The Yale Preoperative Anxiety Scale: How Does It Compare with a “Gold Standard”?

Zeev N. Kain, MD, Linda C. Mayes, MD, Domenic V. Cicchetti, PhD, Andrea L. Bagnall, MSW, Jessica D. Finley, BS, and Maura B. Hofstadter, PhD

Departments of *Anesthesiology, †Pediatics, and ‡Child Study Center, Yale School of Medicine and Yale-New Haven Children’s Hospital, New Haven, Connecticut

Evaluating the effectiveness of interventions directed toward the treatment of preoperative anxiety in children has been hindered by the absence of a statistically valid measurement tool. In a previous investigation, we developed an instrument (Yale Preoperative Anxiety Scale [YPAS]) that can be used to assess anxiety in children undergoing induction of anesthesia. The purpose of the present investigation was to modify and expand the applicability of the instrument to the preoperative holding area and to validate the modified instrument (m-YPAS) against a recognized “gold standard” (State-Trait Anxiety Inventory for Children [STAIC]). Videotapes of children in a preoperative holding area were analyzed by the investigators. The existing five categories of the YPAS were found to reflect most of the behaviors observed. Several items, however, were modified to describe new behaviors observed. Reliability analysis using weighted $\kappa$ statistics revealed that interobserver agreement ranged from 0.68 to 0.86, whereas intraobserver weighted $\kappa$ ranged from 0.63 to 0.90. Concurrent validity between the YPAS and the STAIC was acceptable ($P = 0.01$, $r = 0.79$). Construct validity was high as assessed by increased m-YPAS scores from the preoperative holding area (28 ± 8) to entering the operating room (35 ± 12), to introduction of the anesthesia mask (43 ± 15); $F[1,36] = 0.6$, $P = 0.001$. Showing good to excellent observer reliability and high concurrent and construct validity, the m-YPAS proved to be an appropriate tool for assessing children’s anxiety during the perioperative period. Implications: The absence of a statistically valid measurement tool that can be applied easily in perioperative settings hinders the evaluation of interventions directed toward treatment of preoperative anxiety in children. The authors describe the development of such a tool, the modified Yale Preoperative Anxiety Scale.

Preoperative anxiety is characterized by subjective feelings of tension, apprehension, nervousness, and worry. Although some investigators suggest that mild to moderate preoperative anxiety is associated with improved postoperative behavioral recovery in adults (1), others disagree (2). Most investigators suggest that a high preoperative anxiety state is associated with both psychologically and physiologically adverse outcomes (2–4). In children, preoperative anxiety is reported to result in postoperative negative psychological effects, including nightmares, separation anxiety, eating problems, and increased fear of doctors (3,4). In adults, preoperative anxiety has been suggested to correlate with increased postoperative pain, increased postoperative analgesic requirements, and prolonged recovery and hospital stay (7).

Currently, the “gold standard” for anxiety evaluation is Spielberger’s State-Trait Anxiety Inventory (STAI). This instrument has both an adult version (STAI) (5) and a pediatric version (STAIC) (6) that can be used with children over the age of five years. To date, more than 1000 studies involving research using the STAI and the STAIC have been published in peer-reviewed literature (7). Although most of this research has been conducted by psychologists, these instruments have also been extensively used by investigators from disciplines such as anesthesiology (8,9). Before adopting the STAIC for routine use in investigations related to perioperative anxiety, however, its limitations should be recognized. First, it takes approximately 5–10 minutes to complete the questionnaire. This may not be feasible in busy operating room settings. Second, the STAIC can be used...
only in children over the age of five years, and, in fact, it must be read to most children under the age of eight years. Thus, preschool children who are particularly vulnerable to preoperative anxiety cannot benefit from this instrument.

In a previous investigation, we developed an observational instrument (Yale Preoperative Anxiety Scale [YPAS]) that can be used to assess anxiety in children during induction of anesthesia (10). This instrument can be used for children aged two to six years, can be completed in less than one minute, and has good inter- and intraobserver reliability data ($\kappa_w = 0.66–0.91$). The objectives of the present investigation were (a) to expand the applicability of the YPAS to the preoperative holding area by modifying its original descriptors, (b) to obtain reliability data for the modified instrument in children aged 5–12 years, and (c) to validate the modified instrument against a recognized gold standard for anxiety assessment (STAIC).

**Methods**

**Phase 1: Scale Development**

After receiving institutional review board approval, a research team composed of two anesthesiologists and two psychologists examined videotapes of children taken in a preoperative holding area ($n = 12$). During multiple group sessions, the original YPAS was examined for applicability to the new perioperative situation. The existing five categories (activity, emotional expressivity, state of arousal, vocalization, use of parents) and 22 items of the YPAS were found to reflect most of the behaviors observed in the preoperative holding area. Nine items within the five categories, however, had to be modified to describe new behaviors observed. After the development of the modified YPAS (m-YPAS), reliability and validity analyses were performed.

**Phase 2: Reliability and Validity Analysis**

The study population for this phase was drawn from children aged 5–12 yr who were about to undergo elective surgery ($n = 51$). To avoid blinding issues, children who participated in Phase 1 were excluded from participation in Phase 2 of the study. Also, children with developmental delay or who did not speak English were excluded from the study.

On the day of surgery in the preoperative holding area, demographic data, such as age, gender, and history of previous surgery, were collected, and all children and parents were videotaped for 2 min. Next, all children completed the STAIC. This questionnaire contains two separate, 20-item self-report rating scales for measuring trait (baseline) and state (situational) anxiety. Children were asked to respond on a 3-point scale. Total scores for situational and baseline questions separately range from 20 to 60, with higher scores denoting higher levels of anxiety. The normative data for the STAIC are based on two large samples of elementary school children: 456 male and 453 female elementary school students drawn from three different schools in Bradenton and Manatee Counties, FL (6).

On the day of surgery in the operating room, anesthesia was induced using O$_2$/N$_2$O and halothane administered via a scented mask. The child’s anxiety during induction was assessed using the m-YPAS. The rating was performed at two points: entering the operating room and introduction of the anesthesia mask to the child.

The videotapes ($n = 51$) were analyzed separately and independently by a naive observer (anesthesiologist) and an experienced observer (psychologist). Reliability between and within the two observers was assessed using weighted $\kappa$ statistics for overall chance-corrected agreement ($\kappa_w$). Calculations were performed using the computer program Rater Agreement Categorical Data (RATCAT) (11). Output from this program consists of observed agreement (PO), chance agreement (PC), and chance-corrected agreement ($\kappa_w$). In addition, this program also assigns the appropriate clinical significance to individual weighted $\kappa$ values: $\kappa_w <0.40 =$ poor, $0.40–0.59 =$ fair, $0.60–0.74 =$ good, and $0.75–1.00 =$ excellent (12).

To assess concurrent validity (how a scale compares with a recognized gold standard), we determined the correlation coefficient ($r$) for the m-YPAS against the STAIC. We hypothesized *a priori* that the m-YPAS should correlate highly ($r >0.65$) with the STAIC. To assess construct validity (whether a scale accurately measures what it is intended to measure), the m-YPAS was examined by comparing scores obtained under increasing stress conditions. That is, comparing the anxiety of children in the preoperative holding area (T1) to the anxiety of children during Phase 1 (T2) and Phase 2 (T3) of induction of anesthesia. This method was also used by Spielberger, who evaluated the STAIC by administering it to students under high- and low-stress conditions (5). Statistical analysis over the three time points (i.e., T1–T3) was performed using repeated-measures analysis of covariance with trait anxiety as a co-variable. *Post hoc* analysis was performed using paired t-tests with Bonferroni’s correction for multiple comparisons indicating significance at $P < 0.0125$.

Finally, by using the STAIC as a gold standard, we identified a m-YPAS score that indicates high anxiety in children undergoing surgery. The cutoff score for high anxiety was derived from Spielberger’s manual (6) which provides normative data for children aged 5–17 yr. The mean anxiety score for the normative group was 31 ± 6, and we defined the high-anxiety
Sixty-seven percent of the children had experienced previous surgery or hospitalization (Table 1). The m-YPAS was administered to 51 children under-23

Table 1. Characteristics of Children Studied

<table>
<thead>
<tr>
<th>Study subjects (n = 51)</th>
<th>Age (yr) (mean ± sd)</th>
<th>Gender (F/M) (%)</th>
<th>Race (%)</th>
<th>Parental education (%)</th>
<th>Preparation program (%)</th>
<th>PPIA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 ± 1.5</td>
<td>43/57</td>
<td>Caucasian 84</td>
<td>&gt;High school 67</td>
<td>No 31</td>
<td>Yes 69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African-American 8</td>
<td>&lt;High school 33</td>
<td>Yes 80</td>
<td>No 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other 8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PPIA = parental presence during induction of anesthesia.

state at 1 sd above the normative mean. That is, children who scored more than 37 on the STAIC would be classified in our study as high-anxiety cases. Next, the sensitivity (proportion of correctly identified high-anxiety cases), the specificity (proportion of correctly identified noncases), and the positive predictive value (probability of a high m-YPAS score indicating a high-anxiety case) were examined for different cutoff points on the m-YPAS in relation to the STAIC (gold standard).

Results

Phase 1: Scale Development

The m-YPAS was administered to 51 children undergoing general anesthesia and surgery. The mean age of the 27 male and 24 female patients was 8 ± 1.5 yr. Sixty-seven percent of the children had experienced previous surgery or hospitalization (Table 1). Operations included a wide variety of surgical procedures, such as herniorrhaphy, tonsillectomy and adenoidectomy, and myringotomy tube placement. The m-YPAS consists of 22 items in five categories (activity, emotional expressivity, state of arousal, vocalization, and use of parents) (see Appendix 1). The highest behavioral level observed in each of the five m-YPAS categories is the score for that category. Because each category of the m-YPAS has a different number of items (either four or six), partial weights were calculated and then added to a total score that ranged from 0 to 100. For example, for two categories containing four and six items, with a score of 1 in each category, the calculation is: \((\frac{4}{4} + \frac{6}{6}) \times \frac{1}{102} = \) total adjusted score. In the preoperative holding area, the m-YPAS score ranged from 23 to 59, with a median of 23 and a mean ± sd of 28 ± 8; the mean ± sd STAIC score was 33 ± 7. There were no statistically significant relationships among history of previous surgery, gender, age, type of surgery, and the scores on the m-YPAS or STAIC.

Phase 2: Reliability and Validity Analysis

Table 2 presents the agreement between observers (chance-corrected weighted k) for the m-YPAS. In the preoperative holding area, the agreement between observers ranged from 0.68 (emotional expressivity, good clinical significance) to 0.86 (vocalization, excellent clinical significance). Intraobserver weighted K agreement ranged from 0.63 (state of arousal, good reliability) to 0.90 (vocalization, excellent reliability) in the preoperative holding area (Table 3).

Concurrent validity as assessed by the correlation between the m-YPAS total score and the STAIC was good \((r = 0.79, P = 0.01)\). Construct validity was assessed by comparing the m-YPAS scores at three increasingly stressful points (i.e., preoperative holding, induction Phase 1, and induction Phase 2). There were significant differences in m-YPAS scores over the three stress points \((28 ± 8 vs 35 ± 12 vs 43 ± 15; F[1,36] = 0.6, P = 0.001)\). Post hoc analysis indicated that m-YPAS scores during Phase 2 of induction were significantly higher than m-YPAS scores during Phase 1 \((P = 0.01)\) and in the preoperative holding area \((P = 0.01)\). Similarly, scores obtained in Phase 1 of induction were significantly higher than scores obtained in the preoperative holding area \((P = 0.01)\).

As discussed above, a score of 37 on the STAIC was chosen as a reference point to define high-anxiety cases in the preoperative holding area. Using this reference point, the sensitivity, specificity, and predictive value were calculated at different cutoff points on the m-YPAS. Table 4 shows that the cutoff point of 30 on the m-YPAS leads to a balance in which the sensitivity and specificity are high and the predictive value is 79%. At this score, only five patients were misclassified. three patients (6%) were anxious on the m-YPAS but not on the STAIC (false positives), and two patients (4%) were classified as not highly anxious on the m-YPAS, although they were on the STAIC (false negatives). If we chose a lower cutoff point, sensitivity increased, but because of the lower specificity, the predictive value was lower, resulting in a higher number of patients who are anxious on the m-YPAS but not on the STAIC (false positives) (Table 4). In contrast, when the cutoff score was raised to 35, the specificity increased, but because of the lower sensitivity, the number of false-negative patients increased. Using the score of 37 on the STAIC as a reference point, the prevalence of high-anxiety cases in the preoperative holding area was 24%.
Table 2. Interobserver Agreement on the m-YPAS

<table>
<thead>
<tr>
<th>m-YPAS category</th>
<th>Observed agreement (PO)</th>
<th>Chance agreement (PC)</th>
<th>Weighted $\kappa (k_w)$</th>
<th>Clinical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocalization</td>
<td>0.99</td>
<td>0.95</td>
<td>0.86</td>
<td>Excellent</td>
</tr>
<tr>
<td>Use of parent</td>
<td>0.98</td>
<td>0.87</td>
<td>0.82</td>
<td>Excellent</td>
</tr>
<tr>
<td>Activity</td>
<td>0.97</td>
<td>0.89</td>
<td>0.75</td>
<td>Excellent</td>
</tr>
<tr>
<td>State of arousal</td>
<td>0.96</td>
<td>0.86</td>
<td>0.70</td>
<td>Good</td>
</tr>
<tr>
<td>Emotional expressivity</td>
<td>0.97</td>
<td>0.90</td>
<td>0.68</td>
<td>Good</td>
</tr>
</tbody>
</table>

m-YPAS = modified Yale Preoperative Anxiety Scale.  
*Clinical significance was assigned by using the RATCAT (Rater Agreement Categorical Data) computer program (Heavens, 1978).

Table 3. Intraobserver Agreement on the m-YPAS

<table>
<thead>
<tr>
<th>m-YPAS category</th>
<th>Observed agreement (PO)</th>
<th>Chance agreement (PC)</th>
<th>Weighted $\kappa (k_w)$</th>
<th>Clinical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1: Vocalization</td>
<td>0.99</td>
<td>0.93</td>
<td>0.86</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 2: Vocalization</td>
<td>0.99</td>
<td>0.93</td>
<td>0.86</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 1: Activity</td>
<td>0.99</td>
<td>0.93</td>
<td>0.88</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 2: Activity</td>
<td>0.99</td>
<td>0.93</td>
<td>0.88</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 1: Use of parent</td>
<td>0.98</td>
<td>0.86</td>
<td>0.86</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 2: Use of parent</td>
<td>0.96</td>
<td>0.86</td>
<td>0.78</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 1: State of arousal</td>
<td>0.98</td>
<td>0.93</td>
<td>0.76</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer 2: State of arousal</td>
<td>0.95</td>
<td>0.81</td>
<td>0.74</td>
<td>Good</td>
</tr>
<tr>
<td>Observer 1: Emotional expressivity</td>
<td>0.97</td>
<td>0.87</td>
<td>0.63</td>
<td>Good</td>
</tr>
<tr>
<td>Observer 2: Emotional expressivity</td>
<td>0.94</td>
<td>0.85</td>
<td>0.63</td>
<td>Good</td>
</tr>
</tbody>
</table>

m-YPAS = modified Yale Preoperative Anxiety Scale.  
*Clinical significance was assigned by using the RATCAT (Rater Agreement Categorical Data) computer program (Heavens, 1978).

Table 4. Identifying the High-Anxiety Group

<table>
<thead>
<tr>
<th>Cutoff score on the m-YPAS</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
<th>False positives</th>
<th>False negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>92</td>
<td>65</td>
<td>48</td>
<td>38</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>85</td>
<td>92</td>
<td>79</td>
<td>54</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>78</td>
<td>95</td>
<td>78</td>
<td>14</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

Data are expressed as percentages. Statistical data of the modified Yale Preoperative Anxiety Scale (m-YPAS) at different cutoff points with a score of 37 on the State-Trait Anxiety Inventory for Children as a reference point ($n = 51$).

Discussion

As part of continuing investigations toward evaluating perioperative anxiety, we have modified an anxiety assessment instrument that can be used for children aged two years and older. This instrument has good to excellent observer reliability, and good validity compared with a gold standard.

The m-YPAS can be used to assess the effectiveness of anxiety reduction interventions for children undergoing surgery. To date, available instruments that measure children's anxiety during the perioperative period include self-report (e.g., STAIC) and observational (e.g., YPAS) assessment tools. Although self-report tools are used extensively in the general medical literature, their use in the perioperative settings is limited for several reasons. First, self-report of anxiety is based on the ability of the child to respond verbally to an instrument (13). Young children, however, are unable to communicate verbally because of their developmental level. Second, all structured self-report instruments require at least 5-10 minutes to complete. This may not be feasible in a busy operating room setting. Thus, an observational rather than a self-report measure of anxiety may be the preferred instrument, both because it takes less time to administer and because it can be used for children of all ages. Current studies in the preoperative preparation and anesthesia literature use observational global ratings of 3-7 points or visual analog scales (14,15). Rarely are issues such as inter- and intraobserver reliability reported for these types of scales, and none has been validated against a recognized gold standard such as the STAIC.

In addition to having good reliability and validity data, the m-YPAS has several other important features. First, it can be applied to all children older than two years of age in less than one minute. Second,
because it is a structured instrument that consists of five domains of anxiety, it is much more sensitive to changes in anxiety levels than global instruments. Third, this tool can be applied both in the preoperative holding area and during induction of anesthesia. Most previous investigations that evaluated interventions directed to reduce preoperative anxiety in children assessed the control and treatment groups at baseline and after an intervention. The assessment tools that have been used at these time points, however, may have been different, and the baseline data were usually used only to demonstrate that the groups were similar at the onset of the study. The most common statistical analysis performed to investigate the effect of the intervention is Student’s t-test. We propose that such studies should be analyzed by means of repeated measures. That is, by using a similar assessment instrument pre- and postintervention, investigators may take advantage of the additional statistical power gained with repeated measures in the same subject.

Finally, it is important to note several important methodological issues related to this study. First, although this tool has good reliability data for children aged 2-12 years (10), it was validated against a gold standard only appropriate for children aged 5-12 years. This is because there is no gold standard for anxiety evaluation in the younger child. In our previous investigation, however, we validated the YPAS for the younger age group against other global measures of anxiety with good results. Second, the m-YPAS was validated against a gold standard only for the preoperative holding area and not for induction of anesthesia. Obviously, no self-report instrument can be used during induction of anesthesia, and validation of the m-YPAS against the STAIC at that point is not possible. Again, in our previous investigation, we validated the YPAS during induction of anesthesia against other global measures of anxiety with good results.

In conclusion, showing good to excellent observer reliability and good concurrent and construct validity, the m-YPAS proves to be an excellent tool for assessing children’s anxiety during the perioperative period.

The m-YPAS should be of particular value to anesthesiologists for evaluating new interventions directed toward reducing anxiety in children undergoing anesthesia and surgery.

Appendix 1

The m-YPAS

Activity
1. Looking around, curious, playing with toys, reading (or other age-appropriate behavior); moves around holding area/treatment room to get toys or to go to parent; may move toward operating room equipment
2. Not exploring or playing, may look down, fidget with hands, or suck thumb (blanket); may sit close to parent while waiting, or play has a definite manic quality
3. Moving from toy to parent in unfocused manner, non-activity-derived movements; frenetic/frenzied movement or play; squirming, moving on table; may push mask away or cling to parent
4. Actively trying to get away, pushes with feet and arms, may move whole body; in waiting room, running around unfocused, not looking at toys, will not separate from parent, desperate clinging

Vocalizations
1. Reading (nonvocalizing appropriate to activity), asking questions, making comments, babbling, laughing, readily answers questions but may be generally quiet; child too young to talk in social situations or too engrossed in play to respond
2. Responding to adults but whispers, “baby talk,” only head nodding
3. Quiet, no sounds or responses to adults
4. Whimpering, moaning, groaning, silently crying
5. Crying or may be screaming “no”
6. Crying, screaming loudly, sustained (audible through mask)

Emotional expressivity
1. Manifestly happy, smiling, or concentrating on play
2. Neutral, no visible expression on face
3. Worried (sad) to frightened, sad, worried, or tearful eyes
4. Distressed, crying, extreme upset, may have wide eyes

State of apparent arousal
1. Alert, looks around occasionally, notices or watches what anesthesiologist does (could be relaxed)
2. Withdrawn, sitting still and quiet, may be sucking on thumb or have face turned into adult
3. Vigilant, looking quickly all around, may startle to sounds, eyes wide, body tense
4. Panicked whimpering, may be crying or pushing others away, turns away

Use of parents
1. Busy playing, sitting idle, or engaged in age-appropriate behavior and doesn’t need parent; may interact with parent if parent initiates the interaction
2. Reaches out to parent (approaches parent and speaks to otherwise silent parent), seeks and accepts comfort, may lean against parent
3. Looks to parent quietly, apparently watches actions, doesn’t seek contact or comfort, accepts it if offered or clings to parent
4. Keeps parent at distance or may actively withdraw from parent, may push parent away or desperately clinging to parent and not let parent go

The authors thank Paul G. Barash, MD, for his critical review of this manuscript.

References