Adv. Behav. Res. Ther. Vol. 6, pp. 207–254, 1984.
Printed in Great Britain. All rights reserved.

0146-6402/84 $0.00 + .50

PHOBIC DIMENSIONS: I. RELIABILITY AND GENERALIZABILITY ACROSS SAMPLES, GENDER AND NATIONS

The Fear Survey Schedule (FSS–III) and the Fear Questionnaire (FQ)

Willem A. Arrindell, Paul M. G. Emmelkamp and Jan van der Ende

The Department of Clinical Psychology, Academic Hospital of the State University of Groningen, Oostersingel 59, 9713 EZ Groningen, The Netherlands

SUMMARY

A brief overview is given of results of: (a) factor analytic studies with Fear Survey Schedules of relevance to the cross-sample and cross-national constancy of phobic dimensions; (b) reliability analyses with these dimensions. Employing data of somewhat more than 3000 *S*s from 10 independent clinical and normal samples, a number of analyses with the FSS-III and the FQ showed the dimensions which have been found to underlie these measures (Arrindell, 1980; Marks and Mathews, 1979) to be replicable in similar samples and invariant across distinct ones, and to be sufficiently reliable for purposes of research, in some instances even for applied problems. In addition, the FSS-III and FQ dimensions were shown to be invariant across sex. Of the main conclusions which were drawn three were particularly outstanding: (1) the Agoraphobic cluster of fear is *not* peculiar to samples comprising phobics in which agoraphobic fears occur full-blown in a significant number of *S*s. Such a factor can be interpreted in data of, for instance, (non)student normals as well provided that suitable techniques of multivariate analysis are employed, thus refuting the claim (cf. Hallam, 1978; Hallam and Hafner, 1978) that agoraphobia is a phenomenon that has "all or none" characteristics; (2) the use of total fear scores would appear to lead to invalid results; therefore, this general practice should be discarded; (3) small- to moderate-size decreases are likely to occur on phobic measures during relatively short and long non-treatment periods; such changes should be taken into consideration when drawing conclusions from results obtained in studies on pre- and post-therapy measurement.

A differential use of the FSS-III and the FQ is stressed.

INTRODUCTION

The Fear Survey Schedule (FSS) has been used clinically and experimentally in various forms for over 20 years and has enjoyed considerable prestige among behaviour therapists and researchers as a measure of phobic anxiety (e.g. Farley and Mealiea, 1972; Wade, 1978; Neiger *et al.* 1981). It has received attention from investigators and practitioners from many different countries (cf. Saigh, 1982, 1984). However, its usefulness as an assessment instrument has been criticised on psychometric grounds (Wolff and Merrens, 1974; Franks and Wilson, 1978).

Some of the shortcomings of the FSS as an assessment instrument could have been remedied if a number of reasonable proposals which have been made in the past to enhance its usefulness and psychometric properties had been followed up.

Recently, Marks and Mathews (1979) have introduced a brief one-page standard self-rating form (the Fear Questionnaire) for use as a minimum set of severity and outcome criteria in studies of phobic patients. The Fear Questionnaire (FQ) was derived from earlier versions used in large samples of phobic club members and phobic patients; it yields four scores — main phobia, global phobia, total phobia and anxiety–depression. The total phobia score is composed of Agoraphobic, Social and Blood-Injury subgroups. It has been suggested that the adoption of the FQ for research in clinical populations would facilitate comparison of results across research centres and studies. However, considering the meagre results which are available on psychometric properties for this new scale, the criticism — both methodological and theoretical in nature — which has been launched challenging the usefulness of the FSS may equally well apply to the FQ.

The aim of the present article is to give a critical review of studies which have examined certain psychometric properties of fear questionnaires, and to report the results of a study of factorial invariance characteristics and the reliability of phobic dimensions as captured by the Fear Survey Schedule (FSS-III; Wolpe and Lang, 1964) and the Fear Questionnaire.

FACTOR ANALYSES OF FEAR SURVEY SCHEDULES*

Hersen (1973) and Tasto (1977) have contributed reviews of factor analyses of FSSs administered to both clinical samples and normal student samples.

*Studies describing results obtained with specific fear questionnaires (i.e. scales measuring a narrow range of fears) are not dealt with here. The interested reader is referred to Tasto (1977), Fredrikson (1983) and Johnston *et al.* (1984). Further, the survey presented here is based on data obtained for adult *Ss*; studies describing findings obtained for adolescents can be found in Bamber (1974, 1979) and Ollendick (1983).

These reviews have been updated by Wade (1978) and, more recently, by Granell de Aldaz (1982), who restricted her survey to studies conducted on student samples. A comparison of these reviews indicates that the studies conducted on student samples (mostly American) clearly outnumber those in which subject groups comprised psychiatric individuals, which is surprising since the schedules were primarily constructed for use in clinical samples. This may reflect the established practice of using students as guinea pigs for psychological experiments and factor analytic and other multivariate studies, possibly because it affords a cheap and easily accessible source of labour. Aside from the convenience of such a custom, there is also the assumption that any results obtained from college or university students are applicable to the population at large, including the extreme ends of the normal distribution curve (Simón, 1979). However, the usage of multivariate statistics on a restricted sample could at times give results which apply to similar samples only (e.g. Berrier *et al.*, 1981; Chambless *et al.*, 1982; Galassi *et al.*, 1982).

The problem of generality of factor analytic results across samples has by no means received the attention it deserves. In Eysenck's (1973) words:

> "Factor-analysts usually assume that factors derived from one sample of the population will be very much the same as factors derived from another, and that factor structure is independent of choice of sample (except perhaps when sampling involved restriction of range). . . . factor-analysts ought to pay more attention to the problem, and not assume what remains to be proved" (p. 130).

On the basis of a *visual inspection* of the items and factors across the studies of his survey, Tasto (1977) concludes that they seem to indicate that among college students there is a fair degree of factor stability and that the most frequently occurring clusters are (1) fears related to small animals, (2) fears associated with death, physical pain, and surgery, (3) fears about aggression, and (4) fears concerning interpersonal events. The last is sometimes broken down into the components of social evaluation versus social interaction. With respect to the generality of these findings for different populations, he states that conclusions about the factor structure for self-reported fears for psychiatric populations remain somewhat tentative pending further studies.

The schedule performance of phobic patients in particular has rarely been investigated, despite the extensive use by behaviour therapists of self-ratings on fear schedules in treating such patients (Neiger *et al.*, 1981). A mere four factor analytic studies have been reported in which phobic patients are used as subjects (Meikle and Mitchell, 1974; Hallam and Hafner, 1978; Arrindell, 1980; Arrindell and Zwaan, 1982).

Meikle and Mitchell (1974) used a sample of 115 phobics and described 21 factors on the FSS-III (number of schedule items, $i = 76$), although they concluded that only two of these (Fear of Social Disapproval and Witnessing Physical Assaults) were genuine factors. Hallam and Hafner (1978) factor analysed a combined fear and symptom questionnaire ($i = 70$) completed by a sample of phobic individuals ($N = 171$), and interpreted three of the most commonly described factors: Social, Injury/Disease and Animal fears. In addition, an Agoraphobia factor clearly emerged. In a re-analysis which excluded the agoraphobic *S*s, self-ratings of *S*s diagnosed as socially phobic ($N = 25$), or as having some other specific phobia ($N = 53$), failed to produce an Agoraphobia factor, which was interpreted as pointing to the all-or-none nature of agoraphobic fear. Arrindell (1980) factor analysed the FSS-III ($i = 76$), utilizing a large phobic sample ($N = 703$) comprising predominantly agoraphobics. Five factors were interpreted: (1) Social Fears, (2) Agoraphobia, (3) Fears related to Bodily Injury, Death and Illness, (4) Fear of Display of Sexual and Aggressive Scenes, and (5) Fear of Harmless Animals. In a factor analysis of data of 510 predominantly agoraphobic *S*s on a 95-item fear questionnaire (AZUV), Arrindell and Zwaan (1982) described four conceptually clear factors which they termed: (1) Agoraphobia, (2) Fears related to Contamination, Violence and Sex, (3) Fears concerning Bodily Injury, Illness and Death, and (4) Social Phobia.

Since the appearance of the Tasto (1977) survey only one study which might shed light on the structure of fear responses in normal (*non*-student) samples has been conducted. In a normal sample of South African Indians ($N = 120$), randomly and diversely selected, Spinks (1980) interpreted three fear dimensions after having factor analysed item-scores on a 73-item FSS: (1) Physical Danger, (2) Interpersonal Situations and (3) Animals.

Culture and nationality may be among the significant determinants of FSS's factor structure. Although there has been some basic research into the nature of fear in groups identifiable by cultural and national demarcations (Goldberg *et al.*, 1975b; Farley *et al.*, 1978), hitherto not one study has been carried out with the purpose of determining the cross-national or cross-cultural generalizability of dimensions of irrational fear. Such an endeavour would improve the internal validity of results from cross-national comparisons (in terms of descriptive statistics) of normals, and also enhance the utility of fear measures constructed with the aim of internationally comparing results obtained for patients across research centres. Therefore, we shall now consider results from factor analytic investigations which might shed light on this matter. Since most studies conducted to date have utilized Americans as *S*s, studies describing data of non-American samples would appear to be more interesting in the present context. Besides the Dutch studies by Arrindell (1980) and Arrindell and Zwaan (1982) and the South African study by Spinks (1980) mentioned above, two recently conducted studies among *students* in Venezuela and Britain deserve comment. Granell de Aldaz (1982) developed a FSS ($i = 64$) for university students in Venezuela. It was administered to 871 first-year students and resulted in the identification of seven factors in the female sample ($N = 318$). These factors were: fears relating to (1) Aversive Interpersonal Events, (2) Violence and Physical Assault, (3) Injuries and Surgical Operations, (4) Animals and Insects, (5) Responsibility, (6) Death and Physical Threats, (7) Doctors and Hospitals. The factors of the males ($N = 553$) coincided with the first four of their female counterparts. Additionally, a fifth Loss of Self-Control factor was interpreted. In a recent study, Kartsounis *et al.* (1983) administered an 88-item version of the FSS-III to 547 British university students. Factor analysis produced factors which related to (1) Social Inadequacy, (2) Tissue Damage, (3) Travel, and (4) Animals.

At face value, these factors show a great deal of resemblance to those described for Israeli students in a study by Goldberg *et al.* (1975a). In the data of the males ($N = 129$) the following factors ($i = 97$) were identified: (1) Seeing Blood, (2) Social Competence, (3) Medical Intervention, (4) Social Criticism, (5) Animals and Insects, (6) High Places, (7) Danger Signals, and (8) Sickness. Four of these (1, 2, 4 and 5) were also interpreted in the data of the females ($N = 215$). Additionally, the following factors were described: Isolation (3), Weather and Sudden Noises (6), Dangerous Places (7), and Medical Intervention (8).

Although the names given to the factors across studies (see Hersen, 1973; Tasto, 1977; Wade, 1978; Granell de Aldaz, 1982) are different and could give the impression of great heterogeneity, critical inspection of the factor loadings of the items which constitute each factor in the different studies suggests that a significant proportion of between-study variance can be attributed to the different names investigators have chosen to describe their factors rather than to large differences in factor composition (cf. Tasto, 1977, p. 159). However, the comparability of the findings from the different studies is clearly limited because of the idiosyncratic modifications of the FSS made by investigators and because of the far-from-perfect overlap of items among various FSS forms (Tasto, 1977, p. 157).

At the conclusion of his survey of findings on certain psychometric properties of the FSS, Tasto remarks that data on the validity of the FSS applied to populations other than non-hospitalized phobics suggest that it would be inappropriate to assume its consistent applicability across populations. And further that:

> "More data on pathological populations and normal adult populations (in contrast to the more easily obtained and more frequently utilized college population) would be helpful in expanding the nomological network of the FSS" (1977, p. 174).

Despite the relatively large number of factor analytic studies which have been conducted with FSSs, to date no empirical attempt has been made to examine the degree of generalizability of phobic dimensions across the distinct populations mentioned above. This is peculiar since the execution of studies in which (for whatever purposes) patient groups are compared with non-patient normals is predicated upon the availability of measures evidencing crosspopulation constancy, i.e. the operational (dimensional) definitions of phobic constructs must be defined in a consistent fashion across the populations in question. Indeed, a primary prerequisite for dimensional approaches if they are to achieve continued acceptance and practical application is a demonstration of reliability or generalizability of construct definitions. The utility of a construct, then, is to some degree equated with generalizability, and the range of its constancy or generalizability must be ascertained empirically to identify the valid limits of its application.

One of the aims of this study is to examine the tenability of Tasto's (1977) remark with respect to the most frequently occurring clusters of fears (among students) and the problem as to whether these are generalizable across distinct populations. This will be done empirically by using techniques of *factorial invariance.*

Factorial invariance has come to be recognised as an essential qualtity for valid and meaningful (clinical) assessment. In spite of the appreciation by theorists of its importance, there has been relatively little recognition at the practical clinical and research level of the necessity of dimensional constancy; with the exception of a few investigators, most researchers involved in clinical psychometrics have failed to address themselves to the issue (cf. Derogatis and Cleary, 1977).

In simplest terms, factorial invariance refers to constancy in composition of a dimension as one moves across significant subject parameters such as sex, age, social class, population, or culture. To illustrate, if one operationally defines a measure labelled 'Agoraphobia' by a set of items that intercorrelate in a sample of females, the same set of items must be shown to define the dimension among males if the measure is to have invariance across 'sex'. If one does not perform invariance studies with (new) instruments, then the definitional limits of the scales are indeterminate and it becomes extremely difficult to establish valid limits for their use.

Information on factorial invariance is important for at least three reasons (Derogatis, 1977, p. 35): (1) it provides information on the practical limits of generalizability, and, thereby, reliability of measures; (2) it may be utilized to underscore important differences among various (clinical) groups in their manifestations of psychopathology; (3) if groups under direct comparison show significant variation in their structural definitions of the (clinical) scales used to measure them, then pattern or level difference comparisons between them are rendered problematic at best. A lack of information with respect to this matter may lead to serious misinterpretations of results.

The question of generalizability of measures, if left unanswered, does not only lead to difficulties and complications when one is considering comparisons across distinct populations within the same country/culture; obviously, it also holds when one is dealing with comparisons between countries: are the same dimensions adequate for describing certain areas of

behaviour in the two nations/cultures being compared, i.e. the question of *comparative dimensionality* in the two different populations (Eysenck and Eysenck, 1983).

It its by no means certain or even likely that the dimensions found in one country/culture would be found to be sufficiently identical to those found in another country/culture to warrant a meaningful comparison between the two, or to adopt for one country (in this case, the Netherlands) without alterations a dimensional structure which has emanated from analyses conducted on data from another country (Britain). In fact, the meaning of certain actions, experiences or attitudes canvassed in measures may change completely as one moves from one country, and one culture, to another, so that the meaning of scale scores may not always be apparent and any straightforward interpretation of results may be of doubtful psychological value (e.g. Eysenck *et al.,* 1977, 1980; Iwawaki *et al.,* 1980; Eysenck and Eysenck, 1983). Eysenck and Eysenck (1983) admonish:

> "It cannot be assumed that the same items will be found to share common variance when different cultures are being studied, and the assumption that they do so must be empirically verified" (p. 43).

RELIABILITY

Internal Consistency

Notwithstanding the fact that a multitude of studies have shown several dimensions to underlie FSSs, to date only three attempts (Gulas *et al.*, 1975; Arrindell, 1980; Arrindell and Zwaan, 1982) have been made to investigate the internal consistency reliability of subscales based on empirically established robust dimensions rather than on total fear scales which tend to reflect an intuitive grouping of items.

Studies in which internal consistency reliabilities of general scales are indicated have shown these to attain values well above 0.90 (Geer, 1965; Spiegler and Liebert, 1970; Hersen, 1971; Spinks, 1980).

Geer (1965) reported a *KR–20* coefficient of 0.94 for a sample of 783 students on the 51-item FSS-II (equally strong values for males and females: 0.93), concluding that it is a scale with high internal consistency reliability. Spiegler and Liebert (1970), employing data of 349 *S*s drawn from high school, college and various civic and social organizations, reported virtually identical *alpha's* (0.95) for a 78-item FSS-III and the 67-item Supplementary Fear Questionnaire, concluding that "both subscales possess high internal consistency reliability" (p. 692). A similar conclusion was drawn by Hersen (1971, p. 377), who reports that *KR-20* coefficients for the FSS-III ($i = 76$) are equally high for males, females and the total group of psychiatric inpatients

($N = 351$): 0.98. Finally, employing a 73-item version of the FSS, Spinks (1980) reports a coefficient *alpha* of 0.95 for his sample of 122 normal Natal Indians, suggesting that "the FSS has ... good ... internal consistency and item reliability, in terms of the amount of overlapping variance ..." (p. 365).

Perhaps, inspired by these very high coefficients *solely*, researchers have been inclined to consider these general measures as valid and usable (cf. Geer, 1965, p. 47). However, the tenability of these conclusions is questionable as long as arguments can still be advanced to refute the claim of the utility of general fear measures.

It has become common usage to regard Cronbach's coefficient *alpha* as a direct measure of item-intercorrelations and therefore as a measure of a common underlying factor (Green *et al.*, 1977). The value which the *alpha* coefficient attains, however, is a function of both the extent to which items measure the same factor and the number of items which are contained by the scale in question. Given even minimum common variance among items, *alpha* asymptotically approaches unity as the number of items increases (e.g. Gulliksen, 1950). When *alpha* is used on tests with a large number of items, as is usually the case with the FSS, the size of the item-pool must be adjusted in such a way that the number of items does not significantly inflate the coefficient value (Tyler and Fiske, 1968); this can, for instance, be done by examining the coefficient *alpha* after reducing the size of the item-pool to 1 item. In other words, by inspecting homogeneity among items, i.e. the strength of the mean correlation between items making up the scale under consideration.

Several other findings should also be inspected prior to concluding that it is warranted to employ a general measure; for instance, the degrees of association that are denoted by the correlations among subscales (corrected for attenuation) underlying such a general measure should be considered.

There is evidence to indicate that fear *sub*scales are reliable measures. Gulas *et al.* (1975), for instance, interpreted six factors after factor analysing data of 305 college students (137 males and 168 females) on an 88-item FSS. They reported satisfactory levels of reliability for males (M) and females (F) for each factor: (I) Hostile-Dependent ($i = 13$) — M, 0.91; F, 0.90; (II) Body Assault ($i = 10$) — both sexes, 0.97; (III) Developmental Fear ($i = 7$) — M, 0.83; F, 0.66; (IV) Performance and Evaluation ($i = 5$) — M, 0.80; F, 0.69; (V) Death Evasion ($i = 6$) — M, 0.80; F, 0.73; (VI) Nuisance Animals ($i = 7$) — both sexes, 0.81.

The Arrindell (1980) study in which the FSS-III ($i = 76$) was factor analysed using data of a sample of predominantly agoraphobic *S*s ($N = 703$) described five factors for which highly satisfactory *alpha* values were reported. These are shown in Table 1 for the pooled group of *S*s (151 males plus 552 females). The very high coefficient of 0.95 was produced for the General Fear scale. To enable comparisons between the coefficients obtained for the subscales and that obtained for the General scale, the length of the latter was reduced to that of each of the factored scales, i.e. to 13 items for a comparison with the Social Phobia and Agoraphobia subscales and to 12, 8 and 6 items respectively for comparisons with the Fear of Bodily Injury, Illness and Death, the Fear of Sexual and Aggressive Scenes and the Harmless Animals scales. The values obtained in this re-analysis are also given in Table 1. It is readily apparent from this table that the subscales are more homogeneous than the General scale.

Table 1. Comparisons between the Reliabilities of FSS Subscales and the Total Scale after Adjustment for Scale Length

Subscales	Scale length	Subscale *alpha*	Adjusted* total scale *alpha*
Social Phobia	13	0.91	0.76
Agoraphobia	13	0.88	0.76
Fear of Bodily Injury, Illness and Death	12	0.89	0.75
Fear of Sexual and Aggressive Scenes	8	0.79	0.67
Fear of Harmless Animals	6	0.85	0.62

*Original length of the total scale: 76 items.

There are additional data to indicate the lack of utility of a general fear measure. For instance, it was pointed out (Arrindell, 1980) that "several fear variables do not relate in a uniform degree to the global scores of fearfulness as measured by the total scale. Though the total variance ... accounted for by the first general unrotated factor is three-and-a-half times greater than the variance of the next largest factor, 18 (23.7 percent) out of 76 items, using 0.40 as a cut-off point (for a high loading; Brand-Koolen, 1972), fail to have satisfactory loadings. No loading exceeds 0.70, a value which denotes a very high loading ..." (p. 234). Additionally, only 26.3 percent of the item-remainder correlations for the General scale appeared to be greater than or equal to 0.50 compared to clearly higher percentages obtained for each *sub*scale. Finally, generally moderate inter-subscale correlations were found. Altogether, this indicates that "the FSS-III is a collection of items with largely unique variance and that the state of 'general fearfulness' ... measured by this schedule is at best a weak one" (p. 235).

In the Arrindell and Zwaan (1982) study, factor analysis of data of a sample of predominantly agoraphobics ($N = 510$) on the AZUV ($i = 95$) produced four highly reliable factors. The corresponding scale *alpha's* are shown in Table 2 for the total sample (82 males and 428 females). The very satisfactory *alpha* value of 0.96 was obtained for the total scale. A comparable procedure to that described above for the FSS-III was carried out for the AZUV, which required that the length of the Total scale was adjusted to 22 (Agoraphobia), 16 (Fears related to Contamination, Violence and Sex), 11 (Fears concerning

Bodily Injury, Illness and Death) and 8 (Social Phobia) items respectively. The corresponding values obtained in this re-analysis are also shown in Table 2.

Again, the subscales appear to be more homogeneous than the General scale. Additionally, a re-analysis of the presented item- and factor-analytic findings revealed that: (a) while the first *unrotated* factor explained (1) almost two-and-a-half times the variance of the second largest factor, and (2) somewhat more variance than the sum of variances explained by the remaining two factors, 29 (30.5%) out of 95 items failed to have satisfactory loadings on the first factor, with no item-loadings exceeding 0.70; (b) only 15 (15.8%) out of 95 items were found to correlate 0.50 or higher with the Total scale compared to clearly higher percentages of *sub*scale items meeting this criterion, i.e. 86 (Agoraphobia), 37.5 (Fears related to Contamination, Violence and Sex), 91 (Fears concerning Bodily Injury, Illness and Death) and 87.5 percent (Social Phobia). In addition to these findings, inter-subscale correlations were found to be generally low. Thus, these empirical data restrict the validity and general use of Total fear dimensions in the study of self-reported fear. Further, they underscore Green *et al.*'s (1977) contention that high reliabilities may be obtained even when measures are not uni-dimensional (that is, while homogeneity implies high internal consistency, high internal consistency need not imply homogeneity), and that one should therefore investigate the dimensionality of a measure prior to conducting reliability analyses.

Table 2. Comparisons between the Reliabilities of AZUV Subscales and the Total Scale after Adjustment for Scale Length

Subscales	Scale length	Subscale *alpha*	Adjusted* total scale *alpha*
Agoraphobia	22	0.93	0.85
Fears related to Contamination, Violence and Sex	16	0.85	0.80
Fears concerning Bodily Injury, Illness and Death	11	0.89	0.75
Social Phobia	8	0.85	0.67

*Original length of the total scale: 95 items.

Test–Retest Reliability

A small number of studies have reported test–retest reliabilities of FSSs. Unfortunately, most of these (Cooke, 1966; Braun and Reynolds, 1969; Suinn, 1969; Goldberg *et al.*, 1975a; Wade, 1978) have examined this form of reliability for a General scale only. A survey of the obtained results is given in Table 3. With intervals of 3–10 weeks, satisfactory values ranging from 0.72 to 0.90 have been found.

Table 3. Survey of Studies determining Test–Retest Reliabilities for General Fear Scales

Study	Sample	Schedule (i)	N*	Time-interval	r_{tt}
Cooke (1966)	American university students	FSS Modified**	26	3 weeks	0.90
Braun and Reynolds (1969)	American university students	TFSI (100)	70	10 weeks	0.87***
Suinn (1969)	American university students	FSS (98)	82	5 weeks	0.72
Goldberg, Yinon and Cohen (1975a)	Israeli university students	IFSI (97)	69	4 weeks	0.80
Wade (1978)	American university students	FSS (40)	25	2 months	0.89

Note. TFSI=Temple Fear Survey Inventory; ISFI=Israeli Fear Survey Inventory.
*Males plus females. **Not reported.
***In fact, reliability coefficients were reported for males (N=45, r_{tt}=0.88) and females (N=25, r_{tt}=0.85) separately. The presented coefficient has been calculated for the pooled S sample by applying Fisher's z transformation.

Suinn's (1969) study is worth noting since, besides giving the 0.72 test–retest estimate for his untreated student sample (N = 82) for a 5-week time-span for a 98-item General scale, it also reports the occurrence of an overall decrease in the scale on the second testing. Suinn does not report a test of significance. No figures concerning the internal consistency of the scale or standard error of measurement are given.

In only three studies (Tasto and Suinn, 1972; Marks and Mathews, 1979; Michelson and Mavissakalian, 1983) are test–retest reliabilities reported for relatively substantive dimensions (subscales).

Tasto and Suinn (1972) determined test–retest coefficients (time-interval: 10 weeks) for the five factors that emerged from the Rubin *et al.* (1969) factor analysis of a 122-item FSS (as well as for the Total scale). These coefficients were reported for non-treated students (27 males and 46 females), for the sexes separately. Both initial and retest means were reported for all scales. Table 4 sets out the findings. As can be seen, the coefficients fall in the moderate to high range, which is satisfactory, with one clear exception: the conspicuously low value of 0.27 for the men's ratings of the 'Fear of Isolation and Loneliness' factor. Additionally, the study indicated significant decreases on some of the scales (see Table 4).

Marks and Mathews (1979) report for the Fear Questionnaire, preliminary tests of reliability with a clinical sample (20 phobic patients). Taking the three components' subscores of 'Total Phobia' as well as the General scale itself, the following results were obtained with a short retest interval of 7 days:

Table 4. Pre-Test and Post-Test Means, t Tests for Differences between the Means and Test–Retest Reliabilities (Interval: 10 weeks) from the Tasto and Suinn (1972) Study for Males and Females Separately (N resp., 27 and 46)

Factor	Pre $\bar{X}$	Post $\bar{X}$	t	r_{tt}
Men				
I	26.70	23.92	2.25**	0.45
II	21.87	20.69	2.58**	0.75
III	29.86	27.74	1.34	0.42
IV	14.47	15.77	1.05	0.27
V	28.33	27.73	0.50	0.70
Total	372.30	364.37	1.74*	0.71
Women				
I	23.62	22.92	1.66	0.75
II	21.53	20.13	2.58**	0.63
III	28.78	27.55	2.91***	0.85
IV	16.03	15.30	2.13**	0.70
V	27.33	27.11	0.43	0.76
Total	367.38	363.56	2.40**	0.67

Note. Factors: I. Fears related to small animals
II. Fears of the precipitators of hostility and manifestations of hostility
III. Primitive moralistically related fears and sexual fears
IV. Fears of isolation and loneliness
V. Fears of anatomical destruction and physical pain.

* $p \leqslant 0.10$, ** $p \leqslant 0.05$, *** $p \leqslant 0.01$

Agoraphobia, 0.89; Blood-Injury Phobia, 0.96; Social Phobia, 0.82; Total Phobia, 0.82. Initial means and standard deviations were also reported, but those of the retest were not given.

Michelson and Mavissakalian (1983) examined the temporal reliability of measures encompassing anxiety, phobia, depression and general symptomatology dimensions of functioning. Among the instruments were outcome measures which are widely employed in agoraphobia research. Two of these were the FSS (Wolpe and Lang, 1964) and the FQ. Forty *S*s meeting DSM-III criteria for agoraphobia (one of several criteria) completed, on two occasions, the test battery; the test–retest time periods were separated by 4–, 10– and 16–week intervals. There were no *S*s in treatment during the testing periods. The FSS was used only as a measure of General fearfulness; for the FQ, both the Total phobic and phobic subscores were utilized. The vast majority of assessment instruments were found to possess acceptable temporal stability across the three time intervals in relation to total scores. The FSS revealed high temporal stability irrespective of testing period: 0.82 (1–4 weeks), 0.89 (2–10 weeks) and 0.90 (3–16 weeks) (mean correlation: 0.87). The corresponding figures for the FQ Total score ('A') were 0.69, 0.71 and 0.90 (mean correlation: 0.79) respectively. The Agoraphobia subscale evidenced good temporal stability over both short and longer test–retest intervals as well: 0.86,

0.86 and 0.85 respectively. The corresponding findings for the FQ Blood-Injury subscore were 0.34, 0.73 and 0.86, while those for the FQ Social Phobia subscore were 0.61, 0.81 and 0.84. Thus, compared with total fear scores, subscore reliabilities were lower and less consistent in terms of stability, "possibly related to their having fewer items" (Michelson and Mavissakalian, 1983, p. 696). Although internal consistency reliability figures could have helped clarify this matter these were not reported. Further, as can be seen above, all FQ (sub)scales, except Agoraphobia, evidenced increased test–retest reliability over time. The authors conclude:

> "Not surprisingly, the test–retest reliability of the various scales' subscores were generally inferior to those of the total score. Given their increased vulnerability to spurious fluctuations, both clinicians and researchers need to exercise caution in their use. The only possible exception might be the use of the agoraphobia subscale ..." (1983, p. 697).

On the basis of their results the authors concluded that it appears that the psychometric properties of certain instruments (the FQ included) may change as a function of the duration of the testing intervals.

In the study, initial and retest means were displayed for each time-point (no corresponding *SDs* were given). FSS and FQ (sub)scores increased in a relatively small interval (1–4 weeks) but decreased in time-intervals of 2–10 weeks and 3–16 weeks. (No tests of significance were reported in which initial and retest means were compared.)

An obvious conclusion which can be drawn from the findings by Suinn (1969), Tasto and Suinn (1972) and Michelson and Mavissakalian (1983) is that decreases may occur in self-reported fears which have not been treated and that any decreases following treatment may be, in part, a function of the passage of time. On these grounds, Wolff and Merrens (1974) believed the findings to underscore the lack of usefulness of the FSS, in particular of scaled total scores. Taylor and Agras (1981) considered the Suinn (1969) findings "most alarming" (p. 189), concluding that the reliability of fear surveys is particularly disturbing considering how commonly they are used in studies as outcome measures. However, in our opinion, these conclusions are somewhat premature due to the incompleteness of the data on which they are based. Firstly, the omission in the Suinn (1969) study of internal consistency reliability and/or standard error of measurement (SE_m) of the scale employed is a regrettable one since observed fluctuations in scores from pretest to retest may well be a function of unreliable measurement. A second point concerns the question of how dramatic the observed drops are, which in a sense is related to the SE_m. To analyse this, effect sizes were derived here to estimate the extent of change from pretest to retest for the Tasto and Suinn (1972) study. The measure of Effect Size (*ES*) used was *d* (Cohen, 1977, p. 40), which can be defined as

$$d = \frac{\bar{X}_z}{s_z}$$

where $\bar{X}_z$ is the mean difference in "outcome" at two points in time and s_z is the standard deviation of this difference (see also Nicholson and Berman, 1983). Thus, the *d* measure expresses the difference between means relative to within-group variation. Since in the Tasto and Suinn (1972) study Effect Size estimates could not be obtained directly (*SDs* were not reported), these were retrieved through a different technique (see e.g. Smith *et al.*, 1980, Appendix 7). When reported differences between pretest and retest are given in terms of a *t* ratio, the *ES* can be calculated by means of the inferential statistic

$$d = t\sqrt{\frac{1}{n}}$$

where *n* is the sample size at each time-point.

Effect Size values of 0.43 and 0.49 emerged for the factors of the males relating to 'Small Animals' and 'Hostility' respectively, and 0.39, 0.44 and 0.32 for the factors of the females referring to 'Hostility', 'Sex' and 'Isolation/Loneliness' respectively (see Table 4 for a fuller description of the factors). For purposes of interpretation, Cohen (1977, p. 40) considers a *d* of 0.20 *small*, a value of 0.50 *medium* and a coefficient of 0.80 *large*. Therefore, the obtained values denote small to medium differences and hence a moderate degree of fluctuation on the scales through time. However, the practical implications of the fluctuations found are, as Suinn (1969, p. 206) has noted, that "such decreases must be accounted for in studies relying upon pre- and post-therapy measurement", and further that, for the therapist treating individual clients, "the results provide a frame of reference ... against which he may assess the progress of his clients" (p. 206). We contend that, *d* indices in excess of 0.50 may generally be necessary to consider changes on *sub*dimensions of fear from pre- to post-therapy clinically relevant.

In fact, as far as stability is concerned, measures which on the grounds of their accompanying instructions can be assumed to measure *states* (cf. Allen and Potkay, 1981), fall between stable personality characteristics (such as "intelligence") on the one hand and vacillating states (such as "mood") on the other (Derogatis, 1977, p. 34). Therefore, it is reasonable to expect some vacillation on fear measures. Three decades ago, Windle (1954) reviewed 41 test–retest studies in which psychological questionnaire measures of adjustment had been administered to a total of 51 different control groups or other untreated samples, mainly students. Mean adjustment "improved" in

the second test in 28 of the samples, remained unchanged in 17 and deteriorated in 6, "improvement" being more often found when the interval between test and retest was two months or less (see also Hochstim and Renne, 1971; Henderson *et al.*, 1981, pp. 119–123).

The effect of drops from initial testing to subsequent rounds is not confined to measures of possibly fleeting states of distress; it has been noted in measures of personality *traits* too (see Henderson *et al.*, 1981, pp. 122–123). Henderson *et al.* have argued that this (general) decrease in reported symptoms may be due to an artifact caused by a change in response style to the questionnaires: "respondents may tend to present themselves more favourably on the second and subsequent rounds of testing", which implies that "there is no true change in the number of intensity of symptoms, but instead there is an increase in the threshold for reporting their presence" (1981, p. 121). It would therefore seem advisable to investigate the degree to which scores on subdimensions of any fear schedule are affected by social desirability at different points in time.

STUDY

The present study seeks to assess the generality of phobic dimensions across significant subject parameters. Specifically, attempts will be made to: (1) examine the cross-sample constancy of dimensions previously identified with the FSS (Arrindell, 1980) with similar populations (phobics), (2) ascertain whether and, if so, to what extent phobic dimensions (as identified with the FSS and/or FQ) generalise to populations other than phobic ones (i.e. heterogeneous psychiatric outpatients, non-student normals and non-patient students), (3) determine the degree of constancy of phobic dimensions across gender, and (4) delineate the degree of universality of phobic dimensions across nations by (a) comparing factor analytic findings obtained for phobics in Britain (Marks and Mathews, 1979) with those for their Dutch counterparts on the Fear Questionnaire, and (b) comparing factor analytic data of British university students on the FSS (Kartsounis *et al.*, 1983) with those of their Dutch counterparts. To tackle these problems objective techniques of factorial invariance will be employed. A further aim is to investigate the internal consistency reliability of the two fear measures, and the test–retest reliability of the FQ over a relatively long period of time (13 months).

To summarise, this study has two principal objectives: (1) to determine whether and to what extent dimensions of phobic factors are generalizable across samples, gender and nations; (2) to examine two forms of reliability (internal consistency and test–retest) for phobic subscales.

METHOD

Subjects and Procedures

To address the issues raised here, data from 10 independent samples, comprising a total of 3143 *Ss*, were analysed. Among these samples were four (agora)phobic subsamples. The groups were composed of:

(1) 439 phobics; 92 males and 337 females with a mean age of 45.4 yr (SD=12.1 yr; range, 22–87 yr). These *Ss* also participated in the Arrindell (1980) study which was conducted in 1978. In the present study, which was implemented in 1981, they were again among the participants;

(2) 402 phobics; this subsample comprises 79 males and 323 females with a mean age of 38.7 yr (SD=11.4 yr; range, 16–78 yr);

(3) 119 phobics, 22 males and 97 females with a mean age of 39.1 yr (SD=11.2 yr; range, 23–68 yr);

(4) 15 female clinically diagnosed agoraphobics with a mean age of 27.9 yr (range, 29–63 yr);

(5) 408 (non-selected) psychiatric outpatients; 191 males and 213 females (4 *Ss* failed to reveal information with respect to their gender) with a mean age of 36.1 yr (SD=12.3 yr; range, 16–72 yr);

(6) 250 non-psychiatric partners of psychiatric outpatients; 126 males and 124 females with a mean age of 38.5 yr (SD=12.1 yr; range, 18–83 yr);

(7) 277 (non-student) normal community *Ss*; 137 males and 140 females with a mean age of 41.1 yr (SD=12.5 yr; range, 20–69 yr);

(8) 270 non-patient college students: 144 males, 123 females (N=3: missing data with respect to gender); mean age, 22.7 yr (SD=2.8 yr; range, 18–41 yr);

(9) 416 Dutch non-patient university students, the sample comprising 145 males, 194 females, 77 *Ss* having failed to state their sex; mean age: 20.4 yr (SD=2.6 yr; range, 18–35 yr);

(10) 547 British non-patient university students from various academic departments of Bedford College, University of London. This group contained 201 males, 302 females and 44 *Ss* who withheld information about their sex; no data were available as regards age (see Kartsounis *et al.*, 1983).

Subjects from the first four samples were all members of a society for phobic individuals; samples 1 through 3 contained mainly agoraphobic *S*s. For most samples (1 through 9) there are additional data available describing each *S* in terms of marital status, educational and occupational status, and religious affiliation. However, for the sake of brevity, these figures will not be displayed here. They are available upon request from the senior author.

Subjects were volunteers who, mostly as part of larger mental health projects, participated in large-scale (mail) surveys which were conducted with the purpose of developing and evaluating psychometrically a number of psychopathology measures. In a small number of cases (student samples) the present scales were group-administered. In most groups, *S*s completed their questionnaire sets anonymously. In cases where *S*s could choose between either stating their names or completing their questionnaires anonymously (sample 3), multivariate analysis of variance (MANOVA) comparing the two groups on the relatively large number of other scales, besides those used in the present study (see e.g. Arrindell, van der Ende and Emmelkamp, 1985) indicated that no systematic between-group differences could be detected (MANOVA, $F<1$, $P>0.10$).

In all instances, confidentiality was stressed; also, all measures were completed prior to the *S*s' participation in other studies/experiments. Further, if *S*s were awaiting treatment for their complaints (sample 5), participation in the present study occurred, in all cases, prior to treatment.

Finally, it should be noted that due to the requirement of "complete" anonymity in settings in which scales were group-administered, a relatively small number of *S*s (students) were allowed to withhold (besides their name) information concerning their sex and age, and other demographic data.

Measures

To assess self-rated fears, the FSS-III (Wolpe and Lang, 1964) and the FQ (Marks and Mathews, 1979) were used. Standard instructions were given in which *S*s were requested to indicate their degree of felt *anxiety* with respect to the 76 stimulus situations and/or objects on the FSS-III on 5-point Likert-type scales which ranged from 1 "not at all disturbed" to 5 "very much disturbed".

The FQ contains 20 items; 15 of these items (situations) are rated on 9-point continua indicating the *S*s' degree of *avoidance* (from 0 "would not avoid it" to 8 "would always avoid it"). The remaining five anxiety–depression items are rated on 0–8 scales with respect to the degree to which the *S* is troubled by each problem (0= "hardly at all" to 8 "very severely troublesome").

The subscales which make up these measures have been described above.

Subjects from all groups filled out either the FSS-III or the FQ, with the exception of those from samples 1 and 2, who completed both.

Subjects from the British sample filled out an 88-item variation of the FSS (Wolpe, 1973); 75 of its items are captured by the here utilized FSS-III.

Statistical Analyses

In dealing with the constancy of dimensions of the FSS-III, self-ratings of *S*s from the following subsamples were used: phobics (subsamples 1 and 2 separately), psychiatric outpatients (subsample 5), non-patient partners of psychiatric outpatients (subsample 6), Dutch university students (subsample 9), and British university students (subsample 10).

As regards the FQ, the responses of *S*s from the following subgroups were utilized: phobics (subsamples 1, 2 and 3 pooled, producing an N of 960), non-student normals (subsample 7), and non-patient college students (subsample 8).

Distinct courses were followed for the FSS-III and the FQ in carrying out invariance procedures.

In studying the constancy of the FSS's factor structure, a method was utilized which compares factors from the samples mentioned with a known theoretical structure on the basis of "perfectly-congruent weights" (ten Berge, 1984). The basic aim of this method is to recover components in a second (new) study with the same interpretations as those in a first (original) study, and evaluating the amounts of variance which they explain. The weights obtained in the first study are used to cross-validate hypothetical factors in a second study. Examination of the strength of the uniquely-defined perfectly-congruent counterpart of the hypothetical factor in the new study provides an index of the degree of constancy of factors. Hence, in this approach, the amount of variance explained by any hypothetical factor in a new study is of critical interest. The strength of each (new) factor is expressed in terms of percentage explained variance. This percentage may vary from 0 to 100%. The weights impose correlated factors in the new studies (i.e. samples).

The VARIMAX rotated 5-factor principal-component solution of phobics' self-ratings (N=703) reported in Arrindell (1980) was used as the theoretical structure. The fear factors obtained in this study were, as mentioned earlier, interpreted as (I) Social, (II) Agoraphobic, (III) Bodily Injury, Death and Illness, (IV) Sex and Aggression, and (V) Harmless Animals.

Actual calculations involved were performed with the PCON (Perfect CONgruence) computer programme (ten Berge, 1984).

Consider now the constancy of the dimensions of the FQ. Besides the use of weights from the original study, it is also possible to use hypothetical weights, for instance on the basis of a known theoretical structure.

To do this, the theoretical weight matrix of the four primary dimensions of the FQ, namely Agoraphobia, Blood-Injury Phobia, Social Phobia, and Anxiety–Depression (Marks and Mathews, 1979), was cast as the hypothetical matrix **H** in binary form. This simply means that all items rationally defined as comprising a particular dimension (e.g. Anxiety–Depression) were assigned a 1 for that dimension and a 0 for all other dimensions, which resulted in a matrix in which each row had one non-zero entry only; the column of weights conformed to item sets theoretically assigned to each FQ dimension. Next, the binary weight matrix enabled computation of a structure matrix from the correlation matrices of FQ items for each subsample. Although this procedure is identical to the multiple-group method of confirmatory factor analysis

(Gorsuch, 1974, pp. 73–84; Nunnally, 1978, pp. 394–400), distinct routes are followed in interpreting the results which they produce: in the latter, attention is primarily focussed on the matrix of item-loadings, whereas in the analysis employing perfectly-congruent weights the strength of each factor is of prime interest. This is not to say, however, that utilization of one method rules out the use of the other. In fact, in the present article, both were used conjointly.

As in the analyses involving the FSS-III, the weights impose correlated factors on the data matrices. In making interpretations, the correlations among the factors (i.e. their degree of relative separate existence) should also be considered. These factors are considered *standardised weighted summations of item-scores.*

A final remark concerns the sizes of the (sub)samples utilized in conducting the present analyses.

Following Gorsuch (1974, p. 296), it has been argued that to have some confidence in the results obtained from factor analyses of fear variables an absolute minimum ratio of five individuals to every variable is necessary (Wade, 1978, pp. 923–925). Researchers conducting these types of analyses in which this necessity is stressed have indeed followed this advice (Neiger *et al.*, 1981; Granell de Aldaz, 1982; Kartsounis *et al.*, 1983). However, this requirement which was based on research experience of factorists of repute has recently been submitted to empirical testing and has been found to be unnecessarily stringent; both the need for a minimum ratio of observations to variables and for an absolute minimum of observations for stable factor solutions has been challenged (Barrett and Kline, 1981; Arrindell and van der Ende, 1985). These authors have shown empirically that small sample factoring is justifiable for the purpose of replicating a given number of true factors. Arrindell and van der Ende (1985), for instance, have shown that in order to be able to have confidence in results obtained from sample factoring with the aim of replicating a supposed factor structure, the number of *S*s should equal at least 20 times the number of factors (i.e. in this study, a minimum N of approximately 100). On this ground the present samples were deemed large enough for addressing the questions raised.

RESULTS AND DISCUSSION

Inspection of a number of descriptive item-statistics indicated that these were such that no "difficulty factors" (Gorsuch, 1974, p. 260) could be expected to emerge from the analyses to be performed below.

The FSS-III

Table 5 sets out the results of comparison (percent explained variance by each factor and the factors jointly/totally) by subsample for the sexes separately and the sexes combined.

The factors in the Arrindell (1980) study explained 9.7% (I Social), 9.1% (II Agoraphobic), 8.2% (III Bodily Injury, Death and Illness), 7.3% (IV Sex and Aggression), and 7.0% (V Harmless Animals) of the total variance, which summated produces the figure of 41.3%.

From Table 5, it will be seen that there were hardly any differences between the original figures from the 1980 study and those found for the new studies with phobics (subsamples 1 and 2). Factor correlations for these subsamples were mostly very close to zero. Hence, since the factors in the 1980 study were uncorrelated as an artificial product of the principal-component method of factor analysis, the factor correlations were replicated as well.

Examination of results for the sexes separately gave identical conclusions; however, as can be seen, the 'new' factors appeared to be somewhat stronger in *males* in both phobic subsamples, there being a difference in strength of 14% between the sexes in the present study as opposed to half this difference (6%) in that reported in 1980.

The percentage-explained variance for each of the factors in the subsamples of psychiatric outpatients, non-patient partners of psychiatric outpatients, and Dutch normal students were either comparable to, or even *higher* than, those given for the original phobic sample. Again, the factor correlations were mostly of indifferent magnitude, the highest values being at most low. However, a trend was noted mirroring slightly higher correlations between factors with decreasing phobic pathology: for instance, in the non-patient normal and student samples some correlations were stronger than the corresponding figures in either of the phobic subsamples. Nonetheless, the associations were all low enough not to obstruct the relative separate existence of the entities denoted by the factors. Further, the sex difference noted above was clearly less apparent in these subsamples. The strength of the factors for each of the sexes in these subsamples was also comparable to those found in the original subsample. Once more, factor correlations in each of the subgroups were mostly very low. Thus, contrary to expectations (e.g. Hallam and Hafner, 1978), an Agoraphobic factor emerged in both psychiatric outpatient, non-patient normal and student samples and even appeared to be (slightly) stronger than its counterpart in the phobic sample.

Omission of one item (no. 22) from the original phobic sample in order to make comparisons possible between identical item-sets across the British and the Dutch Target ratings gave the following factor strengths: 9.9% (I), 9.2% (II), 8.3% (III), 7.0% (IV), and 7.0% (V). Summated this gave 41.4%. The

Table 5. Invariance of FSS-III Dimensions (Percentage Explained Variance and Factor correlations* by Subsample for the Sexes Separately and the Sexes Combined)

		Males								Females						
		Percentages explained variance		Correlations among factors						Percentages explained variance		Correlations among factors				
				I	II	III	IV	V				I	II	III	IV	V
Phobics subsample 1	I	12.12		100					I	10.09		100				
	II	10.66		06	100				II	7.58		−02	100			
	III	10.08		13	13	100			III	8.58		−09	01	100		
	IV	10.25		09	14	−17	100		IV	8.09		07	03	−05	100	
	V	9.75		−05	−02	00	28	100	V	8.20		03	−01	11	11	100
	Total	46.80	*N*	92					Total	40.59	*N*	337				
Phobics subsample 2	I	17.33		100					I	9.80		100				
	II	16.38		29	100				II	7.37		−01	100			
	III	14.18		12	20	100			III	7.62		−05	−01	100		
	IV	15.62		20	16	27	100		IV	6.70		02	−01	−01	100	
	V	13.24		10	16	21	38	100	V	6.62		10	04	−06	03	100
	Total	51.97	*N*	79					Total	37.76	*N*	323				
Psychiatric outpatients subsample 5	I	12.45		100					I	10.91		100				
	II	15.05		25	100				II	13.62		23	100			
	III	9.38		−05	11	100			III	7.84		−07	03	100		
	IV	7.84		04	26	−02	100		IV	8.38		05	22	−03	100	
	V	8.23		06	06	10	03	100	V	7.46		−03	15	08	07	100
	Total	44.49	*N*	191					Total	41.51	*N*	213				
Non-patient normal partners subsample 6	I	13.36		100					I	11.78		100				
	II	11.08		18	100				II	12.72		28	100			
	III	8.57		01	−01	100			III	10.39		04	−10	100		
	IV	10.17		39	18	−14	100		IV	10.21		03	37	03	100	
	V	6.51		−01	20	15	−16	100	V	8.00		−04	−06	21	−06	100
	Total	41.13	*N*	126					Total	45.50	*N*	124				
Dutch university students subsample 9	I	10.49		100					I	9.63		100				
	II	9.59		17	100				II	9.84		17	100			
	III	8.49		07	09	100			III	6.87		01	−16	100		
	IV	6.32		06	32	−09	100		IV	6.05		04	33	−08	100	
	V	6.92		04	12	07	−06	100	V	6.04		04	32	−04	−22	100
	Total	35.78	*N*	145					Total	34.28	*N*	194				
British university students subsample 10**	I	10.60		100					I	9.91		100				
	II	9.27		26	100				II	7.71		27	100			
	III	7.93		03	−21	100			III	6.24		−06	−07	100		
	IV	8.63		24	36	05	100		IV	5.72		16	30	−09	100	
	V	4.25		00	24	−05	−16	100	V	5.03		−01	10	07	−19	100
	Total	34.05	*N*	201					Total	30.76	*N*	302				

Table 5 (*continued*)

		Males plus Females						
		Percentages explained variance		Correlations among factors				
				I	II	III	IV	V
Phobics subsample 1	I	10.14		100				
	II	8.71		01	100			
	III	9.21		−05	06	100		
	IV	8.54		08	07	−06	100	
	V	9.22		01	04	13	16	100
	Total	41.91	*N*	439				
Phobics subsample 2	I	10.61		100				
	II	8.95		05	100			
	III	8.37		−04	04	100		
	IV	7.70		07	03	03	100	
	V	7.29		05	09	01	06	100
	Total	39.97	*N*	402				
Psychiatric *outpatients* subsample 5	I	11.53		100				
	II	15.55		24	100			
	III	8.71		−05	09	100		
	IV	8.74		06	26	−01	100	
	V	9.29		01	19	11	11	100
	Total	43.34	*N*	408				
Non-patient normal partners subsample 6	I	12.82		100				
	II	13.92		27	100			
	III	10.99		07	−01	100		
	IV	10.34		17	35	−01	100	
	V	10.39		02	11	27	−04	100
	Total	45.66	*N*	250				
Dutch university *students* subsample 9	I	10.20		100				
	II	10.17		20	100			
	III	7.49		06	−05	100		
	IV	5.41		07	30	−09	100	
	V	7.17		08	29	05	−19	100
	Total	34.25	*N*	416				
British university students subsample 10**	I	9.71		100				
	II	8.40		27	100			
	III	6.67		−02	−11	100		
	IV	6.76		18	33	−02	100	
	V	5.39		−01	17	04	−16	100
	Total	31.63	*N*	547				

*Decimal points have been omitted and coefficients multiplied by 100.
**Excluding one item (no. 22) from the FSS-III.
Note. I=Social *phobia*, II=Agoraphobia, III=Bodily Injury, Illness and Death, IV=Sex and Aggression, V=Harmless Animals.

corresponding figures for the British student subgroups were of comparable magnitude, albeit slightly lower, with the male factors being slightly stronger than those for the females. The factor correlations were of comparable magnitude to those found in the other subsamples.

The present findings thus indicated that the FSS factors originally found in Arrindell (1980) with non-institutionalised phobics (*N*=703):

(1) are to a high degree generalizable to other phobic populations;
(2) evidence clear-cut cross-sample constancy (invariance);
(3) show a fair degree of generalizability across cultures (British vs Dutch).

Inspection of correlations among *scales* (Table 6) showed lowest associations for the phobic subsamples and highest for the non-patient normal samples. In a sense this observation seemed to parallel the pattern of correlations found between *factors* across the different groups. This finding appears to indicate that as the degree to which *S*s are phobic decreases, the likelihood of overlap between distinguishable types of phobic fears increases, i.e. there is a *slight* tendency for different fear types to generalize to some general phobic entity in normal groups. Judging from the strengths of both factor and scale correlations, however, this observation does not justify the use of a general phobic scale in either of the samples. The inter-subscale correlations ranged in magnitude from low to moderate, further stressing the possible use of the scales as separate measures within the schedule.

The FQ

Results of analyses of invariance with the FQ are summarised in Table 7. The Table shows each subsample for the sexes separately and the sexes combined: (a) item-loadings for each of the structure matrices; (b) the percentage explained variance by each factor as well as the totally explained variance; (c) the correlations among the factors.

It will be seen that *all* items hypothesized to load on their respective factors were found to do so in a meaningful fashion; most of the items evidenced very high (⩾0.70) loadings on their theoretically related factor and only moderate (0.20–0.39) and in most cases low (<0.20) loadings on their unrelated counterparts.

Of particular interest was the finding, which also emerged for the FSS-III, that the Agoraphobic factor was reproduced in the non-patient normal subgroups; in these subgroups, this factor appeared to be of comparable magnitude to its counterpart in the phobic samples. Also, all factors, except that of Agoraphobia, were found to be comparably strong across the present subgroups, irrespective of gender. But, despite consistently less stronger loadings of Agoraphobic items on their hypothesized factor in the subgroups of college students, the factor was consistently reproduced as a clear-cut one (that is, none of its items had meaningful loadings on theoretically unrelated

Table 6. Correlations among Subscales of the FSS-III for each Subsample for the Sexes Separately and the Sexes Combined.*

		FSS-subscales				
		I	II	III	IV	V
I. Social phobia	♂		58	58	64	53
	♀		24	36	48	39
	♂, ♀		32	40	52	38
II. Agoraphobia	♂	38		62	55	48
	♀	21		38	21	18
	♂, ♀	27		45	31	27
III. Bodily Injury, Illness and Death	♂	50	55		71	65
	♀	35	40		37	34
	♂, ♀	39	46		45	41
IV. Sex and Aggression	♂	61	36	39		76
	♀	48	26	49		46
	♂, ♀	52	32	50		51
V. Harmless Animals	♂	41	25	51	55	
	♀	32	17	48	54	
	♂, ♀	35	24	52	58	

Correlations obtained for *phobics (subsample 1)* are shown in the lower-left triangle; those calculated for the *phobic subsample 2* are given in the upper-right triangle. The *N*s for subsample 1 for males (♂), females (♀), and the total sample (♂, ♀) are 92, 337 and 439 respectively. Those for subsample 2 are 79, 323 and 402 respectively.

		I	II	III	IV	V
I. Social phobia	♂		43	48	70	35
	♀		53	50	49	32
	♂, ♀		55	55	61	41
II. Agoraphobia	♂	50		43	37	38
	♀	50		47	57	21
	♂, ♀	52		53	57	36
III. Bodily Injury, Illness and Death	♂	41	58		51	51
	♀	41	54		65	55
	♂, ♀	43	58		65	61
IV. Sex and Aggression	♂	57	45	50		38
	♀	50	50	51		45
	♂, ♀	54	53	54		50
V. Harmless Animals	♂	40	35	47	42	
	♀	33	35	45	53	
	♂, ♀	38	41	49	55	

Correlations found in the *psychiatric outpatient sample (subsample 5)* are shown in the lower-left triangle; those for their *non-patient normal partners (subsample 6)* are displayed in the upper-right triangle. The *N*s for the patient sample for males, females, and the total sample are 191, 213 and 408 respectively. Those for the partners are 126, 124 and 250 respectively.

Table 6 (*continued*)

		FSS-subscales				
		I	II	III	IV	V
I. Social phobia	♂		69	44	62	35
	♀		60	36	60	31
	♂, ♀		63	40	61	34
II. Agoraphobia	♂	50		48	57	36
	♀	41		39	48	30
	♂, ♀	49		45	54	35
III. Bodily Injury, Illness and Death	♂	44	56		51	43
	♀	40	38		39	40
	♂, ♀	46	47		47	43
IV. Sex and Aggression	♂	58	53	51		39
	♀	54	53	48		40
	♂, ♀	58	56	51		41
V. Harmless Animals	♂	43	35	41	48	
	♀	32	34	35	36	
	♂, ♀	41	40	42	44	

Associations calculated for the *Dutch student sample (subsample 9)* are given in the lower-left triangle; those for their *British counterparts*† (subsample 10) are shown in the upper-right triangle. The *N*s for the Dutch student sample are 145 (males), 194 (females), and 416 (total sample). Those for the British sample are 201, 302 and 547 respectively.

*The coefficients have been multiplied by 100 (Decimal points have been omitted); Pearsonian-type correlations, $P \leqslant 0.05$, two-tailed testing.
†Excluding item no. 22.

factors, nor did items from non-analogous factors have significant loadings on the Agoraphobic factor).

It should be noted that in cases where items had loadings on both their theoretically related factor and on an unrelated one, those on the former were, in general, qualitatively stronger (e.g. 'very high' vs 'high').

Correlations between FQ factors appeared to be mostly of low magnitude, the highest values denoting moderate associations. If one were to correct these associations for attenuation (Nunnally, 1978, p. 219) the obtained figures would not attain such high values that they would challenge a separate use of certain scales (see Table 9 for an overview of scale reliabilities). Therefore, for the time being, these figures justify the separate use of the factors within this brief schedule. Thus, it can be concluded that the empirically established dimensional representation of the FQ found in Britain with phobic *S*s (Marks and Mathews, 1979) appeared:

(i) to be transferable to the present Dutch groups, thus evidencing cross-cultural constancy; and

(ii) to evidence clear-cut cross-sample constancy (phobics vs non-student normals vs students).

Table 7. Results of Invariance of FQ Dimensions: (a) Item-Loadings for each of the Calculated Structure Matrices; (b) the Percentage-Explained Variance by each Factor as well as the Totally-Explained Variance; (c) Correlations among the Factors for each of the Subsamples for each of the Sexes and the Sexes Combined

		Males											
		Phobics (N=193)				Non-patient normals (N=137)				College students (N=144)			
		Ag	Bl	So	A–d	Ag	Bl	So	A–d	Ag	Bl	So	A–d
Scale-item no.													
Bl 2			74	41			74				85		
So 3				77	40	45		70				51	
Bl 4				75		42	69				80		
Ag 5		78				73				51			
Ag 6		89	42			82				65			
So 7				79		57		79				70	
Ag 8		77	47	53		77		49		66			
So 9				83	41	45		81	42		41	73	
Bl 10			69				64				69		
So 11			47	71				69				62	
Ag 12		77				69		53		70			
Bl 13		43	71			41	65				64		
So 14				74				68				68	
Ag 15		79				65				64			
Bl 16		44	67				43				71		
A–d 18					80				84				84
A–d 19				49	65				68				73
A–d 20				41	83			44	78				85
A–d 21					75				81				84
A–d 22					72				70				63
Percentage-explained variance (PEV)		23.3	21.1	24.5	21.2	22.7	15.5	22.4	19.2	14.1	17.6	16.4	18.1
Total PEV					59.3				53.4				51.0
Factor correlations	Ag	100				100				100			
	Bl	47	100			42	100			31	100		
	So	42	47	100		58	40	100		40	39	100	
	A–d	35	31	51	100	33	20	43	100	20	24	35	100

Table 7 (*Continued*)

		Females											
		Phobics (N=757)				Non-patient normals (N=140)				College students (N=123)			
		Ag	Bl	So	A–d	Ag	Bl	So	A–d	Ag	Bl	So	A–d
Scale-item no.													
Bl 2			79				80				74		
So 3				68		40		71				55	
Bl 4			76				80				74		
Ag 5		80				82		42		62			
Ag 6		88				84		42		71			
So 7				79		51		65				73	
Ag 8		77				70		42		75			
So 9				82		45		80				73	
Bl 10			68				62				69		
So 11				69				69				64	
Ag 12		76				75		46		59			
Bl 13			70				71				63		
So 14				73		44		73				75	
Ag 15		74				71		49		56			
Bl 16			64				61				51		
A–d 18					79				78				76
A–d 19					65				71				72
A–d 20					81				78			41	83
A–d 21					77				84				71
A–d 22					72				73			42	72
Percentage-explained variance (PEV)		19.8	18.4	18.5	17.6	23.1	18.8	21.3	19.3	15.1	13.9	16.5	19.9
Total PEV					57.3				57.5				49.1
Factor correlations	Ag	100				100				100			
	Bl	41	100			43	100			22	100		
	So	27	36	100		58	35	100		36	22	100	
	A–d	22	27	35	100	30	33	33	100	42	30	42	100

Table 7 (*Continued*)

	Total											
	Phobics (*N*=960)				Non-patient normals (*N*=277)				College students (*N*=270)			
Scale-item no.	Ag	Bl	So	A–d	Ag	Bl	So	A–d	Ag	Bl	So	A–d
Bl 2		78				77				80		
So 3			69		40		69				51	
Bl 4		76				73				78		
Ag 5	80				78		40		57			
Ag 6	88				83		41		66			
So 7			79		54		72				73	
Ag 8	77		40		71		45		69			
So 9			82		48		81				74	
Bl 10		62				62				69		
So 11			69				69				64	
Ag 12	76				73		50		64			
Bl 13		70			40	68				64		
So 14			73		40		71				73	
Ag 15	75				70		45		60			
Bl 16		65				54				60		
A–d 18				79				81				80
A–d 19				65				70				72
A–d 20				82				78				85
A–d 21				77				82				78
A–d 22				72				72				68
Percentage-explained variance (PEV)	20.4	18.8	19.6	18.2	22.7	17.1	22.2	19.0	14.6	15.9	17.0	18.8
Total PEV				57.5				54.6				49.5
Factor correlations Ag	100				100				100			
Bl	42	100			44	100			29	100		
So	30	39	100		59	38	100		39	33	100	
A–d	25	28	38	100	32	27	38	100	33	28	41	100

Note. The subgroups described here correspond in the following way with those described earlier: phobics=subsamples 1, 2, and 3 pooled; non-patient normals=subsample 7; college students=subsample 8.
Ag=Agoraphobia, Bl=Blood-Injury, So=Social, and A–d=Anxiety–depression.
Both factor loadings and factor correlations have been multiplied by 100 (Decimal points have been omitted.)
Only factor loadings which exceed |±0.39| have been depicted.
Factor correlations are of the Pearsonian product-moment type.

Reliability

Internal consistency

Results of item-analyses conducted for each scale (including Total fear scales) for each subsample (for the sexes separately and the sexes combined) are shown in Table 8 for the FSS-III and Table 9 for the FQ.

Consider first the figures obtained for the FSS-III. Scale reliabilities (Cronbach's coefficient α) for its subscales and General scale were all (very) high and of comparable magnitude across subsamples for each of the sexes. Further, item–remainder correlations for the *sub*scales were all of satisfactory magnitude, mostly hovering between the substantial to high magnitude range (Surwillo, 1980, p. 59). The corresponding figures for the General scale indicated very satisfactory α values, although a number of its items were found to be only very lowly correlated with it. It should be noted that the item–remainder associations were comparable across samples as well. This finding is of particular importance for our comparison between British and Dutch university students since it points to metric equivalence of the scales which, in conjunction with the scales' invariance properties, warrants future mean level and pattern comparisons between the present groups.

Critical inspection of the Table further suggests that the 'Sex and Aggression' scale produced consistently lower αs and lower item-remainder correlations than the remaining scales in all of the subgroups. The figures yielded were all satisfactory for research purposes though. A suggestion to enhance the internal consistency reliability of this subscale will be given below.

As could be expected on the basis of the smaller number of items which are captured by its subscales the α figures obtained for the FQ were smaller than those for the FSS-III. Nevertheless, it can be argued that with scales containing five items only the obtained values are in all subgroups satisfactory for *research* purposes (Nunnally, 1978, p. 245). However, any attempts at using the FQ for applied problems, where a great deal hinges on the exact score made by the *S*, would be obstructed by the fact that the α reliabilities are not high enough for that purpose (cf. Nunnally, 1978, p. 246). In each subgroup, the magnitude of the item–remainder correlations for each subscale was satisfactory (generally falling within the moderate to substantial range). Also, these figures were found to be of comparable magnitude across 'subsamples' and 'the sexes'. Further, it is clearly more visible with the FQ than with the FSS-III that *sub*scales evidence clearly higher α reliabilities than the Total scale when the necessary adjustments for scale length have been carried out (see, for instance, the figures for the student subgroups). Related to this matter was the finding that item–remainder associations found for the Total scale were generally of smaller magnitude than those calculated for the subscales.

To undertake a comparison of homogeneity of *sub*scales vs General scales, mean inter-item correlations were calculated for each scale for the different groups (for the sexes combined). The following values were obtained for the FSS-III (the Roman figures correspond to those given in Table 5; T=Total scale).

	N	I	II	III	IV	V	T
Phobics I	439	0.45	0.31	0.41	0.35	0.49	0.22
Phobics II	402	0.47	0.31	0.39	0.29	0.43	0.21
Psychiatric outpatients	408	0.48	0.38	0.40	0.28	0.45	0.25
Non-patient partners (normals)	250	0.45	0.29	0.42	0.30	0.51	0.27
Dutch university students	416	0.40	0.21	0.33	0.18	0.37	0.19
British university students	547	0.31	0.17	0.32	0.20	0.35	0.17

The corresponding figures for the FQ were (A=Agoraphobia, B=Blood-Injury Phobia, S=Social Phobia, T=Total Phobia).

	N	A	B	S	T
Phobics	960	0.54	0.38	0.44	0.28
Non-patient normals	277	0.45	0.30	0.41	0.28
College students	270	0.25	0.36	0.31	0.20

It is clear from these figures that, on the whole, homogeneity among items from subscales is more apparent than that among items from General scales — mean inter-item correlations for subscales are in some instances of stronger magnitude than those for the General scales; in no instance do the correlations pertaining to the General scales exceed those for the subscales in a meaningful fashion.

These figures further underline the lack of utility and validity of General fear scales and the view that scale unidimensionality is not correctly represented by coefficient *alpha* (Green *et al.*, 1977; Butler, 1983).

Test–retest reliability

Figures with respect to this method of reliability estimation were calculated for the FQ subscales only. Of 32 *S*s who had participated in an earlier study (Arrindell and Emmelkamp, 1985), 15 female *S*s who had not received any treatment during the testing-period considered here, participated in the present study (subsample 4). Selection criteria were given in the cited study. However, it should be noted that the *S*s were recruited because at the time this study was being carried out (13 months after their initial participation in the cited study), they reported that they had not undergone any treatment for

their agoraphobic or other complaints during this testing-period. DSM-III diagnosis of agoraphobia still held for all *S*s.

The test–retest data for the 13-month interval were analysed using Spearman's ρ. The results, presented in Table 10, indicate that all phobic subscales possess acceptable temporal stability across such a long time-interval. However, the Anxiety–Depression subscale evidenced very low stability (ρ=0.09) and this low figure cannot be accounted for by a low level of

Table 8. FSS-Scale Reliabilities (Cronbach's Coefficient α) and Range of Item-Remainder Correlations* by Subsample for each of the Sexes and the Sexes Combined

	Phobics (subsample 1)			Phobics (subsample 2)		
	Males *N*=92	Females *N*=337	Total *N*=439	Males *N*=79	Females *N*=323	Total *N*=402
I	92 (53–78)	91 (39–76)	91 (42–76)	95 (58–83)	91 (46–74)	92 (49–75)
II	87 (26–72)	84 (32–65)	85 (35–68)	91 (38–79)	83 (28–64)	85 (31–67)
III	90 (44–74)	89 (48–76)	89 (50–76)	91 (55–80)	88 (44–73)	89 (44–75)
IV	83 (50–74)	80 (35–64)	81 (38–68)	84 (26–78)	73 (27–59)	76 (33–63)
V	85 (33–79)	83 (44–68)	85 (50–70)	83 (47–76)	81 (46–65)	82 (50–66)
Total	96 (22–74)	95 (09–63)	96 (14–65)	97 (10–74)	94 (16–59)	95 (16–59)
	Psychiatric outpatients (subsample 5)			Non-patient normal partners (subsample 6)		
	Males *N*=191	Females *N*=213	Total *N*=408	Males *N*=126	Females *N*=124	Total *N*=250
I	92 (50–80)	92 (47–77)	92 (48–78)	90 (40–76)	90 (42–76)	91 (45–73)
II	88 (29–75)	87 (33–71)	88 (42–70)	78 (14–77)	81 (38–60)	82 (39–62)
III	89 (52–72)	88 (42–76)	89 (43–75)	85 (41–69)	89 (40–75)	89 (47–74)
IV	73 (18–64)	78 (39–60)	77 (29–63)	75 (16–67)	79 (20–66)	79 (17–67)
V	82 (53–68)	78 (46–62)	83 (52–68)	71 (07–58)	85 (56–71)	86 (62–71)
Total	96 (10–66)	96 (13–64)	96 (12–65)	95 (02–69)	96 (11–70)	97 (12–71)
	Dutch university students (subsample 9)			British university students** (subsample 10)		
	Males *N*=145	Females *N*=194	Total *N*=416	Males *N*=201	Females *N*=302	Total *N*=547
I	89 (32–86)	90 (45–76)	90 (38–81)	85 (21–67)	86 (19–70)	85 (22–69)
II	75 (23–65)	76 (23–57)	76 (24–58)	75 (23–54)	67 (17–41)	71 (20–46)
III	85 (26–70)	84 (20–76)	85 (29–73)	86 (40–71)	82 (28–72)	84 (34–71)
IV	69 (04–58)	63 (22–48)	65 (17–54)	69 (24–50)	64 (18–49)	67 (30–49)
V	76 (30–60)	77 (48–56)	77 (47–56)	68 (32–48)	76 (43–55)	76 (45–57)
Total	95 (15–66)	94 (09–63)	95 (09–62)	94 (09–69)	93 (08–59)	94 (13–62)

*The item-remainder associations have been placed between parentheses. **Excluding one item (no. 22) from subscale III.
All coefficients have been multiplied by 100 (Decimal points have been omitted).
Note. I=Social phobia, II=Agoraphobia, III=Bodily Injury, Illness and Death, IV=Sex and Aggression, V=Harmless Animals.

Table 9. FQ-Scale Reliabilities (Cronbach's coefficient α) and Range of Item–Remainder Correlations* by Subsample for each of the Sexes and the Sexes Combined

	Phobics (subsamples 1, 2, 3)		
	Males N=193	Females N=757	Total N=960
Agoraphobia	86 (63–82)	85 (59–80)	85 (60–80)
Blood-injury	75 (47–57)	76 (42–63)	75 (42–62)
Social	82 (53–72)	79 (48–69)	80 (50–70)
Anxiety–depression	81 (46–72)	80 (45–68)	80 (46–69)
Total phobia	88 (39–71)	84 (31–59)	85 (35–62)
	Non-patient normals (subsample 7)		
	Males N=137	Females N=140	Total N=277
Agoraphobia	78 (44–70)	81 (50–71)	80 (51–71)
Blood-injury	59 (11–50)	71 (34–61)	66 (25–56)
Social	78 (49–67)	75 (42–66)	77 (50–67)
Anxiety–depression	82 (50–73)	83 (56–74)	82 (52–71)
Total phobia	85 (08–66)	85 (20–61)	85 (16–61)
	Non-patient college students (subsample 8)		
	Males N=144	Females N=123	Total N=270
Agoraphobia	62 (22–47)	60 (27–52)	60 (30–46)
Blood-injury	78 (44–72)	68 (22–54)	74 (35–64)
Social	66 (23–53)	72 (30–58)	70 (25–55)
Anxiety–depression	84 (46–75)	80 (52–72)	83 (50–74)
Total phobia	80 (19–55)	75 (22–45)	78 (23–51)

*Item–remainder correlations have been placed between parentheses.
All coefficients have been multiplied by 100 (Decimal points have been omitted).

internal consistency: α=0.76 at retest (the corresponding figures for the phobic subscales were 0.87 for Agoraphobia, 0.79 for Blood-Injury Phobia, and 0.80 for Social Phobia). In the light of findings by Michelson and Mavissakalian showing that measures of Depression evidenced acceptable reliability coefficients and that these tended to improve as the test–retest periods (up to 4 months) increased (i.e. that agoraphobic patients' depression was relatively stable and chronic), a possible explanation for the very low stability of the Anxiety–Depression subscale might be the well-documented episodic nature of panic attacks which might produce greater variance over long periods of time than over shorter temporal spans (Michelson and Mavissakalian, 1983). With respect to the Anxiety component in the FQ subscale under consideration, it is worth mentioning that on a comparable

measure (PANIC) on which patients were requested to rate the 'frequency and intensity' of their panic attacks Michelson and Mavissakalian (1983) found a test–retest r=0.69 at period 1–4 weeks, r=0.41 at 2–10 weeks, and r=0.14 at 3–16 weeks.

Inspection of mean-scale scores on the Anxiety–Depression subscale further indicated that agoraphobics (failing to seek and find treatment for their complaints) reported being more anxious and depressive than described at pretest. The difference from pretest to retest on this subscale was very large — 1.44 *SD*.

Further comparison of scale statistics from the first and second test administration (Table 10) showed some evidence of regression towards the mean on the second testing. Significant drops ($P \leqslant 0.10$, one-tailed) of medium Effect Size are to be expected on the FQ Blood-Injury (d=0.62) and Agoraphobia (d=0.47) subscales over a non-treatment period exceeding one year. (A nonsignificant drop occurs on the Social Phobia subscale in this same period). These figures would be of interest in follow-up studies covering a retest interval of one year.

Test–retest reliabilities for the FQ phobic subscales are high enough to warrant the conclusion that the relative intensity of phobic avoidance experiences in these agoraphobics appears to remain relatively constant over a one year period.

Table 10. Test–Retest Reliabilities (13 Months) for the Scales of the FQ for an Agoraphobic Subsample (N=15), and Corresponding Pretest and Retest Scale Statistics

		Pretest statistics		Retest statistics		
	r_{tt}*	M	SD	M	SD	S_z
Agoraphobia	0.67	24.80	7.60	21.07	10.14	7.91
Blood-injury	0.62	21.40	7.02	16.80	9.40	7.38
Social	0.45	23.79	7.76	21.14	9.54	9.21
Anxiety–depression	0.09	7.14	2.25	17.79	7.01	7.42

*Spearman's ρ.

Thus, in conclusion, the present figures prove that changes in scores over time are minimally affected by the unreliability of the scales and moderately so by regression effects.

GENERAL DISCUSSION

Generality of Phobic Dimensions and its Implication

The present study clearly indicates that phobic factors ('Animal', 'Death, Pain and Illness', 'Aggression' and 'Social') frequently reported for college/university students (Hersen, 1973; Tasto, 1977; Wade, 1978; Granell de Aldaz, 1982) and normal and clinical adolescent samples (Bamber, 1974, 1979; Ollendick, 1983) are replicable in similar populations (phobics) and invariant across distinct ones (e.g. phobics vs normals). In addition, contrary to expectations (e.g. Hallam and Hafner, 1978), Agoraphobia could be interpreted as a separate factor in self-ratings from student and non-student normal samples. Occasionally, this factor re-appeared as an even stronger factor in these samples than it did in those comprising phobic (predominantly agoraphobic) individuals. Further, more than clear evidence was found for cross-national constancy, reliability and validity of all phobic dimensions under consideration.

The appearance of an Agoraphobic factor in a psychiatric outpatient sample is also remarkable. In only one (Dixon *et al.*, 1957) of the few studies conducted to date employing psychiatric patients (mostly inpatients) as *S*s (Bates, 1971; Lawlis, 1971; van der Toorn and Bremer, 1971; Rothstein *et al.*, 1972; Holmes *et al.*, 1975), was a comparable factor produced. In their classic study on patterns of anxiety, Dixon *et al.* (1957) described two factors which were labelled: separation and castration anxiety. They were found in factor analysing those items which referred to phobic anxiety (i=26) in a larger inventory comprising almost 900 statements. These were completed by 250 *S*s who were successive admissions to a psychiatric clinic. The first factor had threat of separation, helplessness and loneliness as a common element in all of its items; hence the label Agoraphobia would have been equally appropriate for its description.

It has been remarked that, in view of the variation in symptomatology of which agoraphobics complain, it may not be valid to recognise the syndrome as a distinct clinical entity; rather, agoraphobic fears should be looked at as occurring in or perhaps as being secondary to a variety of clinical contexts. Indeed, agoraphobia has been found to be related to alcoholism (e.g. Mullaney and Trippett, 1979), depression (Bowen and Kohout, 1979; Jarrett and Schnurr, 1979) and hyperventilation (Garssen *et al.*, 1983). For instance, Hallam (1978) has argued that agoraphobia should not be classified with the phobias since it is a variable feature of patients suffering from anxiety neurosis. He bases this argument on (1) the fact that Hersen's (1973) review of factor analytic studies of FSSs (mainly student data) does not mention the agoraphobic cluster of fears, "which if occurring in a mild form might have been expected to emerge as a common factor" (p. 315), (2) the elusiveness of the syndrome in samples of psychiatric patients with miscellaneous diagnosis, and (3) its emergence as a correlated cluster in data from clinical populations with known affective disorders or phobic anxieties.

Hallam (1978) states that the notion that agoraphobia is yet another phobia would be supported by a demonstration of its existence in the general population. Our own findings, however, which support the validity of the

agoraphobic syndrome, clearly point to the existence of such a factor in populations other than phobic ones; thus underscoring the idea that the symptoms typical of agoraphobia are so closely related to one another (in phobic, miscellaneous psychiatric outpatient, non-student normal and student samples) that they require unitary consideration; i.e. it is possible to recognise agoraphobia as a distinct clinical entity (Marks, 1969). Further support for this view can be found in a study by Torgersen (1979) which was not primarily conducted with the aim of establishing clusters of fears. Torgersen (1979) factor analysed self-report data of 99 same-sexed pairs of non-patient normal twins on a fear questionnaire whose items (i=51) were taken from a number of published fear questionnaires. Factor analysis revealed five conceptually clear factors, which were interpreted as (I) Separation fears or Agoraphobia, (II) Animal fears, (III) Mutilation fears, (IV) Social fears and (V) Nature fears.

Hitherto, the lack of emergence of an Agoraphobic factor in populations other than phobic ones has been interpreted as supporting Roth's description of the clinical picture of agoraphobia, namely, not as a graded affective response, but as a phenomenon that has "all or none" characteristics (1959, cited by Hallam, 1978) and as indicating that "true" agoraphobia is not common in a mild form (Hallam and Hafner, 1978). Clearly, this view can no longer be maintained. Further, it is unlikely that what has been described by Hallam (1978, pp. 315–316) as having produced, for clinical populations with known affective and phobic disorders, an Agoraphobic factor (i.e. the probably full-blown occurrence of — according to Roth's all-or-none hypothesis — the agoraphobic syndrome in a significant proportion of *S*s from these populations) would also apply to our non-patient normal samples or even to our heterogeneous psychiatric outpatient sample. This unlikeliness is discernable from a recent epidemiological study of fears and phobias conducted in Canada among female *S*s (N=449) from the community (Costello, 1982). Costello reported a prevalence figure of 132 per 1000 for Agoraphobia (Separation fears) irrespective of intensity. The fears were described as being *subclinical* forms of agoraphobia: the women had fears of being alone, travelling, and being in crowds, as described in DSM-III (APA, 1980), but none of them had the degree of restriction of normal activities that DSM-III lists as a diagnostic criterion for agoraphobia. As a matter of fact, most of the rated separation fears could be placed under the heading "mild fear" (10%) while only 4% could be categorised under the rubric "intense fear", thus indicating that agoraphobia does occur in mild forms. Further, other prevalence studies suggest an incidence of only 2.5–3.0% of phobic patients among psychiatric patients, while as many as 20% of psychiatric patients have significant phobic symptoms (cf. Thorpe and Burns, 1983, p. 7). Hence, the search for an explanation concerning the occurrence of an Agoraphobic factor in non-phobic samples in this study should perhaps be guided by methodological, statistical and theoretical requirements rather than clinical ones.

Close inspection of the large number of articles on factor analyses of responses to FSSs indicates that most of these (the most recent studies by Wade (1978), Granell de Aldaz (1982) and Kartsounis *et al.* (1983) being notable exceptions) are characterised by extraction of virtually the last remnants of systematic variance from the matrix of intercorrelations, and interpretation of the emerging factor stucture as though it had some inevitable stability, usefulness, and significance.** Usually, a minimum eigenvalue of 1.00 is used as a criterion for factor sufficiency, although frequently the researcher fails to report extraction criteria. The eigenvalue =1.00 criterion generally produces a large number of factors. In practice, however, the strategy of rotating all these factors has proved quite unrewarding (cf. Walkey, 1983), since it generally leads to overextraction of factors and, consequently, produces a distorted solution (e.g. Lee and Comrey, 1979; Revelle and Rocklin, 1979; Zwick and Velicer, 1982): the factor structures which emerge appear to be unstable and difficult to reproduce with any real precision, and, frequently, item-clusters which should provide the basis for the identification of factors are uninterpretable. Walkey (1983) has given a number of empirical illustrations of such a restriction of comparability to relatively few factors. He concludes that, on the whole, any stability of substantial numbers of items in practically reproducible structures appears to be restricted to solutions involving a much more limited number of factors than the eigenvalue $\geqslant$1.00 criterion would give, and even then, the factor structure which finally emerges may, due to underfactoring, obscure the effective distribution of the items into a pattern which clearly and appropriately reflects the scales built into the questionnaire by its developers. One reasonable way to overcome these problems would be to use confirmatory techniques of factor analysis by imposing a structure on the data, and to test the adequacy of various solutions in their adherence to or departure from that structure. In Walkey's (1983) opinion, comparatively little attention appears to have been paid to the use of the theoretical structure of the pattern of scales under examination when determining the number of factors to be rotated to simple structure. We contend that in the area of factor analyses of fear-related items no attention has been paid at all to the use of a theoretical structure (or some rational psychological criterion), the present study being the first to do so.

Walkey (1983) cites a number of examples from the literature to illustrate that analysis based on an appreciation of the developer's intent or on some alternative restricted theoretical structure may be more profitable than the frequency of its use in published studies would suggest, whereas less parsimonious criteria tend to lead to relatively unstable, unreliable, and often uninterpretable results. Walkey further demonstrates empirically that analyses of two psychometrically equivalent matrices containing an imposed-scale structure led to dissimilar solutions, neither of which revealed the imposed structure, when a minimum eigenvalue of 1.00 was used as a criterion for sufficiency of factors. More conservative (two- and three-factor) solutions revealed the imposed-scale structure in both matrices. Comparisons in his study with higher-order solutions showed considerable similarity between two-factor solutions and between three-factor solutions at different levels. These contrasted with substantial dissimilarities found in equivalent four-factor solutions.

Of interest to the present study is Walkey's (1983) illustration of the negative effect of using alternative criteria for the number of factors in effectively equivalent studies in reference to results of two analyses of fear-related items. In the Rubin *et al.* (1968) study, responses of four groups of *S*s to the FSS-II were factor analysed. Rubin *et al.* computed quartimax solutions, and found, without specifying their criterion for the number of factors, four generally identifiable factors. In a subsequent study, Braun and Reynolds (1969) took the 51 items of the FSS-II and, avoiding duplications, added the items of all previously published fear surveys to form a 100-item scale, the Temple Fear Survey Inventory (TFSI). They did not report any attempt to test the possibility that the extended scale might also be resolved into the four areas indicated by Rubin *et al.* (1968), but indicated instead the presence of 28 factors in the responses of 209 females, and 26 factors in those of the males (N=226). In each case only 21 factors were labelled, and there appeared to be 16 with common content. Walkey (1983) suggests, that, contrary to the implication of Braun and Reynolds (1969), the differences in the factor structures did not arise from sex differences, but were largely a function of the inherent unreliability of solutions of this kind. No criterion was given for the number of factors, although it seems most likely that the use of a minimum eigenvalue of 1.00 was responsible for such a proliferation of factors, many of which were defined by high loadings on only one or two of the 100 items (cf. Walkey, 1983).

Another study of interest in this context is that of Hallam and Hafner (1978). The results of their factor analysis of fear and symptom variables were briefly discussed in the Introduction. The authors interpreted the finding that an Agoraphobic factor emerged in both the total sample of phobics (N=171) and the subsample of agoraphobics (N=93) but not in that of the group of miscellaneous specific and social phobics (N=78), as pointing to the all-or-none nature of this fear. In the first instance, the data were analysed for the total sample of *S*s. Thus, to be able to draw a more definite conclusion it would have been advisable to use a methodology in which some hypothetical

**The problems considered here apply to the recently published study by Hafner and Ross (1984) as well.

structure were constructed to serve as point of departure in examining whether and, if so, to what degree the imposed structure was reproduced in the *sub*groups. Instead, separate factor analyses extracting distinct numbers of factors were run for the two diagnostic groups of phobics (of very small sample sizes), thereby increasing the risk of sampling errors.

The factor analytic findings by Hallam and Hafner (1978) are also difficult to interpret for reasons other than the problem of "the sole use of the eigenvalue=1.00 criterion". Although other extraction criteria were carefully applied in their study (e.g. interpretability and intercorrelations of the factors), nevertheless too many factors were extracted, since a number of these contained less than three variables. Such factors are clearly *not* overdetermined (Kline and Barrett, 1983) and would therefore tend to be unstable across samples (Overall and Klett, 1972, p. 108) and perhaps represent fragments of larger factors. Hence, extraction of a smaller number of factors would have seemed a more desirable strategy.

Inspection of the item-content of factors produced in a number of studies conducted among students (Braun and Reynolds, 1969; Merbaum and Stricker, 1972), psychiatric patients (Rothstein *et al.,* 1972) and phobics (Meikle and Mitchell, 1974) reveals (mostly among the overextracted factors) the presence of item-complexes (e.g. "Travel"), which in a sense relate to what has been referred here to as Agoraphobia or Separation fears. It is argued that it is quite possible that a re-analysis extracting a smaller number of factors would, due to a redistribution of the extracted variance across factors and thus of the items across factors (e.g. Gorsuch, 1974, p. 156), shift the nature of the factors and thus produce a distinct structure which may include an Agoraphobic factor, and give greater comparability of results across the distinct studies. Further, it is contended that the use of confirmatory techniques would lead to the rejection of different factors hitherto reported as being of some (clinical) importance.

While Marks and Mathews (1979) offered more than clear indications for an underlying (theoretical) dimensional representation of the FQ, this was not as clearly the case for the FSS-III and even less so for any of its variations. This uncertainty would not have inspired researchers to use confirmatory techniques in attempting to verify their underlying structures. Wolpe and Lang (1964) subclassified each of the seventy-six items into six broad categories: Animal (A); Tissue Damage, Illness, Death or associated stimuli (T); other Classical phobias (C); Social stimuli (S); Noises (N); and Miscellaneous (M). The fact that a few items in fact fit into more than one subclass and that arbitrary decisions were made about their placement may have prevented researchers from verifying this *a priori* structure by means of suitable techniques. However, it should also be noted that Wolpe and Lang acknowledged that the classification was given mainly to facilitate clinical use and that it was in no sense definitive (1964, p. 28).

Improving the Utility and Psychometric Qualities of the FSS-III and FQ

The FSS-III

In a recent study by Arrindell and van der Ende (1985) the 'Sex and Aggression' factor was found to be poorly replicable across split samples of phobics. The encouraging invariance characteristics of this factor across distinct samples found here, in conjunction with its moderate but satisfactory internal consistency reliabilities, however, does not preclude its further use. Rather than discarding the corresponding subscale, some work needs to be done to improve its utility, reliability and replicability. This would imply the selection of new items for inclusion into this subscale and thorough analysis of its construct validity. Since the unique definition of an underlying factor depends (among other aspects) on its degree of overdetermination, one way to increase the replicability of the 'Sex and Aggression' factor is to append new items from a comparable factor to it. The 'Contamination, Violence and Sex' factor, which was interpreted in Arrindell and Zwaan (1982) in their factor analysis of data of phobics on the AZUV, might serve the function of providing a number of non-overlapping items for the purpose described above. Thus, 12 AZUV items relating to Contamination, Violence and Sex were selected for inclusion in the FSS-III in order to examine their interrelationship with FSS items thoroughly in future research. (See the Appendix for the items concerned).

Further, the utility of the questionnaire can be enhanced by dropping those items (a total of 24) which do not form part of any of its subscales. Among these are items like dogs, cats and birds (not loading on the Animals factor). Gulas *et al.* (1975) have noted that the animals which form part of their 'Nuisance Animals' factor (which is identical except for one item, 'spiders', to the Arrindell, 1980, 'Harmless Animals' factor) are all culturally undesirable, unattractive, low in commonly perceived usefulness, and do not relate significantly in their fear-arousing qualities to the more domestic animals listed above. In fact, a fairly recent study by Bennett-Levy and Marteau (1984) found that fear ratings were significantly correlated with ratings of animal characteristics. The authors suggest that preparedness to fear certain animals (e.g. snakes) is not a function of the animals *per se* but of their fear-evoking perceptual properties and their discrepancy from the human form. Further, it was shown that the animals given above were among the less-feared animals. This may well be the reason why these domestic animal items also generally fail to have meaningful loadings on Animal factors in other studies (e.g. Lawlis, 1971; Granell de Aldaz, 1982; Kartsounis *et al.,* 1983).

The test–retest reliabilities of the subscales still have to be examined.

Further research might also focus attention on improving the quality of wording of certain FSS items, in order to reduce the possibility of multiple interpretation of item-content. For instance, the item aeroplanes (no. 62), which forms part of the Agoraphobic subscale, may have different connotations for distinct types of phobics: an aeroplane may be anxiety-arousing because of (a) the enormous noise which it produces, (b) its enclosed property, (c) a certain height at which it is flown, (d) it being a means of travel/transport, or (e) any of these aspects combined. A vaguely-worded item having no reference to specific situations can easily be misconstrued.

The Appendix gives a survey of the items which are captured by the currently used FSS-III, excluding the items which are superfluous and including those which may prove useful in enhancing the psychometric qualities of the 'Sex and Aggression' subscale. Further research with this version of the FSS now seems highly advisable.

The FQ

The test–retest reliability figure obtained here for the Anxiety–Depression subscale was quite low. However, in view of satisfactory internal consistency reliability and invariance characteristics of this scale, we would not suggest its exclusion from the FQ. Rather, the low test–retest figure may indicate that the Anxiety–Depression scale should be designated the predicate of a 'mood' scale rather than one measuring a 'state'. The first step in this direction has been taken by Cobb *et al.* (1984) in their study on 'the spouse as co-therapist in the treatment of agoraphobia'. However, Cobb *et al.* (1984) did not specify any reason for undertaking such a step.

In the light of the excellent invariance properties of the FSS-III factors, it now appears that two substantive fear components ('Sex and Aggression' and 'Animals') are not captured by the FQ, which is a drawback in the study of fear in a general population. These factors are less relevant if the FQ is used for the clinical study of fears. However, the subscale 'Animals' is particularly to be recommended for the investigation of irrational fears among community *S*s, since Costello's (1982) epidemiological study has shown that of five different kinds of fears, that of animals is the most prevalent in the community (the other four being: 'nature', 'social', 'mutilation' and 'separation', in this order). Analyses of the prevalence figures as a function of intensity of fear indicated that Animal fears ranked highest for the "intense fear" and "intense fear with avoidance" categories.

Further, due to higher internal consistency reliabilities, the FSS-III would appear to be a more suitable measure of fear than the FQ (both for clinical and research purposes), although not all scales of the former seem to be as appropriate for applied problems. The FSS Agoraphobic, Social and Bodily Injury factors appear (with reliabilities of $\alpha \simeq .9$) to be very suitable for both purposes mentioned.

A further problem with the FQ is the fact that it contains "one-item scales" (main target phobia and global phobia), which are by definition unreliable (e.g. Epstein, 1979; Zuckerman, 1983).

All phobic subscales may (for research purposes) be maintained in their present forms; increasing the reliability of the subscales would imply an increase in the number of items in each of the subscales. A suggestion in this direction would countermand one of the aims of its constructors, namely that of manageability and quick completion of the measure. Nevertheless, the problems noted should be given some thought since the FQ is emerging as an extensively-used measure and has now even been introduced in automated assessment form — the Computerized Fear Questionnaire (CFQ; Carr and Ghosh, 1983a, 1983b). Furthermore, it has been argued recently (Williams, 1984) that the FQ embodies too few agoraphobic situations to sample adequately the wide range of "typical" fears found in agoraphobic patients.

Some General Recommendations for the Use of the FSS-III and FQ

Total fear scales vs subscales

We feel, on the basis of data presented here and in the Arrindell (1980) study, that it is no longer acceptable to use the total FSS (or FQ) score as a valid measure of fear *per se;* the scales within these measures provide far more reliable information for program evaluation and other types of research than any other presently existing self-report measures of fear.

The use of either specific item-scores or general scores is likely to lead to invalid results. Unfortunately, a large number of studies published in leading journals in the field of behaviour research and therapy has followed such a route (see e.g. Tasto's, 1977, review of studies on the validity of the FSS). Using a total raw score on the FSS-III (or FQ) as an index of change resulting from therapeutic procedures designed to alleviate social anxiety, for instance, would create the problem that the changes implied by many items may simply not be susceptible to such a treatment procedure. Since both the FSS-III and the FQ contain a subscale relating to social matters which is empirically, relatively independent of other dimensions, the change as a result of treatment would be masked by the use of a total score (both by the unscaled items — FSS-III and by the subscales which are unrelated to the target behaviour being treated).

The FSS-III and FQ as indices of therapeutic effectiveness

It might be appropriate to use changes on these instruments from pre-therapeutic to post-therapeutic intervention as an index of therapeutic effectiveness, provided that the therapist and researcher have some indication as to what changes on the different subscales can be expected as a function of time (cf. Tasto, 1977).

In general, findings from this and other studies indicate that small to moderate decreases occur over both relatively short (5–10 weeks) and relatively long (13 months) periods of time (if there have been no interventions). Thus, in outcome studies covering a year's follow-up period, decreases in avoidance behaviour can be expected on phobic dimensions of the FQ merely due to the passage of time. A drop of at least d=0.65 should occur on the Blood-Injury subscale, and of d=0.50 on the Agoraphobic subscale to be indicative of effective therapy intervention.

In making judgements about change in individual cases the standard errors of measurement of the phobic subscales should be taken into account. These can be calculated by means of the reliability figures given above. To exclude measurement error as the primary source of the differences between a first and a second test score, the observed change (the first score minus the second score) should at least exceed the standard error of measurement.

FSS-III or FQ?

For research on irrational fears in the general population the FSS-III appears to be a more suitable tool than the FQ for the reasons mentioned above. A number of FSS scales are so reliable that they are actually more suitable than the FQ for selecting, for instance, analogue *S*s. (See e.g. Emmelkamp *et al.,* 1985, who selected students on the basis of a cut-off score on the 'Social' subscale of the FSS-III in a study on the external validity of analogue outcome research.) The currently used FSS-III is also suitable for the clinical purpose of obtaining a coherent first picture of the stimulus antecedents of a patient's neurotic reactions. For although certain relevant antecedents are often easily discernable, and may in fact be brought up by the patients when presenting their complaints, others may be quite obscure and identifiable only by extensive verbal probing —and sometimes not even then. In this case, the FSS-III can frequently save a great deal of clinical effort, revealing reactions to many stimulus classes in a relatively short time (cf. Wolpe and Lang, 1977, p.2).

On the other hand, the use of the FQ should be restricted, for the time being, to studies with phobic patients (agoraphobics and social phobics), which was in fact one of the reasons why it was introduced. The advantages of the FQ for this purpose have been outlined by Marks and Mathews (1979).

A study describing the construct validity, discriminatory power and norms for the FSS-III and FQ is currently underway.

Acknowledgements — The research reported in this article was supported by a grant (56-153) from the Netherlands Organization for the Advancement of Pure Research (ZWO). We are indebted to Mary Pickersgill (Bedford College, University of London) for making the British data here described available, to Richard van Dijck (Jelgersma Polikliniek, State University of Leiden) for making collection of parts of the data here utilized possible, to Jos M. F. ten Berge (State University of Groningen) for his helpful comments on an earlier draft of this article, and to Iefke Bakker and Boudien Krol, who devoted a substantial amount of their time to data coding. Further, thanks are due to Inger Mees for her help with syntax.

REFERENCES

Allen, B. P. and Potkay, C. R. (1981) On the arbitrary distinction between states and traits. *J. Personality soc. Psychol.,* **41,** 916–928.

American Psychiatric Association (1980) *Diagnostic and Statistical Manual of Mental Disorders (3rd edition).* APA., Washington, D.C.

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. *Behav. Res. Ther.,* **18,** 229–242.

Arrindell, W. A. and Emmelkamp, P. M. G. (1985) Psychological profile of the spouse of the female agoraphobic patient: personality and symptoms. *Br. J. Psychiat.* (in press).

Arrindell, W. A., van der Ende, J. and Emmelkamp, P. M. G. (1985) *Phobic dimensions: II. Construct validity, discriminatory power and norms.* (forthcoming).

Arrindell, W. A. and van der Ende, J. (1985) The need for data in factor analysis: an empirical test of the utility of the 'observation to variable'-ratio employing real world data. *Appl. Psychol. Measmt.* (in press).

Arrindell, W. A. and Zwaan, E. J. (1982) Fobische dimensies, de AZU-Vreesvragenlijst (AZUV) en Wolpe en Lang's Fear Survey Schedule (FSS-III). *Ned. Tijdsch. Psychol.,* **37,** 225–239.

Bamber, J. H. (1974) The fears of adolescents. *J. genet. Psychol.,* **125,** 127–140.

Bamber, J. H. (1979) *The Fears of Adolescents.* Academic Press, London.

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

Bates, H. D. (1971) Factorial structure and MMPI correlates of a Fear Survey Schedule in a clinical population. *Behav. Res. Ther.,* **9,** 355–360.

Bennett-Levy, J. and Marteau, T. (1984) Fear of animals: What is prepared? *Br. J. Psychol.,* **75,** 37–42.

Berrier, G. D., Galassi, J. P. and Mullinix, S. D. (1981) A comparison of matched clinical and analogue subjects on variables pertinent to the treatment of assertion deficits. *J. Consulting clin. Psychol.* **49,** 980–981.

Bowen, R. C. and Kohout, J. (1979) The relationship between agoraphobia and primary affective disorders. *Can. J. Psychiat.,* **24,** 317–322.

Brand-Koolen, M. J. M. (1972) *Factoranalyse in het Sociologisch Onderzoek.* H. E. Stenfert Kroese, N. V., Leiden.

Braun, P. R. and Reynolds, D. J. (1969) A factor analysis of a 100-item fear survey inventory. *Behav. Res. Ther.,* **7,** 399–402.

Butler, Jr. J. K. (1983) A program for computing an index of scale unidimensionality. *Educ. psychol. Measmt.,* **43,** 845–847.

Carr, A. C. and Ghosh, A. (1983a) Response of phobic patients to direct computer assessment. *Br. J. Psychiat.,* **142,** 60–65.

Carr, A. C. and Ghosh, A. (1983b) Accuracy of behavioural assessment by computer. *Br. J. Psychiat.,* **142,** 66–70.

Chambless, D. L., Hunter, K. and Jackson, A. (1982) Social anxiety and assertiveness: a comparison of the correlations in phobic and college student samples. *Behav. Res. Ther.,* **20,** 403–404.

Cobb, J. P., Mathews, A. M., Childs-Clarke, A. and Blowers, C. M. (1984) The spouse as co-therapist in the treatment of agoraphobia. *Br. J. Psychiat.,* **144,** 282–287.

Cohen, J. (1977) *Statistical Power Analysis for the Behavioral Sciences. Revised Edition.* Academic Press, New York.

Cooke, G. (1966) The efficacy of two desensitization procedures: an analogue study. *Behav. Res. Ther.,* **4,** 17–24.

Costello, C. G. (1982) Fears and phobias in women: a community study. *J. abnorm. Psychol.,* **91,** 280–286.

Derogatis, L. R. (1977) *SCL-90. Administration, Scoring and Procedures Manual-I for the R(evised) Version and other Instruments of the Psychopathology Rating Scale Series.* Johns Hopkins University School of Medicine, Clinical Psychometrics Research Unit, Boston, Mass.

Derogatis, L. R. and Cleary, P. A. (1977) Factorial invariance across gender for the primary symptom dimensions of the SLC-90. *Br. J. soc. clin. Psychol.*, **16**, 347–356.
Dixon, J. J., de Monchaux, C. and Sandler, J. (1957) Patterns of anxiety: the phobias. *Br. J. med. Psychol.*, **30**, 34–40.
Emmelkamp, P. M. G., Mersch, P. P. and Vissia, E. (1985) The External Validity of Analogue Outcome Research: Evaluation of Cognitive and Behavioral Interventions. *Behav. Res. Ther.* (in press).
Epstein, S. (1979) The stability of behavior: I. On predicting most of the people much of the time. *J. Personality soc. Psychol.*, **37**, 1097–1126.
Eysenck, H. J. (1973) *The Measurement of Intelligence.* MTP, Lancaster.
Eysenck, S. B. G., Adelaja, O. and Eysenck, H. J. (1977). A comparative study of personality in Nigerian and English subjects. *J. soc. Psychol.*, **102**, 171–178.
Eysenck, H. J. and Eysenck, S. B. G. (1983) Recent advances in the cross-cultural study of personality. In: J. N. Butcher and C. D. Speilberger (Eds), *Advances in Personality Assessment. Volume 2.* LEA, Hillsdale, New Jersey.
Eysenck, S. B. G., Humphery, N., and Eysenck, H. J. (1980) The structure of personality in Australian as compared with English subjects. *J. soc. Psychol.*, **112**, 167–173.
Farley, F. H. and Mealiea, W. L. (1972) Personality and fears. *J. Personality Assess.*, **36**, 451–453.
Farley, F. H., Cohen, A., Goldberg, J. and Yinon, Y. (1978) Fears in American and Israeli women. *J. soc. Psychol.*, **106**, 17–24.
Franks, C. M. and Wilson, G. T. (Eds) (1978) *Annual Review of Behavior Therapy. Theory and Practice, Vol. 6.* Brunner/Mazel, New York.
Fredrikson, M. (1983) Reliability and validity of some specific fear questionnaires. *Scand. J. Psychol.*, **24**, 331–334.
Galassi, J. P., Berrier, G. D. and Mullinix, S. D. (1982) The appropriateness of using statistically-selected college students as analogues to adult clinical subjects: a preliminary comparison on measures of psychological adjustment. *Behav. Therapist*, **5**, 179–180.
Garssen, B., van Veenendaal, W. and Bloemink, R. (1983) Agoraphobia and the hyperventilation syndrome. *Behav. Res. Ther.*, **21**, 643–649.
Geer, J. H. (1965) The development of a scale to measure fear. *Behav. Res. Ther.*, 3, 45–53.
Goldberg, J., Yinon, Y. and Cohen, A. (1975a) A factor analysis of the Israeli Fear Survey Inventory. *Psychol. Rep.*, **36**, 175–179.
Goldberg, J., Yinon, Y. and Cohen, A. (1975b) A cross-cultural comparison between the Israeli and American Fear Survey Inventory. *J. soc. Psychol.*, **97**, 131–132.
Gorsuch, R. L. (1974) *Factor Analysis.* Saunders, Philadelphia.
Granell de Aldaz, E. (1982) Factor analysis of a Venezuelan Fear Survey Schedule. *Behav. Res. Ther.*, **20**, 313–332.
Green, S. B., Lissitz, R. W. and Mulaik, S. A. (1977) Limitations of coefficient alpha as an index of test unidimensionality. *Educ. psychol. Measmt.*, **37**, 827–838.
Gulas, I., McClanahan, L. D. and Poetter, R. (1975) Phobic response factors from the Fear Survey Schedule. *J. Psychol.*, **90**, 19–25.
Gulliksen, H. (1950) *Theory of Mental Tests.* Wiley, New York.
Hafner, R. J. and Ross, M. W. (1984) Agoraphobia in women: factor analysis of symptoms and personality correlates of factor scores in a clinical population. *Behav. Res. Ther.*, **22**, 441–444.
Hallam, R. S. (1978) Agoraphobia: a critical review of the concept. *Br. J. Psychiat.*, **133**, 314–319.
Hallam, R. S. and Hafner, R. J. (1978) Fears of phobic patients: factor analyses of self-report data. *Behav. Res. Ther.*, **16**, 1–6.
Henderson, S., Byrne, D. G. and Duncan-Jones, P. (1981) *Neurosis and the Social Environment.* Academic Press, Australia.
Hersen, M. (1971) Fear scale norms for an in-patient population. *J. clin. Psychol.*, **27**, 375–378.
Hersen, M. (1973) Self-assessment of fear. *Behav. Ther.* **4**, 241–257.
Hochstim, J. R. and Renne, K. S. (1971) Reliability of response in a sociomedical population study. *Public Opinion Q.*, **35**, 69–79.
Holmes, G. R., Rothstein, W., Stout, A. L. and Rosecrans, C. J. (1972) Comparison of four factor analyses of the Fear Survey Schedule. *J. clin. Psychol.*, **31**, 56–61.

Iwawaki, S., Eysenck, S. B. G. and Eysenck, H. J. (1980) The universality of typology: a comparison between English and Japanese schoolchildren. *J. soc. Psychol.*, **112**, 3–9.
Jarrett, F. J. and Schnurr, R. (1979) Phobias and depression: clinical and psychometric aspects. *J. Behav. Ther. expl Psychiat.*, **10**, 167–171.
Johnston, M., Johnston, D. W., Wilkes, H., Burns, L. E. and Thorpe, G. L. (1984) Cumulative scales for the measurement of agoraphobia. *Br. J. clin. Psychol.*, **23**, 133–143.
Kartsounis, L. D., Mervyn-Smith, J. and Pickersgill, M. J. (1983) Factor analysis of the responses of British university students to the Fear Survey Schedule (FSS-III). *Personality individ. Diff.* **4**, 157–163.
Kline, P. and Barrett, P. (1983) The factors in personality questionnaires among normal subjects. *Adv. Behav. Res. Ther.*, **5**, 141–202.
Lawlis, G. F. (1971) Response styles of a patient population on the Fear Survey Schedule. *Behav. Res. Ther.*, **9**, 95–102.
Lee, H. B. and Comrey, A. L. (1979) Distortions in a commonly used factor analytic procedure. *Multivariate behav. Res.*, **14**, 301–321.
Marks, I. M. (1969) *Fears and Phobias.* Heinemann, London.
Marks, I. M. and Mathews, A. M. (1979) Brief standard self-rating for phobic patients. *Behav. Res. Ther.*, **17**, 263–267.
Meikle, S. and Mitchell, M. C. (1974) Factor analysis of the Fear Survey Schedule with phobics. *J. clin. Psychol.*, **30**, 44–46.
Merbaum, M. and Stricker, G. (1972) Factor analytic study of male and female responses to the Fear Survey Schedule. *J. Behav. Ther. expl. Psychiat.*, **3**, 87–90.
Michelson, L. and Mavissakalian, M. (1983) Temporal stability of self-report measures in agoraphobia research. *Behav. Res. Ther.*, **21**, 695–698.
Mullaney, J. A. and Trippett, C. J. (1979) Alcohol dependence and phobias: clinical description and relevance. *Br. J. Psychiat.*, **135**, 565–573.
Neiger, S., Atkinson, L. and Quarrington, B. (1981) A factor analysis of personality and fear variables in phobic disorders. *Can. J. behav. Sci.*, **13**, 336–348.
Nicholson, R. A. and Berman, J. S. (1983) Is follow-up necessary in evaluating psychotherapy? *Psychol. Bull.*, **93**, 261–278.
Nunnally, J. C. (1978) *Psychometric Theory. 2nd edition.* McGraw-Hill, New York.
Ollendick, T. H. (1983) Reliability and validity of the Revised Fear Survey Schedule for Children (FSSC-R). *Behav. Res. Ther.*, **21**. 685–692.
Overall, J. E. and Klett, C. J. (1972) *Applied Multivariate Analysis,* McGraw-Hill, New York.
Revelle, W. and Rocklin, T. (1979) Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behav. Res.*, **14**, 403–414.
Rothstein, W., Holmes, G. R. and Boblitt, W. E. (1972) A factor analysis of the Fear Survey Schedule with a psychiatric population. *J. clin. Psychol.*, **28**, 78–80.
Rubin, B. M., Katkin, E. S. Weiss, B. W. and Efran, J. S. (1968) Factor analysis of a Fear Survey Schedule. *Behav. Res. Ther.*, **6**, 65–75.
Rubin, S. E., Lawlis, G. F., Tasto, D. L. and Namenek, T. (1969) Factor analysis of the 122 item Fear Survey Schedule. *Behav. Res. Ther.*, **7**, 381–386.
Saigh, P. A. (1982) The Libanese Fear Inventory: a normative report. *J. clin. Psychol.*, **38**, 352–355.
Saigh, P. A. (1984) Pre- and postinvasion anxiety in Lebanon. *Behav. Ther.*, **15**, 185–190.
Simón, A. (1979) Effects of selective sampling on a factor analysis. *J. gen. Psychol.*, **101**, 259–264.
Smith, M. L., Glass, G. V. and Miller, T. I. (1980) *The Benefits of Psychotherapy.* The Johns Hopkins University Press, Baltimore.
Spiegler, M. D. and Liebert, R. M. (1970) Some correlates of self-reported fear. *Psychol. Rep.*, **26**, 691–695.
Spinks, P. M. (1980) An item and factor analysis of the FSS (III). *Personality individ. Diff.* **1**, 363–370.
Suinn, R. M. (1969) Changes in non-treated subjects over time: data on a Fear Survey Schedule and the Test Anxiety scale. *Behav. Res. Ther.*, **7**, 205–206.
Surwillo, W. W. (1980) *Experimental Design in Psychiatry. Research Methods for Clinical Practice.* Grune & Stratton, New York.

Tasto, D. L. (1977) Self-report schedules and inventories. In: A. R. Ciminero, K. S. Calhoun and H. E. Adams (Eds), *Handbook of Behavioral Assessment.* Wiley, New York.
Tasto, D. L. and Suinn, R. M. (1972) Fear Survey Schedule changes on total and factor scores due to nontreatment effects. *Behav. Ther.*, **3**, 275–278.
Taylor, C. B. and Agras, S. (1981) Assessment of phobia. In: D. H. Barlow (Ed.), *Behavioral Assessment of Adult Disorders.* The Guilford Press, New York.
ten Berge, J. M. F. (1984) Rotation to perfect congruence and the cross-validation of component weights across populations. *Submitted for publication.*
Thorpe, G. L. and Burns, L. E. (1983) *The Agoraphobic Syndrome.* Wiley, Chichester.
Torgersen, S. (1979) The nature and origin of common phobic fears. *Br. J. of Psychiat.*, **134**, 343–351.
Tyler, T. A. and Fiske, D. W. (1968) Homogeneity indices and test length. *Educ. psychol. Measmt.*, **28**, 767–777.
van der Toorn, W. and Bremer, J. J. C. B. (1971) Vreesvragenlijst-bevindingen bij psychiatrische patiënten. *De Psycholoog*, **6**, 212–216.
Wade, T. C. (1978) Factor analytic approaches to the investigation of common fears: a critical appraisal and reanalysis. *Behav. Ther.*, **9**, 923–935.
Walkey, F. H. (1983) Simple versus complex factor analyses of responses to multiple scale questionnaires. *Multivariate behav. Res.*, **18**, 401–421.
Williams, S. L. (1984). On the nature and measurement of agoraphobia. In: *Progr. Behav. Modification* (in press).
Windle, C. (1954) Test-retest effect on personality questionnaires. *Educ. Psychol. Measmt.*, **14**, 617–633.
Wolff, W. T. and Merrens, M. R. (1974) Behavioral assessment: a review of clinical methods. *J. Personality Assess.*, **38**, 3–16.
Wolpe, J. (1973) *The Practice of Behavior Therapy. (second edition).* Pergamon Press, New York.
Wolpe, J. and Lang, P. J. (1964) A Fear Survey Schedule for use in behaviour therapy. *Behav. Res. Ther.*, **2**, 27–30.
Wolpe, J. and Lang, P. J. (1977) *Manual for the Fear Survey Schedule.* EdITS, San Diego, California.
Zuckerman, M. (1983) The distinction between trait and state scales is *not* arbitrary: comment on Allen and Potkay's "On the arbitrary distinction between traits and states". *J. Personality soc. Psychol.*, **44**, 1083–1086.
Zwick, W. R. and Velicer, W. F. (1982) Factors influencing four rules for determining the number of components to retain. *Multivariate behav. Res.*, **17**, 253–269.

APPENDIX 1

The items contained in the Fear Survey Schedule including the 12 items appended* to the measure for further research purposes.

1. Open wounds (B)
2. Being alone (A)
3. Being in a strange place (A)
4. Dead people (B)
5. Speaking in public (S)
6. Crossing streets (A)
7. Falling (A)
8. Being teased (S)
9. Failure (S)
10. Being touched by others*
11. Entering a room where other people are already seated (S)
12. High places on land (A)
13. People with deformities (B)
14. Worms (An)

15. Receiving injections (B)
16. Strangers (S)
17. Bats (An)
18. Journeys by train (A)
19. Journeys by bus (A)
20. Journeys by car (A)
21. People in authority (S)
22. Flying insects (An)
23. Seeing other people injected (B)
24. Crowds (A)
25. Large Open Spaces (A)
26. One person bullying another (S-A)
27. Sight of deep water*
28. Tough-looking people (S-A)
29. Being watched working (S)
30. Dirt (S-A)
31. Crawling insects (An)
32. Sight of fighting (S-A)
33. Sight of earthworms*
34. Ugly people (S-A)
35. Sight of fire*
36. Sick people (B)
37. Being criticized (S)
38. Strange shapes (S-A)
39. Being in an elevator (A)
40. Witnessing surgical operations (B)
41. Mice (An)
42. Human blood (B)
43. Sight of a rat*
44. Animal blood (B)
45. Enclosed places (A)
46. Feeling rejected by others (S)
47. Sight of knives or other sharp objects*
48. Sight of parasites (lice, fleas etc.)*
49. Aeroplanes (A)
50. Medical odours (B)
51. Feeling disapproved of (S)
52. Harmless snakes (An)
53. Sight of weapons*
54. Sudden noise*
55. Being in a storm*
56. Cemeteries (B)
57. Being ignored (S)
58. Nude men (S-A)
59. Nude women (S-A)
60. Doctors (B)
61. Visiting people who never clean their houses*
62. Making mistakes (S)
63. Looking foolish (S)
64. Working with poisoned material*

Note. The instructions for filling out the questionnaire are analogous to those given by Wolpe and Lang (1964).
S=Social; A=Agoraphobia; B=Blood Injury, Illness and Death; S-A=Sex and Aggression; An=Harmless Animals.
As with the FSS-III, all items are rated on 5-point continua.