# An Empirical Test of the Utility of the Observations-To-Variables Ratio in Factor and Components Analysis

Willem A. Arrindell and Jan van der Ende
State University of Groningen, The Netherlands

Many researchers have proposed a minimum ratio of observations to variables or an absolute minimum of observations in order to obtain stable factor configurations. However, hardly any empirical studies employing real data are available that attest to the tenability of these proposals. A systematic analysis of the problem was undertaken, using self-report data from two large phobic samples on the Fear Survey Schedule-III ($N$ = 1104) and the Fear Questionnaire ($N$ = 960). The data sets were randomly split into subsamples with ratios varying from 1.3:1 to 19.8:1. Neither the observations to variables ratio nor an absolute minimum of observations had any influence on factor stability.

In recent years, the use of multivariate statistical procedures such as factor analysis has increased rapidly. This development has been greatly facilitated by the advent of high speed computer software capabilities. For example, a large number of studies have used several factor analytic methods in the investigation of irrational fears in clinical and student samples (see Granell de Aldaz, 1982, and Wade, 1978, for reviews). Even though multivariate procedures offer distinct advantages in allowing the simultaneous consideration of the associations among multiple variables, their use may be subject to misuse by researchers who do not fully appreciate their complexities; hence a con-

siderable level of expertise is required (e.g., Everitt, 1975).

A number of decisions are involved in conducting, interpreting, and reporting factor analyses and several good textbooks are available to help the investigator avoid possible pitfalls (see Comrey, 1978, pp. 658–659; Skinner, 1980, p. 1101). Some of the common trouble spots that may be encountered relate to the decision points that are utilized in employing factor analysis and include the following: the number of subjects or observations, the factoring method/model, the number of factors to rotate, and the rotation technique (e.g., Comrey, 1978; Rummel, 1970).

The present paper is concerned with the first of these decisions. The remaining decisions have been discussed quite extensively in, for instance, Harman (1967), Rummel (1970), van de Geer (1971), Mulaik (1972), Overall and Klett (1972), Gorsuch (1974), Harris (1975), and Nunnally (1978). In addition, six excellent general surveys of methodological concerns in the use of factor analysis in behavior research have become available recently to the clinically trained researcher (Comrey, 1978; Everitt, 1975; Futch, Scheirer, & Lisman, 1982; Kennard, 1978; Kline & Barrett, 1983; Skinner, 1980). The present article represents a complement to these surveys and was motivated by an unresolved methodological issue pertaining to the observations to variables ratio in factor analysis. This

165

is obviously related to the "economics" of data collection and to the difficulty of obtaining data in particular areas of behavior research, and, consequently, cannot be considered solely from a statistical viewpoint (cf. Everitt, 1975).

Thus, though the established practice of using college or university students as subjects for psychological experiments and multivariate studies affords a cheap and easily accessible source of labor (Simón, 1979), the requirement in clinical research of an overwhelmingly large number of people—say, well over 100—suffering from rarely encountered psychiatric disorders such as claustrophobia, height phobia, or any other form of specific phobia, may make a study impracticable.

Despite the classical notion that an adequate sample size is crucial in helping reduce errors in correlation coefficients and hence factor loadings, that is, in helping to minimize sampling error (e.g., Guilford, 1954, pp. 532–533), its recommendations in the literature are far from consistent. For instance, Guilford noted that "experience seems to show that when Pearson $r$'s are used, a minimum $N$ of 200 is good policy" (1954, p. 533), and further that "factor loadings from samples near 200 have been fairly consistent with loadings in the same factors and tests from samples above 1,000" (p. 533). Guilford further stated that "the upper limits will depend upon the compulsions of the investigator and the circumstances of testing arrangements" (p. 533).

Gorsuch (1974) observed that, as a rule of thumb, the number of observations should be at least five times greater than the number of variables and that this rule should apply to problems when the variables are "excellent." In addition, he noted that " ... the ratio should be larger in many cases. No sample should be less than 100 even though the number of variables is less than 20" (p. 333). However, he also stated that "the rule ... is ... only a rule of thumb" (p. 333).

Cattell (1978, p. 508) suggested a ratio of 3 to 6 times as many observations as variables and also indicated an absolute minimum of about 250 observations. Nunnally (1978, p. 276), on the other hand, stated that it is a good rule to employ at least 10 times as many observations as variables: "then there will be more 'significant' factors than the investigator will care to interpret" (p. 421). Unfortunately, empirical evidence to support these views, based on research experience, is not reported.

Additional vague recommendations in the literature increase the confusion even more. Cureton and D'Agostino (1983), for example, suggested that for dependable results, factor analysis procedures require large samples, "ideally samples of several hundred" (p. 11). Kline (1979, p. 40) argued that given that Guilford's (1954) recommendation—that a sample of 200 is the minimum for ensuring statistically reliable results—is a conservative figure, "there would seem little reason to doubt the reliability of factors derived from samples of 100 subjects," the explanation for this being unclear. Kline further argued that "it is also generally agreed that (for reasons of matrix algebra) there should be at least twice as many subjects as variables in factor-analytic investigations. This means that in any large study on this account alone, one should have to use more than the minimum 100 subjects" (p. 40).

In a cautionary paper, Everitt (1975) argued, on the basis of findings from monte carlo simulation studies, that perhaps 10 individuals for each variable may be the kind of observations to variables ($N{:}p$) ratio to aim for, though even this may be rather optimistic. He added that "in general the conclusion of any analysis performed on data for which $N < 5p$ should be viewed with a certain scepticism" (p. 238).

To date, it appears that only one study (Barrett & Kline, 1981) has had as its major aim the empirical test of the "observations to variables ratio" or the "absolute minimum of observations" on the stability of factor structures employing *real* data. In their systematic analysis of the problem, Barrett and Kline randomly split two sets of data from the Sixteen Personality Factors Questionnaire (16PF, Form A; Cattell, Eber, & Tatsuoka, 1970) and from the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975) into subsamples with $N{:}p$ ratios varying from 1.25:1 ($N$s ranging from 20 to 288 on the 16PF *scales*) to 31:1 ($N$s ranging from 112 to 810 on the EPQ *items*).

The results indicated that the observations to variables ratio did not influence factor stability, the important variable being the absolute number of observations. An $N = 50$ was shown to be the minimum to yield a clear, recognizable factor pattern. However, the measures employed contained good "strong" variables yielding exceedingly clear factor structures from the total samples ($N$s of 491 and 1,198, respectively). In addition, the factorial constructs manipulated in this analysis are assumed (in trait theory) to be general pervasive influences in all individuals. Barrett and Kline (1981) stated that, on these grounds, " … virtually any sample of individuals would suffice to yield the underlying structures," and that therefore, in their study, "statistical error is relatively minimal compared with the errors to be found from bad sampling of target populations" (p. 32).

In discussing the results from this, Kline and Barrett (1983) stressed that "with factors of lesser clarity, it is possible that different findings might be obtained" (p. 145). Also, it should be noted that Barrett and Kline (1981) stated that on the basis of their findings, small-sample factoring should only be carried out when replicating a supposed factor structure (p. 32; see also O'Looney & Barrett, 1983, p. 218).

In addressing the present issue there is the choice between distinct models for representing a large set of observed variables ($p$) by some smaller set ($m < p$) that still preserves the essential original information. The most widely employed methods to accomplish this goal can be generally classified into one of two broad categories: principal *components* analysis (PCA) or principal *factor* analysis (PFA; Velicer, Peacock, & Jackson, 1982).

Most theorists have recommended factor analysis as the method of choice (e.g., Comrey, 1978), though applied researchers have generally not followed this advice. Surveys of educational research journals cited by Velicer et al. (1982) found that PCA was employed twice as often as PFA in the early and mid 1960s, and that this ratio rose to five to one in favor of PCA in the late 1970s. A number of reasons can perhaps be advanced for the preference of PCA above PFA: (1) the factor analysis approach is computationally unwieldly, and typically much more costly in terms of computer time and space; (2) the presence of improper solutions in PFA; and (3) the factor analysis model suffers from factor score indeterminancy, a problem which has become recognized as a serious theoretical flaw (Velicer, 1977; Velicer et al., 1982).

By employing a monte carlo approach Velicer et al. (1982) performed principal components analysis, image components analysis, and maximum likelihood factor analysis on data matrices. Comparisons were made between each of the three methods and to ideal structures, with sample size, saturation, and type of structure systematically varied. The general conclusion reached was that the three methods produced results which were equivalent. However, since theorists do not agree with respect to their choice of a model of factor analysis (see Velicer et al., 1982, p. 371), both PCA and PFA were used in the present study.

## Method

### Measures

Two assessment devices were employed: the Fear Survey Schedule (FSS-III; Wolpe & Lang, 1964) and a self-rating questionnaire for phobic patients, the Marks and Mathews (1979) Fear Questionnaire (FQ). Standard instructions were given in which subjects were requested to indicate their aversion (felt anxiety) to the 76 stimulus situations and/or objects on the FSS-III on 5-point Likert-type scales, which range from 1 = "not at all disturbed" to 5 = "very much disturbed." The FQ contains 20 items; 15 of these (situations) are rated for the individual's degree of avoidance on a 9-point continuum, which ranges from 0 = "would not avoid it" to 8 = "would always avoid it." The remaining 5 anxiety-depression items are rated on scales ranging from 0 to 8 with respect to the degree to which the individual is troubled by each described problem (0 = "hardly at all" to 8 = "very severely troublesome"). The subscales that underlie the instruments are described below.

### Subjects

Two samples comprised of members of a society for phobic individuals (predominantly agorapho-

bics) served as subjects. One sample of 1,104 phobics (20.8% males, 79.2% females) with a mean age of 40.95 years (SD = 12.07 years; range: 16–87 years) completed the FSS-III. Their data were drawn from Arrindell (1980) and Arrindell, Emmelkamp, and van der Ende (1984); members from the latter study who had also participated in the former study were excluded. The FQ responses were obtained from two unrelated phobic samples as described in Arrindell et al. This combined sample was comprised of 960 observations (20% males, 80% females) with a mean age of 41.80 years (SD = 12.12 years; range: 16–87 years; further details of both samples are given in the studies cited).

A different number of subsamples was employed for the FSS-III and the FQ—four and six, respectively. Their composition was determined by means of computerized random sampling *without* replacement. The essential details of these samples, for both questionnaires, are given in Table 1.

## Analysis

As indicated, both PCA and PFA were used. Much confusion exists in actual usage of these models. Following Harman (1967) and Nie, Hull, Jenkins, Steinbrenner, and Bent (1975), the solution of factoring with unities in the diagonals of the correlation matrix is called PCA, and the solution with communalities in the diagonals of the correlation matrix is referred to as PFA. The routines in SPSS (Nie et al., 1975) that correspond to the respective models are the PA1 and PA2 factoring methods. PA2 differs from PA1 in two important respects. First, PA2 automatically replaces the main diagonal elements of the correlation matrix with communality estimates. Initial estimates of the communalities are given by the squared multiple correlation between a given variable and the rest of the variables in the matrix (in cases where the determinant of the matrix is smaller than $10^{-8}$ or the matrix is singular, the absolute value of the largest element in each column is used instead of the squared multiple correlation). In SPSS, if the determinant is negative, factor extraction will *not* be performed.

The second difference from PA1 is that PA2 employs an iteration procedure for improving the estimates of communality. First, the program determines the number of factors to be extracted from the original/unreduced correlation matrix. The program then replaces the main diagonal elements of the correlation matrix with initial estimates of communalities, the $R^2$ estimates. Next, it extracts the same number of factors from this reduced matrix, and the variances accounted for by these factors become new communality estimates. The diagonal elements are then replaced with these new communalities. This process continues until the differences between the two successive communality estimates are negligible. However, if for a partic-

Table 1
Number of Subjects and Subjects to Variables Ratio
in Each Sample for FSS-III and FQ Questionnaires

| Sample | FSS-III | | FQ | |
|---|---|---|---|---|
| | $N$ | Ratio | $N$ | Ratio |
| 1 | 100 | 1.3:1 | 26 | 1.3:1 |
| 2 | 200 | 2.6:1 | 52 | 2.6:1 |
| 3 | 300 | 3.9:1 | 78 | 3.9:1 |
| 4 | 504 | 6.6:1 | 132 | 6.6:1 |
| 5 | – | – | 280 | 14.0:1 |
| 6 | – | – | 392 | 19.8:1 |
| Total | 1104 | 14.5:1 | 960 | 48.0:1 |

ular iteration any of the communalities exceed 1.0, the iteration process will cease and the factors for the previous acceptable iteration will be retained (Nie et al., 1975, p. 480). In the present analyses, no messages were printed indicating that the latter had occurred.

On the basis of theoretical and previous empirical findings, it was quite clear how many factors had to be extracted for rotation for each of the schedules: (1) 5 for the FSS-III [i.e., Social fears (I); Agoraphobic fears (II); fears related to Bodily Injury, Death and Illness (III); fears of Sexual and Aggressive Scenes (IV); and fears of Harmless Animals (V)], as obtained by Arrindell (1980) from factor analyzing data of 703 phobic club members; and (2) 4 factors for the FQ [i.e., Agoraphobia (I), Social Phobia (II), Blood-Injury Phobia (III), and Anxiety-Depression (IV)], on the basis of factor analyses by Marks and Mathews (1979) of data of 1,000 phobic club members and 300 phobic patients in Britain. Both sets of factors are among those that have been repeatedly found in a relatively large number of factor analyses of data of distinct samples (Granell de Aldaz, 1982; Tasto, 1977; Wade, 1978) and judged to reflect the clinical status of (phobic) psychiatric patients (Marks, 1969; Marks & Mathews, 1979).

Following factor analyses of responses to each questionnaire for each subsample, the Varimax rotated matrices of factor loadings were compared with each other by using methods that enabled the interpretation of the similarity between sets of factors derived for identical items from different populations in terms of indices of factor comparisons (Tucker's coefficient of congruence $\phi$, the best measure of factorial invariance available, see ten Berge, 1977, p. 9). Tucker's $\phi$ ranges from $-1$ through 0 (and .70 which denotes poor similarity) to $+1$ (perfect agreement).

The lower bound cutoff for the $\phi$ index was set at .80, as Haven and ten Berge (1977) have shown empirically that sets of factors for which calculated coefficients have been found to be .80 and larger tend to be judged as similar. Actual calculations involved were carried out by means of Rotation of Unequal Matrices (RUM; Brokken, 1981).

In addition, the factorial structures obtained for each subsample were compared with the theoretical structure by using the multiple-group method of confirmatory analysis (Nunnally, 1978, pp. 394–400). Following Nunnally's (1978) recommendations in interpreting results obtained with this method, both the internal consistency reliabilities and the intercorrelations of the factors were considered.

## Results

Prior to conducting factor analyses, statistics describing item properties in each of the samples were critically inspected. The great majority of item statistics that were produced were such that no "difficulty factors" (e.g., Gorsuch, 1974, p. 260) could be expected to occur in any of the analyses performed.

### The FSS-III

A survey of the results that were obtained in comparing the subsamples ($Ns = 100, 200, 300,$ and 504) with each other are displayed in Tables 2 and 3. The upper-right and lower-left rectangular triangles in Table 2 give the indices of factor comparisons for analyses using PCA and PFA, respectively. The diagonal coefficients give an indication of the degree of match between analogous factors across factor analytic method (PCA vs. PFA) by subsample.

As can be seen (Table 2), FSS factor stability is demonstrated with both PCA and PFA across all samples, except for Factor IV (fears of Sexual and Aggressive Scenes). Further, the $\phi$ values in the diagonal clearly indicate that, with the factor analytic methods used here, the results produced are virtually identical ($\phi \geqq .99$).

Since Factor IV did not appear to be consistently stable across group comparisons, this finding was for the time being neglected in the process of making inferences about whether the observations-to-variables ratio affects factor stability.

The remaining four factors all yielded coefficients of invariance in excess of .80, mostly above .90. Fortunately, in neither of the analyses were

there $\phi$ coefficients found for comparisons involving nonanalogous factors that exceeded .74. Inspection of $\phi s > .60$ (an arbitrary cutoff level) indicated that of the 500 coefficients that were

Table 2

*PHI* Coefficients Between Identical Factors Obtained With PCA (Upper Triangle), PFA (Lower Triangle), and Comparisons Between Factor Analytic Method For Identical Factors (Diagonals) For FSS Data

| Sample and Factor/Component | Sample | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Total |
| **Sample 1** | | | | | |
| I | 1.00 | .94 | .93 | .92 | .94 |
| II | 1.00 | .87 | .90 | .89 | .92 |
| III | 1.00 | .83 | .84 | .85 | .87 |
| IV | 1.00 | .67 | .63 | .77 | .82 |
| V | 1.00 | .83 | .86 | .86 | .87 |
| Mean* | .40 | .44 | .44 | .44 | .44 |
| **Sample 2** | | | | | |
| I | .94 | 1.00 | .97 | .95 | .97 |
| II | .88 | 1.00 | .96 | .95 | .98 |
| III | .84 | 1.00 | .90 | .93 | .95 |
| IV | .73 | .99 | .79 | .85 | .90 |
| V | .83 | 1.00 | .89 | .91 | .93 |
| Mean* | .45 | .43 | .42 | .43 | .43 |
| **Sample 3** | | | | | |
| I | .94 | .97 | 1.00 | .96 | .98 |
| II | .91 | .97 | 1.00 | .97 | .99 |
| III | .85 | .92 | 1.00 | .96 | .97 |
| IV | .70 | .82 | .99 | .68 | .73 |
| V | .87 | .90 | 1.00 | .91 | .92 |
| Mean* | .47 | .46 | .43 | .44 | .43 |
| **Sample 4** | | | | | |
| I | .92 | .96 | .96 | 1.00 | .99 |
| II | .89 | .95 | .97 | 1.00 | .99 |
| III | .85 | .94 | .96 | 1.00 | .99 |
| IV | .81 | .89 | .75 | 1.00 | .98 |
| V | .87 | .92 | .92 | 1.00 | .98 |
| Mean* | .47 | .47 | .47 | .45 | .43 |
| **Total** | | | | | |
| I | .95 | .98 | .99 | .99 | 1.00 |
| II | .92 | .98 | .99 | .99 | 1.00 |
| III | .88 | .96 | .98 | .99 | 1.00 |
| IV | .85 | .93 | .79 | .99 | 1.00 |
| V | .88 | .95 | .94 | .99 | 1.00 |
| Mean* | .48 | .47 | .47 | .48 | .46 |

*Mean of the off-diagonal coefficients.

Table 3
Correlations (Pearson's $r$) Among FSS-III Factors, Internal Consistency Reliability
(Cronbach's $Alpha$) for Each Scale and Range of Item-Total Score Correlations by Sample

| Sample and Factor/Component | Factor | | | | | $Alpha$ | range $r_{it}$ |
| | I | II | III | IV | V | | |
|---|---|---|---|---|---|---|---|
| **Sample 1** | | | | | | | |
| I | 1.00 | | | | | 0.91 | 0.46-0.76 |
| II | 0.28 | 1.00 | | | | 0.87 | 0.30-0.77 |
| III | 0.33 | 0.52 | 1.00 | | | 0.86 | 0.36-0.71 |
| IV | 0.62 | 0.33 | 0.46 | 1.00 | | 0.78 | 0.32-0.70 |
| V | 0.30 | 0.50 | 0.42 | 0.57 | 1.00 | 0.84 | 0.52-0.72 |
| **Sample 2** | | | | | | | |
| I | 1.00 | | | | | 0.90 | 0.46-0.73 |
| II | 0.26 | 1.00 | | | | 0.88 | 0.36-0.72 |
| III | 0.45 | 0.40 | 1.00 | | | 0.90 | 0.52-0.77 |
| IV | 0.56 | 0.21 | 0.48 | 1.00 | | 0.80 | 0.43-0.65 |
| V | 0.38 | 0.17 | 0.47 | 0.45 | 1.00 | 0.86 | 0.58-0.70 |
| **Sample 3** | | | | | | | |
| I | 1.00 | | | | | 0.92 | 0.49-0.77 |
| II | 0.32 | 1.00 | | | | 0.87 | 0.39-0.72 |
| III | 0.40 | 0.49 | 1.00 | | | 0.88 | 0.41-0.78 |
| IV | 0.50 | 0.28 | 0.45 | 1.00 | | 0.77 | 0.30-0.63 |
| V | 0.34 | 0.25 | 0.35 | 0.40 | 1.00 | 0.85 | 0.56-0.68 |
| **Sample 4** | | | | | | | |
| I | 1.00 | | | | | 0.91 | 0.47-0.75 |
| II | 0.35 | 1.00 | | | | 0.87 | 0.36-0.70 |
| III | 0.42 | 0.41 | 1.00 | | | 0.89 | 0.45-0.73 |
| IV | 0.47 | 0.30 | 0.47 | 1.00 | | 0.78 | 0.35-0.64 |
| V | 0.36 | 0.16 | 0.42 | 0.44 | 1.00 | 0.82 | 0.56-0.63 |
| **Total** | | | | | | | |
| I | 1.00 | | | | | 0.91 | 0.47-0.75 |
| II | 0.32 | 1.00 | | | | 0.87 | 0.39-0.70 |
| III | 0.41 | 0.44 | 1.00 | | | 0.88 | 0.45-0.75 |
| IV | 0.51 | 0.28 | 0.47 | 1.00 | | 0.78 | 0.36-0.64 |
| V | 0.35 | 0.22 | 0.41 | 0.44 | 1.00 | 0.84 | 0.57-0.65 |

calculated for nonanalogous factors only 43 (8.6%) exceeded this criterion. Closer inspection of these figures indicated that this "disturbance" was clearly caused by Factor IV: 95.3% (41 out of 43) of the coefficients exceeding .60 were for comparisons between this factor and any of the remaining factors, in particular Factor V (53.4%).

There seemed to be a trend mirroring higher invariance coefficients for comparisons involving higher $N$s. This pattern was more marked for the analysis involving PCA. However, when the strength of each factor was compared across studies, the differences were small (both PCA and PFA), with the strength of PFA factors being consistently less strong than those obtained with PCA. Altogether, the factors explained a comparable amount of variance: 34.0, 33.2, 32.0, and 31.1 (PCA) vs. 31.3, 30.4, 29.1, and 28.2 (PFA) in subsamples for which $N$ equaled, respectively, 100, 200, 300, and 504. The corresponding figures for the total sample were comparably strong: 30.9 and 28.1, respectively.

Further analyses indicated that all factors were internally very consistent, with the aforementioned unstable factor producing consistently lower coef-

ficients (Cronbach's coefficient α) in each subsample. Further, though the items of this factor attained the minimum recommended value of .2 in their correlations (item-remainder) with their corresponding scale, the associations were consistently less strong than the other items' correlations with their related total scales. Incidentally, inspection of the Varimax rotated matrices of factor loadings revealed, in some instances, that the items on Factor IV loaded less strongly on their related factor than the remaining items on their respective factors.

Finally, comparisons of factor correlations across subsamples did not reveal a meaningful number of large departures in the degree of relationships denoted by the coefficients (see Surwillo, 1980, p. 59). The generally low to moderate factor correlations, in conjunction with the corresponding very satisfactory scale reliabilities, demonstrate that the FSS-III factors can be distinguished reliably in self-ratings of irrational fears.

Table 2 gives φs for factor comparisons between subsamples and the total sample ($N = 1,104$). The coefficients clearly indicate that, with a ratio of 1.3:1, there is, irrespective of whether PCA or PFA is used (all φs for identical factors across analytic method = 1.00), virtually no difference in factor structure compared to the one that emerges from the original larger sample.

Some clarification of the status of Factor IV is appropriate before proceeding. The factors obtained in Arrindell (1980) were based on analyses of data of 703 phobic individuals solely and not on repeated replications with different phobic (sub)samples. A number of criteria were used to yield the five scales mentioned above. One of these criteria was that the factorial solution chosen should be invariant across gender. Since a 6-factor solution produced mostly low levels of invariance for the factors across the genders (male $N = 151$; female $N = 552$), a 5-factor solution was chosen, which obtained the following satisfactory φs: Factor I = .97, II = .94, III = .90, IV = .92, and V = .83.

In addition, these factors were found to be highly invariant across age as well ($Ns$: 366, 18-42 years vs. 335, 43-85 years). The corresponding φ coefficients were: Factor I = .97, II = .97, III = .96, IV = .82, and V = .92. Thus, validation across significant subject parameters, instead of split samples or similar populations, was obtained in order to provide some external means of evaluating the replicability prior to accepting the produced factor analytic results. The lack of replicability of Factor IV across split samples found here emphasizes the importance of cross-validation of factor structures with the purpose of replicating the (supposed) structure and maintaining a constant check on the behavior of the variables (e.g., Armstrong & Soelberg, 1968; Barrett & Kline, 1981).

## The FQ

Tables 4 and 5 provide results that were obtained in comparing the subsamples ($Ns = 26, 52, 78, 132, 280$, and $392$) with each other. Also, Table 4 shows the φs for comparisons between the subsamples and the total sample ($N = 960$).

As can be seen (Table 4), factor stability is demonstrated across all subsamples (in both PCA and PFA), except for the one for which $N = 26$. In this sample only one factor (Factor I, Agoraphobia) appeared consistently fairly invariant across the comparisons. The subsamples for which $N \geq 52$ yielded consistently high φ coefficients, mostly above .90. As with the FSS-III, the coefficients in the diagonal (Table 4) clearly indicate that, in all comparisons across subsamples, the results produced are virtually identical for the two factor extraction procedures used here ($\phi \geq .98$). Also, there were no results found for comparisons involving nonanalogous factors that exceeded .74; only 6 out of 588 (1%) of the off-diagonal coefficients were above .60. These were found for comparisons involving Sample 1 ($N = 26$). In addition, with $N = 52$ (Sample 2) the off-diagonal coefficients produced were mostly higher than those obtained for comparisons involving $Ns \geq 78$.

A fairly clear trend was observed in that higher coefficients of invariance were yielded for comparisons involving higher $Ns$ (both PCA and PFA). However, inspection of the strength of the factors across studies suggested a clear trend in comparable eigenvalues for $Ns \geq 78$. When totaled, the

FQ factors explained a comparable amount of variance across the largest four samples (both PCA and PFA). The corresponding figures for the six subgroups were: 12.2, 13.3, 11.9, 11.7, 11.9, and 11.8 (PCA) vs. 10.8, 11.8, 10.1, 9.9, 10.0, and 9.8 (PFA), respectively. The corresponding figures

Table 4

PHI Coefficients Between Identical Factors Obtained With PCA (Upper Triangle), PFA (Lower Triangle), and Comparisons Between Factor Analytic Method For Identical Factors (Diagonals) For FQ Data

| Sample and Factor/Component | Sample | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| **Sample 1** | | | | | | | |
| I | .99 | .88 | .91 | .81 | .83 | .83 | .85 |
| II | .98 | .59 | .48 | .48 | .59 | .59 | .58 |
| III | .99 | .75 | .78 | .77 | .84 | .85 | .85 |
| IV | 1.00 | .76 | .87 | .79 | .86 | .86 | .86 |
| Mean* | .18 | .28 | .23 | .23 | .23 | .23 | .23 |
| **Sample 2** | | | | | | | |
| I | .83 | 1.00 | .93 | .89 | .92 | .89 | .92 |
| II | .58 | 1.00 | .84 | .84 | .87 | .87 | .89 |
| III | .77 | 1.00 | .92 | .95 | .91 | .91 | .93 |
| IV | .72 | .99 | .93 | .85 | .89 | .89 | .90 |
| Mean* | .31 | .48 | .38 | .40 | .41 | .40 | .41 |
| **Sample 3** | | | | | | | |
| I | .89 | .93 | 1.00 | .94 | .95 | .93 | .96 |
| II | .48 | .87 | 1.00 | .93 | .95 | .94 | .95 |
| III | .77 | .93 | 1.00 | .92 | .96 | .96 | .97 |
| IV | .85 | .93 | 1.00 | .93 | .97 | .97 | .98 |
| Mean* | .27 | .43 | .31 | .30 | .31 | .30 | .31 |
| **Sample 4** | | | | | | | |
| I | .81 | .91 | .95 | 1.00 | .96 | .96 | .97 |
| II | .48 | .85 | .93 | .99 | .97 | .95 | .97 |
| III | .74 | .96 | .93 | 1.00 | .93 | .93 | .96 |
| IV | .76 | .89 | .94 | .99 | .97 | .97 | .98 |
| Mean* | .26 | .44 | .35 | .33 | .32 | .31 | .32 |
| **Sample 5** | | | | | | | |
| I | .82 | .94 | .96 | .97 | 1.00 | .98 | .99 |
| II | .58 | .91 | .95 | .96 | 1.00 | .98 | 1.00 |
| III | .82 | .93 | .96 | .93 | 1.00 | .97 | .99 |
| IV | .84 | .92 | .98 | .97 | 1.00 | .99 | 1.00 |
| Mean* | .27 | .46 | .37 | .38 | .36 | .33 | .33 |
| **Sample 6** | | | | | | | |
| I | .83 | .91 | .94 | .97 | .98 | 1.00 | .99 |
| II | .59 | .91 | .94 | .93 | .98 | 1.00 | .99 |
| III | .84 | .93 | .96 | .93 | .97 | 1.00 | .99 |
| IV | .85 | .91 | .98 | .97 | .99 | 1.00 | 1.00 |
| Mean* | .26 | .44 | .35 | .36 | .37 | .32 | .31 |
| **Total** | | | | | | | |
| I | .84 | .94 | .97 | .98 | .99 | .99 | 1.00 |
| II | .57 | .92 | .96 | .96 | .99 | .99 | .99 |
| III | .83 | .95 | .97 | .96 | .99 | .99 | 1.00 |
| IV | .84 | .92 | .98 | .99 | .99 | 1.00 | 1.00 |
| Mean* | .27 | .46 | .36 | .37 | .39 | .37 | .35 |

*Mean of the off-diagonal coefficients.

Table 5

Correlations (Pearson's $r$) Among FQ Factors, Internal Consistency Reliability (Cronbach's *Alpha*) for Each Scale and Range of Item-Total Score Correlations by Sample

| Sample and Factor/Component | Factor | | | | Alpha | range $r_{it}$ |
|---|---|---|---|---|---|---|
| | I | II | III | IV | | |
| **Sample 1** | | | | | | |
| I | 1.00 | | | | 0.76 | 0.41-0.71 |
| II | 0.55 | 1.00 | | | 0.70 | 0.32-0.63 |
| III | 0.42 | 0.44 | 1.00 | | 0.71 | 0.29-0.62 |
| IV | -0.16 | 0.10 | -0.01 | 1.00 | 0.82 | 0.52-0.70 |
| **Sample 2** | | | | | | |
| I | 1.00 | | | | 0.90 | 0.68-0.88 |
| II | 0.58 | 1.00 | | | 0.78 | 0.22-0.79 |
| III | 0.63 | 0.51 | 1.00 | | 0.83 | 0.42-0.79 |
| IV | 0.42 | 0.51 | 0.52 | 1.00 | 0.76 | 0.36-0.66 |
| **Sample 3** | | | | | | |
| I | 1.00 | | | | 0.85 | 0.60-0.86 |
| II | 0.46 | 1.00 | | | 0.74 | 0.43-0.68 |
| III | 0.25 | 0.30 | 1.00 | | 0.79 | 0.45-0.72 |
| IV | 0.20 | 0.25 | 0.40 | 1.00 | 0.81 | 0.48-0.69 |
| **Sample 4** | | | | | | |
| I | 1.00 | | | | 0.85 | 0.56-0.82 |
| II | 0.42 | 1.00 | | | 0.69 | 0.32-0.56 |
| III | 0.29 | 0.31 | 1.00 | | 0.79 | 0.37-0.72 |
| IV | 0.42 | 0.20 | 0.32 | 1.00 | 0.80 | 0.37-0.75 |
| **Sample 5** | | | | | | |
| I | 1.00 | | | | 0.85 | 0.61-0.77 |
| II | 0.39 | 1.00 | | | 0.76 | 0.45-0.63 |
| III | 0.31 | 0.35 | 1.00 | | 0.80 | 0.49-0.67 |
| IV | 0.27 | 0.30 | 0.43 | 1.00 | 0.84 | 0.55-0.73 |
| **Sample 6** | | | | | | |
| I | 1.00 | | | | 0.85 | 0.60-0.80 |
| II | 0.41 | 1.00 | | | 0.77 | 0.42-0.62 |
| III | 0.24 | 0.42 | 1.00 | | 0.81 | 0.53-0.71 |
| IV | 0.19 | 0.28 | 0.37 | 1.00 | 0.78 | 0.45-0.67 |
| **Total** | | | | | | |
| I | 1.00 | | | | 0.85 | 0.60-0.80 |
| II | 0.42 | 1.00 | | | 0.75 | 0.43-0.62 |
| III | 0.30 | 0.39 | 1.00 | | 0.80 | 0.50-0.70 |
| IV | 0.25 | 0.28 | 0.38 | 1.00 | 0.80 | 0.46-0.69 |

for the total sample were 11.6 and 9.7, respectively. In addition, further analyses indicated that the factor correlations were found to stabilize across samples with $N \geqq 78$, that is, the coefficients were comparable in terms of the degree of relationships which they denoted (Table 5). Further, with $N = 26$ or 52, a significant number of items were shown to correlate low with their theoretically related scale.

The satisfactory internal consistency reliabilities of the scales (Table 5) in conjunction with the generally low factor correlations, point to the possibility of reliable distinction between the FQ dimensions in the self-report of irrational fears and of complaints of anxiety and depression.

Comparison between subsamples and the total sample clearly indicated that, with a ratio of 3.9:1,

there is, once more irrespective of whether PCA and PFA was used, virtually no difference in factor structure compared to the one which was found for the original larger sample. Note that, again, the off-diagonal coefficients were highest for an $N:p$ ratio of 2.6:1 or $N = 52$. Further, in the comparisons involving $N$s $\geqq 78$ there are hardly any differences in the $\phi$ values obtained with increasing $N$s (in both PCA and PFA). All $\phi$s for identical factors across factor analytic method equal 1.00.

As stated earlier, the results presented here are not influenced by the factor analytic methods applied. However, it should be noted that (1) *off-diagonal* coefficients are, on the average, somewhat higher in PFA than in PCA; and (2) the sum of squares obtained in PCA are, on the average, 1.1 higher than in PFA in studies with the FSS-III. The corresponding figure for the FQ studies equals 1.2. Hence, the loadings obtained in PCA are, on the average, $\sqrt{1.1}$ and $\sqrt{1.2}$, respectively, higher than those produced with PFA. Put somewhat differently: loadings of .80 and .30 in PCA correspond with loadings of .76 and .29, respectively, in PFA in regards to the studies conducted with the FSS-III. The corresponding figures for the FQ are .80 vs. .73 and .30 vs. .27, respectively.

## Discussion and Conclusions

Mathematically, it is desirable to factor analyze cross-products, covariance, or correlation matrices computed with data matrices where the number of observations $N$ is greater than the number of variables $p$. Although there is some evidence that analysis of $N < p$ association matrices can produce somewhat similar results to those from $N > p$ matrices, it is not a methodology to be recommended (Kline & Barrett, 1983, p. 145). Nunnally (1978) provided an excellent summary of the arguments against numerically and statistically undersampling the variable domain. However, the matter of just how many observations should be made on each variable has remained unresolved. Indeed, how great this ratio should be for reliable factor loadings is a matter of dispute among factor analysts.

The present study provides information that is useful in refuting one commonly stated rule of thumb

in factor analysis, such as the trend emerging in *clinical* research that $N:p$ should equal 5:1. The minimum ratio required to yield a clear, recognizable factor solution with the FSS-III was 1.3:1, whereas the corresponding figure for the FQ was three times larger: 3.9:1. The respective minimum $N$s were 100 and 78.

What is indicated in the present study is that none of the suggestions outlined earlier is supported: neither the 10:1 ratio recommended by Nunnally (1978), nor the 5:1 ratio proposed by Gorsuch (1974) and Everitt (1975), nor the 3:1 to 6:1 ratio suggested by Cattell (1978). In addition, no support was found for the view that $N$s of $\cong$ 200 (Cattell, 1978; Guilford, 1954) would be necessary for performing sound factor analyses. It thus appeared that neither the observations to variables ratio nor an absolute number of observations had any effect on factor stability. With the present measures, even smaller numbers of observations than suggested by Guilford and Cattell appeared to be sufficient to yield reliable results.

It is suggested that, for the purpose of establishing a given number of true factors, sample size should be related to the number of *factors* drawn: The present results confirm that stable factor solutions may be obtained when sample size is approximately 20 times the number of factors (FSS-III: $5 \times 20 = 100$; FQ: $4 \times 20 \cong 78$).

The unique definition of an underlying factor depends on the precision of estimates of correlation coefficients (or the degree of sampling error, which depends on $N$) *and* on the degree of factor overdetermination (i.e., the number of measurement variables related to each factor in the data set). Clear factor overdetermination is good policy since it favors better interpretation and replicability of factors (e.g., Gorsuch, 1974, p. 295; Kline & Barrett, 1983). Since the more redundancy (variables) there is in a given factor domain, the more stable the factor structure will be for any set of "number of subjects," both the number of observations and the number of variables should exceed the number of factors by a substantial multiple. Obviously, it is advisable that external means are provided to permit an evaluation of the replicability of factor analytic results prior to their final acceptance, since

interpretable factors can be easily found in random data (e.g., Armstrong & Soelberg, 1968; Horn, 1967).

The results obtained in this study make the possibility of conducting factor analyses more feasible than when, say, an $N:p$ ratio of 10:1 (which is unnecessarily stringent) has to be used. However, the following points should be noted. In the present study, the number of extracted components/factors was set in advance. This was done on the basis of knowledge of a clear a priori factor structure, which knowledge is a valid criterion in considering a selection of the number of components or factors to rotate (Futch et al., 1982, p. 31). With the present instruments, extraction of larger numbers of factors than were actually extracted (e.g., all those with eigenvalues exceeding unity) would have led to gross overfactoring and, consequently, would have produced distorted factor solutions (e.g., Zwick & Velicer, 1982). On a number of grounds mentioned earlier, it was judged that the a priori structures were clear and reliable enough to form a baseline against which to assess the stability of factor solutions from small to large samples.

However, future studies (employing different variable sets than were used here) should critically examine the influence of the number of factors extracted in examining the question raised in the present study, so as to rule out the possibility that the results may be due to the problem of not extracting enough factors (i.e., underfactoring). Related to this, the influence of the ratio of factors extracted to the number of variables in the data set should be considered in a repetition of this study as well. However, it should be borne in mind that, in this study, different ratios (FSS-III, 76:5 vs. FQ, 20:4) produced similar conclusions. Also, the question as to whether the results obtained might be due to an artifact of the type of variables (fear items) used should be addressed in a new study.

Another important point of discussion concerns the standard used in the present study for an acceptable $\phi$ coefficient ($\geq .80$). This criterion was based on empirical work by Haven and ten Berge (1977) indicating that $\phi$ is highly correlated with objective judgments of similarity between factors.

In addition, the rule of thumb of accepting factors as identical when their congruence is in the .80s or .90s (Nesselroade & Baltes, 1970) was supported by their data. By contrast, L. Tucker (personal communication, 1984) proposed more conservative standards for interpretation of the coefficient of congruence of factors. According to Tucker, congruence coefficients of .98 to 1.00 are considered ''excellent,'' coefficients of .92 to .98 as ''good,'' .82 to .92 as ''borderline,'' .68 to .82 as ''poor,'' and coefficients below .68 as ''terrible.'' The use of this information—for which empirical support is lacking—in reinterpreting the present findings produces mixed results, $N$s of 200 (FSS-III) and 78 (FQ) appearing to be sufficient to obtain reliable results. Thus, the use of Tucker's standards does *not* refute the claim made above that neither an observations to variables ratio nor an absolute minimum of observations would be required to obtain stable factor solutions. Further, it should be borne in mind that, on the whole, a poor match was found between sets of nonanalogous factors, both with the FSS-III and the FQ.

Of more than mere incidental interest are the additional conclusions that can be drawn from this study. First, comparisons of results from PCA and PFA (with real data) supported findings from the Velicer et al. (1982) study (which employed simulated data) that ''the major distinction between the two approaches ... is the slightly higher loadings produced by component analysis'' (p. 387). Velicer et al. used this finding to illustrate why factor analysis resulted in more Type II errors and component analysis in more Type I errors in their decision tables. This small difference runs parallel with the findings obtained in the present study: $\phi$ coefficients comparing the degree of match between nonanalogous factors obtained in analyses involving PFA were somewhat higher than those obtained in comparisons of unlike factors produced in PCA. Thus, the present study provides additional evidence to support the Velicer et al. (1982) conclusion that, beyond the very small difference noted above, no support is found for the viewpoint that factor analysis produces results which are superior to or even different for any practical pur-

poses from component analysis (p. 387).

Secondly, support was found for the replicability of four out of five dimensions of the FSS-III and for cross-national reliability and validity of the FQ dimensions. Thirdly, this study has offered a means of evaluating the internal validity of gathered data on the generalizability of phobic dimensions across different samples (phobics vs. unselected psychiatric outpatients vs. nonpatient spouses of psychiatric outpatients vs. nonstudent normals vs. students), which are currently underway (Arrindell, Emmelkamp, & van der Ende, 1984).

## References

Armstrong, J. S., & Soelberg, P. (1968). On the interpretation of factor analysis. *Psychological Bulletin, 70,* 361–364.

Arrindell, W. A. (1980). Dimensional structure and psychopathology correlates of the Fear Survey Schedule (FSS-III) in a phobic population: A factorial definition of agoraphobia. *Behaviour Research and Therapy, 18,* 229–242.

Arrindell, W. A., Emmelkamp, P. M. G., & van der Ende, J. (1984). Phobic dimensions: I. Reliability and generalizability across samples, gender and nations. *Advances in Behaviour Research and Therapy, 6,* 207–254.

Barrett, P. T., & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study and Group Behaviour, 1,* 23–33.

Brokken, F. B. (1981). *RUM: Rotation of unequal matrices.* Groningen, The Netherlands: State University of Groningen.

Cattell, R. B. (1978). *The scientific use of factor analysis.* New York: Plenum Press.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16 PF).* Champaign IL: Institute of Personality and Ability Testing.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46,* 648–659.

Cureton, E. E., & D'Agostino, R. B. (1983). *Factor analysis: An applied approach.* Hillsdale NJ: Lawrence Erlbaum.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry, 126,* 237–240.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Eysenck Personality Questionnaire manual.* San Diego CA: Educational and Industrial Testing Service.

Futch, E.J., Scheirer, C. J., & Lisman, S. A. (1982). Factor analyzing a scale of assertiveness. *Behavior Modification, 6,* 24–43.

Gorsuch, R. L. (1974). *Factor analysis.* Philadelphia: W. B. Saunders.

Granell de Aldaz, E. (1982). Factor analysis of a Venezuelan Fear Survey Schedule. *Behaviour Research and Therapy, 20,* 313–332.

Guilford, J. P. (1954). *Psychometric methods* (2nd Ed.). New York: McGraw-Hill.

Harman, H. H. (1967). *Modern factor analysis* (2nd Ed., Revised). Chicago: The University of Chicago Press.

Harris, R. J. (1975). *A primer of multivariate statistics.* New York: Academic Press.

Haven, S., & ten Berge, J. M. F. (1977). *Tucker's coefficient of congruence as a measure of factorial invariance. An empirical study* (Heymans Bulletins, HB 77-290-EX). Groningen, The Netherlands: State University of Groningen, Department of Research in Personality.

Horn, J. L. (1967). On subjectivity in factor analysis. *Educational and Psychological Measurement, 27,* 811–820.

Kennard, D. (1978). Identifying dimensions of perceived patient change: Some problems in the use of factor analysis. *British Journal of Medical Psychology, 51,* 41–52.

Kline, P. (1979). *Psychometrics and psychology.* London: Academic Press.

Kline, P., & Barrett, P. (1983). The factors in personality questionnaires among normal subjects. *Advances in Behaviour Research and Therapy, 5,* 141–202.

Marks, I. M. (1969). *Fears and phobias.* London: Heinemann.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy, 17,* 263–267.

Mulaik, S. A. (1972). *The foundation of factor analysis.* New York: McGraw-Hill.

Nesselroade, J. R., & Baltes, P. B. (1970). On a dilemma of comparative factor analysis: A study of factor matching based on random data. *Educational and Psychological Measurement, 30,* 935–948.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). *Statistical package for the social sciences* (2nd Ed.). New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory* (2nd Ed.). New York: McGraw-Hill.

O'Looney, B. A., & Barrett, P. T. (1983). A psychometric investigation of the Multidimensional Health Locus of Control Questionnaire. *British Journal of Clinical Psychology, 22,* 217–218.

Overall, J. E., & Klett, C. J. (1972). *Applied multivariate analysis.* New York: McGraw-Hill.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston: Northwestern University Press.

Simón, A. (1979). Effects of selective sampling on a factor analysis. *The Journal of General Psychology, 101,* 259–264.

Skinner, H. A. (1980). Factor analysis and studies on alcohol. *Journal of Studies on Alcohol, 41,* 1091–1101.

Surwillo, W. W. (1980). *Experimental design in psychiatry. Research methods for clinical practice*. New York: Grune and Stratton.

Tasto, D. L. (1977). Self-report schedules and inventories. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 153–193). New York: Wiley.

ten Berge, J. M. F. (1977). *Optimizing factorial invariance,* Groningen, The Netherlands: VRB.

van de Geer, J. P. (1971). *Introduction to multivariate analysis for the social sciences*. San Francisco: W. H. Freeman.

Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *The Journal of Multivariate Behavioral Research, 12,* 3–22.

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A monte carlo approach. *Multivariate Behavioral Research, 17,* 371–388.

Wade, T. C. (1978). Factor analytic approaches to the investigation of common fears: A critical appraisal and reanalysis. *Behavior Therapy, 9,* 923–935.

Wolpe, J., & Lang, P. J. (1964). A fear survey schedule for use in behaviour therapy. *Behaviour Research and Therapy, 2,* 27–30.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17,* 253–269.

## Author's Address

Send requests for reprints or further information to Willem A. Arrindell, Department of Clinical Psychology, Academic Hospital of the State University of Groningen, Oostersingel 59, 9713 EZ Groningen, The Netherlands.