

Beyond internal consistency reliability: Rationale and user's guide for Multitrait Analysis Program on the microcomputer

RON D. HAYS

The RAND Corporation, Santa Monica, California

and

TOSHI HAYASHI

California State Assembly, Office of Research, Sacramento, California

A methodology for evaluating Likert-type scales is presented. Multitrait scaling is a straightforward approach to scale analysis that focuses on items as the unit of analysis and utilizes the logic of convergent and discriminant validity. Multitrait scaling is illustrated with the Multitrait Analysis Program, using patient satisfaction data from the Medical Outcomes Study.

The process of validating psychological measures involves the accumulation of evidence of many different types that indicates the degree to which the measures represent what they were intended to represent. Construct validity (Cronbach & Meehl, 1955) is evaluated by hypothesizing how measures should "behave" and confirming or disconfirming these hypotheses. "It is theory which specifies the properties of a concept, its relations to other concepts, and its implications for behavior, including test behavior" (Jessor, Graves, Hanson, & Jessor, 1968, p. 146). Hypotheses are stated regarding the direction (and sometimes the strength) of relationship that might be expected, and validity is supported when the associations are consistent with hypotheses. Construct validation is iterative by its very nature, with empirical results feeding into revision of measures, retesting, and further revisions, if necessary (Ware, 1984).

Convergent and discriminant validity are two fundamental aspects of construct validity. Convergent validity refers to the extent to which different ways of measuring the same trait intercorrelate with one another. Discriminant validity involves the demonstration that a measure does not correlate too strongly with measures that are intended to indicate other traits. For example, the correlation between a measure of loneliness and a measure of health locus of control would not be expected to be very large (Hays & DiMatteo, 1987). When more than one method of data collection or scale construction has been used, con-

vergent and discriminant validity can be assessed according to multitrait-multimethod (MTMM) analytic methods.

Procedures for implementing the MTMM strategy have been based on zero-order correlations among measures (Hayashi & Hays, 1987), partial correlations (Schriesheim, 1981), analysis of variance (Boruch & Wolins, 1970; Stanley, 1961), generalized proximity function analysis (Hubert & Baker, 1979), smallest space analysis (Levin, Montag, & Comrey, 1983), exploratory factor analysis (Jackson, 1969), path analysis (Alwin, 1974), and confirmatory factor analysis (Schmitt & Stults, 1986; Widaman, 1985). Of the available procedures, confirmatory factor analysis offers the greatest flexibility and advantages.

When multiple traits are assessed according to a single method, the MTMM analytic strategy is not applicable. However, the general principles of convergent and discriminant validity can be applied. The purpose of this paper is to extend the logic of MTMM analysis to scale construction and validation when one method of evaluation is employed. *Multitrait scaling* is introduced as a methodology for evaluating Likert-type scales.

Multitrait Scaling Analysis Method

It is common practice in scale development to evaluate the internal consistency reliability of multi-item scales by estimating Cronbach's (1951) alpha coefficient. Alpha provides an indication of the degree of convergence between different items hypothesized to represent the same construct or trait. In multitrait scaling, *item discrimination* across scales is also evaluated. Items are examined with respect to how well they represent a particular trait relative to other traits. Item discrimination extends the MTMM ideas from the level of traits to the level of items or indicators of traits.

Item-scale correlations are the fundamental elements of multitrait scaling (Ware, Snyder, Wright, & Davies,

Preparation of this article was supported by a grant from the RAND Corporation as part of its program of public service. The opinions expressed are those of the authors and do not necessarily reflect the views of the RAND Corporation or the California State Assembly Office of Research. Requests for reprints may be sent to Ron D. Hays, RAND Corporation, 1700 Main Street, Santa Monica, CA 90406-2138.

1983) and constitute the multitrait/multi-item (MTMI) correlation matrix. A synthetic MTMI matrix is shown in Table 1. Three different traits are defined by three items each: Trait 1 (Items 1-3), Trait 2 (Items 4-6), and Trait 3 (Items 7-9). Each row of the MTMI matrix contains correlations between scores for one item and all hypothesized traits (defined by the sum of items comprised by each trait). Each column contains correlations between the scores for one trait and all items in the analysis, including those hypothesized to be parts of that trait and those hypothesized to be parts of other traits. Correlations between items hypothesized to define a given trait and the trait itself are corrected for overlap (using the technique recommended by Howard & Forehand, 1962), so that estimates of the item-trait relationships are not spuriously inflated.

Item convergence is supported if an item correlates substantially (a corrected correlation of about 0.40 or above is recommended as a rule of thumb) 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, a more stringent convergent validity criterion (i.e., larger correlation) can be applied. Each of the indicators shown in the synthetic example in Table 1 demonstrates item convergence, correlating 0.80 with the hypothesized trait (corrected for item overlap).

If the highest correlation in a row of the MTMI matrix is the correlation between the item and the trait that it is hypothesized to measure, and if this correlation is significantly larger than the other correlations in the row, item discrimination is supported. The significance of the difference of these correlations can be evaluated with Steiger's (1980) *t*-test for dependent correlations. Item discrimination is shown for all nine items given in Table 1. Each item-convergent correlation is significantly larger than the other correlations in the same row. The correlations between traits are also evaluated in multitrait scaling: The zero-order correlations between scales and the correlations adjusted for unreliability of measurement (Guilford, 1954) should be less than unity to support the distinctiveness of the traits.

Multitrait scaling was used routinely in scale development for the RAND Health Insurance Experiment (see,

e.g., Donald & Ware, 1982; Eisen, Donald, Ware, & Brook, 1980) and the Medical Outcomes Study (see, e.g., Hays & DiMatteo, 1987; Stewart, Hays, & Ware, 1988). Despite its simplicity, very few investigators at other institutions have utilized the technique directly (cf. McCusker, 1984, for an exception). However, similar methods were advocated two decades ago by Jackson (1970), and they have recently been recommended in the context of exploratory factor analysis by Comrey (1988).

Application of Multitrait Scaling Analysis

For the present paper, data from the Medical Outcomes Study (MOS) have been analyzed to illustrate the multitrait scaling methodology. In the MOS, patient satisfaction with medical care was measured every 6 months using a periodic satisfaction survey, the PSQ-III. The PSQ-III consists of 50 Likert-type items that assess general satisfaction and satisfaction with six specific dimensions of medical care: technical quality, interpersonal quality, communication, financial aspects, time spent with provider, and access/availability/convenience.

Patient satisfaction ratings obtained at baseline of the MOS from 1,192 patients of medical providers were analyzed. For this analysis, ratings of satisfaction with the provider's technical quality (6 items), interpersonal quality (6 items), and communication skills (5 items), as well as ratings of satisfaction with financial arrangements (8 items), were examined. Each of these four dimensions of satisfaction with care was hypothesized to be distinct, as has been the case in previous research (Ware et al., 1983).

Table 2 provides the MTMI correlation matrix for the 25 patient satisfaction items hypothesized to represent the four dimensions of satisfaction with medical care. Item-scale correlations for hypothesized traits, corrected for item overlap, are designated with an asterisk. These item-convergent coefficients are reasonably large, ranging from 0.48 to 0.73. Internal consistency reliability, as estimated by Cronbach's alpha coefficient, was 0.80 or higher for each scale. Thus, the items representing the four traits are internally consistent.

Item discrimination is examined through comparison of the convergent correlations with the other correlations in the same row of the MTMI matrix. Correlations within two standard errors of the corresponding convergent correlations (denoted by a footnote) indicate lack of item discrimination. For example, each item hypothesized to measure satisfaction with technical quality correlates about as high (within two standard errors) with the satisfaction with communication scale as it does with the sum of the other items measuring technical quality. Similarly, all of the satisfaction with communication items correlate about as highly with the technical quality and interpersonal aspects scales as they do with the other communication items. In summary, items hypothesized to assess technical quality, interpersonal aspects, and communication lack discriminant validity.

The correlations between the scales formed by summing items designed to measure each of the four dimensions

Table 1
Synthetic Multitrait/Multi-Item Correlation Matrix

Item	Trait 1	Trait 2	Trait 3
1	0.80*	0.20	0.20
2	0.80*	0.20	0.20
3	0.80*	0.20	0.20
4	0.20	0.80*	0.20
5	0.20	0.80*	0.20
6	0.20	0.80*	0.20
7	0.20	0.20	0.80*
8	0.20	0.20	0.80*
9	0.20	0.20	0.80*

*Item-scale correlation, corrected for overlap.

Table 2
Multitrait/Multi-Item Correlation Matrix for Patient Satisfaction Ratings

	Technical	Interpersonal	Communication	Financial
Technical				
1	0.66*	0.63†	0.67†	0.28
2	0.55*	0.54†	0.50†	0.25
3	0.48*	0.41	0.44†	0.26
4	0.59*	0.53	0.56†	0.26
5	0.55*	0.60†	0.56†	0.16
6	0.59*	0.58†	0.57†	0.23
Interpersonal				
1	0.58	0.68*	0.63†	0.24
2	0.59†	0.58*	0.61†	0.18
3	0.62†	0.65*	0.67†	0.19
4	0.53†	0.57*	0.60†	0.32
5	0.54	0.62*	0.58†	0.18
6	0.48†	0.48*	0.46†	0.24
Communication				
1	0.58†	0.59†	0.61*	0.26
2	0.47†	0.50†	0.50*	0.25
3	0.58†	0.66†	0.63*	0.23
4	0.66†	0.66†	0.67*	0.25
5	0.66†	0.71†	0.70*	0.25
Financial				
1	0.35	0.35	0.35	0.72*
2	0.17	0.14	0.15	0.65*
3	0.25	0.23	0.23	0.61*
4	0.18	0.15	0.16	0.67*
5	0.31	0.27	0.29	0.70*
6	0.24	0.23	0.22	0.73*
7	0.25	0.23	0.25	0.55*
8	0.34	0.31	0.31	0.64*
Cronbach's alpha	0.80	0.82	0.82	0.88

Note—Standard error of correlation is 0.03. Technical = satisfaction with technical quality; Interpersonal = satisfaction with interpersonal aspects; Communication = satisfaction with communication; Financial = satisfaction with financial arrangements. *Item-scale correlations for hypothesized scales (corrected for item overlap). †Correlation within two standard errors of the correlation of the item with its hypothesized scale.

of care are presented in Table 3. These correlations provide further indication of the lack of discrimination (at the level of the traits) between technical quality, interpersonal aspects, and communication; adjusted correlations (for unreliability of measurement) between scales were 0.93 or higher.

How to Use the Multitrait Analysis Program

To use the Multitrait Analysis Program (MAP), the user must supply two input files: a raw data file and an input specification file (map.in). When MAP is executed, the user is asked to supply the name of the file where the raw data is stored. On the basis of 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.

The raw data input file is an ASCII file. (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, including TITLES, JOB-OPTIONS, HOWREAD, ITEMS, CRITERIA, SCALE, and HOWWRITE. Following each keyword, the user enters appropriate specifications as described below. The

Table 3
Correlations Between Patient Satisfaction Scales

	Technical	Interpersonal	Communication	Financial
Technical	1.00	0.75	0.76	0.34
Interpersonal	0.93	1.00	0.80	0.31
Communication	0.94	0.98	1.00	0.32
Financial	0.41	0.36	0.38	1.00

Note—Zero-order correlations are provided above the diagonal; correlations adjusting for unreliability of measurement are given below the diagonal.

user needs to specify the 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 an example of how to use the following keywords, see Table 4.

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.

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

Table 4
Example of map.in File

```

TITLES
'MOS PAQ00 APRIL 13 1988'
'PATIENT SATISFACTION AND SDRS'
JOBOPTIONS
'HALFSCALE'
'WRITESCA'
/
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
/
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)'

```

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 (see below), **HOWWRITE** must be specified as well.

For the second type of option, one of two keywords, **HALFSCALE** or **COMPLETE**, may be optionally 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 **HALFSCALE** (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. Normally, this option is not recommended, but it must be used if the **CRITERIA** (see below) keyword 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, or blank, 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**.) Because of a bug in **FORTRAN**, the end of each input line should be anchored with a constant (e.g., 9).

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 by 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 subscales 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 were 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** are 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 the number of items in the scale are given after the keyword. If indicators of a scale are to be standardized, standardization is indicated 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 value 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 to 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 only be named in one scale.

HOWWRITE. This option is used only if one desires to write item and scale scales 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 and left parenthesis, and end with a right parenthesis and single quote. In setting up the **HOWWRITE** field, one should describe the subject ID 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`.

Debugging Input Errors

MAP prints out some information to `map.out` that is useful for debugging input errors, particularly a section described as "program control input processing." This section lists the beginning and ending of processing for each of the keywords in the input file. If no errors are found in the input file, this section ends with the statement **END OF PROGRAM CONTROL INPUT**. If an error occurs involving a keyword, this section will point to the location of the error by indicating where it ends. For example, if an error occurs in the input of the first scale, the end of processing for this keyword will not be printed.

The example listed below illustrates an instance in which an input error is present. **MAP** begins processing the second scale that it reads, but execution is terminated before

Table 5
Example of map.out File: Selected Portions

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

Item	Begins	---	Ends
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

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 5 (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

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

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 SCALE keyword (for the third scale) before finishing the reading of items in the second scale.

PROGRAM CONTROL INPUT PROCESSING STATUS—

```
BEGIN PROCESSING KEYWORD TITLES
END PROCESSING KEYWORD TITLES
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
```

Example of Using MAP on the Microcomputer

An example map.in file is given in Table 4. The example is an analysis of patient satisfaction scales for a small subsample of participants ($n = 73$) in the Medical Outcomes Study. The analysis specifies an assessment of general satisfaction, satisfaction with technical care and socially desirable response set (SDRS). The TITLE section states that the analysis is of patient satisfaction and SDRS. The file name, abaaraw, is given when the program asks for the name of the input raw data file. 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 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 (i.e., this option is rarely used).

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 5 SDRS scale items are dichotomized using the 'C' recode option. The SDRS items are recoded into two levels: For three of the 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 5 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 en-

sure 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 SDRS. 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 5 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; and (2) general satisfaction partly subsumes technical satisfaction conceptually, so we 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 given.

Availability

MAP is available for IBM (and compatible) microcomputers and can be obtained from the National Collegiate Software Clearinghouse (Duke University Press, 6697 College Station, Durham, NC 27708; phone: 919-684-6837, extension 54).

REFERENCES

- ALWIN, D. (1974). An analytic comparison of four approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H. L. Costner (Ed.), *Sociological methodology, 1973-1974* (pp. 79-105). San Francisco: Jossey-Bass.
- BORUCH, R. F., & WOLINS, L. (1970). A procedure for estimation of trait, method, and error variance attributable to a measure. *Educational & Psychological Measurement*, *30*, 547-574.
- COMREY, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting & Clinical Psychology*, *56*, 754-761.
- CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.
- CRONBACH, L. J., & MEEHL, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.
- DONALD, C. A., & WARE, J. E. (1982). *The quantification of social contacts and resources* (Rep. No. R-2937-HHS). Santa Monica, CA: The RAND Corporation.
- EISEN, M. B., DONALD, C. A., WARE, J. E., & BROOK, R. H. (1980). *Conceptualization and measurement of health for children in the Health*

- Insurance Study* (Rep. No. R-2313-HEW) Santa Monica, CA The RAND Corporation.
- GUILFORD, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- HAYASHI, T., & HAYS, R. D. (1987). A microcomputer program for analyzing multitrait-multimethod matrices. *Behavior Research Methods, Instruments, & Computers*, **19**, 345-348.
- HAYS, R. D., & DiMATTEO, M. R. (1987). A short-form measure of loneliness. *Journal of Personality Assessment*, **51**, 69-81.
- HOWARD, K. I., & FOREHAND, G. G. (1962). A method for correcting item-total correlations for the effect of relevant item inclusion. *Educational & Psychological Measurement*, **22**, 731-735.
- HUBERT, L. J., & BAKER, F. B. (1979). A note on analyzing the multitrait-multimethod matrix. An application of a generalized proximity function comparison. *British Journal of Mathematical & Statistical Psychology*, **32**, 179-184.
- JACKSON, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin*, **72**, 30-49.
- JACKSON, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61-96). New York: Academic Press.
- JESSOR, R., GRAVES, T. D., HANSON, R. C., & JESSOR, S. L. (1968). *Society, personality, and deviant behavior: A study of a tri-ethnic community*. New York: Krieger.
- LEVIN, J., MONTAG, I., & COMREY, A. L. (1983). Comparison of multitrait-multimethod, factor, and smallest space analysis on personality scale data. *Psychological Reports*, **53**, 591-596.
- McCUSKER, J. (1984). Development of scales to measure satisfaction preferences regarding long-term and terminal care. *Medical Care*, **22**, 476-493.
- SCHMITT, N., & STULTS, D. M. (1986). Methodology review. Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, **10**, 1-22.
- SCHRIESHEIM, C. A. (1981). Leniency effects on convergent and discriminant validity for grouped questionnaire items. A further investigation. *Educational & Psychological Measurement*, **41**, 401-411.
- STANLEY, J. C. (1961). Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, **26**, 205-219.
- STEIGER, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, **87**, 245-251.
- STEWART, A. L., HAYS, R. D., & WARE, J. E. (1988). The MOS Short-form General Health Survey: Reliability and validity in a patient population. *Medical Care*, **26**, 724-735.
- WARE, J. E. (1984). Methodological considerations in the selection of health status assessment procedures. In N. K. Wenger, M. E. Mattson, C. D. Furberg, & J. Elinson (Eds.), *Assessment of quality of life in clinical trials of cardiovascular therapies*. New York: LeJacq.
- WARE, J. E., SNYDER, M. R., WRIGHT, R., & DAVIES, A. R. (1983). Defining and measuring patient satisfaction with medical care. *Evaluation & Program Planning*, **6**, 247-263.
- WIDAMAN, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, **9**, 1-26.