SCIENTIFIC ARTICLES

Development of an Interval Scale of Anxiety Response

Norman L. Corah, Ph.D.,* Maria A. Zielezny, Ph.D.,† Robert M. O'Shea, Ph.D.,† Terrance J. Thines, D.D.S.‡ and Pauline Mendola*

*Department of Behavioral Sciences, School of Dental Medicine, †Department of Social and Preventive Medicine, School of Medicine, ‡Department of Oral Medicine, School of Dental Medicine, State University of New York at Buffalo

Summary

This paper reports the development of an interval scale of anxiety response. Magnitude estimation procedures were used with three different groups of subjects to develop a suitable scale of seven anxiety descriptors. The ratio of the highest to lowest descriptor magnitudes was 21 to 1. Analyses of the descriptor sets in the various groups indicated high reliability of meaning and high objectivity. In addition, high agreement on meaning was shown for groups of differing education and socioeconomic status. A preliminary study using the scale indicates appropriate preliminary construct validity. Further reliability and validity research is needed. This scale may be useful for assessing anxiety response changes in a variety of contexts.

One of the major problems confronting the researcher in anxiety studies is the assessment of situational anxiety response or state anxiety. This assessment has been conducted through the use of behavioral, physiological and self-report techniques. The most widely used techniques have been some form of verbal self-report — especially with adult subjects. A variety of self-report measures have been used with a reasonable degree of success. A common type of measure is the item response scale in which the total of item scores answered in the "anxious" direction is the measure of anxiety.[1] Adjective check lists utilizing the total number of appropriately checked terms is a similar measure.[2] Corah and his co-workers[3] have used a seven-point graphic rating scale derived from the work of Lazarus.[4] All of these verbal self-report measures are ordinal in character but are usually treated as interval measures for purposes of analysis. However, the actual psychological distance between points or scores on these scales is typically unknown.

In recent years, investigators in pain research have developed psychophysical ratio scales of pain response.[5,6] These scales were developed by cross-modality matching which involved linkage of pain descriptors with subjects' response to calibrated pain stimuli. No such physical stimulus dimension exists, however, for anxiety response. Recent authors have called for the development of anxiety scales which are at least ordinal in character, carefully tied to conceptual referents and demonstrating the usual characteristics of good scales — reliability, validity and usefulness.[7,8] It is reasonable to argue than an equal interval scale may offer a more sensitive assessment of anxiety response than an ordinal scale since the perceptual distance between categories is defined in the interval scale. Heft and Parker[9] have argued this point effectively for the development of graphic scales of pain response. This paper reports the development of an interval scale of anxiety descriptors by the psychophysical method of magnitude estimation.[10] Our intent was to develop a graphic rating scale because of our previous experience in using this type of scale.

Development of the Scale

Initial Selection and Scaling of the Descriptors

The conceptual basis for our approach to the assessment of anxiety was derived from the work of Lazarus.[4] This view holds that anxiety is a dysphoric

Accepted for publication August 15, 1986.

Address reprint requests to Dr. Norman L. Corah, Department of Behavioral Sciences, School of Dental Medicine, Squire Hall, Buffalo, NY 14214.

response which is a cue of danger and a motivator of instrumental behaviors to avoid or reduce the fear. Operationally, we consider anxiety to be expressed verbal distress to an impending threat. The empirical basis for selection of descriptors was derived from the two-factor structure of affect developed by Watson and Tellegen[11] from reanalysis of studies of self-reported mood. The dimension they called "negative affect" included many of the descriptors already utilized in a number of anxiety measures.[1,3,12] This list of descriptors also included some that were indicative of anger (e.g., "hostile" and "scornful") that were eliminated on our first trial where three raters were asked to eliminate those words which did not belong in the list. A list of 19 descriptors was used in the first scaling procedure.

The 28 subjects, 13 men and 15 women between the ages of 22 and 51 (mean, 32), were staff members (faculty and research assistants), graduate students and fourth-year dental students from the university. Standard magnitude estimation instructions were given including the observance of proportionality among stimuli by asking subjects to assign numbers. Two tasks were presented. The first was a large page with 7 line lengths ranging from 1.3 cm to 33 cm in equal log steps presented in random order. The following instructions were given: "Several lines are shown below. Your task is to assign a number to each line such that it is considered in relation to the other lines in a proportional fashion. A line that appears to be twice as long as another line would be assigned a number twice as large. Similarly, a line that is half as long would be assigned a number one-half that of the other. Carefully consider the following lines and assign an appropriate number at the bottom of each." This task served two functions. It provided subjects with practice in making proportional judgments of length or metric stimuli. This practice helped to make subsequent judgments of the descriptors (non-metric stimuli) somewhat easier. This task also permitted the calculation of a power function for group magnitude estimation regression effects.[13]

The second task was the assignment of magnitude estimates to the 19 anxiety descriptors presented in a randomly ordered list. The subjects were given the following instructions: "A number of words are listed below. Your task is to assign a number to each word such that the word is considered in relation to the other words in a proportional fashion. A word that is twice as great as another would be assigned a number twice as large. Similarly a word that is half as great would be assigned a number one-half that of the other. For example, consider the words Mild, Moderate, Strong. If Strong were considered twice as great as Moderate, you would give it a number twice as large as that you would assign to Moderate. If Mild were considered one-half as great as Moderate, it would receive half the numerical value. Carefully consider the following words and assign an appropriate number to each. Use whole numbers."

Mean log magnitude estimates for line lengths and descriptors were obtained for the group. Since the choice of modulus (or number set) was arbitrary, the responses were standardized to reduce inter-subject variability.[14] The predicted power function exponent for line length judgments was 1.0; the obtained power function exponent was .93. The difference represents "regression bias" caused by *S*s contracting the response continuum they were using.[13] This contraction in estimating line lengths was then used to adjust the magnitude estimates for descriptors as follows:

$$Rw = mw^{1/m}$$

where Rw = the relative magnitude of a given word; mw = the geometric mean magnitude estimate for the word; and m = the power function exponent obtained from the line estimation task. Such correction results in "bias-free" relative magnitudes.[13] The relative magnitudes of the 19 descriptors are given in Table 1 where the ratio of largest to smallest magnitudes is 27 to 1.

The list of 19 descriptors was reduced to 11 by combining two sets and eliminating 6 other descriptors. The criteria for reduction were a function of closeness of adjacent descriptors and our own judgments of the clarity of the terms. The revised list of 11 descriptors were again presented to the 28 subjects who repeated the magnitude estimation task two

TABLE 1. Relative Magnitude Estimates of the Descriptors from the Initial Scaling.

	Group 1 (N=28)		
Descriptor	Session 1	Session 2	Group 2 (N=29)
Terrified	82.6	80.8	32.3
Panicked	66.9	73.2	30.1
Scared	50.1	—	—
Afraid	49.3	41.4	14.0
Fearful	42.5	39.1	12.2
Distressed	38.1	40.1	14.2
Upset	31.1	22.6	11.8
Nervous	27.4	22.5	11.4
Tense	27.4		
Worried	27.0	—	—
Jittery	22.7	—	—
Anxious	22.0	21.5	11.1
Somewhat nervous	16.6	—	—
A little nervous	15.8	12.8	5.5
Comfortable	6.7	4.4	2.5
At ease	5.6	—	—
Relaxed	4.9	3.2	1.5
Calm	4.0		
Placid	3.0	—	—

weeks after the first session. The results of this second set of ratings are also presented in Table 1. This time the ratio of largest to smallest magnitudes was 25 to 1.

Both the line length task and the 11 descriptor list were presented to a new group of 29 subjects from the same population as the first group, 18 men and 11 women between the ages of 20 and 61 (mean, 35), and the relative magnitudes were calculated. The power function exponent obtained from the line length judgments for this group was .99. The results for this second group are also presented in Table 1 and the ratio of highest to lowest magnitudes was 22 to 1.

Further Steps in Development of the Measure

The reliability of descriptor scales of this type is difficult to determine in the conventional fashion. While test-retest reliability might be obtained in an experimental situation, it is not very meaningful with a clinical response measure.[8] However, it is possible to determine the reliability of the meaning of the list of descriptors both within and between groups. The correlation between the relative magnitudes of the 11 descriptors for Group 1 from the first to the second sessions was .99. The correlation between the two groups was also .99. These results indicate high group reliability.

An issue related to reliability is that of objectivity of the descriptors defined as the extent of agreement among raters.[15] Operationally, objectivity may be defined in relation to how closely individual estimates match those of the group as well as the stability of individual estimates over time. Two sets of correlations were obtained for the 11 descriptors in Group 1: individual-individual correlations for the two sessions and individual-group correlations with the individual's data removed from the group means. Each set of correlations was transformed to *z* scores, corrected for size or number of descriptors, and mean corrected *z*s obtained which were converted to correlations and will be termed "average" *r*s.[16] Individual-individual correlations over time ranged from .81 to .99 with an average *r* of .93 (median *r* = .93). The individual-group correlations ranged from .77 to .99 with an average *r* of .93 (median *r* = .92). The individual-group correlations indicate a high degree of correspondence between individual estimates of the word values and those made by the group as a whole. The fact that both the between and within correlations are the same indicates a high degree of objectivity for the magnitude estimates. Furthermore, the individual-group correlations obtained from Group 2 ranged from .83 to .97 with an average *r* of .93 (median r = .94). These data indicate that the method of magnitude estimation produced a reliable and objective descriptor scale.

Final Item Selection and Descriptor Scaling

Again, judgments were made concerning the redundancy of some adjacent descriptors. Four descriptors (fearful, distressed, anxious and nervous) were eliminated; "tense" was combined with "upset" because of their similar magnitudes and the fact that they had been used together as a single category in previous work.[3] Since we intended to use the descriptors in a graphic rating scale, we were concerned about the relatively large distance between "afraid" and "panicked." Consequently, two modified descriptors were also added to the list — "very afraid" and "a little panicked."

The line judging task and the new list of 9 descriptors were presented to another group of 53 subjects, 28 men and 25 women between the ages of 22 and 61 (mean, 33), who were from the same population of university staff and students as the previous groups. The power function exponent obtained from the line judging task by this group was .93. The relative magnitudes of the descriptors are given in Table 2. The ratio of highest to lowest magnitudes was 21 to 1. The group modulus (or number set) is arbitrary since it depends on the numbers used by subjects in the "free scaling" approach to magnitude estimation used in the present study. Comparison of group values for the same descriptors in Tables 1 and 2 indicate this arbitrary nature. In order to standardize the range and easily round to the nearest whole number, each magnitude was divided by the smallest magnitude, multiplied by 2 and rounded to the nearest whole number. These revised relative magnitude values obtained from the group of 53 *S*s are also given in Table 2. These revised values are the ones used for the final scale.

Again, because of the close proximity between adjacent descriptors, "comfortable" and "a little panicked" were eliminated from the descriptor list. This elimination resulted in a final seven-descriptor

TABLE 2. Relative Magnitude Estimates of the Final Set of Nine Descriptors (N = 53).

Descriptors	Relative Magnitudes	Revised Relative Magnitudes*
Terrified	52.0	42
Panicked	41.4	33
Very afraid	32.2	26
Afraid	22.6	18
A little panicked	21.0	17
Tense, upset	15.7	13
A little nervous	9.8	8
Comfortable	3.6	3
Calm, relaxed	2.5	2

*Relative magnitudes each divided by smallest relative magnitude (2.5), multiplied by 2 and rounded to the nearest whole number.

scale. Each of the 53 subjects' 7 descriptor values were correlated with those of the group with the subject's values excluded from the group data. The individual-group correlations obtained ranged from .68 to .999 with only three falling below .91. The average r determined from the mean corrected z_r was .98 (median r = .97). These results indicate that individual and group norms for the anxiety descriptor magnitudes were highly related.

This final set of descriptor values was also checked for sex differences. The magnitude estimates for the 28 men and 25 women were compared for each of the seven descriptors. No significant (p = .05) differences were found. These results suggest that men and women ascribed the same relative values to the descriptors.

Generalization and Validity

It has been suggested that educational level or socioeconomic status (SES) may influence the meanings associated with verbal descriptors and also subjects' ability to quantify these meanings.[6] One of the criteria which we used in discarding one of two closely adjacent descriptors was meaning difficulty. When one of two descriptors appeared to be a more difficult vocabulary item, it was eliminated.

Nevertheless, we planned to use this scale with hospital outpatient dental patients who were clearly lower in educational attainment and SES than those subjects with whom the scale was developed. We did a further small investigation to check for an educational bias in our scale. The 7 descriptors were typed individually on small cards and presented in random order on a board to patients waiting for dental treatment in a hospital outpatient dental clinic waiting room. There were 34 patients, 14 blacks and 20 whites, 13 men and 21 women between the ages of 18 and 66 (mean, 36). Each patient was asked to rank the words in order from the least distressing to the most distressing. The mean ranks for the descriptors were obtained for this group and correlated with the relative magnitudes obtained from the group of 53 subjects yielding an r of .95. This result suggests that the descriptors had the same relative meaning for the patient group.

Finally, an attempt was made to provide a test of construct validity for the scale. Previous research had indicated that patients presenting for exodontia were more anxious than patients undergoing other procedures.[3] Consequently, the scale was used to compare a group of exodontia patients with those presenting for other treatments (mostly restorations and extensive cleaning). Each group was comprised of 31 patients, 13 men and 18 women each, who came to two private hospital outpatient dental clinics. The mean ages of the patients were 37 (s.d. = 14.74) for the exodontia group and 43 (s.d. = 14.82) for the "other" group. The difference in age was not significant.

The scale was presented to the patient immediately after completion of treatment. It was given as a vertical graphic rating scale 90 mm high with scale units equivalent to 2 mm each. The descriptors were appropriately listed next to the scale. See Fig. 1. Two illustrations were presented with instructions to "mark each of the thermometers below to indicate *how you felt today.*" One scale was labelled "Just before treatment" and the other was labelled "During treatment."

Only three of the 62 patients marked a response between descriptors on the scale. The raw score distributions for tha groups were either J-curves or rectangular. Because of this marked deviation from a normal curve distribution, Mann-Whitney U-tests were used to test differences between the groups. The group medians and Us (Table 3) show that exodontia patients rated themselves as more anxious than patients undergoing other procedures both before and during treatment.

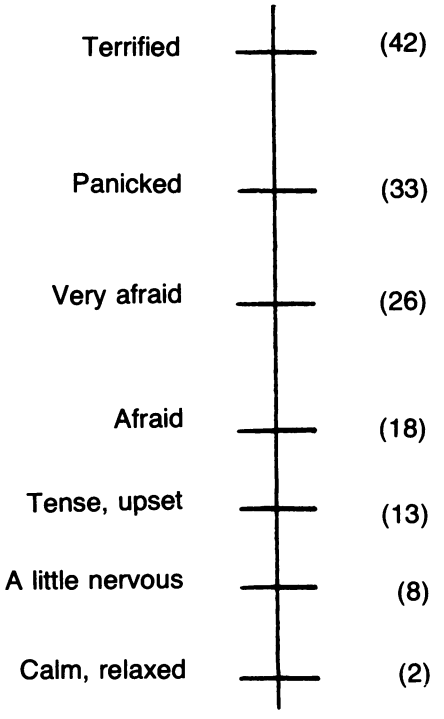

Fig. 1 — The seven-descriptor anxiety scale in graphic form. The numbers in parentheses indicate the scale values of the descriptors but were *not* shown on the scale as presented to subjects.

TABLE 3. Medians and U-test Values for the Anxiety Measures of the Two Patient Groups.

Anxiety Measure	Group		U
	Exodontia	Other Treatment	
Before treatment	13	8	331.0*
During treatment	8	2	299.0**

*$p < .05$
**$p < .01$

Since one of the major uses of an anxiety response scale would be to assess change in response, we conducted analyses on the two sets of measures as change scores. The differences between the before and during treatment anxiety measures were calculated for each of the two groups of patients. These "difference scores" more closely approximated normal distributions but were still somewhat skewed. Student *t*-tests were calculated for each group separately and also for the groups combined. The mean difference for the exodontia group was 6.74 ($s.d.$ = 15.18, t = 2.47, df = 30, p = .02) which was significantly different from zero. The mean difference for the "other procedures" group was 1.84 ($s.d.$ = 8.45, t = 1.21, df = 30, p = .23) which was *not* significantly different from zero. When both groups were combined, the mean difference between before and during treatment measures was 4.29 ($s.d.$ =12.43, t = 2.72, df =61, p = .009) which was significant. These results suggest that the anxiety response scale can detect differences in patient response under conditions in which variability of response is high. That variability would likely be reduced in carefully controlled experimental designs and would increase sensitivity of the scale.

Discussion and Conclusions

The anxiety response scale developed here appears to be reliable in the sense that the descriptors have approximately the same relative magnitudes when obtained from three groups of subjects. The relative meaning appears to be consistent in two widely divergent population samples. Finally, an initial attempt to demonstrate construct validity seemed successful. The fact that most subjects responding to the scale chose to place their ratings at the descriptor points is consistent with previous research using graphic rating scales.[9,17] These findings also suggest that the scale *may be* just as useful when presented as a decriptor list in which subjects are asked to pick the "one" that best describes their feelings.

The results obtained in this research support the development of a practical and versatile scale that shows promise as a reliable and valid measure for anxiety response. The findings for the various groups indicate that the relative meaning of the descriptors is stable and more closely approximates a real perceptual metric than previously used anxiety scales. However, a number of issues remain to be evaluated. The context of scale presentation is one such issue. Does the scale provide similar results when administered following the anxiety provoking event (retrospectively) as it might if given during the event? Such an issue is relevant in terms of interrupting treatment procedures, for example, to make an assessment.

Normative data is needed to assess the likelihood of differences in response by different population subgroups in a wide variety of clinical applications. Further studies of reliability and validity in a variety of experimental and clinical contexts are also needed.

Finally, it may be noted that the anxiety scale developed in this research has demonstrated consistency of descriptor meaning in several groups of subjects. Group descriptor magnitudes predicted those of most individuals with a high degree of accuracy. The scale appears to have the appropriate properties of an interval scale. The potential increase in measurement sensitivity afforded by such a scale remains to be demonstrated. Hopefully, the scale will prove to be useful in assessing the effects of behavioral and pharmacological interventions designed to reduce anxiety in patients.

Acknowledgement

This research was supported, in part, by Research Grant No. DE 04494 from the National Institute of Dental Research. We thank the staff and patients of the dental clinics at Buffalo General and Deaconess Hospitals for their cooperation in this study.

References

1. Spielberger CD, Gorsuch RL, Lushene RE: Manual for the State-Trait Anxiety Inventory, Palo Alto, California, Consult Psychol Press, 1970.
2. Zuckerman M and Lubin B: Manual for the Multiple Affect Adjective Check List. San Diego, California: Educa and Industr Test Serv, 1965.
3. Corah NL, O'Shea RM, Bissell GD: The dentist-patient relationship: perceptions by patients of dentist behavior in relation to satisfaction and anxiety. J Am Dental Assoc 111:443-446, 1985.
4. Lazarus RS: Psychological Stress and the Coping Process, New York, McGraw-Hill, 1966.
5. Tursky B: The development of a pain perception profile: a psychophysical approach. In: Weisenberg M and Tursky B, eds., Pain — New Perspect in Ther and Res. New York, Plenum Press, pp. 171-194, 1976.
6. Gracely RH, McGrath PA, Dubner R: Ratio scales of sensory and affective verbal pain descriptors. Pain 5:5-18, 1978.
7. McGrath PA: Measurement issues in research on dental fears and anxiety. Anesth Progr 33:43-46, 1986.
8. Corah NL: Methodological needs and behavioral research with adult dental patients. Anesth Progr 33:46-49, 1986.
9. Heft MW and Parker SR: An experimental basis for revising the graphic rating scale for pain. Pain 19:153-161, 1984.
10. Stevens SS: On the psychophysical law. Psychol Rev 64:153-181, 1957.
11. Watson DW and Tellegen A: Toward a consensual structure of mood. Psychol Bull 98:219-235, 1985.
12. Izard CE: Anxiety: a variable combination of interacting fundamental emotions. In: Spielberger CD, ed., Anxiety: Current Trends in Theory and Research, Vol. I, New York, Academic Press, pp. 51-106, 1972.
13. Stevens SS and Greenbaum HB: Regression effect in psychophysical judgment. Percept Psychophys 1:439-446, 1966.
14. Lane HL, Catania HC, Stevens SS: Voice level: autophonic scale, perceived loudness, and effects of sidetone. J Acoust Soc Amer 33:160-167, 1961.
15. Guilford JP: Psychometric Methods (2nd ed.). New York, McGraw-Hill, pp. 251-253, 1954.
16. Du Bois PH: An Introduction to Psychological Statistics, New York, Harper & Row, pp. 339-340, 1965.
17. Scott J and Huskisson EC: Graphic representations of pain. Pain 2:175-184, 1976.