An Observation Scale for Measuring Children's Distress During Medical Procedures

Charles H. Elliott
Department of Psychiatry, University of New Mexico

Susan M. Jay
Department of Pediatrics, University of Southern California, School of Medicine

Patricia Woody
Children's Hospital of Los Angeles

Received March 18, 1986; accepted August 18, 1986

This study is on the development and validity of the Observation Scale of Behavioral Distress (OSBD), a scale developed to measure children's behavioral responses to painful medical procedures. Subjects were 55 pediatric cancer patients, ages 3 to 13, who were observed during bone marrow aspirations (BMAs). To investigate validity, OSBD scores were correlated with nurse ratings, children's self-report ratings of pain and anxiety, and physiological measures taken before, during, and after the BMA. Analyses were also conducted to investigate whether the use of interval recording and severity weights for distress behaviors increased the validity of the OSBD. Results indicate that the OSBD is a valid scale and that the additional interval record-

1This study was supported by a grant from the National Cancer Institute, National Institutes of Health, to Susan Jay and Charles Elliott. We gratefully acknowledge the assistance of Michael Dolgin and Myra Saltoun for data collection, and Lynn Dahlquist for editorial comments.

2All correspondence should be sent to Susan M. Jay, Behavioral Sciences Program, Division of Hematology-Oncology, Children's Hospital of Los Angeles, 4650 Sunset Boulevard, Los Angeles, California 90027.
ing and severity weights increased the validity coefficients on certain variables such as nurse rating and blood pressure scores.

The current state-of-the-art in the area of assessing pain and anxiety experienced by children undergoing painful medical procedures remains at a rudimentary level. Early studies in the preparation literature included global Likert-type rating scales of children's distress and cooperative behaviors (Visintainer & Wolfer, 1975; Wolfer & Visintainer, 1979). More recent studies have focused on the development of more objective and operationalized behavioral observation scales which have been used specifically to measure distress in pediatric cancer patients undergoing bone marrow aspirations (BMAs) and lumbar punctures (LPs) (Jay, in press; Jay & Elliott, 1984; Jay, Ozolins, Elliott, & Caldwell, 1983; Katz, Kellerman, & Siegel, 1980; LeBaron & Zeltzer, 1984). For example, Katz et al. (1980) developed the Procedure Behavior Rating Scale (PBRS) which consisted of 13 operationally defined behaviors that were recorded as present or absent during bone marrow aspirations. The PBRS was found to be a reliable instrument and preliminary validity data (i.e., nurse ratings) were also encouraging. LeBaron and Zeltzer (1984) presented additional validity data on an eight-item version of the PBRS which they labeled the Procedure Behavior Check List.

Jay and Elliott (1984) reported on the development of the Observation Scale of Behavioral Distress (OSBD), an 11-item revision of the PBRS which included two methodological refinements of the PBRS: (a) continuous behavioral recording in 15-second intervals, and (b) a weighting score of severity of distress for each behavioral category in the scale. It was hypothesized that these refinements would result in a more highly sensitive and valid measure of children's distress. Rather surprisingly, initial data suggested relatively small, if any, increases in the sensitivity and validity of the OSBD with these revisions. The purpose of the current study was to replicate and expand the original investigation of validity of the OSBD by (a) expanding the sample size with an emphasis on younger children (since they are most in need of intervention for their distress); (b) expanding the range of variables used to assess the validity of the instrument; and (c) once again assessing the effects of methodological refinements that have been made. An additional purpose was to conduct an item analysis of the OSBD for the purpose of further refining and revising the instrument.
METHOD

Subjects

Subjects were a consecutive sample of 55 children with leukemia who entered the study over a 3-year period (1982-1985). The sample included 35 males and 20 females, a ratio that reflects the higher incidence of childhood leukemia in boys than girls. The age range of the patients was 3 to 13 years with a mean of 6.7 years. Seventy-one percent of the subjects were younger than 8 years of age reflecting the higher incidence of childhood leukemia in younger children. The ethnic composition of the sample was as follows: 55% white, 25% Hispanic, 13% black, 7% Asian.

Procedure

All subjects were part of a longitudinal study designed to evaluate the effects of various preprocedure conditions on distress during BMAs (Jay, Elliott, Katz, & Siegel, in press). The children in this study watched 30 minutes of cartoons prior to being observed during their BMA thus giving them a standardized preobservation experience that was considered unlikely to exert a significant effect on the various validity measures.

Assessment

Observational Scale of Behavioral Distress (OSBD). Two major methodological revisions of the PBRS (Katz et al., 1980) were made in developing the OSBD: (a) behaviors are recorded in continuous 15-second intervals within each of three phases of the BMA procedure rather than simple, single event recording of occurrence over an entire phase; and (b) each behavioral category in the OSBD is weighted according to severity. Other revisions of the PBRS included the addition of a category, "nervous behavior," and the combining of several categories. The behavioral categories and their definitions are listed in Table I.

Observation Procedures. An observer watched the bone marrow aspiration and recorded the occurrence of distress-related behaviors during

---

1The Observation Scale of Behavioral Distress (OSBD) is available from Susan M. Jay and includes information on development, reliability, validity, and scoring procedures.
Table 1. Behavioral Definitions of Categories for the Observation Scale of Behavioral Distress

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information seeking</td>
<td>Any questions regarding medical procedure</td>
<td>&quot;Is the needle in?&quot;</td>
</tr>
<tr>
<td>Cry</td>
<td>Onset of tears and/or low-pitched nonword sounds of more than 1-second duration</td>
<td></td>
</tr>
<tr>
<td>Scream</td>
<td>Loud, nonword, shrill vocal expressions at high pitch intensity</td>
<td></td>
</tr>
<tr>
<td>Physical restraint</td>
<td>Child is physically restrained with noticeable pressure and/or child is exerting bodily force and resistance in response to restraint.</td>
<td></td>
</tr>
<tr>
<td>Verbal resistance</td>
<td>Any intelligible verbal expression of delay, termination, or resistance</td>
<td>&quot;Stop&quot;</td>
</tr>
<tr>
<td>Seeks emotional support</td>
<td>Verbal or nonverbal solicitation of hugs, physical or verbal comfort from parents or staff</td>
<td>&quot;I don't want it&quot;</td>
</tr>
<tr>
<td>Verbal pain</td>
<td>Any words, phrases, or statements in any tense which refer to pain or discomfort</td>
<td>&quot;Ouch&quot;</td>
</tr>
<tr>
<td>Flail</td>
<td>Random gross movements of arms, legs, or whole body</td>
<td>&quot;My leg hurts&quot;</td>
</tr>
<tr>
<td>Verbal fear*</td>
<td>Any intelligible verbal expression of fear of apprehension</td>
<td>&quot;That hurts&quot;</td>
</tr>
<tr>
<td>Muscular rigidity*</td>
<td>Noticeable contraction of observable body part</td>
<td>&quot;I'm scared&quot;</td>
</tr>
<tr>
<td>Nervous behavior*</td>
<td>Physical manifestations of anxiety or fear. Consist of repeated, small physical actions</td>
<td>Clinched fists; gritted teeth; facial contortions. Legs bent tightly upward off R₄ table</td>
</tr>
</tbody>
</table>

*Eliminated after item analysis.

15-second intervals (indicated on an audiotape played to the observer over headphones). Observers were trained by the second author over a 6- to 8-week period until a minimum criterion of 75% reliability over 6 consecutive BMA procedures was established.

Self-Report Measures of Fear. Children were asked to rate "how scared" they felt about their upcoming BMA, at the beginning of the session. A "faces" scale was used with three faces ranging from a "happy" face indicating "not at all scared," to a "sad" face indicating "very scared."
Self-Report Measures of Pain. A “Pain Thermometer” (Katz et al., 1982) was used to measure children's anticipated and experienced pain. The Pain Thermometer is a graphic depiction of a large thermometer graded on a 0 to 100 scale, with 0 representing “no pain at all” and 100 representing “the worst pain possible.” At the beginning of the session, children were asked to rate how much they thought their upcoming BMA would hurt (Pain I), as a measure of anticipated pain. After the BMA, they were asked to rate how much the BMA did hurt as a measure of actual experienced pain (Pain II).

Physiological Measures. Children’s heart rate and blood pressure were taken using a Dinamap Model 845 at three points in time: (a) upon arrival to the clinic, but prior to the cartoon viewing; (b) when the child got on the treatment table, just before the BMA occurred; and (c) 3 to 5 minutes after the BMA.

Nurse Ratings. After the procedure was over, the nurse who assisted the person conducting the BMA was asked to rate the degree of behavioral distress exhibited by the child on a 5-point Likert-type scale, with 1 representing “no distress” and 5 representing “extreme distress.”

RESULTS

Reliability

Pearson product-moment correlations analyses conducted between total OSBD scores for two observers for 20% of the procedures yielded an $r$ of .98. In addition, the number of agreements between two observers as to whether or not each of the 11 behaviors occurred within each 15-second interval were divided by the total number of agreements plus disagreements. The mean agreement score obtained was 84%.

Validity

Analyses were conducted to determine the validity of the OSBD and to investigate the effects of continuous interval coding and severity weights on the validity of the OSBD. Therefore, (a) original OSBD scores were correlated with all of the validity measures and (b) the OSBD data were rescored without using the continuous 15-second intervals or the severity weights and these recalculated scores were then also correlated with the validity measures. Thus, rather than scoring behaviors in 15-second intervals, be-
Table II. Correlation Coefficients Between Validity Measures and Original and Recalculated OSBD Scores

<table>
<thead>
<tr>
<th>Validity measure</th>
<th>Original OSBD scores</th>
<th>Recalculated OSBD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse ratings of distress*</td>
<td>.69*</td>
<td>.55*</td>
</tr>
<tr>
<td>Fear ratings</td>
<td>.38*</td>
<td>.37*</td>
</tr>
<tr>
<td>Anticipated pain</td>
<td>.24*</td>
<td>.26*</td>
</tr>
<tr>
<td>Experienced pain</td>
<td>.20</td>
<td>.24</td>
</tr>
<tr>
<td>Heart rate I</td>
<td>.38*</td>
<td>.41*</td>
</tr>
<tr>
<td>Heart rate II</td>
<td>.55*</td>
<td>.47*</td>
</tr>
<tr>
<td>Heart rate III</td>
<td>.33*</td>
<td>.36*</td>
</tr>
<tr>
<td>Systolic blood pressure I</td>
<td>.32*</td>
<td>.04</td>
</tr>
<tr>
<td>Diastolic blood pressure I</td>
<td>.32*</td>
<td>.21</td>
</tr>
<tr>
<td>Systolic blood pressure II</td>
<td>.38*</td>
<td>.11</td>
</tr>
<tr>
<td>Diastolic blood pressure II</td>
<td>.38*</td>
<td>.20</td>
</tr>
<tr>
<td>Systolic blood pressure III</td>
<td>-.13</td>
<td>.03</td>
</tr>
<tr>
<td>Diastolic blood pressure III</td>
<td>.03</td>
<td>-.04</td>
</tr>
</tbody>
</table>

*On these variables, the differences between the two sets of correlation were significant at the .5 level.

Second, the differences between the validity correlation coefficients for the two sets of OSBD scores (original scores with severity weights and interval coding, and recalculated scores without these refinements) were analyzed using an r to z transformation for dependent samples (Glass & Stanley, 1970) to determine whether the continuous interval recording and the

behaviors were rescored merely for occurrence of nonoccurrence over the entire phase and severity weights were not used.

Table II contains the Pearson product-moment correlation coefficients between the various validity measures and both original and recalculated OSBD scores. The results indicate, when one views the correlations between original OSBD scores and the validity measures, that the OSBD is significantly related to almost all of the measures of distress including Nurse Ratings ($r = .69, p < .0001$), Fear Ratings ($r = .38, p < .01$), Anticipated Pain ($r = .24, p < .05$), Heart Rate I ($r = .38, p < .01$), Heart Rate II ($r = .55, p < .0001$), Heart Rate III ($r = .33, p < .01$), Systolic Blood Pressure I and II ($r = .32, p < .01$ and $r = .38, p < .01$, respectively), and Diastolic Blood Pressure II ($r = .38, p < .01$). The only measures that were not significantly correlated with original OSBD scores were Experienced Pain and Systolic and Diastolic Blood Pressure III.

*The interval coding data sheets for the OSBD allows for scoring of behaviors with or without the severity weights and with or without continuous interval recording.
severity weights strengthened the validity of the OSBD. The two sets of validity coefficients for Fear Ratings, Anticipated Pain, and Experienced Pain were virtually identical, with no significant differences. Nor were the differences significant between the two sets of correlations for Heart Rate I, II, and III.

By contrast, the correlation between the original OSBD scores and Nurse Ratings ($r = .69$) was higher than the correlation between the recalculated OSBD scores and Nurse Ratings ($r = .55$), and this difference was significant ($z = -2.5, p < .05$). Significant differences (favoring the original OSBD scores) also were found between the two sets of correlations for Systolic Blood Pressure I ($z = -2.3, p < .05$), Systolic Blood Pressure II ($z = -4.1, p < .05$), and Diastolic Blood Pressure II ($z = -2.7, p < .05$). The difference between the validity correlations for the original and recalculated OSBD scores were not significant for the other blood pressure measures. Thus the OSBD, when recalculated without interval data and severity weights, yielded distress scores that were significantly correlated with a number of the validity measures, but the validity was not quite as strong as when the original distress scores were used.

Item Analysis

The 11 behavioral categories of the OSBD were subjected to an item analysis in which individual category scores were (a) scored for frequency of occurrence, (b) intercorrelated, and (c) correlated with total OSBD scores. Item analyses were conducted for the total sample and for each age group separately. The purpose of the item analysis was to eliminate any categories that were of very low frequency and those that were not correlated with other behavioral categories or to total OSBD scores.

The criteria for retaining a behavioral category were as follows: (a) category scores had to occur for at least 10% of the subjects, and (b) category scores had to have an item-total correlation coefficient of +.3 or more for the total sample and/or for at least one age group. One exception to these criteria was made for the category Emotional Support because it correlated .28 for the young age group and it was a very high frequency item, that is, it occurred in over half the sample.

Results indicated that 8 of the 11 OSBD categories met the criteria and three were eliminated. Verbal Fear was eliminated because it occurred in only 5% of the total sample and never in children above the age of 6 years. Nervous Behavior was eliminated because it correlated .07 with the total score for the entire sample, -.28 for children aged 4 to 6 years, and -.11 for children aged 7 to 14 years. Muscular Rigidity was eliminated because it cor-
related — .20 with the total score for the total sample, — .17 for the children aged 4 to 6 years, and — .37 for the children aged 7 to 14 years.

Cronbach's alpha test of internal consistency was conducted before and after elimination of Nervous Behavior, Verbal Fear, and Muscular Rigidity. Results before the categories were eliminated indicated an alpha internal consistency coefficient of .68 and a coefficient of .72 was obtained after the behaviors were eliminated.

DISCUSSION

The results of this study provide support for the reliability and validity of the Observational Scale of Behavioral Distress. The validity of the instrument is supported by significant correlations between OSBD total scores and a wide range of validity measures including fear, anticipated pain, heart rate, and blood pressure at two time periods, and nurse ratings.

OSBD total scores were not correlated with self-reports of actual experienced pain for the total sample. This finding may be due to the fact that the instrument is a better overall measure of anxiety than actual experienced pain sensations. However, since self-reported experienced pain did correlate significantly with OSBD scores of subjects older than 7 ($r = .51, p < .05$), it may be, that younger children are simply unable to rate experienced pain with a high degree of reliability.

The effects of the methodological refinements (interval recording and severity weights) appear to be significant in this somewhat larger sample. It should be noted, however, that even when scored without the refinements, the OSBD has reasonably good validity. Since these refinements greatly complicate the task of scoring and training of observers, one must consider the purpose for using the instrument before deciding whether to include the refinements. These findings suggest that the increased sensitivity derived from the refinements is largely restricted to improving the assessment of relationships between OSBD scores and measures with relatively weak predictive power (e.g., nurse ratings and blood pressure). Thus, studies investigating the influence of subtle individual difference variables such as coping styles and parental anxiety may wish to use a maximally sensitive instrument, whereas clinical applications involving the reduction of overall distress may find that the OSBD is sufficient without including the refinements.

The item analysis has allowed the instrument to be reduced to eight items. This reduction from the original 11 items should make observer training somewhat easier and may increase reliability as well. In its present form, the OSBD appears to be a robust scale which is likely to be equally useful for measuring children's distress associated with a variety of other painful medical procedures.
REFERENCES


