

A RAND NOTE

**User's Guide for the Multitrait
Analysis Program (MAP)**

**Ron D. Hays, Toshi Hayashi,
Sally Carson, John E. Ware, Jr.**

November 1988

This publication was supported by The RAND Corporation as part of its program of public service.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.

A RAND NOTE

N-2786-RC

**User's Guide for the Multitrait
Analysis Program (MAP)**

**Ron D. Hays, Toshi Hayashi,
Sally Carson, John E. Ware, Jr.**

November 1988

PREFACE

This Note represents a how-to-use manual for the Multitrait Analysis Program (MAP), a FORTRAN-77 program that performs multitrait scaling analysis. No previous experience with the program is assumed in this manual, but a basic understanding of the multitrait scaling methodology is supposed. MAP was developed for the Medical Outcomes Study (MOS) and has been used extensively for analysis of data from that project as well as many others. This manual is designed for the version of MAP that runs on the VAX-780 at RAND, although most of the documentation is equally applicable to the microcomputer version of MAP (available from National Collegiate Software Clearinghouse, Box 8101, North Carolina State University, Raleigh, NC 27695).

MAP had its roots in the Analysis of Item Test Homogeneity (ANLITH) program developed by Thomas Gronck and Thomas Tyler (Schnittjer and Cartledge, 1976). ANLITH was further developed and refined during the Measuring Health Concepts Research Project conducted at Southern Illinois School of Medicine, 1972-1975, sponsored by the National Center for Health Services Research. At the start of the RAND Health Insurance Experiment, Patricia Camp and William Rogers created a PL/1 program that takes SAS input, converts it to ANLITH FORTRAN-readable format, runs the ANLITH program, and outputs a SAS-derived variable file.

For MOS, we developed an ANLITH-like program compatible with the MOSVAX IBM PC (Microsoft FORTRAN) environments. Our goals were to improve the user-friendliness of the output of the program, make the program IBM PC- and FORTRAN 77-compatible such that it could be used on both a VAX-780 and on IBM-compatible microcomputers, to make user input easier, to add some additional statistics and a multitrait summary table, and to segment the output so that the main output contained only essential material.

The development of MAP was made possible by grants to MOS from The Robert Wood Johnson Foundation, the Henry J. Kaiser Family Foundation, and the Pew Charitable Trusts.

SUMMARY

Multitrait scaling is a straightforward yet elegant methodology for scale analysis. The procedure involves examining item frequencies; item and scale descriptive statistics (e.g., mean, standard deviation, variance); scale internal consistency estimates; item-scale correlations (corrected for overlap); and correlations among scales. The Multitrait Analysis Program (MAP) provides the statistics needed for multitrait scaling analysis.

To use MAP on the VAX, the user must supply two input files, a raw data file and an input specification file (*map.in*) containing a series of keywords. This user's guide specifies how to set up the MAP input file, provides an example of using MAP on the VAX-780, and summarizes published examples of multitrait scaling.

ACKNOWLEDGMENTS

We would like to thank Anita Stewart, Cathy Sherbourne, Elizabeth McGlynn, Kenneth White, Sharon Arnold, and Haya Rubin for serving as guinea pig users of MAP. The excellent secretarial support of Kim Lavoie and Nelie Gill was greatly appreciated. Comments on an earlier draft of this document provided by Cathy Sherbourne and editorial suggestions by Judith Westbury were especially helpful.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	vii
TABLES	xi
Section	
I. MULTITRAIT SCALING	1
II. WHAT MAP PROVIDES	5
III. HOW TO USE MAP ON THE VAX-780	9
EXAMPLE OF USING MAP ON THE VAX-780	14
IV. PUBLISHED EXAMPLES OF MULTITRAIT SCALING	22
REFERENCES	25

TABLES

1. Example of <i>map.in</i> file for VAX-780	15
2. Example of partial <i>map.out</i> file	18
3. Example of partial <i>mapscale.out</i> file	21

I. MULTITRAIT SCALING

Multitrait scaling is a straightforward yet elegant methodology for scale analysis. When responses to multiple questionnaire items are summed into a single scale score, it is generally termed a summated or a "Likert-type" scale. Summated scales are constructed by summing the items in each hypothesized scale and assigning equal weights to the items. There are a number of analyses that must be performed to determine whether a set of items can be appropriately combined into a summated ratings scale. In multitrait scaling, construction of summated rating scales from item groupings follows four steps. These steps add several scaling criteria to those usually associated with Likert scaling. These steps are designed to determine whether:

- (1) each item in a hypothesized grouping is substantially linearly related to the total score computed from other items in that group (a traditional convergent validity criterion usually expressed in terms of internal-consistency),
- (2) each item correlates higher with the construct it is hypothesized to measure than it correlates with other constructs (item discriminant validity criterion),
- (3) items in the same scale contain the same proportion of information about the construct or should be given different weights (test for equal item-total correlations),
- (4) items measuring the same construct have equal variances and therefore do not need to be standardized before combining them in the same scale (equal variances criterion).

If items in each hypothesized grouping satisfy these assumptions, simple summation of items to derive a scale score is appropriate. If the first two scaling criteria are not satisfied, item groupings should be revised. If the third scaling criterion is not satisfied, unequal weights may be appropriate for different items. Items should be standardized prior to combining whenever their variances differ (fourth

scaling criterion). Equality of variance may be assessed using multiple range tests (Levy, 1975).

Multitrait scaling involves examining item frequencies; item and scale descriptive statistics (e.g., mean, standard deviation, variance); scale internal consistency estimates; item-scale correlations (corrected for overlap); and correlations among scales. Multitrait scaling goes beyond traditional tests of internal-consistency primarily because it tests item discrimination across scales (step two above). Thus, items are evaluated with respect to how well they represent a particular construct *relative* to other constructs.

Item-scale correlations are the fundamental elements of multitrait scaling (Ware, Snyder, Wright, and Davies, 1983). These item-scale correlations constitute what we refer to as the multitrait/multi-item (MTMI) correlation matrix. The first two steps in multitrait scaling analysis involve examining the MTMI matrix. Each row of the matrix contains correlations between scores for one item and all hypothesized item groupings (constructs defined by scales). Each column contains correlations between the scores for one scale and all items in the analysis, including those hypothesized to be part of that scale and those hypothesized to be part of other scales. Correlations between items hypothesized to be in a given scale and the scale itself are corrected for overlap (using the technique recommended by Howard and Forehand, 1962) so that estimates of the item-scale relationships are not spuriously inflated.

Item convergent validity is supported if an item correlates substantially (corrected correlation of 0.30 or above) with the scale it is hypothesized to represent. This is the traditional internal-consistency criterion. For scales that have a previous history of development, and for which analyses are intended to refine and add finishing touches rather than develop from scratch, we advocate using a more stringent convergent validity criterion of 0.40.

Satisfaction of the second multitrait scaling criterion is obtained if the highest correlation in a row of the MTMI matrix is the correlation between the item and the scale defining the construct it is hypothesized to measure. This is a test of discriminant validity

(following Campbell and Fiske, 1959).¹ Item discriminant validity is supported and a scaling "success" counted whenever the correlation between an item and its hypothesized scale is more than two standard errors higher than its correlation with other scales. When a correlation between an item and its hypothesized scale is more than two standard errors below its correlation with another scale, a "definite" scaling error is counted. When the correlation between the item and another scale in the same row is within two standard errors of its correlation with its hypothesized scale, a "probable" scaling error is counted.²

Items that consistently account for definite scaling errors should usually be excluded from the scale in question. Inclusion or exclusion of items associated with probable scaling errors depends on several factors, including the number of subjects in the analysis, the number of items in the scale, and the strength of associations between the constructs involved.

There are no cut-and-dry rules for deciding what to do when probable scaling errors are present. Nonetheless, probable scaling errors should not be a cause for undue concern if the analysis sample is less than 100 (standard error of correlation > 0.10) and the correlation between an item and its hypothesized scale is larger than its correlation with other scales. (However, sample sizes of 100 or larger are recommended to achieve reasonably accurate estimates of item-scale correlations.) As the number of items in a scale increases, the tolerance of probable errors may increase because a sufficient number of items may be available to anchor the construct distinctly from other constructs. For example, an item in a 5-item scale that displays a

¹The multitrait/multimethod (MTMM) approach developed by Campbell and Fiske (1959) is based on convergent and discriminant validity. However, their technique is more thorough than multitrait scaling because it involves using more than one measurement method (cf. Hayashi and Hays, 1987). In multitrait scaling, we use only one method but evaluate more than one trait in a matrix. Multitrait scaling can be regarded as a cousin of the MTMM methodology.

²The significance of the difference can be evaluated using Steiger's (1980) t-test for dependent correlations. We employ the two-standard-errors rule described here for ease of application and interpretation.

probable error may more easily be tolerated than an item in a 2-item scale. If measures are being constructed that are known to be related theoretically (e.g., loneliness and alienation), probable errors associated with items in the two scale groupings are more likely to be tolerated, at least early in the process of scale development refinement.

II. WHAT MAP PROVIDES

From raw data input, MAP provides:

- number of subjects in the analysis
- number of subjects omitted for missing data
- item descriptive statistics (frequency distribution, mean, standard deviation)
- scale descriptive statistics (mean, standard deviation, lowest observed score, highest observed score, range)
- item-scale correlations (corrected for overlap)
- item-scale multitrait summary table (indicating scaling successes, failures, probable failures)
- Cronbach's alpha (R_{TT})
- alpha estimated for a 10-item scale (R_{10})
- intraclass correlation for items (R_{II})
- correlations among scales (uncorrected and adjusted for attenuation or lack of perfect reliability)
- Scott's homogeneity ratio (SCOTT)
- coefficient of homogeneity of persons (R_{GG})
- intraclass correlation for persons (R_{PP})

The number of subjects included and excluded in the analysis is provided to give the user a handle on the influence of missing data on the representativeness of the analysis sample. Item descriptive statistics are given so that the user can determine whether items are behaving as expected and for evaluation of item discrimination across respondents. Scale descriptive statistics provide similar information for the hypothesized scales. Item-scale correlations are the fundamental elements of the MAP program and are used for multitrait scaling decisions. The multitrait summary table is an aid for the user that helps identify scaling successes, failures, and probable failures.

Cronbach's alpha is a standard measure of internal consistency reliability. It is estimated for each scale and also estimated assuming that each scale had 10 items and 1 item (intraclass correlation). The latter estimates provide a way of comparing scales of different lengths by controlling for the number of items. These estimates are especially useful if one is interested in evaluating how much "bang-for-the-buck" the average item in a scale contributes to internal consistency of the scale. Scott's homogeneity ratio is a "weighted average inter-item correlation, in which the correlation between every pair of items is weighted by the geometric mean of their variances" (Scott, 1968, p. 254). If items are standardized (means centered at zero, standard deviations fixed at 1.0), Scott's homogeneity ratio and the intraclass correlation for items are identical. For unstandardized items, these indexes are similar but not identical to one another. Correlations among scales are provided as supplementary information to that given by item-scale correlations.

The coefficient of homogeneity for persons and the intraclass correlation for persons are analogous to alpha and the intraclass correlation for items, but these indexes are person-based rather than item-based. The coefficient of homogeneity for persons is not very useful because it depends on the number of subjects in the analysis. The intraclass correlation for persons can be "interpreted as the average correlation between persons" (Tyler and Fiske, 1968, p. 774). These indexes have no role in traditional multitrait scaling analyses, but they may be useful for special applications (cf. Tyler and Fiske, 1968). They are included in MAP for those who wish to compare item-level with person-based statistics.

The following paragraphs describe the computation of Cronbach's alpha, the intraclass correlation for items, Scott's homogeneity ratio, the coefficient of homogeneity of persons, and the intraclass correlation for persons.

Cronbach's alpha is calculated using the analysis of variance approach to reliability (Guilford, 1954). The analysis of variance model is a one-way repeated measures design with items functioning as

the repeated measures. Hoyt's (1941) formula for reliability is employed:³

$$\alpha = 1 - \frac{MS_{RI}}{MS_R} = \frac{MS_R - MS_{RI}}{MS_R}$$

Calculation formulas for the intraclass correlation for items (R_{II}), Scott's homogeneity ratio (SCOTT), the coefficient of homogeneity of persons (R_{GG}), and the intraclass correlation for persons (R_{PP}) are provided below (K = number of items in scale, S = standard deviation, N = number of respondents):

$$R_{II} = \alpha / (K + \alpha - (K * \alpha))$$

$$SCOTT = (MS_R - MS_I) / ((\epsilon S_I)^2 - MS_I)$$

$$R_{GG} = 1 - (MS_{RI} / MS_I)$$

$$R_{PP} = R_{GG} / (N + R_{GG} - (N * R_{GG}))$$

The computation of item-scale correlations is presented below. Uncorrected item-scale correlations are not printed, but they are computed first, using the following formula (N = number of respondents, J = item, W = scale with "J" in it, S = standard deviation):

$$r_u = \frac{N * \epsilon X_{JW} - (\epsilon X_J)(\epsilon X_W)}{N^2 S_W S_J}$$

³ MS_{RI} = mean square for respondents x items; MS_R = mean square for respondents; MS_I = mean square for items.

Finally, corrected item-scale correlations are computed using the following formula:

$$r_c = \frac{r_u S_w - S_J}{[S_w^2 + S_J^2 - 2(r_u)(S_w)(S_J)]^{1/2}}$$

III. HOW TO USE MAP ON THE VAX-780

To use MAP on the VAX, the user must supply two input files, a raw data file and an input specification file (this file must be named *map.in*). The raw data file is an ASCII file. Because zeros are treated as missing data by the MAP program, it is important that legitimate values be coded as something other than zero. The file should not contain any out-of-range values because MAP stops execution if such is the case, unless the user invokes the "OVERRIDE" job option. In the raw data file, the subject ID should appear first before other variables. Following the subject ID should be any variables that will be used to create subsets of observations (see "IFVARS and IF" subsection below). All analysis variables can then follow in any order. Based on the input files, MAP generates the program output (stored in *map.out*) and writes scale scores to a file (*mapscale.out*), if so desired. MAP writes information about imputation of scores for missing data and out-of-range values to a separate file, *mapdata.out*. Program processing information, useful for debugging, is written to yet another file, *debug*.

The *mapdata.out* file lists the subject identification number, variable name, and recoded value for imputed data. Subject identification number and the scale that leads to omission are listed for omitted subjects. An excerpt from a *mapdata.out* file is shown below.

```
79, Q8      MISSING, SET TO  3.000 (MEAN NON-MISSING IN SCALE OVHQ9  )
79, Q43     MISSING, SET TO  4.000 (MEAN NON-MISSING IN SCALE DISCHAR3)
79, Q53     MISSING, SET TO  4.000 (MEAN NON-MISSING IN SCALE DISCHAR3)
79, OMITTED, ALL ITEMS IN SCALE VALUE1 MISSING (OTHERS MAY BE MISSING ALSO)
```

The *debug* file provides a listing of the beginning and end of each keyword as it is processed by MAP. If no errors are found in the input file, *debug* ends with the statement END OF PROGRAM CONTROL INPUT. An example *debug* file is given below. The example listed below illustrates an instance in which an input error is present. MAP begins processing

the second scale it reads, but execution is terminated before the end of processing (see below). In this example, the user omitted one of the items in the second scale, and MAP detected this when it read the keyword SCALE (for the third scale) before finishing the reading of items for the second scale.

PROGRAM CONTROL INPUT PROCESSING STATUS --

```
BEGIN PROCESSING KEYWORD TITLES
  END PROCESSING KEYWORD TITLES
BEGIN PROCESSING KEYWORD RAWDATA
  END PROCESSING KEYWORD RAWDATA
BEGIN PROCESSING KEYWORD JOBOPTIONS
  END PROCESSING KEYWORD JOBOPTIONS
BEGIN PROCESSING KEYWORD HOWREAD
  END PROCESSING KEYWORD HOWREAD
BEGIN PROCESSING KEYWORD CRITERIA
  END PROCESSING KEYWORD CRITERIA
BEGIN PROCESSING KEYWORD SCALE
  END PROCESSING KEYWORD SCALE
BEGIN PROCESSING KEYWORD SCALE
```

The raw data input file contains the raw data that MAP works with. (Because FORTRAN does not distinguish zeros from missing values, all zeros that are meaningful values should be recoded to nonzero values.) The specification file, *map.in*, consists of a series of keywords that include TITLES, RAWDATA, JOBOPTIONS, HOWREAD, ITEMS, CRITERIA, SCALE, HOWWRITE, IFVARS, and IF. Following each keyword the user enters appropriate specifications as described below. The user needs to specify keywords in the order provided here. The contents of each line of *map.in* can be up to 250 columns wide; however, we recommend the use of 80 columns or fewer for everything except the HOWREAD section. Variable names, for example, are limited to 8 characters when printed out, so it does not help to read in extra characters. For examples of how to use the following keywords, see Table 1 below.

TITLES. Two title lines that describe the MAP run must be provided by the user. Each title line is surrounded by single quotes, with *END* appearing before the final single quote on both lines.

RAWDATA. The name of the raw data input file must be given within single quotes.

JOBOPTIONS. Up to three types of job options may be specified: (1) for writing out data--WRITESCA, (2) for estimating missing data--HALFSCAL or COMPLETE, and (3) for terminating the program--OVERRIDE.

For the first option type, one keyword may be optionally specified: WRITESCA. Specification of this keyword causes derived variable scale scores to be written to a file called *mapscale.out*. Scores are written as sums of items in each scale. If WRITESCA is specified, HOWWRITE (see below) must be specified as well.

For the second type of option, one of two keywords, HALFSCAL or COMPLETE, may optionally be selected to control estimation of missing data. By default, respondents are retained in the analysis if they answer at least one item for every scale in the analysis. This default option can be changed by specifying HALFSCAL (respondents must answer at least half the items in each scale) or COMPLETE (respondents must have complete data for all items in the analysis).

The final type of option can be implemented using the OVERRIDE keyword. OVERRIDE, as its name implies, overrides the default setting of immediate termination of MAP when an out-of-range value is encountered. This option is normally not recommended, but it must be used if CRITERIA (see below) is specified (due to a bug in the out-of-range routine that appears when CRITERIA is given).

A slash mark designates the end of JOBOPTIONS.

HOWREAD. The input format specification is listed on a single line, enclosed by left and right parentheses. The subject ID is referenced first in the HOWREAD line using the FORTRAN "A" character format (up to 10 characters may be used), but the ID does not have to be located in the input file (the program does not require a real ID--a "dummy" ID may be read in). Using the FORTRAN "T" pointer format, one can jump to any desired column location on the first card of data. In theory, the subject ID could be located after the first card of data; however, the user will be unable to jump back to read input variables located on cards preceding the subject ID. Following the respondent ID, indicators (items) to be analyzed are read in using FORTRAN "F" format. Next, criteria variables are read in. (Zeros in input data are treated as missing values by MAP.)

ITEMS. Items (indicators) to be input and processed by MAP are named after the ITEMS keyword. Up to 150 items and criteria (see below) may be included, and only items that are used in the job should be specified. Item names are surrounded in single quotation marks. If less than 150 items are input, a slash mark must follow the item input stream.

CRITERIA. MAP calculates item-scale correlations and reliability estimates for a "total" scale which includes all items hypothesized to be in subscales. This total scale is only meaningful if all the scales represent a common higher-order construct. Criteria in MAP are single indicators (not multi-item) such as age or gender that the user wants to keep separate from the total scale score that MAP produces. For example, if one was analyzing dimensions of mental health and there was an interest in an overall mental health score, criteria variables could be separated from the total score by specifying them under the CRITERIA keyword rather than the SCALE keyword. Because of a bug in the out-of-range checking routine when CRITERIA is specified, the OVERRIDE option must be used before MAP will run with CRITERIA.

SCALE. Scales are denoted by the SCALE keyword. Up to 19 scales may be defined. Each scale is preceded by the SCALE keyword. The scale name, surrounded with single quotation marks, and number of items in the scale are given after the keyword. If indicators of a scale are to be standardized, this is denoted by 'STND' followed by a slash; if indicators of a scale are not to be standardized, the number of items is followed by a slash. On lines following the scale name, the user needs to specify (1) the indicators of each scale (names enclosed within single quotes); (2) the action to be performed (P = preserve, no recoding; R = reverse score recoding; C = other categorization recoding); and (3) the minimum and maximum values the indicator can assume (after recoding, if applicable).

If recoding other than reverse coding is desired, it is specified by up to 10 sets of lower bound, higher bound, and recoded categories. The range of recoded values determines the number of sets that must be specified. For example, if the recoded values range from 1-4, four

recode sets are expected by MAP. Thus, if one wants to recode 1-3 to 1 and keep 4 as is, it is necessary to specify four sets as follows: 1 1 1, 2 2 1, 3 3 1, 4 4 4.

Each indicator can be named only in one scale.

HOWWRITE. This option is used only if the user desires to write item scale scores to an output file. The output format section begins with the HOWWRITE keyword. A maximum of six 80-character (including quotation marks) lines are available to describe the output format for the *mapscale.out* file. The output format, like the input format, must begin with a single quote, left parentheses, and end with right parentheses, single quote. In setting up the HOWWRITE field, the subject ID should be described first. Next, output of the items is specified (items recoded in MAP are output in their recoded form). Then, scale score output is specified. The HOWWRITE keyword should appear last in *map.in*.

IFVARS and IF. Analyses of selected subpopulations are possible using the IFVARS and IF keywords. The IFVARS keyword is used to supply the selection variables and their input format. The IF keyword is used to specify selection conditions. The IFVARS and IF sections must be ended with a slash. The syntax of these sections is described below.

The IFVARS section consists of a selection of variable names separated with one or more blanks, followed by their input format enclosed in parentheses. The input format should use the FORTRAN style of edit descriptors delimited by a comma. There are three strict requirements concerning the input format, the location of selection variables, and the location of the IFVARS section. First, the FORTRAN-style input format must start with the T descriptor immediately followed by the absolute column position of the first selection variable. Second, the selection variables must be located, in terms of their column positions, between the ID variable and other analysis variables. In addition, if one observation consists of multiple cards or lines, the selection variables must be on the first card or line.

The IF section consists of selection conditions in a SAS-like format, specifying the disposition of selected observations (i.e., to KEEP or DELETE them), preceded by THEN. The only allowed options of

disposition are KEEP and DELETE. The conditions and disposition are allowed to be on multiple lines separated by one or more blanks. The conditions are specified using three comparison operators, > = <, alone or in combination, and four logical operators, AND & OR and |. But only one layer of parenthesis is accepted in specifying the conditions. The comparison operators are defined as follows.

A = B	A is equal to B
A > B	A is greater than B
A < B	A is smaller than B
A >= B	A is greater than or equal to B
A => B	
A <= B	A is smaller than or equal to B
A =< B	
A <> B	A is not equal to B
A >< B	

EXAMPLE OF USING MAP ON THE VAX-780

An example *map.in* file is given in Table 1. The example is an analysis of patient satisfaction scales in the Medical Outcomes Study. The analysis specifies an assessment of general satisfaction, satisfaction with technical care, and SDRS. The TITLE section states that the analysis is of patient satisfaction and SDRS. The RAWDATA specification indicates that input data are stored in a file called *abaaraw*. The HALFSCALE and WRITESCA job options are specified, indicating that respondents providing data for half or more of the items in each scale will be included and that scale scores will be written to a file, *mapscale.out*. The IFVARS keyword is used to limit the analysis to the subsample of respondents based on age (PAGE) and marital status (SP1MARRY). The locations of the two IFVARS are columns 15-17 and columns 20-21. The HOWREAD section tells MAP to read the subject identification variable from the first 7 columns, skip to the 7th card, move to column 8, read the first 16 items, skip to the 10th card, move to column 44, and read the last 5 items. The ITEMS section lists the 21 item names used in this example. The CRITERIA section is left blank, as is typical when using MAP (this option is rarely used). Following the CRITERIA section, the condition associated with the IFVARS is given. The IF statement given in the example is used to limit the analysis to respondents who are over 64 and divorced.

Table 1

EXAMPLE OF *map.in* FILE FOR VAX-780

```
TITLES
'MOS PAQ00 APRIL 13 1988*END*'
'PATIENT SATISFACTION AND SDRS*END*'
RAWDATA
'abaaraw'
JOBOPTIONS
'HALFSCALE'
'WRITESCA'
/
IFVARS
PAGE SP1MARRY (T15,F3.0,2X,F2.0)
/
HOWREAD
(A7////////T8,16F1.0///T44,5F1.0)
ITEMS
'AA1PSQ01' 'AA1PSQ02' 'AA1PSQ03' 'AA1PSQ04'
'AA1PSQ05' 'AA1PSQ06' 'AA1PSQ07' 'AA1PSQ08'
'AA1PSQ09' 'AA1PSQ10'
'AA1PSQ11' 'AA1PSQ12' 'AA1PSQ13' 'AA1PSQ14' 'AA1PSQ15'
'AA1PSQ16'
'SDRS1' 'SDRS2' 'SDRS3' 'SDRS4' 'SDRS5'
/
CRITERIA
/
IF
PAGE>64 AND SP1MARRY=3 THEN KEEP
/
SCALE
'PSQGS' 6/
'AA1PSQ01' 'R' 1 5
'AA1PSQ02' 'P' 1 5
'AA1PSQ03' 'R' 1 5
'AA1PSQ04' 'P' 1 5
'AA1PSQ05' 'R' 1 5
'AA1PSQ06' 'P' 1 5
SCALE
'PSQTS' 10/
'AA1PSQ07' 'R' 1 5
'AA1PSQ08' 'P' 1 5
'AA1PSQ09' 'R' 1 5
'AA1PSQ10' 'P' 1 5
'AA1PSQ11' 'R' 1 5
'AA1PSQ12' 'P' 1 5
'AA1PSQ13' 'R' 1 5
'AA1PSQ14' 'P' 1 5
'AA1PSQ15' 'R' 1 5
'AA1PSQ16' 'P' 1 5
```

```
SCALE
'SDRS' 5/
'SDRS1' 'C' 1 2
           1 4 1
           5 5 2
'SDRS2' 'C' 1 2
           1 1 2
           2 5 1
'SDRS3' 'C' 1 2
           1 4 1
           5 5 2
'SDRS4' 'C' 1 2
           1 1 2
           2 5 1
'SDRS5' 'C' 1 2
           1 4 1
           5 5 2
HOWWRITE
'(A7,T8,21F2.0,1X,4F6.2)'
```

Next, three scales are specified in the SCALE section: PSQGS containing 6 items, PSQTS containing 10 items, and SDRS containing 5 items. The patient satisfaction items are all either scored as is ('P') or reverse-scored ('R') so that larger numbers always indicate satisfaction with care. The "1" indicates the minimum possible score and "5" represents the maximum possible. The five SDRS scale items are dichotomized using the 'C' recode option. The SDRS items are recoded into two levels: for three of the raw SDRS items (SDRS1, SDRS3, SDRS5) values of 1 to 4 are recoded to 1 and values of 5 are recoded to 2; for the other two SDRS items (SDRS2, SDRS4) values of 1 are recoded to 2 and values of 2 to 5 are recoded to 1. The final section shown, the HOWWRITE section, specifies the desired output format for the 21 items, 3 scales, and 1 overall scale.

Table 2 provides selected portions of the *map.out* file produced by the example *map.in* file. Shown first is the specification of the input items location in the raw data file. This output is extremely useful in checking to ensure that the FORTRAN HOWREAD input specification was done as intended. Next, item-scale correlations for the example are provided. Note that asterisks are used to designate correlations (corrected for overlap) between items and their hypothesized scales.

Item-scale correlations range from 0.66 to 0.75 for general satisfaction, 0.43 to 0.76 for satisfaction with technical quality, and 0.25 to 0.50 for socially desirable response set. If scaling decisions were to be made on the basis of this analysis, one might be concerned about the relatively low item-scale correlations of 0.25 and 0.31 for SDRS4 and SDRS1, respectively.

Following the item-scale correlations in Table 2 is the multitrait summary. This piece of the output summarizes scaling successes and failures in the analysis. In this example a total of 20 probable scaling failures is shown. This relatively large number of failures is attributed to two factors: (1) the small number of subjects ($N = 64$) leads to large standard errors; (2) general satisfaction partly subsumes technical satisfaction conceptually and we therefore expect overlap in respective item-scale correlations. Zeros in the general satisfaction and satisfaction with technical quality columns indicate that item-scale correlations of the general satisfaction items with the satisfaction with technical quality scale and the item-scale correlations of the satisfaction with technical quality items with the general satisfaction scale are all within two standard errors of their correlations with their hypothesized scales.

Correlations between scales, scale and sample internal consistency statistics, and correlations between scales adjusting for attenuation (i.e., unreliability of measurement) are given next. Finally, the number of subjects included in the analysis and the number of subjects omitted for missing data are provided.

Table 3 presents a portion of *mapscale.out* for the example. The full file would contain the subjects' identification numbers in the first 7 columns, followed by the item scores, individual scale sums, and overall scale score.

Table 2

EXAMPLE OF PARTIAL *map.out* FILE

According to your input format, items appear in the input data set as follows:

Item	Begins	---	Ends
MM			

Card # 1

ID	1	7
----	---	---

Card # 7

AA1PSQ01	8	8
AA1PSQ02	9	9
AA1PSQ03	10	10
AA1PSQ04	11	11
AA1PSQ05	12	12
AA1PSQ06	13	13
AA1PSQ07	14	14
AA1PSQ08	15	15
AA1PSQ09	16	16
AA1PSQ10	17	17
AA1PSQ11	18	18
AA1PSQ12	19	19
AA1PSQ13	20	20
AA1PSQ14	21	21
AA1PSQ15	22	22
AA1PSQ16	23	23

Card # 10

SDRS1	44	44
SDRS2	45	45
SDRS3	46	46
SDRS4	47	47
SDRS5	48	48

Table 2 (continued)

ITEM-SCALE CORRELATIONS						
ITEM	ITEMEAN	STNDEV	PSQGS	PSQTS	SDRS	TOTAL
AA1PSQ01	4.068	0.937	0.66*	0.61	0.10	0.65*
AA1PSQ02	3.169	0.978	0.71*	0.65	0.08	0.69*
AA1PSQ03	4.094	0.879	0.68*	0.72	0.13	0.73*
AA1PSQ04	3.203	1.063	0.70*	0.70	0.26	0.75*
AA1PSQ05	3.828	0.945	0.75*	0.73	0.15	0.77*
AA1PSQ06	3.527	1.188	0.74*	0.62	0.08	0.67*
AA1PSQ07	3.824	1.006	0.72	0.76*	0.08	0.76*
AA1PSQ08	3.436	1.198	0.57	0.57*	-0.08	0.56*
AA1PSQ09	4.104	0.889	0.63	0.68*	0.24	0.70*
AA1PSQ10	3.324	1.042	0.60	0.66*	0.19	0.67*
AA1PSQ11	3.980	0.715	0.63	0.70*	0.25	0.71*
AA1PSQ12	3.641	0.958	0.42	0.54*	-0.02	0.49*
AA1PSQ13	4.166	0.686	0.52	0.58*	0.22	0.59*
AA1PSQ14	4.135	0.693	0.70	0.76*	0.17	0.76*
AA1PSQ15	3.719	0.874	0.40	0.43*	0.15	0.44*
AA1PSQ16	3.609	1.055	0.65	0.56*	0.15	0.63*
SDRS1	1.656	0.475	-0.05	0.02	0.31*	0.02*
SDRS2	1.270	0.440	0.24	0.18	0.45*	0.26*
SDRS3	1.594	0.491	0.08	0.11	0.50*	0.15*
SDRS4	1.344	0.475	0.10	0.05	0.25*	0.10*
SDRS5	1.344	0.475	0.17	0.20	0.43*	0.24*

Table 2 (continued)

MULTITRAIT SUMMARY TABLE
 1 IS SCALING SUCCESS FOR ROW ENTRY
 0 IS PROBABLE FAILURE FOR ROW ENTRY
 -1 IS DEFINITE FAILURE FOR ROW ENTRY
 ** REPRESENTS CONVERGENT CORRELATION

ITEM-SCALE CORRELATIONS

ITEM	ITEMEAN	STNDEV	PSQGS	PSQTS	SDRS	TOTAL
AA1PSQ01	4.068	0.937	**	0	1	**
AA1PSQ02	3.169	0.978	**	0	1	**
AA1PSQ03	4.094	0.879	**	0	1	**
AA1PSQ04	3.203	1.063	**	0	1	**
AA1PSQ05	3.828	0.945	**	0	1	**
AA1PSQ06	3.527	1.188	**	0	1	**
AA1PSQ07	3.824	1.006	0	**	1	**
AA1PSQ08	3.436	1.198	0	**	1	**
AA1PSQ09	4.104	0.889	0	**	1	**
AA1PSQ10	3.324	1.042	0	**	1	**
AA1PSQ11	3.980	0.715	0	**	1	**
AA1PSQ12	3.641	0.958	0	**	1	**
AA1PSQ13	4.166	0.686	0	**	1	**
AA1PSQ14	4.135	0.693	0	**	1	**
AA1PSQ15	3.719	0.874	0	**	1	**
AA1PSQ16	3.609	1.055	0	**	1	**
SDRS1	1.656	0.475	1	1	**	**
SDRS2	1.270	0.440	0	1	**	**
SDRS3	1.594	0.491	1	1	**	**
SDRS4	1.344	0.475	0	0	**	**
SDRS5	1.344	0.475	1	0	**	**
TOTAL SCALING SUCCESSES=			22			
TOTAL SCALING FAILURES=			0			
TOTAL PROBABLE FAILURES=			20			

Table 2 (continued)

SCALE-SCALE CORRELATIONS							
SCALE/CRIT	MEAN	STNDEV	1	2	3		
1 PSQGS	21.89	4.818					
2 PSQTS	37.94	6.410	0.83				
3 SDRS	7.21	1.499	0.16	0.17			
4 TOTAL	67.03	11.123	0.94	0.96	0.31		

SCALE AND SAMPLE INTERNAL CONSISTENCY STATISTICS							
SCALE	K	RTT	R10	RII	SCOTT	RGG	RPP
PSQGS	6	0.89	0.93	0.57	0.58	0.96	0.27
PSQTS	10	0.88	0.88	0.42	0.44	0.91	0.14
SDRS	5	0.63	0.77	0.26	0.26	0.91	0.14
TOTAL	21	0.92	0.84	0.34	0.37	0.99	0.69

SCALE-SCALE CORRELATIONS, ATTENUATION ADJUSTED							
SUBSCALE	RTT	1	2	3			
1 PSQGS	0.888						
2 PSQTS	0.879	0.94					
3 SDRS	0.631	0.22	0.23				
***** SCALE-TOTAL R CORRECTED FOR OVERLAP *****							
4 TOTAL	0.916	1.02	1.04	0.31			
NUMBER OF SUBJECTS = 64							
NUMBER OF SUBJECTS OMITTED FOR MISSING DATA = 9							

Table 3

EXAMPLE OF PARTIAL *mapscale.out* FILE

002897B5.4.5.4.5.1.5.5.5.4.5.4.5.5.4.4.2.2.2.2.2.	24.00	46.00	10.00	80.00
020043Q3.3.3.2.4.2.4.2.4.3.4.3.4.4.4.4.2.1.2.1.2.	16.80	36.00	8.00	60.80
025142J5.3.5.4.5.5.5.5.4.3.5.5.5.5.4.5.1.2.1.1.1.	27.00	46.00	6.00	79.00
049190H5.4.4.4.4.4.4.2.5.3.4.2.4.4.2.4.2.2.2.1.2.	25.00	34.00	9.00	68.00
08359074.2.3.2.2.2.4.2.4.2.4.4.4.4.3.2.1.1.1.1.	15.00	35.00	6.00	56.00
107205G4.2.4.4.4.4.4.4.4.4.5.3.4.4.4.4.2.1.1.2.1.	22.00	40.00	7.00	69.00
...				
.				
.				
.				

IV. PUBLISHED EXAMPLES OF MULTITRAIT SCALING

To give a sense of different possible uses of multitrait scaling analysis, we summarize a sample of published studies that have made use of the technique.

Ware, Johnston, Davies-Avery, and Brook (1979) employed multitrait scaling to evaluate measures of general well-being in a sample of about 1209 RAND Health Insurance Experiment (HIE) participants from Dayton, Ohio. The convergent and discriminant validity of the measures was soundly supported in these analyses (percentage of scaling successes given in parentheses): anxiety (96.0 percent), depression (86.7 percent), general health (93.3 percent), positive well-being (90.0 percent), self-control (73.3 percent), vitality (85.0 percent).

Eisen, Donald, Ware, and Brook (1980) used multitrait scaling analysis in a validation of proxy (usually mother) health status measures for children aged 0-4 (N = 679) and 5-13 (N = 1473) in the HIE. These analyses revealed good convergent and discriminant validity for measures of current health, resistance/susceptibility, prior health, anxiety, depression, positive well-being, social relations, and satisfaction with development. Among these measures, anxiety, depression, and positive well-being had somewhat fewer scaling successes.

Davies and Ware (1981) employed multitrait scaling to analyze health perceptions measures in the HIE. They published an item-scale correlation matrix for six multi-item scales: current health, prior health, health outlook, resistance, health worry/concern, and sickness orientation. In addition, they noted the proportion of item discriminant validity successes for each scale at each of four separate sites and for all sites combined. Results of this analysis provided strong support for the validity of their scales: an overall success ratio of 97.3 percent was reported.

A modified multitrait scaling analysis was applied to the HIE measures of group participation, social contacts, and social well-being by Donald and Ware (1982). Voluntary group membership and level of group activity were hypothesized indicators of *group participation*; visits with friends/relatives, home visits by friends, and visits to homes of friends were used to represent *social contacts*; nine indicators, including the group participation and social contacts indicators, were used to define *social well-being*. This hypothesized structure is different from the usual application of multitrait scaling because indicators were allowed to define more than one scale. Results of the analysis provided strong support for the *group participation* and *social contacts* scales. The *social well-being* index, which represents a combination of group participation and social contacts, had item-scale correlations (corrected for overlap) ranging from 0.26 to 0.47.

Ware, Snyder, Wright, and Davies (1983) reported multitrait scaling results for the development of the Patient Satisfaction Questionnaire. Eighteen satisfaction subscales were evaluated in four sites. Ware, Snyder, Wright, and Davies found strong support for item convergent and discriminant validity: "Only 11 correlations...between items and their hypothesized scales were below 0.30 in 220 tests across four sites. Of 3,740 tests of item discriminant validity criterion...approximately 98 percent were favorable" (p. 256).

McCusker (1984) published matrices of item-scale correlations for patient satisfaction measures administered to three distinct groups: patients and caretakers (two groups) in a randomized controlled trial comparing a health-care-team treatment approach with an existing community services approach to home care for chronically ill patients, and relatives of cancer patients asked retrospectively about care during the last six months of the patient's life. Based on her assessment of convergent and discriminant validity, McCusker concluded that items measuring general satisfaction and satisfaction with physician availability performed adequately, but items in three physician behavior scales (technical care, interpersonal care, and communication) could not be distinguished from one another.

Hays and DiMatteo (1987) used MAP to evaluate item discriminant validity of the UCLA Loneliness Scale (ULS-20) and two short-form measures (ULS-8, ULS-4) relative to measures of life satisfaction, alienation, and social anxiety. The ULS-20 exhibited a total of 46 out of 60 (76.7 percent) scaling successes; the ULS-8 had 18 scaling successes out of 24 (75.0 percent); and the ULS-4 had 4 scaling successes out of 12 (33.3 percent). Hays and DiMatteo interpreted these results as favoring the ULS-8 over the ULS-4.

Stewart, Hays, and Ware (1988) used MAP to assess the Medical Outcomes Study (MOS) Short-Form General Health Survey. Item-scale correlations (corrected for overlap) for measures of physical functioning, role functioning, mental health, and current health ranged from 0.45 to 0.79, with a median of 0.68. All items in each hypothesized scale also exceeded the discriminant validity criterion. Reliability coefficients for the multi-item health scales ranged from 0.81 to 0.88. Estimates for the four multi-item scales were similar for depressed patients (0.82 to 0.87) and for other subgroups analyzed: congestive heart failure (0.77 to 0.87), diabetes (0.83 to 0.87), myocardial infarction (0.77 to 0.88), less than a high school education (0.86 to 0.88), and over age 75 (0.84 to 0.89). Stewart et al. (1988) concluded that the analysis demonstrated excellent item discrimination among hypothesized scales.

A telephone-administered version of the MOS Short-Form General Health Survey was evaluated by Ware, Sherbourne, Davies, and Stewart (in press). Item-scale correlations (corrected for overlap) ranged from 0.41 to 0.74 for physical functioning, 0.54 to 0.65 for mental health, and 0.57 to 0.72 for current health. The intercorrelation among the role functioning items was 0.63. Alpha reliability estimates for these scales ranged from 0.76 to 0.88.

REFERENCES

- Campbell, D. T., and D. W. Fiske (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Davies, A. R., and J. E. Ware, Jr. (October 1981). *Measuring health perceptions in the Health Insurance Experiment*. Santa Monica, CA: The RAND Corporation, R-2711-HHS.
- Donald, C. A., and J. E. Ware, Jr. (October 1982). *The quantification of social contacts and resources*. Santa Monica, CA: The RAND Corporation, R-2937-HHS.
- Eisen, M. B., C. A. Donald, J. E. Ware, Jr., and R. H. Brook (May 1980). *Conceptualization and measurement of health for children in the Health Insurance Study*, Santa Monica, CA: The RAND Corporation, R-2313-HEW.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Hayashi, T., and R. D. Hays (1987). A microcomputer program for analyzing multitrait-multimethod matrices. *Behavior Research Methods, Instruments, and Computers*, 19, 345-348.
- Hays, R. D., and M. R. DiMatteo (1987). A short-form measure of loneliness. *Journal of Personality Assessment*, 51, 69-81.
- Howard, K. I., and G. G. Forehand (1962). A method for correcting item-total correlations for the effect of relevant item inclusion. *Educational and Psychological Measurement*, 22, 731-735.
- Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6, 153-160.
- Levy, K. J. (1975). Some multiple range tests for variances. *Educational Psychological Measurement*, 35, 599-604.
- McCusker, J. (1984). Development of scales to measure satisfaction preferences regarding long-term and terminal care. *Medical Care*, 22, 476-493.
- Schnittjer, C. J., and C. M. Cartledge (1976). Item-analysis programs: A comparative investigation of performance. *Educational Psychological Measurement*, 36, 183-187.

Scott, W. A. (1968). Attitude measurement. In G. Lindzey and E. Aronson (eds.), *Handbook of social psychology, vol. 2* (pp. 204-273). Reading, MA: Addison-Wesley.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.

Stewart, A. L., R. D. Hays, and J. E. Ware, Jr. (1988). The MOS Short-Form General Health Survey: Reliability and validity in a patient population. *Medical Care, 26*, 724-735.

Tyler, T. A., and D. W. Fiske (1968). Homogeneity indices and test length. *Educational Psychological Measurement, 28*, 767-777.

Ware, J. E., Jr., S. A. Johnston, A. Davies-Avery, and R. H. Brook (December 1979). *Conceptualization and measurement of health for adults in the Health Insurance Study: Vol. III, mental health*. Santa Monica, CA: The RAND Corporation, R-1987/3-HEW.

Ware, J. E., Jr., C. D. Sherbourne, A. R. Davies, and A. L. Stewart (in press). A short-form general health survey. *Health Services Research*.

Ware, J. E., Jr., M. R. Snyder, R. Wright, and A. R. Davies (1983). Defining and measuring patient satisfaction with medical care. *Evaluation and Program Planning, 6*, 247-263.

