PEDIATRIC DENTISTRY/Copyright© 1981 by
The American Academy of Pedodontics/Vol. 3, No. 3

Measuring selected disruptive behaviors of the 36- to 60-month-old patient. Part I: Development and assessment of a rating scale

William L. Chambers, DDS, MS
Henry W. Fields, DDS, MS, MSD
J. Bernard Machen, DDS, MS, PhD

Abstract

The purpose of Part I of the study is to develop an objective method for assessing the amount of disruptive child behavior for children 36-60 months of age undergoing dental treatment. Using videotaped dental visits and recording behavior with an Esterling-Angus event recorder, an initial attempt to utilize a scale with eight behaviors suggested by the literature was evaluated. The final scale focused on four behaviors (high-hand and leg movement; crying; and oral physical resistance) which occurred more than 5% of the time and exhibited interrater reliability from 83-91%. The method was stable across different raters who were trained independently. Correlations between the behaviors and the Frankl scale were in the expected directions.

Introduction

Assessment of child behavior in the dental setting has been attempted by numerous investigators and has resulted in several clinical research tools for measuring child behavior, most of which were developed to assess a specific treatment procedure or condition. Some investigators used physiologic measures[1,2] but these have not been subsequently correlated with observed behavior. Many authors used general systems of classification for observed behavior but these approaches did not quantify in any detail the behavior and rarely was reliability reported.[3-10]

There have been a few investigators who quantified observed behavior and reported reliability.[11-16] While these latter approaches have helped make behavioral quantification more objective, there were several factors that reduced their usefulness as clinical research tools. These complicating factors included too much complexity,[11,14] selectivity in recording behavior,[15] insufficient behavior quantification,[12] subjective weighting,[13] and requiring judgmental interpretation.[16]

Even with these methodological problems the quantitatively oriented studies included many of the attributes that should be included in any scale designed to produce clinically relevant data. Negative behavior was easier to identify than positive behavior and was useful for discriminating between children in the dental setting. Since total reaction to the dental environment includes verbal and non-verbal behavior, both should be quantified. A scale should be uncomplicated for easy implementation and to facilitate high inter- and intrarater reliability. Validity of the scale should be established. The scale should be consistent when used by different dentists under similar conditions. No existing scale contained all of these attributes.

It was the purpose of Part I of this study to develop an objective method of quantifying child behavior in the dental setting that conformed to these attributes. This was accomplished by identifying and defining negative behavior based on previous literature; obtaining a sample of representative behavior; quantifying this behavior; and assessing the reliability and validity of this quantification.

Methods and Materials

Scale Category Identification. The categories used in this study and their operational definitions are shown in Figure 1. This scale is identified as the North Carolina Behavior Rating Scale (NCBRS). Only observable disruptive behavior was chosen as the focus of observation for several reasons. First, disruptive behavior impedes efficient treatment since it demands the attention of the dentist. Negative behavior has

Accepted: January 14, 1981

This study was supported by NIH Grant #05333 from the Division of Research Facilities and Resources.

also been shown in previous investigations to discriminate cooperative from noncooperative children and to provide an objective behavior profile.[12,13] Negative behavior is easier to define and therefore contributes to high reliability.[12-14] Finally, a positive relationship has been demonstrated between disruptive behavior and the Frankl scale.[14] The original eight negative behavior categories in Figure 1 have been consistently mentioned in the literature.

Figure 1. North Carolina Behavior Rating Scale.

Behavior	Operational Definition
Head Movement	Movements of the head such as lifting, turning or twisting.
High-Hands	Hands or hand above level of arm pits or arm's extension angle less than 90° angle (on or off body) — physical restraint used.
Low-Hands	Hands or hand below arm's extension angle greater than 90° angle (only off body or chair arm).
Torso-Trunk Movement	Movement from the torso-trunk area or from the waist up. *Movements* include lifting, turning, twisting, or arching or torso-trunk area — physical restraint used.
Leg Movement	*Any* movement of legs such as kicking, lifting, postural change, etc., to get to prone position — physical restraint used.
Crying Protest	*Pure* crying, or screaming, whining, or sobbing; *or* — crying, etc. while asking for parents, asking to stop, etc. (If any doubt over fact that a verbal protest was crying or noncrying, it should be scored as crying protest.)
Verbal Protest (Noncrying)	Noncrying verbalizations such as requests to stop, calling for Mom, defiant verbalizations, asking questions, etc.
Oral-Physical Resistance	Choking, gagging, coughing, mouth closing, spitting, vomiting, etc. Mouth closing — rate by oral request to open in any treatment segment Refusal to open mouth — rate by oral request to open in any treatment segment Physical restraint used (i.e., mouth prop)

Subjects. Forty children, ages 36 to 60 months, who met the following six requirements were selected from an available sample.

Each child:

1. had no previous dental experience,
2. was mentally and physically healthy so that no unusual treatment procedures were necessary,
3. spoke and understood English,
4. had no siblings in the study, and
5. needed at least two Class I restorations.

These requirements provided a sample of predominantly middle-class children as determined by the Hollingshead analysis. The sample consisted of 22 males and 18 females: 37 Caucasian and 3 minority patients, with a mean age of 46.9 months. This age group requires more attention than any other by the dentist and much of the research on child behavior concentrates on this group.

Appointment Procedures. Each patient was treated at three dental visits approximately one week apart in the same operatory by one of three dentists and a trained dental assistant. The same dentist completed all treatment for a patient. All children used a reception room and operatory not associated with the teaching clinics. The three-bay operatory was used exclusively by this project during the treatment periods in order to simulate a more realistic private practice environment. During the initial appointment an examination of hard and soft tissues, bitewing radiographs, a rubber cup prophylaxis and topical fluoride treatment utilizing two disposable foam trays were performed. Bitewing radiographs were simulated if no interproximal contacts existed. At each of two subsequent restorative visits one Class I amalgam restoration was placed utilizing local anesthesia and a rubber dam.

All appointments were videotaped from behind a curtain with a black and white three quarter inch system. Each appointment followed a treatment outline that incorporated the "tell-show-do" approach and a standardized dialogue. Voice control and minimal restraint were used only when treatment progress was repeatedly hindered. Any further behavior management measures eliminated the patient from the study. This was necessary in order to preserve the behavior observed as a function of the child and environmental interaction. Aggressive behavior management techniques would introduce another variable into the equation and make it more difficult to determine if the behavior is the result of the management technique or that usually encountered by the dentist.

Behavior Rating. Patient behavior was quantified using the Esterline Angus Event Recorder which was

connected to a keyboard with labeled keys. By pressing appropriate keys, raters recorded the duration of negative behavior and marked rating intervals on time-scaled chart paper. Analysis of the negative behaviors was accomplished from these paper tape records. Each rating session was limited to approximately two hours with at least one tape being used to monitor interrater reliability.

Rater Training. Two raters, both dentists, were trained to utilize the scale. Rater training consisted of three steps. First, the scale was explained and examples of the categories of negative behavior were viewed. The second step familiarized raters with the Esterline Angus Event Recorder, provided actual rating practice, and determined areas of disagreement between raters. The third step established interrater reliability. During this step the rater quantified the behavior of patients recorded on videotape but not included in the study. These three steps required approximately 30 minutes, 60 minutes, and 120 minutes respectively. Raters were separated by a screen so neither could observe the other while rating.

Assessment of Interrater Reliability. One rater was randomly selected as the primary rater and these observations were used to report all behavior data. To monitor interrater reliability during the study, 18 of 120 appointments were randomly selected for evaluation by both raters. If a decrease in interrater reliability occurred, a retraining session was instituted. This was done to maintain the high reliability initially found during training: not to create high reliability.

Using the time-scaled recordings of behavior, the percent of rater agreement for each scale category was calculated by dividing the number of eight-second interval agreements by the number of eight-second intervals recorded for that behavior during the appointment. Any category of behavior observed less than 5% of the appointment was eliminated because of low frequency of occurrence. Initial assessment of reliability was determined at the end of the training session.

Intrarater Reliability. To further evaluate the NCBR scale, the primary rater re-rated ten of the tapes. A paired t-test was used to compare total seconds of each negative behavior and by a Wilcoxon matched pair-sign rank test to compare the percent of behavior occurrence.

Assessment of Validity. Since reliability is a necessary consideration for validity, those behaviors with poor intrarater reliability and low frequency of occurrence were eliminated from considerations of validity, and from the NCBR scale.

Validity of the NCBR scale was assessed by three methods. To assess the ability to generalize across raters, two additional raters rated sixteen appointments from the study. These "validity" raters achieved initial training levels of interrater reliability similar to the "study" raters. The primary "study" rater and the primary "validity" rater were compared for interrater reliability. Total seconds of each behavior occurrence were compared using a paired t-test. Percent behavior occurrence was compared using the Wilcoxon matched pair-sign rank test.

In the second method of validity assessment 40 appointments from the study were rated using the Frankl scale. The percent occurrence of each behavior was correlated with the Frankl scale using a Spearman correlation.

Finally, the stability of negative behavior occurrence across dentists was measured by examining the occurrence of negative behavior in three separate samples of patients treated by three different dentists under similar conditions. The stability was assessed using the Kruskal-Wallis test. Significant results were further analyzed using a Wilcoxon rank sum test.

Results

Interrater Reliability. Interrater reliability for the entire study was calculated for 18 appointments randomly chosen throughout the study. Table 1 reports the total number of eight-second intervals rated, the number of agreements between raters and the interrater reliability for the eight categories of negative behavior initially evaluated in the study. Those behaviors with low frequency of occurrence, and interrater reliability less than 0.75 were eliminated from the scale and from further consideration. The scale now consists of high-hands, legs, crying and oral-physical resistance.

Intrarater Reliability. The total seconds of occurrence and percent occurrence for the ten appointments rated twice are shown in Table 2. There was no significant difference at the $p < 0.05$ level.

Scale Validity. Results of the three methods of scale validity assessment were consistent. Comparison of primary raters (study versus validity) for time and percent of negative behavior occurrence is shown in Table 3. There were no significant differences at the $p < 0.05$ level for either total seconds or percent occurrence.

Two negative behaviors, crying and oral-physical resistance, were significantly correlated with the total Frankl score (Table 4). The other observed negative behaviors were negatively correlated with the total Frankl scores or demonstrated essentially no relationships.

The mean percent occurrence of negative behaviors for the three dentists are shown in Table 5. The Kruskal-Wallis test which assessed the stability of the behavior ratings across the three dentists treating 40 patients revealed no differences for high-hands, crying and oral-physical resistance. Leg movement on visit three was the only significant behavior (Table 6).

Table 1. Interrater reliability for eight behaviors.

	Total Number Intervals Rated	Number Agreements	Percent Agreement
Head	26	19	73
Torso-trunk	32	22	68
High-hands	495	447	90
Low-hands	138	101	73
Legs	607	558	91
Crying	439	391	89
Oral-physical resistance	276	231	83
Non-crying resistance	63	42	66

Discussion

Originally, there were eight negative behaviors included in the scale: crying, oral-physical resistance, high-hands, low-hands, torso-trunk movement, leg movement, head movement, and non-crying protests. The categories of head, low-hands, torso-trunk and non-crying protests that were eliminated from the scale were difficult to define and to delineate their beginning and end points. Green et al.[14] reported similar problems. These problems were compounded by a relatively low frequency of occurrence.

Interrater reliability is also critical to useful scale development. The four behaviors eliminated from the scale exhibited poor reliability — probably a result of low frequency of occurrence and the inability to provide accurate definitions.

Table 3. Comparison of primary "study" and "validity" raters.

Behavior	"Study" Rater	"Validity" Rater
High Hands		
Total Seconds	1068	1004
% Occurrence	11.9%	11.5%
Legs		
Total Seconds	2737	2890
% Occurrence	32.8%	34.4%
Crying		
Total Seconds	712	628
% Occurrence	6.3%	5.6%
Oral-Physical		
Total Seconds	185	197
% Occurrence	1.9%	2.1%

Table 2. Intrarater reliability for ten appointments.

Behavior	First Rating	Second Rating
High Hands		
Total Seconds	1068	1114
% Occurrence	11.7%	12.3%
Legs		
Total Seconds	2737	2907
% Occurrence	32.8%	34.8%
Crying		
Total Seconds	712	713
% Occurrence	6.3%	6.4%
Oral-Physical		
Total Seconds	185	189
% Occurrence	1.9%	2.1%

The four well-defined behaviors are complemented by the nature of the recording and scale systems. Videotaped appointments provide permanent records which can be analyzed several ways and times by the same or different observers. The event recorder, which generates permanent behavior records on time scale paper, further insures quantiative data, although recent advances in computer technology provide alternatives to the event recorder. The ease with which the scale can be used is indicated by the fact that training time amounted to less than two hours and raters needed to retrain only once during the entire study.

Intrarater reliability provides evidence of the scale's stability over time. Approximately eight weeks separated the initial and follow-up ratings. Long-term intrarater reliability should be a goal of future research.

The validity assessments of the scale suggest that the instrument is valid. Similar frequencies were obtained by different groups of raters. Behaviors are correlated with the Frankl scale in the appropriate di-

Table 4. Spearman correlations of total Frankl score with percentage occurrence for each behavior.

	n	High-Hands	Legs	Crying	Oral-Physical
Visit 1 Frankl	7	.03	-.33	-.77*	-.09
Visit 2 Frankl	11	-.19	-.06	-.89*	-.50
Visit 3 Frankl	18	-.42	-.13	-.90*	-.77*

*Statistically significant at $p < 0.05$ level

rection for the correlations since increased negative behavior results in a lower Frankl score. The significant negative correlations of crying and oral-physical resistance supports the results of Green et al.[14] and Smith.[15] Observers of dental patients are usually very sensitive to orally-generated behavior and these areas are attended to by the Frankl scale. The specific nature of data recorded by NCBRS expands and complements the general classification schemes such as that developed by Frankl et al.

The stability of the mean percentage behavior for high-hands, crying and oral-physical resistance involving three pedodontists and 40 patients suggests that the scale is not operator specific.

Differences in leg movement between dentists on visit three remain unexplained. It was not possible to discriminate between postural changes and disruptive behavior in this category. The best method to increase the validity of this behavior observation is not to make disruptive versus non-disruptive judgments, but to record postural changes only during the active movement phase until a new stable position is achieved. This is important because postural changes can be disruptive.

Table 5. Mean percent behavior occurrence for three dentists.

	Dentist 1 Mean (S.E.)	Dentist 2 Mean (S.E.)	Dentist 3 Mean (S.E.)
High-Hands			
V1	36 (14)	26 (11)	19 (5)
V2	31 (10)	29 (6)	13 (3)
V3	32 (11)	24 (6)	19 (4)
Legs			
V1	10 (3)	27 (14)	9 (3)
V2	14 (5)	28 (11)	14 (5)
V3	36 (12)	33 (14)	14 (5)
Crying			
V1	10 (7)	36 (18)	6 (2)
V2	19 (13)	23 (14)	21 (5)
V3	19 (9)	21 (13)	17 (5)
Oral-Physical			
V1	4 (1)	2 (1)	3 (1)
V2	3 (1)	2 (0)	2 (0)
V3	4 (1)	2 (1)	2 (0)

Table 6. P-values of difference between operators for percentage occurrence for each behavior at each visit.

	High-Hands	Legs	Crying	Oral-Physical
Visit 1	.5883	.3387	.3130	.1495
Visit 2	.0525	.0985	.8873	.8975
Visit 3	.3688	.0070*	.8492	.0716

*Statistically significant $p < 0.05$ level

Conclusions

This data demonstrates that the four category rating scale is a reliable tool which requires little time for training and implementation. The final scale is defined in Figure 2. It can be used for:

1. Videotaping child behaviors,
2. Training raters,
3. Recording the frequency and duration of each behavior, and
4. Converting data for percent occurrences.

Figure 2. North Carolina Behavior Rating Scale.

Behavior	Operational Definition
High-Hands	Hands or hand above level of arm pits or arm's extension angle less than 90° angle (on or off body) — physical restraint used.
Leg Movement	*Any* movement of legs such as kicking, lifting, postural change, etc., to get to prone position — physical restraint used.
Crying Protest	*Pure* crying, or screaming, whining, or sobbing; *or* — crying, etc. while asking for parents, asking to stop, etc. (If any doubt over fact that a verbal protest was crying or noncrying, it should be scored as crying protest).
Oral-Physical Resistance	Choking, gagging, coughing, mouth closing, spitting, vomiting, etc. Mouth closing — rate by oral request to open in any treatment segment Refusal to open mouth — rate by oral request to open in any treatment segment Physical restraint used (i.e., mouth prop)

This scale should be useful as a tool for research involving the behavior of children in the dental setting. It can be used to compare different treatment conditions and techniques. Varying the conditions outlined in this research should be done with caution. Further assessment of this scale with a larger sample size is planned.

Acknowledgments

The authors would like to recognize Mr. William Sollecito for his statistical consultation and Dr. William Ayer for his review and constructive criticism of the manuscript.

Dr. Chambers is in private practice in Ashville, North Carolina, and clinical assistant professor of pedodontics, Dr. Fields is assistant professor of pedodontics and orthodontics, and Dr. Machen is professor of pedodontics, School of Dentistry, University of North Carolina, Chapel Hill, North Carolina, 27514. Requests for reprints should be sent to Dr. Fields.

References

1. Rosenberg, H. M. and Katcher, A. H.: Heart rate and physical activity of children during dental treatment, J Dent Res, 55:648-651, 1976.
2. Simpson, W. J.; Ruzicka, R. L.; and Thomas, N. R.: Physiologic responses of children to initial dental experience, J Dent Child, 41:465-470, 1974.
3. Wilson, C. W.: Child management, JADA, 20:890-897, 1933.
4. Sands, R. A.: The mental aspects of pedodontia, Dent Items Interest, 55:927-931, 1933.
5. Lampshire, E. L.: Balanced medication, J Dent Child, 26:25-28, 1959.
6. Frankl, S. N.; Shiere, F. R.; and Fogels, H. R.: Should the parent remain with the child in the dental operatory? J Dent Child, 29:150-163, 1963.
7. Lenchner, V.: The effect of appointment length on behavior of the pedodontic patient and his attitude toward dentistry, J Dent Child, 33:61-74, 1966.
8. Bohuslav, G. F.: Effects of "psychological preparation" upon behavior of children in a pedodontic practice, Texas Dent J, 88:18-25, 1970.
9. Sarnat, H.; Peri, J. N.; Nitzan, Es.; and Perlberg, A.: Factors which influence cooperation between dentist and child, J Dent Ed, 36:9-15, 1972.
10. Wright, G. Z.: Behavior Management in Dentistry for Children, Philadelphia: W. B. Saunders Co., 1975, pp 55-56.
11. Bell, J. O.: Psychological aspects of dental treatment of children, J Exp Education, 86, Madison.
12. Sawtell, R. O.; Simon, J. F.; and Simeonsson, R. J.: The effects of five preparatory methods on child behavior during the first dental visit, J Dent Child, 41:367-375, 1974.
13. Melamed, B. G.; Weinstein, D.; Hawes, R.; and Borland, M.: Reduction of fear related dental management problems using filmed models, JADA, 90:822-826, 1975.
14. Green, R. V.; Meilman, P.; Routh, D. K.; and McIver, F. T.: Preparing the preschool child for a vist to the dentist, J Dent, 5:231-236, 1977.
15. Smith, R. C.: Chloral hydrate sedation in dentistry for children, Thesis, University of Iowa, 1977.
16. Venham, L.; Bengston, D.; and Cipes, M.: Children's response to sequential dental visits, J Dent Res, 56:454-459, 1977.

Quotable Quotes

Creative and imaginative people are often not recognized by their contemporaries. In fact, often they are not recognized in school by their teachers either. History is full of illustration. Consider some of these:

Einstein was four years old before he could speak and seven before he could read.

Isaac Newton did poorly in grade school.

Beethoven's music teacher once said of him, "As a composer, he is hopeless."

When *Thomas Edison* was a boy his teachers told him he was too stupid to learn anything.

A newspaper editor fired *Walt Disney* because he had, "no good ideas."

Caruso's music teacher told him, "You can't sing, you have no voice at all."

The director of the Imperial Opera in Vienna told *Madame Schermann Heink* that she would never be a singer and advised her to buy a sewing machine.

Leo Tolstoy flunked out of college.

Werner von Braun flunked 9th grade algebra.

Louis Pasteur was rated as mediocre in chemistry when he attended the Royal College.

Fred Waring was once rejected from high school chorus.

Winston Churchill failed the sixth grade.

From: "Who Are the Gifted," by Milton E. Larson.