 (cf. Tables 5.20 and 5.21). The interpretation of the form effect is that there is a subject-by-form interaction--i.e., individual differences in response to Understanding subscales of differing form, such that some persons tend to have higher scores on the self-descriptive subscales than on the attitude subscales, while others have the reverse pattern. The form component was positively correlated with the content component for both groups, implying that the self-descriptive subscales would have slightly larger variances than the corresponding attitude subscales. Because the form components for the Understanding subscales were very small, and uncorrelated with the form components for Play, they have more theoretical than practical importance.

2. Self-descriptive and attitude items. The hypothesis under this category was that self-descriptive scales would show more endorsement than agreement responding, while the attitude scales would have the reverse pattern. Since there was no endorsement in the data, using ACOVS models, this hypothesis is not confirmed. Note that the ACOVS models being fitted constrain all eight subscales to have equal components $\hat{\sigma}_{\alpha}^{2}$ and $\hat{\sigma}_{\eta}^{2}$ (in the metric of A, if scaling factors are used), so the hypothesis requires the presence of a form x agreement x endorsement variance component in the data. There is no theoretical reason

why we could not observe such an "interaction" component in the absence of the corresponding "main effect" components. For the data at hand, however, there was no evidence that any of the interaction components would materially improve the fit of ACOVS models to the data.

3. *The two-factor theory of acquiescence*. The hypothesis here was that the three *a priori* measures of agreement and the two *a priori* measures of endorsement would have a 2-factor simple structure. Although an endorsement component could not be found in ACOVS models for the separate content areas, it was possible to fit 2-factor models to the five measures, as shown at the end of Chapter 5. There is no doubt about the presence of an agreement component in the data. The best evidence for the presence of an endorsement component is that the *a priori* measures of endorsement have low, nonzero correlations ($r_{PE,UE}$ = .22 for the Technical group and .34 for the Academic group). These correlations can be interpreted as alternate-form reliabilities for the 56-item measures of endorsement, and imply that the best measure of endorsement in the data--the 112-item sum of the scores on the PE and UE scales--would have a reliability of .36 for the Technical group and .51 for the Academic group.

4. *Construct validity of agreement and endorsement*. The hypothesis here was that agreement would be related to verbal ability, and endorsement to impulsivity and speed. The conceptual model for this set of hypotheses was expressed by the diagram in Figure 4.1, and the most direct test of the hypothesis was given by the canonical analyses reported in Figure 5.10. For the Academic

group, neither of the canonical components was significant, and the only variable having a significant relationship with the AGR composite was the F-scale content measure FC. No variables had a significant relationship with the END composite for the Academic group. For the Technical group, there was evidence for the construct validity of agreement. Agreement has the predicted negative relationship with Vocabulary, together with an unpredicted positive relationship with Age and a positive relationship with Desirability. So far as the data indicate, Speed, Impulsivity and the reflected Cognitive Structure variable have a negative relationship with agreement rather than the predicted positive relationship with endorsement. The only variable having a relationship with endorsement for the Technical group is the Understanding content measure UC.

With a little imagination, one can "see" the right half of Figure 5.10, for the Academic group, as a rotated version of the left half, for the Technical group. The variables which have similar relationships with the AGR and END axes of the mentally rotated plots are FC, COG, UC, PC and VOC, and the variables having the most dissimilar relationships are AGE, SPD, INF and perhaps DES. It would have been theoretically preferable to have found the strong agreement component and the negative relationship with Vocabulary in the Academic group rather than in the Technical group. The Academic group more nearly resembles undergraduate groups which are likely to be obtained in the course of research, while the Technical group is unusual in several respects, as discussed in Chapter 5. Subjects in the Technical group are more heterogeneous than subjects in the Academic group. They are more variable on most of the measures used in the study, and particularly on the measure of agreement. They tend to be older, slower, have lower verbal ability, more infrequent responses, fewer desirable responses, and so on. I have no way of knowing the ethnic and cultural backgrounds of the subjects in my sample at this point, since confidentiality of subjects' identities was preserved in collecting the data, but it is known that the student body at the Technical college is predominately made up of American minority groups and foreign students. Now, it makes sense that subjects with a poor command of English, cultural differences, and probably little experience with personality tests would show more response inconsistency and more agreement response tendency, but if agreement is to be considered an important variable in personality research, it ought to be possible to study it in student groups likely to be met in practice. Certainly there is evidence for agreement response style in the Academic group: Referring to Table 6.1, we can see that the content and agreement components are about the same size for the F-scale, the personality variable where acquiescence was first discovered, and the agreement component is about half the size of the content component for the Play and Understanding scales. There is little evidence bearing on the construct validity of agreement for the Academic group, however, which is disappointing. One factor bearing on the present data as a test of the relationship between agreement and verbal ability is that the Vocabulary test appears to have been too easy for many of the students in the Academic group, as noted earlier. The Academic subjects averaged 13 out of 18 items correct, after correction for guessing. With the advantage of hindsight, I would choose a Vocabulary test with a wider range if I were re-doing the study, since topping out on the Vocabulary measure would attenuate agreement-vocabulary relationships for the Academic group. Another factor bearing on the test of a relationship between agreement and verbal ability is that splitting the overall sample into two groups has weakened the power of the analysis, making it more difficult to detect small effects. The rationale for splitting the groups has been discussed earlier, and may be taken up again in connection with the next topic.

5. Linear and nonlinear relationships with acquiescence. The hypothesis here was that agreement might have nonlinear relationships with some of the validity measures in the study, and I was particularly interested in the possibility of a nonlinear (heteroscedastic) relationship with verbal ability. What I had in mind was dichotomizing or perhaps trichotomizing the subjects in terms of Vocabulary and some of the other variables in the study, and looking for differences in variability on the *a priori* measure of agreement. When the data were collected, and it became apparent that I was getting group differences on so many of the measures, the plan of dividing the group by high and low Vocabulary scores was abandoned. Certainly any attempt to divide the overall sample by Vocabulary scores would result in a split which was confounded with group differences.

The data amply justify the conclusion that agreement is more of a measurement problem in the Technical group than in the Academic group. There was no reason to suspect that either group would be more or less variable on content, and the content components are about the same size for both groups, but the agreement components are about twice as large in the Technical group as they are in the Academic group. The measurement operation for getting agreement scores involves counting content-inconsistent responses, and low verbal ability and perhaps unfamiliarity with the test materials can be expected to increase the number of content-inconsistent responses. Whether the responses are inconsistent in the yeasaying or the naysaying direction then depends on personality factors.

It is important to note that the differences between the groups do not simply result in greater error of measurement for the Technical group. If there were simply more error in the the Technical group, we would expect correlations to be lower in the Technical group; in fact, they are generally higher than the correlations for the Academic group. Moreover, we would not expect the experimentally independent measures of agreement from the three content areas to correlate as highly as they do (.70 to .79 for the Technical group). Also, if error of measurement led to a spurious agreement effect, we might expect it to yield an endorsement effect as well, but it is only agreement which seems present as a strong response style.

There are other nonlinear relationships in the data, involving the Age and Speed variables. Both show relationships with agreement in the Technical group, but none in the Academic group. Age is a puzzle: Unexpectedly, it is the strongest single predictor of agreement in the Technical group. My best guess about the Age effect is that Age is correlated with unmeasured variables in the Technical group. Foreign students at the college tend to be older than the American students, and Age may be serving as a proxy for foreign-domestic student status. Speed, Impulsivity and reflected Cognitive Structure have a negative relationship with agreement in the Technical group--the faster and more impulsive students tend to have lower agreement scores. There is no evidence in the study for an acquiescence (agreement or endorsement) dimension which is positively associated with speed, tempo, quickness and impulsive stimulus acceptance.

Apart from Age, the strongest predictors of agreement are the Vocabulary and Desirability variables. I expressed my reservations about the Desirability variable in Chapter 5. The _true_ and _false_ parts of the scale seem to measure different things, and the _false_ part has a noticeably higher variance and coefficient-α than the _true_ part. The direction of the relationship between agreement and Desirability is such that subjects responding in the _false_ or desirable direction on the _false_ part of the scale tend to be naysayers on the agreement dimension. Since the _false_ part of the Desirability scale dominates the behavior of the total scale, the direction of the relationship between agreement and Desirability is consistent with the interpretation that nay-saying is contributing to the relationship. However, my best guess about the presence of the Desirability variable on the negative end of the agreement dimension is that, along with Vocabulary, it is functioning primarily as a measure of ability. Personality items can serve as an indirect measure of ability, as Gough (1953) showed with his "nonintellectual intelligence test" (the Intellectual Efficiency or *Ie* scale of the California Psychological Inventory). PRF Desirability correlates .73 with the *Ie* scale, according to the PRF manual (Jackson, 1974, p. 28), so it is reasonable to interpret Desirability as a proxy for ability in the present data.

In conclusion, the present data are dominated by a large group difference in agreement, ability, and several other variables. Agreement is negatively correlated with ability overall and within the low-ability Technical group. Agreement is also a large variance component of scores in the low-ability group, and contributes to the greater variability of scores in that group. Agreement components for the high-ability Academic group are about half as large as the content components, but there is little evidence that agreement is correlated with other variables within the Academic group.

Implications for Research and Testing Practice

Personality scales should be routinely written with equal numbers of *true* and *false* keyed items, in order to minimize the confounding of content and agreement. The weight of evidence for the presence of agreement acquiescence in test data is so great, that there is hardly any excuse for the use of measures consisting of predominately *true* or predominately *false* items. Such measures arise in practice when content and agreement are correlated, and items are selected in order to maximize internal consistency, as in the case of the F-scale, or empirical validity, as in the case of the MMPI. With a positive correlation between content and agreement, an all-*true* scale will tend to have a higher internal consistency and a higher correlation between the scale and its criterion, but it will also have spuriously high correlations with other measures developed without regard for balancing the numbers of *true* and *false* items.

With scales balanced for *true* and *false* items, the confounding of content and agreement is reduced, but it is not entirely eliminated unless the simplest kind of component model holds--namely, the model which asserts that the score of the i-th subject on the j-th test may be written:

$$[6.1] \qquad x_{ji} = \zeta_i + \alpha_i + \epsilon_{ji} .$$

Such a model holds reasonably well for the Play data in this study, but for the Understanding data we obtained better fits using models resembling

$$[6.2] \qquad x_{ji} = \gamma_j(\zeta_i + \alpha_i) + \epsilon_{ji} ,$$

where the γ_j term is a scaling factor applying to the j-th measure; and for the F-scale data we obtained better fits using models resembling

$$[6.3] \qquad x_{ji} = \zeta_i + \delta_j \alpha_i + \epsilon_{ji} ,$$

where the content components may be considered equal for all scales, but the j-th measure has a scaling factor δ_j applying to its agreement component. When model [6.2] or [6.3] holds, estimates of content unconfounded by agreement can only be obtained when the weights γ_j or δ_j are known or can be estimated from the data. With any real data, the actual state of affairs is probably even more complicated than the one given by [6.2] or [6.3], and resembles the model:

$$[6.4] \qquad x_{ji} = \gamma_j \zeta_i + \delta_j \alpha_i + \epsilon_{ji} .$$

The model [6.4] asserts that neither the content nor the agreement components have a constant variance from scale to scale. Models such as [6.4] will not generally be estimable; the hope, therefore, is that models such as [6.2] or [6.3] or, preferably, [6.1] will provide a reasonable fit to the data. Analysis of covariance structure is the method of choice for exploring the kinds of models which will fit the data.

The simplest model [6.1] implies that content and agreement are estimated by equally-weighted composites of the measures. In practice, such equally-weighted composites are probably adequate for obtaining content and agreement measures, and may even be more robust than the weighted composites obtained with [6.2] and [6.3]. With balanced measures of content, the <u>a priori</u> measure of agreement--the equally-weighted composite of all the true-false differences--should be routinely obtained and analyzed along with the content measures.

The relationships between agreement and verbal ability in the present study were particularly interesting. It is my belief that verbal ability should be routinely measured in personality research involving pencil-and-paper tests, except perhaps when it is known that the subjects are homogeneous in ability. Verbal ability is relatively easy to measure, its construct validity is not in doubt, and it is informative about personality. In the present study, the low-ability group had a higher mean on the agreement measure, replicating the common finding of a negative correlation between agreement and verbal ability, but was also markedly more variable in agreement response tendency. I attribute the greater variability of the low-ability group to the fact that difficulty in understanding test items will lead to more content-inconsistent responses, and thus exaggerate the yea-saying or nay-saying tendencies of the subjects. In the high-ability group, the agreement component was relatively small, and apparently unrelated to any of the other variables used in the study. This suggests to me that agreement acquiescence may not be a serious measurement problem in testing bright, highly selected undergarduate groups, but can be quite important for research involving the general population or groups with large numbers of low-ability subjects.

Appendix A

INSTRUCTIONS FOR RESEARCH QUESTIONNAIRE

Attached to this instruction sheet are an answer sheet and an item booklet. The item booklet contains statements referring to opinions or characteristics. Do not look at them until the instructions are clear to you.

Read each statement and answer either YES or NO.

YES indicates that you agree with the statement, or that it is a statement about something which is characteristic of you.

NO indicates that you disagree with the statement, or that it is a statement about something which is uncharacteristic of you.

It is important that you answer every item, in order. If neither YES nor NO seems like an appropriate answer to a particular question, choose the one which seems more appropriate. Here are two examples:

1. Old age assistance programs should be abandoned.

2. I like to read stories about the sea.

A person who agrees with the statement that old age assistance programs should be abandoned would mark the first box next to item 1 on the answer sheet (under Y, for "YES"). If he feels that liking to read stories about the sea is not characteristic of him, he would mark the second box next to item 2 (under N, for "NO"). His answer sheet would show:

	Y	N
1.	[x]	[ ]
2.	[ ]	[x]

Do not spend a lot of time on any one question. Remember that your first response is probably best.

If you have any questions, please ask them now.

When the examiner says to start, begin answering the questions.

PLEASE DO NOT MAKE ANY MARKS ON THE TEST BOOKLETS.

Appendix B

Research Questionnaire Scales

Content Area	Type of Scale	Items
Cognitive Structure	Marker	16
Desirability	"	20
Exhibition*	Experimental	56
F-scale	"	48
Impulsivity	Marker	16
Infrequency	"	12
Play	Experimental	56
Succorance*	"	56
Understanding	"	56

In the listings of scales which follow, columns 1-3 indicates item numbers in the final Research Questionnaire. Column 4 is a line number within item. For the experimental scales, columns 7-8 represent the stem number and column 9 represents the variant number (1 or 2), but for the marker scales columns 7-9 may be ignored. Columns 12-15 represent the subscale mnemonic.

For the experimental scales, the proportion of true responses is reported, separately for the Technical and Academic groups. Starred (*) scales are experimental scales which were not used in the final form of the Research Questionnaire.

COGNITIVE STRUCTURE (16 Items, Source: PRF-E)

161 9 101 COGT When I go on a trip I prepare a timetable before hand.

441 9 201 COGT Before I ask a question, I decide exactly what it is I
442 9 202 COGT need to find out.

561 9 301 COGT Often when I telephone someone, I make a list of things
562 9 302 COGT to discuss.

881 9 401 COGT When I make something I want to know exactly what it
882 9 402 COGT will look like when finished.

1281 9 501 COGT I try to plan my future so that I can tell what I will
1282 9 502 COGT be doing at any given time.

1561 9 601 COGT I don't like to go into a situation without knowing
1562 9 602 COGT what I can expect from it.

1681 9 701 COGT When I talk to a doctor, I want him to describe in
1682 9 702 COGT detail any illness I have.

2001 9 801 COGT I don't like to start a project until I know the best
2002 9 802 COGT way to proceed.

41 9 901 COGF I seldom make careful plans.

311 91001 COGF I like to be with people who change their minds often.

681 91101 COGF I tend to start right in on a new task without thinking
682 91102 COGF about the best way to do it.

1041 91201 COGF I rarely consider the daily weather report when
1042 91202 COGF deciding what to wear.

1161 91301 COGF I live from day to day without trying to fit my
1162 91302 COGF activities into a pattern.

1431 91401 COGF I can feel comfortable even when I have a number of
1432 91402 COGF unanswered questions in mind.

1801 91501 COGF When I take a vacation I like to go without detailed
1802 91502 COGF plans.

2161 91601 COGF I often start work on something when I have only a very
2162 91602 COGF hazy idea of what the end result will be.

DESIRABILITY (2 Items, Source: PRF-AA)

81 9 101 DYAT Most of my teachers were helpful.

321 9 201 DYAT I am always prepared to do what is expected of me.

721 9 301 DYAT I always try to be considerate of the feelings of my
722 9 302 DYAT friends.

841 9 401 DYAT I often take some responsibility for looking out for
842 9 402 DYAT newcomers in a group.

1001 9 501 DYAT I am seldom ill.

1201 9 601 DYAT My memory is as good as other people's.

1441 9 701 DYAT In the long run, humanity will owe a lot more to the
1442 9 702 DYAT teacher than to the salesman.

1841 9 801 DYAT My life is full of interesting activities.

1961 9 901 DYAT Rarely, if ever, has the sight of food made me ill.

2121 91001 DYAT I am able to make correct decisions on difficult
2122 91002 DYAT questions.

201 91101 DYAP Many things make me feel uneasy.

471 91201 DYAP I often have the feeling that I am doing something
472 91202 DYAP evil.

601 91301 DYAP Nothing that happens to me makes much difference one
602 91302 DYAP way or the other.

921 91401 DYAP I have a number of health problems.

1081 91501 DYAP I believe people tell lies any time it is to their
1082 91502 DYAP advantage.

1321 91601 DYAP I almost always feel sleepy and lazy.

1591 91701 DYAP I am not willing to give up my own privacy or pleasure
1592 91702 DYAP in order to help other people.

1721 91801 DYAP I find it very difficult to concentrate.

2041 91901 DYAP I often question whether life is worthwhile.

2201 92001 DYAP We ought to let the rest of the world solve their own
2202 92002 DYAP problems and just look out after ourselves.

EXHIBITION (56 Items, Source: Experimental)

1 1 111 ESPT Others think I am lively and witty.

1 1 121 ESPF Others think I am somewhat dull.

1 1 211 ESPT When my friends are reluctant, I will often be the
2 1 212 ESPT first one to try something new.

1 1 221 ESNT Even when my friends are reluctant, I will usually not
2 1 222 ESNT wait to try something new.

1 1 311 ESPT I would like to have a flashy car that will make others
2 1 312 ESPT stop and look as I drive by.

1 1 321 ESNF I would like to have an ordinary car that does not
2 1 322 ESNF attract a lot of attention.

1 1 411 ESPT If I wanted to be a good teacher, I would act a little
2 1 412 ESPT bit like an actor in the classroom.

1 1 421 EAPT A successful teacher is a little bit like an actor.

1 1 511 ESPT I occasionally say something shocking just to draw
2 1 512 ESPT attention to myself.

1 1 521 EAPF It is silly to say shocking things just to draw
2 1 522 EAPF attention to oneself.

1 1 611 ESPT I like to give speeches.

1 1 621 EANT People should not mind giving speeches.

1 1 711 ESPT I would like to sing for people if I had the talent.

1 1 721 EANF Singing for people is probably not much fun, even if
2 1 722 EANF you have the talent.

1 1 811 ESPF If I had the musical talent, I would use it for my own
2 1 812 ESPF amusement, and not to entertain others.

1 1 821 ESNT Musical talent is not something which should be used
2 1 822 ESNT just for one's own private amusement.

1 1 911 ESPF If I were an artist, I would be primarily interested in
2 1 912 ESPF the private work of creating.

1 1 921 ESNF If I were an artist, I would not be concerned about
2 1 922 ESNF getting the public to notice my work.

1 11011 ESPF I am too shy to tell jokes.

1 11021 EAPT People should enjoy telling jokes.

1 11111 ESPF When I'm in a conversation, I try to keep other people
2 11112 ESPF talking more than I do.

1 11121 EAPF In a conversation, one should let others talk more than
2 11122 EAPF oneself.

1 11211 ESPF I feel embarrassed whenever I am with loud people.

1 11221 EANT A person should not feel embarrassed just because he is
2 11222 EANT with loud people.

1 11311 ESPF I am a pretty ordinary person, easy to overlook.

1 11321 EANF People do not usually like or approve of those who are
2 11322 EANF readily noticed.

1 11411 ESNT I don't like being in the background at parties.

1 11421 ESNF I wouldn't be caught dead doing some of the things
2 11422 ESNF people do at parties.

1 11511 ESNT I don't like to be in the background, while others get
2 11512 ESNT the glory.

1 11521 EAPT The people who get the glory have usually earned it.

1 11611 ESNT I was not a very quiet child.

1 11621 EAPP Children should mind their manners and be quiet.

1 11711 ESNT I don't mind seeking the spotlight, even when it works
2 11712 ESNT against me.

1 11721 EANT It's not wrong to seek the spotlight.

1 11811 ESNT I don't mind being conspicuous.

1 11821 EANP People don't usually like to be conspicuous.

1 11911 ESNP I don't like having my picture taken.

1 11921 EAPT Most people like to have their picture taken.

1 12011 ESNP I don't like to be the topic of conversation even among
2 12012 ESNP friends.

1 12021 EAPP Even among friends, most people find it embarrassing to
2 12022 EAPP be the topic of conversation.

1 12111 ESNP The idea of acting in front of a large group does not
2 12112 ESNP appeal to me.

1 12121 EANT Acting in front of a large group should not make people
2 12122 EANT feel uncomfortable.

1 12211 ESNP I would never write about myself, because it would look
2 12212 ESNP like bragging.

1 12221 EANP A person should not write about himself since it
2 12222 EANP usually looks as if he were bragging.

1 12311 EAPT Great athletes deserve the fame that goes with their
2 12312 EAPT accomplishments.

1 12321 EAPP Fame has been the ruin of many great athletes.

1 12411 EAPT Most people like to have one person do most of the
2 12412 EAPT talking, as long as they are entertained.

1 12421 EANT Most people don't care if one person does most of
2 12422 EANT talking, as long as they are entertained.

1 12511 EAPT There is something wrong with being just one of the
2 12512 EAPT crowd.

1 12521 EANP There is nothing wrong with being just one of the
2 12522 EANP crowd.

1 12611 EAPP Trying to be the center of attention is a sign of bad
2 12612 EAPP taste.

1 12621 EANT Trying to be the center of attention is not a sign of
2 12622 EANT bad taste.

1 12711 EAPP Most jokes should be forgotten.

1 12721 EANP Most jokes are not worth telling.

1 12811 EANT Tight clothes are not embarrassing to wear.

1 12821 EANP A person should not wear tight clothes.

P-SCALE (48 Items, Source: Experimental)

Item	Text	P(TRUE) TECH	P(TRUE) ACAD
591 5 111 PSRT 592 5 112 PSRT	Our way of life is disappearing so fast that force might be necessary to preserve it.	.473	.156
1711 5 121 PSRF 1712 5 122 PSRF 1713 5 123 PSRF	People should probably pay more attention to new ideas, even if they seem to go against the American way of life.	.834	.930
31 5 211 PSRT 32 5 212 PSRT	A sane, normal, decent person would probably find it hard to think about hurting a close friend or relative.	.731	.695
1151 5 221 PSRF 1152 5 222 PSRF	A sane, normal, decent person may sometimes think about hurting a close friend or relative.	.412	.636
231 5 311 PSRT 232 5 312 PSRT 233 5 313 PSRT	Many of our social problems would probably be solved if we could somehow get rid of the immoral, crooked, and feebleminded people.	.634	.294
1351 5 321 PSRF 1352 5 322 PSRF 1353 5 323 PSRF	Most of our social problems would probably not be solved even if we could somehow get rid of the immoral, crooked, and feebleminded people.	.572	.637
1071 5 411 PSRT 1072 5 412 PSRT	The businessman and the manufacturer tend to be more important to society than the artist and the professor.	.592	.372
2191 5 421 PSRF 2192 5 422 PSRF	The artist and the professor tend to be more important to society than the businessman and the manufacturer.	.271	.292
711 5 511 PSRT 712 5 512 PSRT	Seldom do weaknesses or difficulties hold us back if we have enough willpower.	.824	.794
1831 5 521 PSAF 1832 5 522 PSAF	Weakness or difficulty always hold us back even though we have enough willpower.	.556	.284
671 5 611 PSRT 672 5 612 PSRT	Someday it will probably be shown that astrology can explain a lot of things.	.597	.440
1791 5 621 PSAF 1792 5 622 PSAF	It is absurd to suppose that astrology will ever explain anything.	.391	.362
2071 5 711 PSRT 2072 5 712 PSRT	Nowadays some people are prying into matters that should probably remain personal and private.	.793	.744
951 5 721 PSAF 952 5 722 PSAF	Nowadays not enough people are prying into matters which are kept personal and private.	.474	.156
1311 5 811 PSRT 1312 5 812 PSRT	When a person has a problem or worry, it is probably best for him to try not to think about it.	.468	.136
191 5 821 PSAF 192 5 822 PSAF 193 5 823 PSAF	When a person has a problem or worry, he should always drop everything and concentrate on it until a solution appears.	.322	.077
1191 5 911 PSRF 1192 5 912 PSRF	The findings of science may eventually show that many of our most cherished beliefs are wrong.	.790	.714
71 5 921 PSAT 72 5 922 PSAT 73 5 923 PSAT	Science has its place, but there are many important things that can never possibly be understood by the human mind.	.790	.626
1871 51011 PSRF 1872 51012 PSRF	Insults to our honor are usually not important enough to bother about.	.401	.454
751 51021 PSAT	An insult to our honor should always be punished.	.422	.167
2151 51111 PSRF 2152 51112 PSRF 2153 51113 PSRF	There have been attempts to divide people into two distinct classes--the weak and the strong--but such attempts are probably doomed to failure.	.642	.820
1031 51121 PSAT 1032 51122 PSAT	People can be divided into two distinct classes: the weak and the strong.	.511	.207
1951 51211 PSRF 1952 51212 PSRF	There are probably few people who have learned anything important through suffering.	.402	.146
831 51221 PSAT 832 51222 PSAT	Nobody ever learned anything really important except through suffering.	.298	.068

111	51311	PSRT	Some youths would probably benefit [illegible]	[illegible]	[illegible]
112	51312	PSRT	discipline, rugged determination, and the will to work		
113	51313	PSRT	and fight for family and country.		
1231	51321	PSAT	What the youth needs most is strict discipline, rugged	.720	.296
1232	51322	PSAT	determination, and the will to work and fight for		
1233	51323	PSAT	family and country.		
351	51411	PSRT	Wars and social troubles may someday be ended by an	.457	.508
352	51412	PSRT	earthquake, flood, or other natural disaster destroying		
353	51413	PSRT	all or most of the world.		
1471	51421	PSAT	Wars and social problems will only be ended when the	.468	.489
1472	51422	PSAT	whole world is destroyed.		
911	51511	PSRT	A person who has bad manners, habits, and breeding	.793	.499
912	51512	PSRT	would probably find it hard to get along with decent		
913	51513	PSRT	people.		
2031	51521	PSAT	A person who has bad manners, habits, and breeding can	.720	.459
2032	51522	PSAT	hardly expect to get along with decent people.		
271	51611	PSRT	What this country probably needs, rather than laws and	.638	.519
272	51612	PSRT	political programs, is a few courageous, tireless,		
273	51613	PSRT	devoted leaders who can be trusted.		
1391	51621	PSAT	What this country needs most, much more than laws and	.607	.470
1392	51622	PSAT	political programs, is a few courageous, tireless,		
1393	51623	PSAT	devoted leaders in whom the people can put all of their		
1394	51624	PSAT	faith.		
871	51711	PSRP	It is probably foolish to believe that our lives could	.488	.293
872	51712	PSRP	be controlled, in some way, by plots hatched in high		
873	51713	PSRP	places.		
1991	51721	PSAP	It is ridiculous to believe that our lives are	.540	.450
1992	51722	PSAP	controlled by plots hatched in high places.		
991	51811	PSRP	The urge to jump from high places is probably learned	.621	.630
992	51812	PSRP	rather than inborn.		
2111	51821	PSAP	It is known that the urge to jump from high places is	.699	.430
2112	51822	PSAP	learned, not inborn.		
1911	51911	PSRP	Most mature people tend to outgrow their feelings of	.309	.068
1912	51912	PSRP	submissive respect, gratitude, and love for their		
1913	51913	PSRP	parents.		
791	51921	PSAP	Every truly mature person outgrows his feelings of	.247	.019
792	51922	PSAP	submissive respect, gratitude, and love for his		
793	51923	PSAP	parents.		
1631	52011	PSRP	Young people sometimes get rebellious ideas, which	.432	.421
1632	52012	PSRP	should probably be encouraged and developed in order to		
1633	52013	PSRP	promote mature citizenship in adulthood.		
511	52021	PSAP	Young people sometimes get rebellious ideas, which	.514	.372
512	52022	PSAP	should always be developed in order to guarantee mature		
513	52023	PSAP	citizenship in adulthood.		
1551	52111	PSAT	Every person should have complete faith in some	.236	.088
1552	52112	PSAT	supernatural power whose decisions he obeys without		
1553	52113	PSAT	question.		
431	52121	PSAP	Every person should have complete faith in his own	.772	.676
432	52122	PSAP	judgment, and not rely on some supernatural power.		
2231	52211	PSAT	Human nature being what it is, there will always be war	.796	.807
2232	52212	PSAT	and conflict.		
1111	52221	PSAP	Because human nature is improving, war and conflict	.315	.120
1112	52222	PSAP	will eventually be eliminated.		
1511	52311	PSAT	Obedience and respect for authority are the most	.751	.286
1512	52312	PSAT	imporant virtues children should learn.		
391	52321	PSAP	A love of freedom and distrust for authority are the	.246	.118
392	52322	PSAP	most important virtues children should learn.		
1671	52411	PSAT	If people would talk less and work more, everybody	.515	.126
1672	52412	PSAT	would be better off.		
551	52421	PSAP	If people talked thinos over more and didn't work so	.371	.489
552	52422	PSAP	hard, everybody would be better off.		

PLAY (56 Items, Source: Experimental)

				P (TRUE) TECH	ACAD
1131	2 111	PSPT	Once in a while I enjoy acting as if I were tipsy.	.422	.553
11	2 121	PSPF	I always act very sober, even if I have had a few	.473	.237
12	2 122	PSPF	drinks.		
171	2 211	PSPT	I take advantage of every excuse to have a party.	.256	.214
1291	2 221	PSNT	I try not to miss an excuse to have a party.	.298	.185
1451	2 311	PSPT	I try to make my work into a game.	.401	.420
331	2 321	PSNF	I never try to make my work into a game.	.577	.234
491	2 411	PSPT	I enjoy children's games.	.603	.725
1611	2 421	PAPT	Adults should enjoy playing children's games.	.603	.597
1771	2 511	PSPT	I always have more fun than other people do.	.246	.138
651	2 521	PAPF	Some people have more fun than they should.	.627	.296
811	2 611	PSPT	I always enjoy a joke, even when it is on me.	.741	.450
1931	2 621	PANT	People who cannot enjoy a joke on themselves are much	.793	.705
1932	2 622	PANT	too serious.		
2091	2 711	PSPT	I love to tell and listen to funny jokes and stories.	.883	.832
971	2 721	PANF	People don't like funny jokes and stories as much as	.504	.352
972	2 722	PANF	they pretend to.		
531	2 811	PSPF	I usually try to see the serious side of every	.717	.460
532	2 812	PSPF	situation.		
1651	2 821	PSNT	I rarely try to see the serious side of a situation.	.186	.068
1811	2 911	PSPF	I realize that I often forget to relax when I am	.551	.587
1812	2 912	PSPF	supposed to be enjoying myself.		
691	2 921	PSNF	Many times when I am supposed to be enjoying myself, I	.582	.567
692	2 922	PSNF	don't remember to relax.		
1011	21011	PSPF	I prefer to read worthwhile books rather than spend my	.550	.326
1012	21012	PSPF	spare time playing.		
2131	21021	PAPT	Spare time is meant for play rather than reading books.	.269	.177
1971	21111	PSPF	I would find designing children's toys a very dull	.205	.127
1972	21112	PSPF	activity.		
851	21121	PAPF	Designing children's toys must be a very dull job.	.133	.038
371	21211	PSPF	Even if something gives me pleasure, I need some other	.504	.264
372	21212	PSPF	reason for doing it.		
1491	21221	PANT	If something gives pleasure, one does not need some	.607	.715
1492	21222	PANT	other reason for doing it.		
1331	21311	PSPF	I usually work or study during my spare time.	.577	.431
211	21321	PANF	Most people don't do enough work and studying during	.845	.715
212	21322	PANF	their spare time.		
51	21411	PSNT	I must admit, I never have any trouble forgetting my	.488	.372
52	21412	PSNT	work when I have a chance to have some fun.		
1171	21421	PSNF	I must admit, I am never able to forget my work when I	.478	.431
1172	21422	PSNF	have a chance to have some fun.		
2011	21511	PSNT	It is not often that I turn down a chance to have a	.680	.646
2012	21512	PSNT	good time.		
891	21521	PAPT	People should take advantage of every chance to have a	.731	.665
892	21522	PAPT	good time.		

411	21611	PSNT	I find it almost impossible to talk to someone who	.473	.323
412	21612	PSNT	doesn't have a good sense of humor.		
1531	21621	PAPP	People with a poor sense of humor are often easy to	.174	.127
1532	21622	PAPP	talk to.		
2171	21711	PSNT	I don't like adults who have forgotten how to play.	.585	.474
1051	21721	PANT	Adults who have forgotten how to play are not very	.548	.454
1052	21722	PANT	likable.		
731	21811	PSNT	I don't like to stay at home in the evening when I have	.700	.617
732	21812	PSNT	a chance to go out.		
1651	21821	PANP	Most people don't really like going out in the evening.	.350	.048
1371	21911	PSNP	My worries don't disappear when I get into a	.514	.558
1372	21912	PSNP	boisterous, fun-loving crowd.		
251	21921	PAPT	Joining a boisterous, fun-loving crowd is an effective	.576	.372
252	21922	PAPT	way to make one's worries disappear.		
1091	22011	PSNP	I don't like to make up fanciful stories.	.518	.484
2211	22021	PAPP	Making up fanciful stories is a waste of time.	.353	.171
1691	22111	PSNP	I can't imagine myself playing with children's toys.	.384	.118
571	22121	PANT	It is not difficult to imagine an adult who likes to	.603	.593
572	22122	PANT	play with children's toys.		
931	22211	PSNP	I have never been so fascinated with a trivial game	.360	.224
932	22212	PSNP	that I have played it for hours.		
2051	22221	PANP	Trivial games are never worth playing for hours.	.494	.234
1211	22311	PAPT	Most entertainment serves a worthwhile purpose.	.885	.862
91	22321	PAPP	Most entertainment is a waste of time.	.215	.048
771	22411	PAPT	One should celebrate ordinary events as well as special	.525	.744
772	22412	PAPT	ones.		
1891	22421	PANT	Celebrations should not be saved just for special	.494	.754
1892	22422	PANT	events.		
1571	22511	PAPT	People should go "out on the town" as often as they	.593	.474
1572	22512	PAPT	can.		
451	22521	PANP	People should not go "out on the town" except on rare	.145	.078
452	22522	PANP	occasions.		
291	22611	PAPP	Most people prefer the company of others who are	.535	.284
292	22612	PAPP	relatively serious.		
1411	22621	PANT	Most people do not prefer the company of others who are	.638	.401
1412	22622	PANT	relatively serious.		
1731	22711	PAPP	Practical jokes are a waste of time.	.339	.244
611	22721	PANP	Practical jokes aren't funny.	.473	.332
131	22811	PANT	One should never take matters seriously just because	.710	.465
132	22812	PANT	someone else does.		
1251	22821	PANP	One should never make light of matters which someone	.669	.524
1252	22822	PANP	else takes seriously.		

IMPULSIVITY (16 Items, Source: PRF-E)

```
 241 9 101 IMPT Often I stop in the middle of one activity in order to
 242 9 102 IMPT start something else.

 481 9 201 IMPT I often say the first thing that comes into my head.

 631 9 301 IMPT I like to live dangerously.

 961 9 401 IMPT Many of my actions seem to be hasty.

1361 9 501 IMPT I have often broken things because of carelessness.

1601 9 601 IMPT Most people feel that I act impulsively.

1751 9 701 IMPT Sometimes I get several projects started at once
1752 9 702 IMPT because I don't think ahead.

2081 9 801 IMPT I find that thinking things over very carefully often
2082 9 802 IMPT destroys half the fun of doing them.

 121 9 901 IMPF I am careful to consider all sides of an issue before
 122 9 902 IMPF taking action.

 361 91001 IMPF I am pretty cautious.

 761 91101 IMPF Rarely, if ever, do I do anything reckless.

1121 91201 IMPF Emotion seldom causes me to act without thinking.

1241 91301 IMPF I have a reserved and cautious attitude toward life.

1481 91401 IMPF My thinking is usually careful and purposeful.

1881 91501 IMPF I am not one of those people who blurt out things
1882 91502 IMPF without thinking.

2241 91601 IMPF I generally rely on careful reasoning in making up my
2242 91602 IMPF mind.
```

INFREQUENCY (12 Items, Source: PRF-AA)

```
 281 9 101 INFT I have never seen an apple.

 401 9 201 INFT I am more than eighty years old.

 801 9 301 INFT I have never ridden in an automobile.

1401 9 401 INFT I often sit and stare directly at the sun for hours on
1402 9 402 INFT end.

1521 9 501 INFT I have never talked with anyone by telephone.

1921 9 601 INFT I have never felt sad.

 151 9 701 INFF If I were exploring a strange place at night, I would
 152 9 702 INFF want to carry a light.

 521 9 801 INFF I have attended school for at least six years during my
 522 9 802 INFF life.

 641 9 901 INFF I would have a hard time keeping my mind a complete
 642 9 902 INFF blank.

1271 91001 INFF I usually prefer to have meat cooked before eating it.

1641 91101 INFF Things with sugar in them usually taste sweet to me.

1761 91201 INFF Sometimes I see birds near my home.
```

SUCCORANCE (56 Items, Source: Experimental)

1 3 111 SSPT I like to have a superior who will offer to help me at
2 3 112 SSPT work.

1 3 121 SSPF I would be embarrassed if a superior offered to help me
2 3 122 SSPF at work.

1 3 211 SSPT I usually seek other people's advice before doing
2 3 212 SSPT anything important.

1 3 221 SSNT I seldom do anything important unless I have sought
2 3 222 SSNT other people's advice.

1 3 311 SSPT I sometimes act more upset than I really am, in order
2 3 312 SSPT to gain sympathy from others.

1 3 321 SSNF I have never tried to gain sympathy from others by
2 3 322 SSNF acting more upset than I really was.

1 3 411 SSPT I like people who seem willing to take care of me.

1 3 421 SAPT People who are willing to take care of others are very
2 3 422 SAPT likable.

1 3 511 SSPT If I have to make a selection in a store, I often ask
2 3 512 SSPT the clerk for advice.

1 3 521 SAPF When making a selection in a store, the clerk's advice
2 3 522 SAPF is generally useless.

1 3 611 SSPT As a child, I liked being dependent on grown-ups.

1 3 621 SANT Most children do not like being dependent on grown-ups.

1 3 711 SSPT I would like to be married to a protective and
2 3 712 SSPT sympathetic person.

1 3 721 SANF It is not especially important for a marriage partner
2 3 722 SANF to be protective and sympathetic.

1 3 811 SSPF I am more self-sufficient than most people.

1 3 821 SSNT I am not more self-sufficient than most people.

1 3 911 SSPF I expect that I will always be able to take care of
2 3 912 SSPF myself.

1 3 921 SSNF I rarely worry about being alone and helpless.

1 31011 SSPF I usually make decisions by myself.

1 31021 SAPT People should seek the help and advice of others when
2 31022 SAPT they make decisions.

1 31111 SSPF I feel that I am as well-qualified as any expert to
2 31112 SSPF solve my problems.

1 31121 SAPF Most people are as well-qualified as any expert to
2 31122 SAPF solve their problems.

1 31211 SSPF If I feel sad, I prefer to be left alone.

1 31221 SANT If a person feels sad, it is not a good idea to leave
2 31222 SANT him alone.

1 31311 SSPF I only go to a doctor when I am seriously ill.

1 31321 SANF People should not go to a doctor unless they are
2 31322 SANT seriously ill.

1 31411 SSNT If I have injured myself, I don't hesitate to tell
2 31412 SSNT others about it.

1 31421 SSNF If I have injured myself, I don't like to tell others
2 31422 SSNF about it.

1 31511 SSNT I never avoid sharing my burdens with someone who can
2 31512 SSNT help me.

1 31521 SAPT People should try to share their burdens with someone
2 31522 SAPT who can help them.

1 31611 SSNT I don't like to face problems by myself.

1 31621 SAPP People should try to face their problems by themselves.

1 31711 SSNT I don't hesitate to accept favors from others.

1 31721 SANT People should not hesitate to accept favors from
2 31722 SANT others.

1 31811 SSNT When I was a child, I seldom got along without the
2 31812 SSNT teacher's help.

1 31821 SANP A teacher's help is seldom very important to a child.

1 31911 SSNP I rarely feel unable to face the responsibilities of
2 31912 SSNP adulthood.

1 31921 SAPT The responsibilities of adulthood are often difficult
2 31922 SAPT to face.

1 32011 SSNP I don't want anyone to take care of me when I grow old.

1 32021 SAPP Most people want to take care of themselves when they
2 32022 SAPP grow old.

1 32111 SSNP When I go to a doctor, I never ask unnecessary
2 32112 SSNP questions.

1 32121 SANT When people see a doctor, they should not be afraid to
2 32122 SANT ask questions, no matter how unnecessary they might
3 32123 SANT seem.

1 32211 SSNP I prefer not to talk about how I feel when I am ill.

1 32221 SANP Generally, sick people should not talk about their
2 32222 SANP illnesses.

1 32311 SAPT Most people want others to know how they feel.

1 32321 SAPP Most people prefer to keep their feelings to
2 32322 SAPP themselves.

1 32411 SAPT It is a good idea to get other people's opinions
2 32412 SAPT concerning difficult problems.

1 32421 SANT It is not a good idea to try to solve difficult
2 32422 SANT problems without getting other people's opinions about
3 32423 SANT them.

1 32511 SAPT Offering advice is an essential part of friendship.

1 32521 SANP Offering advice is not an essential part of friendship.

1 32611 SAPP A person should be ashamed to accept charity when he
2 32612 SAPP needs it.

1 32621 SANT A person should not be ashamed to accept charity.

1 32711 SAPP Most people like to be left alone when they are sick.

1 32721 SANP Most people do not like to be waited on when they are
2 32722 SANP sick.

1 32811 SANT Most older people don't like having to be independent
2 32812 SANT of others.

1 32821 SANP Most older people don't like having to be dependent on
2 32822 SANP others.

UNDERSTANDING (56 Items, Source: Experimental)

				P(TRUE) TECH	ACAD
1141 1142	4 111 4 112	USPT USPT	I envy scholars who can spend as much time as they want thinking.	.422	.333
21 22	4 121 4 122	USPF USPF	I feel sorry for scholars who have to spend so much of their time thinking.	.350	.195
181 182	4 211 4 212	USPT USPT	I like to read several books on a topic at the same time.	.380	.352
1301	4 221	USNT	I seldom read only one book on a topic at a time.	.566	.313
1461 1462	4 311 4 312	USPT USPT	When I take a walk, I try to identify the trees and flowers I see on the way.	.298	.421
341 342	4 321 4 322	USNF USNF	When I take a walk, I do not try to identify the trees and flowers I see on the way.	.535	.411
501 502 503	4 411 4 412 4 413	USPT USPT USPT	After I have read a newspaper story, I look for additional facts that will make the story more complete.	.566	.382
1621 1622	4 421 4 422	UAPT UAPT	A reader usually requires additional facts to make a newspaper story more complete.	.720	.558
1781 1782	4 511 4 512	USPT USPT	I often try to grasp the relationships between different things that happen.	.803	.920
661 662	4 521 4 522	UAPF UAPF	Most people ignore the relationships between different things that happen.	.824	.734
821 822	4 611 4 612	USPT USPT	When I was a child, I read almost every book in my house.	.301	.284
1941 1942	4 621 4 622	UANT UANT	Children should never be prevented from reading as much as they like.	.843	.852
2101 2102	4 711 4 712	USPT USPT	Occasionally I have been able to relate a historical trend to my everyday life.	.645	.696
981 982	4 721 4 722	UANF UANF	Knowledge of historical trends is never useful in everyday life.	.204	.068
541	4 811	USPF	There are many activities that I prefer to reading.	.739	.803
1661	4 821	USNT	There aren't many activities that I prefer to reading.	.332	.126
1821	4 911	USPF	When I was a child I was timid about exploring.	.329	.156
701	4 921	USNF	When I was a child I did not like to explore.	.216	.087
1021 1022	41011 41012	USPF USPF	When I see a new invention, I generally take it for granted.	.404	.385
2141 2142	41021 41022	UAPT UAPT	Examining new inventions to find out how they work is well worth one's time.	.880	.960
1981	41111	USPF	I hate solving riddles and puzzles.	.174	.147
861	41121	UAPF	Solving riddles and puzzles is a waste of time.	.081	.058
381	41211	USPF	I was really glad when I graduated from school.	.820	.773
1501	41221	UANT	Going to school is never boring.	.526	.146
1341 1342	41311 41312	USPF USPF	If I were an inventor, I would only be satisfied with inventions having an obvious practical use.	.720	.322
221 222	41321 41322	UANF UANF	A new invention is not worth while unless it has an obvious practical use.	.638	.254
61 62	41411 41412	USNT USNT	I can't think of many things I wouldn't enjoy reading about.	.545	.513
1181	41421	USNF	There aren't many things I really enjoy reading about.	.318	.187
2021	41511	USNT	I don't avoid intellectual discussions.	.835	.781

901	41521	UAPT	Intellectual discussions are very enjoyable.	.835	.860
421 422	41611 41612	USNT USNT	I never seem to run out of questions on topics which interest me.	.721	.675
1541 1542	41621 41622	UAPP UAPP	Once a topic is of interest to a person, it rapidly becomes boring.	.112	.028
2181 2182	41711 41712	USNT USNT	I cannot automatically accept most current ideas and theories.	.621	.686
1061 1062	41721 41722	UANT UANT	Most current ideas and theories should not be accepted without serious questioning.	.804	.767
741 742	41811 41812	USNT USNT	I liked to study in school even when it wasn't required of me.	.504	.489
1861 1862	41821 41822	UANP UANP	Most people would not study at all if it was not required of them.	.741	.470
1381	41911	USNP	I do not have much use for abstract ideas.	.384	.137
261	41921	UAPT	Abstract ideas are often very useful.	.728	.862
1101 1102	42011 42012	USNP USNP	I really don't know what is involved in any of the latest developments.	.290	.254
2221	42021	UAPP	The latest cultural developments are often very boring.	.188	.152
1701	42111	USNP	I seldom read extensively on any one subject.	.442	.197
581 582	42121 42122	UANT UANT	Reading extensively on a single topic is not a bad way to spend one's time.	.493	.583
941	42211	USNP	When I was a child, I showed no interest in books.	.257	.117
2061	42221	UANP	An interest in books is not important for a child.	.154	.117
1221 1222	42311 42312	UAPT UAPT	Men's attempts to predict the future are very important.	.822	.597
101 102	42321 42322	UAPP UAPP	We have enough to think about without trying to predict the future.	.385	.284
781 782	42411 42412	UAPT UAPT	People should continue to read widely after leaving school.	.896	.960
1901 1902	42421 42422	UANT UANT	After leaving school, people should not stop reading widely.	.783	.803
461 462	42511 42512	UAPT UAPT	Someone will eventually figure out why society functions the way it does.	.720	.405
1581 1582	42521 42522	UANP UANP	No one will ever figure out why society functions the way it does.	.442	.514
301	42611	UAPP	Philosophical discussions are a waste of time.	.092	.088
1421	42621	UANT	Philosophical discussions are not a waste of time.	.772	.841
1741	42711	UAPP	Most intellectuals live rather useless lives.	.207	.048
621	42721	UANP	Most intellectuals are not very useful to society.	.291	.077
141 142	42811 42812	UANT UANT	News about the progress made in space technology should never be dull.	.597	.430
1261 1262	42821 42822	UANP UANP	Most people don't care about the progress made in space technology.	.639	.620

REFERENCES

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. The authoritarian personality. New York: Harper, 1950.

Armstrong, J. S., & Soelberg, P. On the interpretation of factor analysis. Psychological Bulletin, 1968, 70, 361-364.

Bargmann, R. E. A study of independence and dependence in multivariate normal analysis. University of North Carolina Institute of Statistics Mimeographed Series, No. 186, Chapel Hill, 1957.

Bartlett, M. S. Properties of sufficiency and statistical tests. Proceedings of the Royal Society, 1937, 160, 268-282.

Bartlett, M. S. Internal and external factor analysis. British Journal of Psychology (Statistical Section), 1948, 1, 73-81.

Bartlett, M. S. Tests of significance in factor analysis. British Journal of Psychology (Statistical Section), 1950, 3, 77-85.

Bartlett, M. S. A further note on tests of significance in factor analysis. British Journal of Psychology (Statistical Section), 1951, 4, 1-2. (a)

Bartlett, M. S. The effect of standardization on a X^2 approximation in factor analysis. Biometrika, 1951, 38, 337-344.

Bass, B. M. Authoritarianism or acquiescence? Journal of Abnormal and Social Psychology, 1955, 51, 616-623.

Bass, B. M. Reply to Messick and Jackson's comments on authoritarianism or acquiescence. Journal of Abnormal and Social Psychology, 1957, 54, 426-427.

Bendig, A. W. An inter-item factor analysis of the California F (Authoritarianism) Scale. Journal of Psychological Studies, 1959, 11, 22-26.

Bendig, A. W. A further factor analysis of the California F (Authoritarianism) Scale. Journal of Psychological Studies, 1960, 11, 248-252.

Bentler, P. M. CLUSTRAN, a program for oblique transformation. Behavioral Science, 1971, 16, 183-185.

Bentler, P. M. An analysis of responses to adjectives: A reply to Samelson. Psychological Bulletin, 1973, 80, 133-134.

Bentler, P. M., Jackson, D. N., & Messick, S. The identification of content and style: A two-dimensional interpretation of acquiescence. Psychological Bulletin, 1971, 76, 186-204.

Bentler, P. M., Jackson, D. N., & Messick, S. A rose by any other name. Psychological Bulletin, 1972, 77, 109-133.

Block, J. The challenge of response sets. New York: Appleton-Century Crofts, 1965.

Block, J. Remarks on Jackson's "Review" of Block's Challenge of response sets. Educational and Psychological Measurement, 1967, 27, 497-502.

Block, J. On further conjectures regarding acquiescence. Psychological Bulletin, 1971, 76, 205-210.

Block, J. The shifting definitions of acquiescence. Psychological Bulletin, 1972, 78, 10-12.

Bock, R. D. Components of variance analysis as a structural and discriminal analysis for psychological tests. British Journal of Psychology, 1960, 13, 151-163.

Bock, R. D. Components of variance due to content and acquiescence in the Hy and Pt scales of the MMPI (Res. Memo 21). Chapel Hill, N. C.: University of North Carolina Psychometric Laboratory, 1964.

Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Bock, R. D., & Bargmann, R. E. Analysis of covariance structures. Psychometrika, 1966, 31, 507-534.

Bock, R. D., Dicken, C., & Van Pelt, J. Methodological implications of content-acquiescence correlation in the MMPI. Psychological Bulletin, 1969, 71, 127-139.

Box, G. E. P. A general distribution theory for a class of likelihood criteria. Biometrika, 1949, 36, 317-346.

Bramble, W. J. Sequential testing of models for the analysis of covariance structures. Unpublished doctoral dissertation. University of Chicago, 1971.

Bramble, W. J., & Wiley, D. E. Estimating content-acquiescence correlation by covariance structure analysis. Multivariate Behavioral Research, 1974, 9, 179-190.

Broverman, D. M., & Klaiber, E. L. Negative relationships between abilities. Psychometrika, 1969, 34, 5-20.

Browne, M. W. A comparison of factor-analytic techniques. Psychometrika, 1968, 33, 267-334.

Burt, C. Factor analysis and analysis of variance. British Journal of Psychology (Statistical Section), 1947, 1, 3-26.

Byrne, D. An introduction to personality: A research approach. Englewood Cliffs, N. J.: Prentice-Hall, 1966.

Camilleri, S. F. A factor analysis of the F-scale. Social Forces, 1959, 37, 316-323.

Cattell, R. B. Psychiatric screening of flying personnel: Personality structure in objective tests--A study of 1,000 Air Force students in basic pilot training (Project No. 21-0202-0007, Report No. 9). Randolph Field, Tex.: Air University, USAF School of Aviation Medicine, 1955.

Cattell, R. B., Dubin, S. S., & Saunders, D. R. Personality structure in psychotics by factorization of objective clinical tests. Journal of Mental Science, 1954, 100, 154-176. (a)

Cattell, R. B., Dubin, S. S., & Saunders, D. R. Verification of hypothesized factors in one hundred and fifteen objective personality test designs. Psychometrika, 1954, 19, 209-230. (b)

Cattell, R. B., & Gruen, W. The primary personality factors in 11-year-old children, by objective tests. Journal of Personality, 1955, 23, 460-478.

Cattell, R. B., & Peterson, D. R. Personality structure in four and five year olds in terms of objective tests. Journal of Clinical Psychology, 1959, 15, 355-369.

Cattell, R. B., & Scheier, I. H. Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire, and physical factors. Journal of General Psychology, 1959, 61, 287-315.

Cattell, R. B., & Scheier, I. H. Stimuli related to stress, neuroticism, excitation, and anxiety response pattern: Illustrating a new multivariate experimental design. Journal of Abnormal and Social Psychology, 1960, 60, 195-204.

Chanda, K. C. A note on the consistency and maxima of the roots of likelihood equations. Biometrika, 1954, 41, 56-61.

Chapman, L. J., & Bock, R. D. Components of variance due to acquiescence and content in the F-scale measures of acquiescence. Psychological Bulletin, 1958, 55, 328-333.

Chapman, L. J., & Campbell, D. T. Response set in the F-scale. Journal of Abnormal and Social Psychology, 1957, 54, 129-132.

Chapman, L. J., & Campbell, D. T. The effect of acquiescence response-set upon relationships among the F-scale. Sociometry, 1959, 22, 153-161.

Christie, R., Havel, J., & Seidenberg, B. Is the F-scale reversible? Journal of Abnormal and Social Psychology, 1958, 56, 143-159.

Clayton, M. B., & Jackson, D. N. Equivalence range, acquiescence, and overgeneralization. Educational and Psychological Measurement, 1961, 21, 371-382.

Cochran, W. G., & Cox, G. M. Experimental designs. (2nd ed.) New York: Wiley, 1957.

Cofer, C. N., Chance, J., & Judson, A. J. A study of malingering on the MMPI. Journal of Psychology, 1949, 27, 491-499.

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.

Cooley, W. W., & Lohnes, P. R. Multivariate data analysis. New York: Wiley, 1971.

Cornfield, J., & Tukey, J. W. Average values of mean squares in factorials. Annals of Mathematical Statistics, 1956, 27, 907-949.

Couch, A., & Keniston, K. Yeasayers and naysayers: Agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 1960, 60, 151-174.

Cronbach, L. J. Studies of acquiescence as a factor in the true-false test. Journal of Educational Psychology, 1942, 33, 401-415.

Cronbach, L. J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475-494.

Cronbach, L. J. Further evidence on response sets and test design. Educational and Psychological Measurement, 1950, 10, 3-31.

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

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley, 1972.

Cuadra, C. A., & Reed, C. F. An introduction to the MMPI. Downey, Ill.: Veteran's Administration Hospital, 1954.

Damarin, F., & Messick, S. Response styles as personality variables: A theoretical integration of multivariate research (ETS RB 65-10). Princeton, N. J.: Educational Testing Service, 1965.

Davidon, W. C. Variable metric method for minimization. AEC Research and Development Report, ANL-5990 (Rev.) Washington D. C.: U. S. Government Printing Office, 1959.

Dicken, C. "Acquiescence" in the MMPI: A method variance artifact? Psychological Reports, 1967, 20, 927-933.

Dicken, C., & Van Pelt, J. Further evidence concerning acquiescence and the MMPI. Psychological Reports, 1967, 20, 935-941.

Draper, N. R., & Smith, H. Applied regression analysis. New York: Wiley, 1966.

Duncan, O. D. Path analysis: Sociological examples. American Journal of Sociology, 1966, 72, 1-16.

Ebel, R. L. How to write true-false test items. Educational and Psychological Measurement, 1971, 31, 417-426.

Edwards, A. L. The social desirability variable in personality assessment and research. New York: Dryden, 1957.

Fletcher, R., & Powell, M. J. D. A rapidly convergent descent method for minimization. The Computer Journal, 1963, 2, 163-168.

Frederiksen, N., & Messick, S. Response set as a measure of personality. Educational and Psychological Measurement, 1959, 19, 137-157.

Fricke, B. G. Response set as a suppressor variable in the OAIS and MMPI. Journal of Consulting Psychology, 1956, 20, 161-169.

French, J. W., Ekstrom, R. B., & Price, L. A. Manual for kit of reference tests for cognitive factors. (Rev. ed.) Princeton, N. J.: Educational Testing Service, 1963.

Fritz, M. F. Guessing in a true-false test. Journal of Educational Psychology, 1927, 18, 558-561.

Gage, N. L., & Chatterjee, B. B. The psychological meaning of acquiescence set: Further evidence. Journal of Abnormal and Social Psychology, 1960, 60, 280-283.

Gage, N. L., Leavitt, G. S., & Stone, G. C. The psychological meaning of acquiescence set for authoritarianism. Journal of Abnormal and Social Psychology, 1957, 55, 98-103.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H. B., & Spence, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. Psychological Issues, 1959, 1, No. 4 (Monogr. 4).

Gardner, R. W., Jackson, D. N., & Messick, S. Personality organization in cognitive controls and intellectual abilities. Psychological Issues, 1960, 2, No. 4 (Monogr. 8).

Gorsuch, R. L. Factor analysis. Philadelphia: Saunder, 1974.

Gough, H. G. A nonintellectual intelligence test. Journal of Consulting Psychology, 1953, 17, 242-246.

Graybill, F. A. Introduction to matrices with applications in statistics. Belmont, Calif.: Wodsworth, 1969.

Gruvaeus, G. T., & Jöreskog, K. G. A computer program for minimizing a function of several variables (ETS RB 70-14). Princeton, N. J.: Educational Testing Service, 1970.

Guilford, J. P. Psychometric methods. (2nd ed.) New York: McGraw-Hill, 1954.

Guilford, J. P., & Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Guttman, L. Some necessary conditions for common-factor analysis. Psychometrika, 1954, 19, 149-161.

Hakstian, A. R., & Muller, V. J. Some notes on the number of factors problem. Multivariate Behavioral Research, 1973, 8, 461-475.

Hathaway, S. R., & McKinley, J. C. MMPI manual. (Rev. ed.) New York: Psychological Corporation, 1967.

Helmstadter, G. C. Procedures for obtaining separate set and content components of a test score. Psychometrika, 1957, 22, 381-383.

Hocking, R. R. A discussion of the two-way mixed model. The American Statistician, 1973, 27, 148-151.

Horn, J. L. On subjectivity in factor analysis. Educational and Psychological Measurement, 1967, 27, 811-820.

Horn, J. L., & Knapp, J. R. On the subjective character of the empirical base of Guilford's structure-of-intellect model. Psychological Bulletin, 1973, 80, 33-43.

Householder, A. S. Principles of numerical analysis. New York: McGraw-Hill, 1953.

Humphreys, L. G., Ilgen, D., McGrath, D., & Montanelli, R. Capitalization on chance in the rotation of factors. Educational and Psychological Measurement, 1969, 29, 259-272.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. Journal of the American Statistical Association, 1970, 65, 1582-1589.

Jackson, D. N. Acquiescence response styles: Problems of identification and control. In I. A. Berg (Ed.), Response set in personality assessment. Chicago: Aldine, 1967. (a)

Jackson, D. N. Balanced scales, item overlap, and the stables of Augeas. Educational and Psychological Measurement, 1967, 27, 503-507. (b)

Jackson, D. N. Review of J. Block, The challenge of response sets. Educational and Psychological Measurement, 1967, 27, 207-219. (c)

Jackson, D. N. Personality Research Form manual. Goshen, N. Y.: Research Psychologists Press, 1974.

Jackson, D. N., & Lay, C. H. Homogeneous dimensions of personality scale content. Multivariate Behavioral Research, 1968, 3, 321-338.

Jackson, D. N., & Messick, S. A note on "ethnocentrism" and acquiescent response sets. Journal of Abnormal and Social Psychology, 1957, 54, 132-134.

Jackson, D. N., & Messick, S. Content and style in personality assessment. Psychological Bulletin, 1958, 55, 243-252.

Jackson, D. N., & Messick, S. Acquiescence and desirability as response determinants on the MMPI. Educational and Psychological Measurement, 1961, 21, 771-790.

Jackson, D. N., & Messick, S. Response styles on the MMPI: Comparison of clinical and normal samples. Journal of Abnormal and Social Psychology, 1962, 65, 285-299. (a)

Jackson, D. N., & Messick, S. Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), Measurement in personality and cognition. New York: Wiley, 1962. (b)

Jackson, D. N., & Messick, S. Acquiescence: The nonvanishing variance component. American Psychologist, 1965, 20, 498 (Abstract).

Jackson, D. N., Messick, S., & Solley, C. M. How "rigid" is the "authoritarian"? Journal of Abnormal and Social Psychology, 1957, 54, 137-140.

Jackson, D. N., & Morf, M. E. An empirical evaluation of factor reliability. Multivariate Behavioral Research, 1973, 8, 439-459.

Jackson, D. N., & Morf, M. E. Testing the null hypothesis for rotation to a target. Multivariate Behavioral Research, 1974, 9, 303-309.

Johnston, J. Econometric methods. (2nd ed.) New York: McGraw-Hill, 1972.

Jordan, L. A. Canonical component analysis: A new twist for interpreting interbattery relationships. Paper presented at the annual meeting of the American Psychological Association, Chicago, September, 1975.

Jöreskog, K. G. Testing a simple structure hypothesis in factor analysis. Psychometrika, 1966, 31, 165-178.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1967, 32, 443-482.

Jöreskog, K. G. A general approach to confirmatory maximun likelihood Psychometrika, 1969, 34, 183-202.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Jöreskog, K. G. Statistical analysis of sets of congeneric tests. Psychometrika, 1971, 36, 109-133.

Jöreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology. Vol. 2. Measurement, psychophysics, and neural information processing. San Francisco: Freeman, 1974.

Jöreskog, K. G., Gruvaeus, G. T., & van Thillo, M. ACOVA—A general computer program for analysis of covariance structures (ETS RB 70-15). Princeton, N. J.: Educational Testing Service, 1970.

Jöreskog, K. G., & Lawley, D. N. New methods in maximum likelihood factor analysis. British Journal of Mathematical and Statistical Psychology, 1968, 21, 85-96.

Jöreskog, K. G., & van Thillo, M. New rapid algorithms for factor analysis by unweighted least squares, generalized least squares and maximum likelihood (ETS RM 71-5). Princeton, N. J.: Educational Testing Service, 1971.

Jöreskog, K. G., & van Thillo, M. LISREL--A general computer program for estimating a linear structural equation involving multiple indicators of unmeasured variables (ETS RB 72-56). Princeton, N. J.: Eductional Testing Service, 1972.

Jöreskog, K. G., van Thillo, M., & Gruvaeus, G. T. ACOVSM--A general computer progra for analysis of covariance structures including generalized MANOVA (ETS RB 71-1). Princeton, N. J.: Educational Testing Service, 1971.

Kaiser, H. F. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 1960, 20, 141-151.

Kaiser, H. F. A note on Guttman's lower bound for the number of common factors. British Journal of Statistical Psychology, 1961, 14, 1-2.

Kaiser, H. F. A second generation little jiffy. Psychometrika, 1970, 35, 401-415.

Kassebaum, G. G., Couch, A. S., & Slater, P. E. The factorial dimensions of the MMPI. Journal of Consulting Psychology, 1959, 23, 226-236.

Kerlinger, F. N., & Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston, 1973.

Kleinmuntz, B. Personality measurement: An introduction. Homewood, Ill.: Dorsey, 1967.

Krug, R. E. An analysis of the F scale. I. Item factor analysis. Journal of Social Psychology, 1961, 53, 285-291.

Krug, R. E., & Moyer, K. E. An analysis of the F scale. II. Relationship to standard personality inventories. Journal of Social Psychology, 1961, 53, 292-301.

Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society in Edinburgh, 1940, 60, 64-82.

Leavitt, H. J., Hax, H., & Roche, J. H. Authoritarianism and agreement with things authoritative. Journal of Psychology, 1955, 40, 215-221.

Lentz, T. F. Acquiescence as a factor in the measurement of personality. Journal of General Psychology, 1939, 35, 659 (Abstract).

Lichtenstein, E., & Bryan, J. H. Acquiescence and the MMPI: An item reversal approach. Journal of Abnormal Psychology, 1965, 70, 290-293.

Mauchly, J. W. Significance test for sphericity of a normal n-variate distribution. Annals of Mathematical Statistics, 1940, 11, 204-209.

McDonald, R. P. The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. British Journal of Mathematical and Statistical Psychology, 1970, 23, 1-20.

Messick, S. Separate set and content scores for personality and attitude scales. Educational and Psychological Measurement, 1961, 21, 915-923.

Messick, S. Response style and content measures from personality inventories. Educational and Psychological Measurement, 1962, 22, 41-56.

Messick, S. The psychology of acquiescence: An interpretation of research evidence. In I. A. Berg (Ed.), Response set in personality assessment. Chicago: Aldine, 1967.

Messick, S., & Frederiksen, N. Ability, acquiescence, and "authoritarianism." Psychological Reports, 1958, 4, 687-697.

Messick, S., & Jackson, D. N. Authoritarianism or acquiescence in Bass's data. Journal of Abnormal and Social Psychology, 1957, 54, 424-426.

Messick, S., & Jackson, D. N. The measurement of authoritarian attitudes. Educational and Psychological Measurement, 1958, 18, 241-253.

Messick, S., & Jackson, D. N. Acquiescence and the factorial interpretation of the MMPI. Psychological Bulletin, 1961, 58, 299-304.

Montanelli, R. G. Jr. The goodness of fit of the maximum-likelihood esitmation procedure in factor analysis. Educational and Psychological Measurement, 1974, 34, 547-562.

Morf, M. E. An analysis of two response sets: True responding and item endorsement. Unpublished doctoral dissertation. University of Western Ontario, 1968.

Morf, M. E., & Jackson, D. N. An analysis of two response styles: True responding and item endorsement. Educational and Psychological Measurement, 1972, 32, 329-354.

Mulaik, S. A The foundations of factor analysis. New York: McGraw-Hill, 1972.

Peabody, D. Two components in bipolar scales: Direction and extremeness. Psychological Review, 1962, 69, 65-73.

Peabody, D. Models for estimating content and set components in attitude and personality scales. Educational and Psychological Measurement, 1964, 24, 255-269.

Peabody, D. Authoritarianism scales and response bias. Psychological Bulletin, 1966, 65, 11-23.

Rao, C. R. Linear statistical inference and its implications. (2nd ed.) New York: Wiley, 1973.

Rorer, L. G. The great response-style myth. Psychological Bulletin, 1965, 63, 129-156.

Rorer, L. G., & Goldberg, L. S. Acquiescence and the vanishing variance component. Journal of Applied Psychology, 1965, 49, 422-430. (a)

Rorer, L. G., & Goldberg, L. R. Acquiescence in the MMPI? Educational and Psychological Measurement, 1965, 25, 801-817. (b)

Roy, S. N. p-statistics or some generalizations in analysis of variance appropriate to multivariate problems. Sankhya, 1939, 4, 381-396.

Samelson, F. Response style: A psychologist's fallacy? Psychological Bulletin, 1972, 78, 13-16.

Schaefer, E. S. A circumplex model for maternal behavior. Journal of Abnormal and Social Psychology, 1959, 59, 226-235.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Schönemann, P. H. A generalized solution of the orthogonal Procrustes problem. Psychometrika, 1966, 31, 1-10.

Searle, S. R. Linear models. New York: Wiley, 1971.

Shapiro, D. Neurotic styles. New York: Basic Books, 1965.

Sörbom, D. A general method for studying group differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 1974, 27, 229-239.

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational measurement. (2nd ed.) Washington D. C.: American Council on Education, 1971.

Sullivan, H. S. Clinical studies in psychiatry. New York: Norton, 1956.

Tellegen, A. Direction of measurement: A source of misinterpretation. Psychological Bulletin, 1965, 63, 233-243.

Thurstone, L. L. Multiple-factor analysis. Chicago: University of Chicago Press, 1947.

Titus, H. E., & Hollander, S. P. The California F scale in psychological research: 1950-1955. Psychological Bulletin, 1957, 54, 47-64.

Tobias, S., & Carlson, J. E. Brief report: Bartlett's test of sphericity and chance findings in factor analysis. Multivariate Behavioral Research, 1969, 4, 375-377.

Tryon, R. C., & Bailey, D. E. Cluster analysis. New York: McGraw-Hill, 1970.

Tucker, L. R., Koopman, R. F., & Linn, R. L. Evaluation of factor analytic research procedures by means of simulated correlation matrices. Psychometrika, 1969, 34, 421-459.

Tucker, L. R., & Lewis, C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 1973, 38, 1-10.

van Thillo, M., & Jöreskog, K. G. SIFASP--A general computer program for simultaneous factor analysis in several populations (ETS RB 70-62). Princeton, N. J.: Educational Testing Service, 1970.

Walker, H. M., & Lev, J. Elementary statistical methods. 3rd ed. New York: Holt, Rinehart & Winston, 1969.

Welsh, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: University of Minnesota Press, 1956.

Wesman, A. G. The usefulness of correctly spelled words in a spelling test. Journal of Educational Psychology, 1947, 37, 242-246.

Wiggens, J. S. Substantive dimensions of self-report in the MMPI item pool. Psychological Monographs, 1966, 80 (22), Whole No. 630.

Wiley, D. E., Schmidt, W. H., & Bramble, W. J. Studies of a class of covariance structure models. Journal of the American Statistical Association, 1973, 68, 317-323.

Wilks, S. S. Certain generalizations in analysis of variance. Biometrika, 1932, 24, 471.

Wilks, S. S. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. Annals of Mathematical Statistics, 1946, 17, 257-281.

Winer, B. J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.