

## SHORTENING THE STATE-TRAIT ANXIETY INVENTORY

F. C. E. VAN KNIPPENBERG, H. J. DUIVENVOORDEN, B. BONKE and J. PASSCHIER

Department of Medical Psychology and Psychotherapy, Faculty of Medicine, Erasmus University,  
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

(Received in revised form 25 July 1989)

**Abstract**—Questionnaires that are used in studies with severely ill patients should be as short as possible. Abridged versions of existing inventories are very practical in these instances. The answers of 444 subjects in three samples (cancer patients, medical students, surgical patients) were used to investigate the possibility of constructing short and reliable versions of the scales of the State and Trait Anxiety Inventory. A stepwise regression procedure showed the possibility to reliably predict the total score of the unabridged versions by means of weighted sums of eight items for each scale. Omission of weights did not lead to substantial loss of information. Cronbach's  $\alpha$  of the State-scale decreased from 0.93 to about 0.85 for different combinations of items and from 0.91 to about 0.82 for combinations of eight items of the Trait-scale. The relationship between both scales was only slightly modified by the shortening procedure.

Anxiety assessment    Test length    Psychometry    Reliability

### INTRODUCTION

The number of publications on the psychological impact of diseases is increasing steadily. In research of this kind standardized psychological tests (and scales) are administered on a regular basis. However, these tests are quite often too extensive for severely ill or handicapped patients. Moreover, in many studies interrelationships between several psychological constructs, operationalized in a variety of ways, are investigated, which requires the completion of several tests. Shortened versions of valid tests are, therefore, desirable, provided they remain reliable and valid without causing changes in the interrelationships with other variables. Furthermore, in specific situations fluctuating states have to be assessed more than once in a short period of time. Repeated measurements with the same instruments lead to recall effects. More than one abridged scale of an original scale can, in these instances, be of use.

The State-Trait Anxiety Inventory (STAI) [1] is a widely used instrument in psychological research on patients [e.g. 2, 3]. It consists of two scales of 20 items each: State Anxiety (S-Anx) and Trait Anxiety (T-Anx). Recently, Ramanaiah *et al.* [4] suggested a reduction in the number of items of the two scales. In one sample ( $n = 212$ ) the reliability (Cronbach's  $\alpha$ ) of S-Anx decreased from 0.92 in the original 20-item version to 0.87 in a 7-item version, and in a second sample ( $n = 205$ ) it did so from 0.90 to 0.85. In both samples the reliability of T-Anx (original vs 5-item versions) also decreased: from 0.92 to 0.75 in the former and from 0.88 to 0.62 in the latter sample. The purpose of the Ramanaiah *et al.* study was to reduce the number of items with low content saturation. High "content saturation" refers to "the extent that the item's correlation with its own scale is greater than its correlation with the irrelevant scale" [4, p. 531]. Low content saturation leads to higher intercorrelations with irrelevant scales.

S-Anx and T-Anx were considered irrelevant to each other and deleting items with low content saturation leads to greater independency of both scales. In the first sample the correlation between the two scales decreased from 0.69 to 0.43, and in the second sample from 0.64 to 0.45. Spielberger and Vagg [5] disagreed on the relevance of independency.

In the present study attention was focused on shortening the two STAI scales while retaining their interrelationship following Spielberger and Vagg. Specifically formulated, the research question is as follows: Is it possible to reliably predict scores on the 20-item version with fewer items and meanwhile retain the correlation between S-Anx and T-Anx?

## METHOD

### Subjects

The possibility of shortening the STAI was tested on data from three different samples.

*Sample 1.* 132 cancer patients filled out S-Anx and T-Anx some 3 weeks to 3 months after having completed a radiation therapy course. Three patients were omitted because of incomplete answers (more than 5 items were missing on both scales together). The remaining group ( $n = 129$ ; mean age: 64.5 yr; range: 31–82 yr) consisted of 42 males with lung cancer, 28 males with cancer of the prostate, 17 males and 4 females with carcinoma of the bladder, 34 females with uterine cancer (cervix or corpus), and 4 women with other malignancies.

*Sample 2.* 121 medical students (66 males and 55 females; mean age: 21.3 yr; range: 19–43 yr), in their second year, completed S-Anx and T-Anx during a psychology course.

*Sample 3.* 194 patients (79 males and 115 women; mean age: 38.7 yr; range: 18–80 yr) awaiting plastic or reconstructive surgery completed both scales one day before surgery.

All subjects completed a Dutch version of the STAI [6]. Each of the 20 items receives a score from 1 to 4. This Dutch translation (DY) contains a balanced number of "anxiety absent" (e.g. "I feel calm") and "anxiety-present" items (e.g. "I feel tense"). The DY is a translation of an earlier form of the American STAI (X-version) and translations of some items from the Y-version were added [6]. Spearman rank correlations between scores on the original American Y version and DY, for Dutch students with a reasonable command of English, ranged from 0.97 to 0.99 [6].

### Method of analysis

In the shortening procedure, items of each scale were selected for inclusion on the basis of their contributions to the prediction of the 20 item scores. Validity of the 20 item versions was taken for granted. For each of the samples, and for both scales separately, items were selected by means of stepwise linear regression analysis, as a tool for efficient prediction [7]. Nonetheless, it is important to examine the degree of multicollinearity [8]. Three indicators for multicollinearity were investigated: (1) high inter-item correlation; (2) opposite signs for each pair of bivariate (item-total) correlations and regression weights; and (3) relative value of the standard errors for regression weights of selected items.

Only two out of all Pearson inter-item correlations (a total of 1140; for each sample: 380) were above 0.80 (0.84 and 0.86) and only 1.5% lay between 0.70 and 0.80. 85% were below 0.50. All regression weights were positive as were the bivariate correlations. None of the standard errors of the regression weights was too high: all  $P$ 's  $\leq 0.01$ . Thus, it can be concluded that multicollinearity has not occurred, and that application of the method of linear regression analysis was justified.

Cross validation was carried out in two ways: by calculating  $MR^2$  (squared multiple correlations) for shrinkage [9], and by the procedure of estimating regression weights in one sample and applying these to the two other samples. This latter procedure led to 6 different linear combinations of weighted item scores, i.e. 3 for S-Anx and 3 for T-Anx.

Cronbach's  $\alpha$  was calculated for the abridged scales. For research purposes 0.80 is adequate [10].

## RESULTS AND DISCUSSION

In Table 1, values for MR and percentages of accounted variance ( $MR^2$ ) are presented for 1–20 items on S-Anx and T-Anx for the 3 samples. Linear combinations of weighted scores on 8 items were mostly sufficient to account for 95% or more of the variance in the sum of 20 item scores. In sample 3 a weighted sum of 8 selected items accounted only for 94% of the variance in the original 20-item T-Anx scores. The number of 8 items was arbitrarily chosen, as the main purpose of the study was to considerably reduce the size of the scales. The

Table 1. Order of selected items in regression analysis, multiple  $R$  and  $MR^2$  for State (section A) and Trait (section B) Anxiety scales in three samples. Numbers of items refer to the standard order of items of the DY form

Step in analysis	Sample 1 (cancer patients)			Sample 2 (medical students)			Sample 3 (surgical patients)		
	Item added	MR	MR <sup>2</sup>	Item added	MR	MR <sup>2</sup>	Item added	MR	MR <sup>2</sup>
<i>(A) State</i>									
1	15	0.83	0.70	15	0.79	0.62	3	0.77	0.59
2	16	0.92	0.84	18	0.86	0.75	11	0.89	0.79
3	7	0.95	0.90	11	0.90	0.82	17	0.93	0.87
4	5	0.96	0.92	17	0.93	0.87	20	0.95	0.90
5	10	0.97	0.94	4	0.95	0.90	15	0.96	0.93
6	12	0.98	0.95	2	0.96	0.92	4	0.97	0.95
7	19	0.98	0.96	8	0.97	0.94	19	0.98	0.96
8	4	0.99	0.97	12	0.97	0.95	12	0.98	0.96
9	1	0.99	0.98	7	0.98	0.96	2	0.98	0.97
10	8	0.99	0.98	5	0.98	0.96	8	0.99	0.97
11	17	0.99	0.99	10	0.99	0.97	7	0.99	0.98
12	11	1.00	0.99	14	0.99	0.98	16	0.99	0.98
13	20	1.00	0.99	3	0.99	0.98	18	0.99	0.99
14	9	1.00	0.99	16	0.99	0.99	13	0.99	0.99
15	18	1.00	0.995	19	0.99	0.99	5	1.00	0.99
16	6	1.00	1.00	9	0.995	0.99	6	1.00	0.995
17	3	1.00	1.00	20	1.00	0.99	1	1.00	1.00
18	14	1.00	1.00	1	1.00	0.995	9	1.00	1.00
19	2	1.00	1.00	13	1.00	1.00	10	1.00	1.00
20	13	1.00	1.00	6	1.00	1.00	14	1.00	1.00
<i>(B) Trait</i>									
1	14	0.82	0.67	18	0.74	0.54	14	0.70	0.49
2	2	0.91	0.83	14	0.86	0.74	11	0.83	0.68
3	1	0.93	0.87	17	0.91	0.83	6	0.88	0.77
4	11	0.95	0.90	16	0.94	0.88	20	0.91	0.83
5	7	0.96	0.92	7	0.96	0.91	19	0.93	0.87
6	12	0.96	0.93	15	0.97	0.94	12	0.95	0.91
7	15	0.97	0.95	8	0.97	0.95	1	0.96	0.92
8	6	0.98	0.96	20	0.98	0.96	17	0.97	0.94
9	20	0.98	0.96	11	0.98	0.96	10	0.97	0.95
10	16	0.98	0.97	2	0.98	0.97	4	0.98	0.95
11	17	0.99	0.97	6	0.99	0.97	15	0.98	0.96
12	4	0.99	0.98	5	0.99	0.98	7	0.99	0.97
13	3	0.99	0.98	12	0.99	0.98	18	0.99	0.98
14	19	0.99	0.99	4	0.99	0.98	3	0.99	0.98
15	10	0.99	0.99	10	0.99	0.98	9	0.99	0.98
16	5	1.00	0.99	19	0.99	0.99	5	0.99	0.99
17	8	1.00	0.99	3	1.00	0.99	13	1.00	0.99
18	9	1.00	1.00	9	1.00	0.99	8	1.00	0.99
19	18	1.00	1.00	13	1.00	1.00	16	1.00	1.00
20	13	1.00	1.00	1	1.00	1.00	2	1.00	1.00

adjusted values ( $MR^2$  for shrinkage) for the shortened scales were slightly lower than those presented in Table 1. These differences were, on the average, about 0.3% of the accounted variance, never exceeding 0.5%, which indicates that the  $MR^2$  values appear to be stable.

The second method of cross-validation was carried out by transferring the selected items and the corresponding weights from each sample to the other two samples. For instance, the linear combination of 8 S-Anx items and weights of sample 1 was used for calculating an S-Anx score for the subjects in samples 2 and 3. This yielded the WS1S (weighted sum sample 1 S-Anx) scores for the three studies. The same

procedure was carried out in calculating the WS2S and WS3S, and, similarly, the T-Anx scores: WS1T, WS2T and WS3T.

These anxiety scores were then correlated with the original 20-item scores (Table 2), for each of the three samples and for all subjects together. As can be seen from this table, the correlations are high, and, as expected, highest for the weighted combination scores extracted from the sample itself (WS1S and WS1T for sample 1, WS2S and WS2T for sample 2, and WS3S and WS3T for sample 3). Table 3 shows that the choices of items do not really influence the correlations, as the intercorrelations of three corresponding linear combinations (WS1S,

Table 2. Correlations between linear combinations of 8 items and the sum of the original 20 item scores for State and Trait Anxiety Scales for each sample and for the total group

	State				Trait				
	Sample				Sample				
	1	2	3	Total	1	2	3	Total	
WS1S	[0.99]	0.96	0.95	0.97	WS1T	[0.98]	0.94	0.94	0.95
WS2S	0.96	[0.98]	0.96	0.97	WS2T	0.97	[0.98]	0.95	0.96
WS3S	0.97	0.96	[0.98]	0.97	WS3T	0.96	0.93	[0.97]	0.96

The values in square brackets are correlations between original score and linear combinations from that sample. WS0S means: weighted sum state (from samples 1, 2 and 3); WS0T: weighted sum trait (from samples 1, 2 and 3).

WS2S, WS3S and WS1T, WS2T, WS3T) are all very high too. There is, however, some item overlap (see Table 1) but this overlap is limited: 3 between WS1S and WS2S, 4 between WS1S and WS3S, 5 between WS2S and WS3S, 3 between WS1T and WS2T, 5 between WS1T and WS3T, and 3 between WS2T and WS3T.

It is possible to correctly predict the 20-item Anxiety scores by means of a weighted sum of 8 items. The weights can vary, without many consequences for the predictive power. In practice, however, weights are not very convenient. The correlations between weighted and unweighted linear combinations are 0.99 or higher for both S-Anx and T-Anx. Consistent with this, the correlations between weighted linear combinations and the original sum of 20 items, are almost identical to the corresponding correlations between unweighted combinations and the original 20 item sums. Consequently, weights can be omitted.

The research question also concerns the reliabilities of the shortened versions. In Table 4 reliabilities (Cronbach's  $\alpha$ ) of the original scales and the shortened versions of both scales are presented. Reliabilities were computed for weighted (*W*) and unweighted (*U*) versions. Shortening leads to a decrease in reliability, but the reliabilities for the unweighted combinations

are, consistently, slightly higher than those of the combinations of weighted items. Furthermore, the reliabilities are higher than those found by Ramanaiah *et al.*, as mentioned in the introduction, and the values in Table 4 clearly meet the standards for research purposes put forward by Nunnally [10]. These values can be increased by adding more items. Internal consistencies have also been calculated for combinations of 10 items as an example (unweighted: Table 4, last column).

The second part of the question concerns the possibility of changes in correlations between the anxiety-scales and other variables. As a test for this possible change the correlations between both scales were used. In Table 5 the correlations between State and Trait measured by the shortened versions were compared with the values obtained via the original 20 item versions (top row). The decrease in correlation was smallest for the values obtained by the shortened version extracted from the sample itself (in Table 5, enclosed in square brackets). In the most extreme case (sample 3) the consequence is a difference in explained variance of 9.4% ( $0.52 * 0.52 - 0.42 * 0.42$ ).

Table 4. Internal consistencies (Cronbach's  $\alpha$ ) for the original 20-item versions and abridged 8 and 10 items versions. Indices were calculated for the total sample\*

	Version	Reliability	2 items added (see text)
<i>State:</i>	20 items	0.93	
	WS1S	0.85	
	WS2S	0.84	
	WS3S	0.86	
	US1S	0.86	US1S:0.89
	US2S	0.85	US2S:0.88
	US3S	0.87	US3S:0.89
<i>Trait:</i>	20 items	0.91	
	WS1T	0.81	
	WS2T	0.82	
	WS3T	0.81	
	US1T	0.84	US1T:0.86
	US2T	0.83	US2T:0.86
	US3T	0.81	US3T:0.81

\* $n = 438$ , due to pairwise deletion of missing values.

Table 3. Intercorrelations between the weighted linear combinations WS1S, WS2S, WS3S and WS1T, WS2T, WS3T for samples 1, 2 and 3 and the total sample

Sample	State		Trait			
	WS1S	WS2S	WS1T	WS2T		
1	WS2S	0.95	—	WS2T	0.94	—
	WS3S	0.95	0.95	WS3T	0.97	0.95
2	WS2S	0.93	—	WS2T	0.95	—
	WS3S	0.92	0.95	WS3T	0.91	0.91
3	WS2S	0.90	—	WS2T	0.90	—
	WS3S	0.93	0.95	WS3T	0.94	0.92
Total	WS2S	0.92	—	WS2T	0.92	—
Sample	WS3S	0.94	0.95	WS3T	0.95	0.93

Table 5. Correlations between abridged versions of State and Trait compared with those of the original 20-item versions (first 4 columns)

	Sample			Total sample	Abridged versions with original 20-items
	1	2	3		
S-Anx with T-Anx (20 items)	0.88	0.73	0.52	0.68	
WS1S with WS1T	[0.86]	0.66	0.42	0.58	(0.63/0.63)
US1S with US1T	[0.85]	0.67	0.43	0.59	(0.63/0.62)
WS2S with WS2T	0.83	[0.71]	0.48	0.65	(0.68/0.65)
US2S with US2T	0.83	[0.71]	0.48	0.65	(0.68/0.65)
WS3S with WS3T	0.84	0.69	[0.53]	0.66	(0.67/0.66)
US3S with US3T	0.82	0.68	[0.52]	0.65	(0.66/0.66)

Correlations between abridged State and original Trait, and between abridged Trait and original State for the total sample (in parentheses in the last column). For the meaning of square brackets: see text.

It is, however, more correct to test this part of the question by correlating the shortened scale with the original of the other scale, because not all scales used in studies will be shortened. In Table 5, last column, the correlations of abridged versions of S-Anx with the 20-item version of T-Anx and of abridged versions of T-Anx with the 20-item version of S-Anx are presented for the total sample. These coefficients are slightly higher than the ones presented in the fourth column.

In analyzing the items for content, it should be noted that half of the 20 items in the original version include references to anxiety, fear, worry, etc. ("Anxiety-present" items,  $A^+$ ), while the other half do not ("Anxiety-absent" items,  $A^-$ ). The item selections from the 3 samples were divided according to this  $A^+$  and  $A^-$  distinction, as this could be of some theoretical relevance [11]. There are many more  $A^+$  items than  $A^-$  items in the selections of T-Anx (ratio of 5:3 in all 3 samples). The  $A^+$  and the  $A^-$  items are more evenly distributed in the State selections (3:5, 4:4 and 4:4 in samples 1, 2 and 3, respectively).

#### CONCLUSION

Shortening the State-Trait Anxiety Inventory seems justified and the reliabilities remain adequate. The method of analysis chooses predictor-variables with interdependence as low as possible and correlations with the criterion (i.e. the sum of scores on 20 items to be predicted) as high as possible. However, when the correlations between predictor-variables are low, a low internal consistency is to be expected. In this study this paradox appeared of minor importance because the internal consistency of the shortened versions still remained high. Starting from a pragmatic point of view (to predict efficiently) the final set of selected items was

effective in predicting the scores on the 20 item version without any weighting.

Choosing items at random or choosing items with high intercorrelations will, generally, result in higher values for the reliability. For example, a 6-item selection from sample 1, derived from factor analysis (3 items loading high on the first and 3 items loading high on a second factor) yielded an  $\alpha$  of 0.80, a 6-item selection of T-Anx yielded an  $\alpha$  of 0.88. The reliability was also calculated for the 10 first items of each scale (total sample):  $\alpha$  (S-Anx) = 0.87,  $\alpha$  (T-Anx) = 0.82. In these 6 and 10 item versions the numbers of  $A^+$  and  $A^-$  items were balanced: 3:3 and 5:5.

The analysis of 3 samples from 3 different populations showed comparable results suggesting that the results can be generalized to other groups. Considerable shortening of both scales is possible with a minimum loss of information. In contrast with our approach, the shortening proposed by Ramanaiah *et al.* probably leads to more loss of information. An incidentally large decrease in reliability and intercorrelation can be prevented by using 10 or 12 items, which is still a considerable reduction in number of items. Using a shortened version of the STAI combines the gain of 5 or 6 minutes in time needed to fill out a questionnaire, with the disadvantage of not being able to compare the test results with norms based on scores of large samples. The time saved must be balanced against this disadvantage.

Abridged versions should include both  $A^+$  and  $A^-$  items, as such selections reflect the content of the original versions.

#### REFERENCES

1. Spielberger CD, Gorsuch RL, Lushene R. *Test Manual for the State-Trait Anxiety Inventory*. Palo Alto, Calif.: Consulting Psychologists Press: 1970.

2. Boeke S, Duivenvoorden H, Bonke B. Agreement for surgical patients on two situations for the Trait Anxiety Inventory. *Psychol Rep* 1984; 54: 278.
3. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the functional Living Index-Cancer: development and validation. *Clin Oncol* 1984; 2: 472-483.
4. Ramanaiah NV, Franzen F, Schill T. A Psychometric study of the State-Trait Anxiety Inventory. *Pers Assess* 1983; 47: 531-535.
5. Spielberger CD, Vagg PR. Psychometric properties of the STAI: a reply to Ramanaiah, Franzen, and Schill. *Pers Assess* 1984; 48: 95-97.
6. Ploeg HM van der, Defares PD, Spielberger CD. *Handleiding bij Zelfbeoordellingsvragenlijst (Z.B.V.)* Lisse, The Netherlands: Swets and Zeitlinger; 1980.
7. Lord FM, Novick MR. *Statistical Theory of Mental Test Scores*. Reading, Mass.: Addison-Wesley; 1968.
8. Kerlinger FN, Pedhazur JP. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston; 1973.
9. Wiggins JS. *Personality and Prediction: Principles of Personality Assessment*. Menlo Park, Calif.: Addison-Wesley; 1973.
10. Nunnally JC. *Psychometric Theory*. New York: McGraw-Hill; 1978.
11. Bonke B, Smorenburg JMJ, Ent CK van der, Spielberger CD. Denial and item-intensity specificity in the State-Trait Anxiety Inventory. *Pers Individ Diff* 1987; 8: 185-191.