PEDIATRIC DENTISTRY/Copyright © 1980 by
The American Academy of Pedodontics/Vol. 2, No. 3

Interval rating scales for children's dental anxiety and uncooperative behavior*

Larry L. Venham, D.D.S., M.S., Ph.D.
Elise Gaulin-Kremer, A.B., Ph.D.
Edward Munster, B.S., M.S.
Dana Bengston-Audia, B.A.
Jan Cohan, B.A.

Biographical Data:

Larry L. Venham is Associate Professor, Department of Pediatric Dentistry, School of Dental Medicine, University of Connecticut Health Center, Farmington, CT.

Elise Gaulin-Kremer is Research Associate, Department of Pediatric Dentistry, School of Dental Medicine, University of Connecticut Health Center, Farmington, CT.

Edward Munster is a Statistician at Pfizer Corporation, Groton, CT.

Dana Bengston-Audia is a Budget Management Officer, University of Hartford, West Hartford, CT.

Jan Cohan is a Research Assistant II, Department of Pediatric Dentistry, School of Dental Medicine, University of Connecticut Health Center, Farmington, CT.

Accepted: March 12, 1980

* This research was supported by USPHS Grant DE04409 from the National Institute of Dental Research.

Requests for reprints may be sent to Larry L. Venham, Department of Pediatric Dentistry, School of Dental Medicine, University of Connecticut Health Center, Farmington, CT 06032.

Abstract

This report describes two rating scales, an Anxiety Scale and an Uncooperative Behavior Scale, developed to assess children's responses to dental stress. The rating procedure is reliable, valid, and easily integrated into clinical or research activities. In the present study, the Method of Paired Comparisons was used to further validate the scales, and to establish interval scaling properties. The rating scales potentially provide a much needed standard tool for behavioral assessment in pedodontic research.

Introduction

The child's emotional and behavioral response in the dental chair is a matter of serious concern to both practitioners and researchers in the pedodontic field. A youngster's response to dental treatment may greatly facilitate or impede the course and quality of treatment provided. Even more significantly, the reactions of young children may presage emerging perceptions of and attitudes toward dental care which affect

their later propensity to follow preventive routines and to accept restorative care. It is therefore appropriate that clinicians and researchers are intensifying their efforts to understand and optimize the young child's response to dental treatment.[1] Pedodontic researchers are exhibiting a growing interest in evaluating techniques for managing the young child's negative behavior and reducing adverse emotional responses to treatment.[2]

One essential component of pedodontic research program addressing these concerns is a reliable and well-validated tool for assessing the child's response to dental treatment. Unfortunately, few approaches have yet emerged to objectively observe and quantify the child's response in the dental setting. Therefore, the development of valid and reliable child assessment techniques is a major prerequisite for refining pedodontic behavioral research and ultimately for improving the clinical management of children's dental anxiety and behavior.

Behavioral rating scales have been the most commonly used indices of children's responses to dentistry. An example is the widely used Frankl Scale,[3] in which the child's reaction to dental treatment is rated on a four-point scale ranging from definitely negative to definitely positive. The advantage of rating scales include ease of administration and conceptualization.[4] The rater uses the trait as an organizing concept which allows him or her to select relevant cues and to superimpose a dimension on the subject's behavior. Thus, the overall impression afforded by a rating may bring out a quality or unity to the child's behavior that a tabulation of discrete behaviors may be unable to reveal. In assigning ratings, the rater is able to take account of individual response styles in behavior and to consider infrequent but significant behaviors. Therefore, the rating represents a high degree of abstraction from the basic observational process.

In assessing the usefulness of a rating technique, several features are important including reliability, validity and measurement properties. Reliability reflects the extent to which a scale is consistent and repeatable in assessing a trait. A highly reliable scale will provide similar results when used to assess the same sample of behavior at different times (test-retest reliability) or by different raters (inter-observer agreement). Validity reflects the degree to which a scale actually measures what it purports to measure. A valid scale therefore is one which accurately and specifically measures the unique trait it was designed to measure. Measurement properties[5] refer to the nature of scale categories and the relationships among scale points.

Most rating scales exhibit an ordinal level of measurement. Ordinal measurement is essentially an ordering or ranking technique, in which subjects assigned to a particular scale category are judged "greater" or "higher" on the relevant dimension than subjects assigned to the next lower scale point. With an ordinal scale, the distance between adjacent scale points remains unspecified, so that it cannot be determined how much "higher" one scale point is than another.

To achieve truly quantitative measurement, an interval scale is required. In such a scale, the rank-order relationship exists; additionally the distances between any two neighboring scale points are of known size. It then becomes possible to specify whether a five-category scale is best treated as a zero, one, two, three, four distribution or, for example, a zero, one, 100, 200, 1000 distribution. Reliability, validity, and measurement level of a rating technique can only be determined through empirical study of the scale's performance in trial situations.

Difficulties which have arisen in using rating scales typically involve problems related to reliability, validity, and measurement level. The major drawbacks of the rating procedure lie in possible undetected bias and distortion of data. The scorer weighs the evidence on which the rating is based on a complex manner which is not easily specified, standardized, or objectified. Subjective factors, such as the ego-involvement or expectations of the rater may lead to scoring bias. To the degree that the rating process remains unstandardized and subjective, reliability and validity may be compromised. The possibility of rater bias, and the associated reliability and validity problems, are maximized when scaling categories are not defined in clearcut behavioral terms. In the commonly used Frankl Scale, an effort was made to define scale points behaviorally. Nonetheless, the definitions remain sufficiently subjective that latitude exists for each rater to interpret the scale categories somewhat idiosyncratically.

A final drawback of rating scales has historically been a dearth of appropriate powerful and flexible data analytic techniques. Parametric statistical techniques, such as the t-test, analysis of variance and multivariate analysis, have rigorous assumptions underlying their use. Assumptions requiring continuous, normally distributed data mitigate against the use of parametric statistics with rating scale data. Unfortunately, the non-parametric techniques traditionally recommended for rating data have sometimes proven less efficient or flexible; with complex multivariate designs, appropriate non-parametric techniques have often been unavailable.

Fortunately, efficient, flexible and sophisticated statistical techniques, such as multidimensional contingency table analysis for categorical data, have re-

Table 1. Rating scales for anxiety and uncooperative behavior

Anxiety rating scale
0. Relaxed, smiling, willing and able to converse.
1. Uneasy, concerned. During stressful procedure may protest briefly and quietly to indicate discomfort. Hands remain down or partially raised to signal discomfort. Child willing and able to interpret experience as requested. Tense facial expression, may have tears in eyes.
2. Child appears scared. Tone of voice, questions and answers reflect anxiety. During stressful procedure, verbal protest, (quiet) crying, hands tense and raised, (not interfering much — may touch dentist's hand or instrument, but not pull at it). Child interprets situation with reasonable accuracy and continues to work to cope with his/her anxiety.
3. Shows reluctance to enter situation, difficulty in correctly assessing situational threat. Pronounced verbal protest, crying. Using hands to try to stop procedure. Protest out of proportion to threat. Copes with situation with great reluctance.
4. Anxiety interferes with ability to assess situation. General crying not related to treatment. More prominent body movement. Child can be reached through verbal communication, and eventually with reluctance and great effort he or she begins the work of coping with the threat.
5. Child out of contact with the reality of the threat. General loud crying, unable to listen to verbal communication, makes no effort to cope with threat. Actively involved in escape behavior. Physical restraint required.
Behavior rating scale
0. Total cooperation, best possible working conditions, no crying or physical protest.
1. Mild, soft verbal protest or (quiet) crying as a signal of discomfort, but does not obstruct progress. Appropriate behavior for procedure, i.e., slight start at injection, "ow" during drilling if hurting, etc.
2. Protest more prominent. Both crying and hand signals. May move head around making it hard to administer treatment. Protest more distracting and troublesome. However, child still complies with request to cooperate.
3. Protest presents real problem to dentist. Complies with demands reluctantly, requiring extra effort by dentist. Body movement.
4. Protest disrupts procedure, requires that all of the dentist's attention be directed toward the child's behavior. Compliance eventually achieved after considerable effort by dentist, but without much actual physical restraint. (May require holding child's hands or the like to start). More prominent body movement.
5. General protest, no compliance or cooperation. Physical restaint is required.

cently become available and provide a powerful unified approach for analyzing ratings.[6] Because of the generality of the approach, the researcher has considerably more latitude in developing a research design and in testing hypotheses precisely geared to specific data than would be possible using traditional non-parametric techniques. This statistical methodology enables contingency table analysis to be performed within any type of analysis of variance or covariance as well as within multivariate designs. These techniques can be used with ordinal data; however, the additional information, such as linear trends, can be obtained when the scale has interval measurement properties.

Efforts in our laboratory have been directed toward developing child assessment scales that surmount the difficulties related to reliability, validity and measurement properties. Two scales were developed to evaluate the child's response to dental treatment, an Anxi-

Table 2. Paired-comparison judgments of dental anxiety

Proportion matrix for six stimulus children judged on dental anxiety

	0	1	2	3	4	5
0	0.500	0.929	1.000	1.000	1.000	1.000
1	0.071	0.500	0.929	1.000	1.000	1.000
2	0.000	0.071	0.500	0.929	1.000	1.000
3	0.000	0.000	0.071	0.500	0.929	0.929
4	0.000	0.000	0.000	0.071	0.500	1.000
5	0.000	0.000	0.000	0.071	0.000	0.500

Transformed matrix for six stimulus children judged on dental anxiety

	0	1	2	3	4	5
0	.00	1.47	—	—	—	—
1	-1.47	0	1.47	—	—	—
2	—	-1.47	0	1.47	—	—
3	—	—	-1.47	0	1.47	1.47
4	—	—	—	-1.47	0	—
5	—	—	—	-1.47	—	0

ety Rating Scale and an Uncooperative Behavior Rating Scale. Each is a six-point scale, with scale points anchored in objective, specific and readily-observable behavior. These scales have been used in several major research projects in our laboratory.[7-9] Consistently high inter-rater agreement has been demonstrated.[8] Three judges were asked to independently view and rate videotapes of children's dental visits. Inter-observer reliability coefficients, computed by correlating the three judges' ratings, ranged from .78 to .98. Validity has been assessed by comparing judges' ratings of anxiety and uncooperative behavior to several self-report and physiological indices of children's responses to dental treatment. Significant correlations emerged which supported the validity of our rating technique; that is, ratings indicating greater anxiety and uncooperative behavior were associated with more highly stressed physiological and self-report measures.

The present study was undertaken to further validate the rating system and to establish interval measurement properties for the scales.

Methods and Materials

The six-point scales used to rate anxiety and uncooperative behavior are shown in Table 1. The rating scales were evaluated using the Method of Paired Comparisons, described by Guilford.[10] In this method,

Matrix deriving distances between neighboring stimuli judged for dental anxiety

	Distance 0-1	Distance 1-2	Distance 2-3	Distance 3-4	Distance 4-5
0	1.47	—	—	—	—
1	1.47	1.47	—	—	—
2	—	1.47	1.47	—	—
3	—	—	1.47	1.47	0
4	—	—	—	1.47	—
5	—	—	—	—	—
X distance	1.47	1.47	1.47	1.47	0

a series of stimuli, presumably differing on some psychological dimension, is chosen to be evaluated on a psychological scale. These stimuli are presented in all possible pairs to an observer, and the observer is asked to produce a comparative judgment for each pair (i.e., to judge which stimulus is "higher" in some specific dimension defined by the scale). The judgments are tabulated into a matrix indicating the proportion of times each stimulus is judged higher on the scale than each other stimulus. Statistical transformations permit estimations to be made of the linear psychological distances between stimuli. Estimated distances between stimuli can then be used to assign each stimulus a single empirically-established numerical value on a linear scale.

The stimuli were chosen from videotapes of children undergoing dental treatment. The videotapes had been recorded as part of a longitudinal study of children's responses to sequential dental visits.[8] Using the rating scales shown in Table 1, scores for anxiety and uncooperative behavior had earlier been assigned to the videotapes by highly trained raters. These raters reviewed the videotapes to select stimuli for the present study. Each stimulus was a 45-second videotape segment depicting a child's response during a dental procedure (mirror-and-explorer examination, prophylaxis, cavity preparation or fluoride application). Six segments were chosen for paired comparisons of anxiety, and six different segments for paired comparisons of uncooperative behavior. The criterion for selecting a segment was the raters' judgment that the segment well represented the scale point which had been assigned to it. No child was depicted in more than one of the 12 segments. The six "anxiety" segments were then transcribed onto a videotape. Each stimulus was paired with every other stimulus, providing a total of 15 paired comparisons. Similarly, a videotape was prepared for the 15 paired comparisons of uncooperative behavior.

Naive observers were recruited to view the videotape segments and make paired comparison judgments. Fourteen observers provided paired comparisons of anxiety, while 13 independent observers made paired comparisons of uncooperative behavior. The observers were personnel employed in diverse teaching, laboratory and clerical positions at the University Health Center.

Observers judging the children's relative anxiety were given the Anxiety Rating Scale to read. This reading gave all observers a common set of criteria on which to base anxiety judgments. The rating scale was then removed and the observer shown the first pair of videotape segments. He or she was then asked "Which child appeared more anxious, the first or the second?" The remaining fourteen pairs were then shown, and relative anxiety judgments obtained for each pair.

A similar procedure was followed for the "Uncooperative behavior" judgments. After reading the Uncooperative Behavior Rating Scale, the first stimulus pair was presented. The observer was asked to

Table 3. Paired-comparison judgments of uncooperative behavior

Proportion matrix for six stimulus children judged on dental anxiety

	0	1	2	3	4	5
0	0.500	1.000	1.000	1.000	1.000	1.000
1	0.000	0.500	0.923	1.000	1.000	1.000
2	0.000	0.077	0.500	0.692	1.000	1.000
3	0.000	0.000	0.308	0.500	0.846	0.923
4	0.000	0.000	0.000	0.154	0.500	0.692
5	0.000	0.000	0.000	0.077	0.308	0.500

Transformed matrix for six stimulus children judged for uncooperative behavior

	0	1	2	3	4	5
0	0.0	—	—	—	—	—
1	—	0	1.43	—	—	—
2	—	-1.43	0.0	.50	—	—
3	—	—	-.50	0.0	1.02	1.47
4	—	—	—	-1.02	.00	.50
5	—	—	—	-1.47	-.50	0.0

judge "Which child behaved more cooperatively, the first or the second?" The remaining fourteen pairs were then presented and judgments obtained. During each presentation, the observer was required to view the entire 45-second stimulus segment.

The ordinal position within stimulus pairs was counterbalanced to control for possible response sets; that is, the child judged more anxious (or more uncooperative) in the original ratings were presented first in approximately half the pairs and second in approximately half the pairs. Additionally, the order of presentation of the 15 paired comparisons was counterbalanced across observers.

Results

A matrix was constructed to summarize all the observers' relative anxiety judgments (Table 2, upper matrix). The six columns and rows represent the six "anxiety" stimuli; the row and column headings reflect the rating scale values originally assigned to the respective stimuli. Each cell entry indicates the proportion of times the column stimulus was judged "more anxious" than the row stimulus. Cell entries on the diagonal were assigned the expected proportion of 0.500. A large number of 0.000 and 1.000 values were obtained. These proportions reflect a precise corre-

Matrix deriving distances between neighboring stimuli judged for dental anxiety

	Distance 0-1	Distance 1-2	Distance 2-3	Distance 3-4	Distance 4-5
0	—	—	—	—	—
1	—	1.43	—	—	—
2	—	1.43	.5	—	—
3	—	—	.5	1.02	.45
4	—	—	—	1.02	.50
5	—	—	—	.97	.50
X distance	—	1.43	.5	1.01	.48

spondence between the observers' ranking judgments and the original ratings made by highly trained judges.

The center matrix was produced by transforming the proportions into corresponding deviate values, using a normal curve table. Numerical deviate values could not be assigned for proportions which reached 0.000 or 1.000, since these proportions theoretically reflect highly improbable events whose deviate value is indeterminately large. Therefore, all 0.000 and 1.000 proportions had to be omitted from the matrix, leaving a number of cell vacancies.[11]

The lower matrix documents the derivation of linear distance scores. Each cell entry is a difference score derived by subtracting the deviate value of a particular column and row in the center matrix from the deviate value of the right adjacent column in the same row. Each difference score for a neighboring pair of stimuli provides an estimate of the linear psychological distance between the respective stimuli.

By averaging the distance estimates within a column, a mean is derived which provides the best possible estimate of the psychological distance between the respective neighboring scale categories. Because the number of distance estimates was substantially attenuated by the large number of cell vacancies, means were based on only one or two estimates and probably suffer some artifactual instability. Nonetheless, the empirical distances between neighboring pairs of stimuli are remarkably constant. Furthermore, the distances between adjacent scale categories correspond reasonably closely to the one-point difference arbitrarily assumed in the original zero through five rating scale. These findings justify the treatment of the original zero through five scale as an interval scale whose neighboring categories are approximately equidistant.

Similar matrices were constructed to summarize the observers' relative judgments of uncooperative behavior. These matrices are illustrated in Table 3. The large number of 0.00 and 1.00 proportions again attest to the high degree of correspondence between the original rating scale scores and the untrained observers' judgments. The bottom row of the lower matrix shows the empirically-derived linear psychological distances between neighboring stimuli. Again the adequacy of the estimates is attenuated by the large number of vacant cells created by the uniform accuracy of our observers' judgments. In fact, one distance cannot even be estimated, due to an empty column. Nonetheless, distances that can be estimated exhibit adequate consistency and satisfactorily approximate the one-point difference between scale categories assumed in the original zero through five scale. These data again suggest that operationally the Uncooperative Behavior Rating Scale can be considered an interval scale whose neighboring points are approximately equidistant.

Discussion

Two valid and reliable rating scales have been developed in our research laboratory and have proven useful for assessing children's responses to dental stress. In the present report, a paired comparison technique was used to establish interval level properties and accurate numerical scaling for the rating procedures.

The paired comparison data provided estimates of the linear psychological distances between neighboring scale points. The paired comparisons indicate that both the Anxiety and Uncooperative Behavior Rating Scales provide interval level measurement. The data suggest that adjacent scale points can operationally be considered equidistant and that the original zero through five scale labels satisfactorily approximate the empirical scale values.

Close and highly consistent agreement was found between the original scores assigned to the videotape segments and the comparative judgments empirically derived from the observers. The high degree of inter-observer agreement is particularly notable since the observers were essentially untrained in assessing children's responses in the dental setting. This impressive consensus enhances our confidence that the scale points and their behavioral labels accurately capture the essence and variable manifestations of "anxiety" and "uncooperative behavior" in young children.

Using these scales, recent statistical approaches[6] enable data obtained to be entered into multidimensional contingency tables and analyzed using powerful parametric techniques, such as the t-test, analysis of variance and planned contrast procedures. Data can also be entered into more sophisticated analyses involving multivariate techniques, such as analysis of covariance, multiple regression, trend analysis, and multivariate analysis of variance. Important ramifications can include an enhanced flexibility of experimental design and hypothesis testing, a reduction in the required number of subjects, and an increase in the power and efficiency of statistical analysis. In particular, access to multivariate techniques can greatly expand the range of hypotheses that can be explored and the experimental protocols that can be exploited.

Conclusion

In summary, the rating scales present a number of significant advantages. Raters can be readily trained to use the scales with a high degree of inter-observer reliability. Because the rating procedure is simple, quick, and non-intrusive, it is easily integrated into ongoing clinical activities or research designs. The data generated using these two scales can appropriately be analyzed using powerful and flexible statistical techniques. In a field sorely lacking well-standardized behavioral assessment tools, the emergence of behavior rating scales with documented reliability, validity, and measurement properties is noteworthy. Such scales potentially provide standard tools which might enhance the comparability of findings from diverse pedodontic research laboratories.

References

1. Chambers, D.: "Behavioral Management Techniques for Pediatric Dentists: An Embarassment of Riches," *J Dent Child,* 44:30-34, 1977.
2. Melamed, B.: "Behavioral Approaches to Fear in the Dental Setting," *Progr in Behav Modifi,* 7:171-203, 1979.
3. Frankl, S., Shiere, F., and Fogels, H.: "Should the Parent Remain with the Child in the Dental Operatory?" *J Dent Child,* 29:150-163, 1962.
4. Lytton, H.: "Three Approaches to the Study of Parent-Child Interaction. Ethological, Interview and Experimental," *J Child Psychol Psychiat,* 14:1-17, 1973.
5. Siegal, S.: *Non-parametric Statistics for the Behavioral Sciences,* New York: McGraw-Hill, 1956, pp. 21-30.
6. Grizzle, J., Starmer, C., and Koch, G.: "Analysis of Categorical Data by Linear Models," *Biometrics,* 25:489-504, 1969.
7. Venham, L., Bengston-Audia, D., and Cipes, M.: "Parent's Presence and the Child's Response to Dental Stress," *J Dent Child,* 45:213-217, 1978.
8. Venham, L., Bengston-Audia, D., and Cipes, M.: "Children's Response to Sequential Dental Visits," *J Dent Res,* 56:454-459, 1977.
9. Sonnenberg, E. and Venham, L.: "Human Figure Drawings as a Measure of the Child's Response to Dental Visits," *J Dent Child,* 44:438-442, 1977.
10. Guilford, J.: *Psychometric Methods,* New York: McGraw-Hill, 1954, pp. 154-177.
11. Torgerson, W.: *Theory and Methods of Scaling,* New York: John Wiley and Sons, Inc., 1958, pp. 191-192.