Journal of Consulting and Clinical Psychology
1978, Vol. 46, No. 6, 1246-1257

An Observational Approach to the Assessment of Anxiety in Young Children

Blair Glennon
Harvard University

John R. Weisz
University of North Carolina at Chapel Hill

Clinical research on anxiety has long relied on assessment techniques that may be inappropriate with young children (e.g., self-report inventories). The present article describes an alternative to such techniques—a scale using observational methodology. To assess the reliability and validity of this instrument, the Preschool Observational Scale of Anxiety (POSA), preschoolers were observed and scored on the scale during two test sessions. Session 1, with mothers absent, was expected to provoke relatively high anxiety; Session 2, with mothers present, was expected to provoke minimal anxiety. Total POSA scores assigned by two independent judges correlated .78 ($p < .001$), with highly significant interjudge correlations for most of the scale items. Regarding the validity of the instrument, it was found that (a) POSA scores were significantly correlated with teachers' and parents' inventory ratings of children's anxiety (all ps $< .01$), and (b) children's POSA scores were significantly higher in Session 1 than Session 2 ($p < .01$). The findings suggest that the POSA may provide a means of assessing situationally induced anxiety in children who are too young to accurately report their internal states.

Anxiety has long been a topic of central importance in clinical research and practice. Achenbach (1974) noted that among personality traits emerging from trait theories, "anxiety is perhaps the most frequently inferred and measured (p. 574)"; yet, demonstrably valid and reliable measurement of anxiety has been difficult to achieve, particularly among children. Especially acute is the need for accurate measures that can be used to assess specific situational effects on anxiety states in children (see Spielberger et al., 1972). Although many investigators recognize the need for such measures, particularly in research with children who are too young to accurately report their own internal states, most researchers also recognize the difficulties that inhere in applying previous anxiety assessment techniques to children.

Historically, the principal methods for measuring anxiety have been physiological measurements, projective techniques, self-reports, and behavior ratings by observers. Through direct gauging of autonomic activity, physiological measures bypass the problems of subjective judgments. However, agreement among physiological measures is frequently poor, since people have different styles of autonomic responding (Phillips, Martin, & Meyers, 1972), and the physiological instruments may tap emotions such as anger or joy rather than anxiety (Lazarus, 1966). Furthermore, the unusual instruments used for physiological measures may distract subjects and make naturalistic observations difficult, particularly when the subjects are children.

Projective techniques have also been used to obtain ratings of generalized anxiety (McReynolds, 1968), but most projective techniques require individual administration as well as substantial time and expertise on the

The study was supported in part by a research grant from the College of Human Ecology, Cornell University.

The authors are grateful to Barbara Cesario, John Doris, Laura Harrison, Robert Loheyde, Linda Mack, and the staff, parents, and children of the Cornell Nursery School for their assistance in the conduct of this study.

Requests for reprints should be sent to John R. Weisz, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina 27514.

part of the examiner, thus limiting their usefulness. Furthermore, McReynolds (1968) has reported that projective techniques "are generally found to be unrelated to inventory essays of anxiety" (p. 258); and Achenbach (1974), after reviewing the massive literature on projective techniques, has concluded that "there is little evidence for the reliability or validity of most of the interpretations made from them" (p. 604).

Three main types of self-reports have been used in assessing anxiety. First, direct self-ratings involve asking subjects specifically how anxious or nervous they feel. McReynolds (1968) has suggested that such self-ratings are most useful for measuring current levels of anxiety or changes in anxiety from one situation to another. A second type of self-report is the Adjective Check List, on which subjects indicate which of several adjectives (e.g., "jittery") best characterize their mood. Adjective Check Lists have been used to assess either current anxiety level or characteristic anxiety (McReynolds, 1968). The third type of self-report is the inventory method, with questionnaires designed to determine how subjects feel in a variety of situations. This method is usually aimed at assessing characteristic rather than current anxiety level. Inventories may focus on generalized anxiety (i.e., across a wide variety of situations) or anxiety related to particular types of situations (such as separations). Several investigators have noted drawbacks of the various self-report methods (see Spielberger, 1972); however, probably the most important drawback in the present context is that such methods assume both the ability and the willingness of subjects to correctly describe their own feelings. The first assumption is clearly tenuous with young children. And, with regard to the second assumption, there is evidence that many children give inaccurate reports about their anxiety due to defensiveness or social pressures (Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960). Consequently, Sarason (1966) has stated that "the verbal response to our [children's self-report anxiety] scales may be telling us more about the self than the affect" (p. 79).

Behavior ratings of anxiety by observers may involve global, subjective judgments about the subjects' apparent nervousness or ratings on many specific, concrete behaviors thought to indicate anxiety (e.g., stuttering). Such approaches appear to offer distinct advantages relative to those discussed above (see Spielberger et al., 1972). For instance, unlike physiological measures, observers' behavior ratings can be made unobtrusively; unlike projective techniques, they do not necessarily require trained clinicians; and unlike self-report measures, they do not rest on tenuous assumptions about the abilities and attitudes of the subjects. In addition, behavioral observations may be particularly useful with children, since they appear to disguise expression of their feelings less effectively than adults (Sarason, 1966).

The observers for the behavior rating scales of children's anxiety are sometimes teachers or parents, with the scales including items relevant to school or family situations (e.g., cries at bedtime). However, these particular scales entail a risk of distortion due to the parent's or teacher's lack of care or objectivity in observing, or to their bias or defensiveness about reporting on a child whom they know personally (Sarason et al., 1960). The observers for the behavior rating scales, however, may be trained observers (see Buss, Wiener, Durkee, & Baer, 1955), and the scales may comprise behaviors indicative of anxiety across situations, thus being more widely applicable and less subject to distortion by observers. The latter type of behavior rating scale would, in principle, permit detection of subtle relationships between specific events and ensuing anxiety in a way that satisfies methodological requirements and is developmentally appropriate as well (see Weisz, 1978).

It is true that different individuals may reveal their anxiety through different behaviors; also, different observers may have difficulty achieving high agreement in their behavior observations. But these problems of behavior ratings should be surmountable. Clearly defining the behaviors to be observed and increasing the training time of observers should reduce problems of low interrater agreement. And total frequency scores on rating scales including the gamut of behaviors suggestive of anxiety should give good indi-

cations of the relative anxiety of individuals with widely varying anxiety manifestations.

The most accurate behavior rating scale would require observers who are able to concentrate fully on the occurrence of the target behaviors and to record the occurrence of these behaviors for later frequency analysis. It should be noted that in the typical clinical interview situation, the clinician would not be able to meet these requirements. However, a second observer could independently record the indicators in clinical situations in which precise anxiety measurements are desired, or the clinicians themselves could use the indicators of behavior rating scales as an aid in making less exact, more global ratings of the client's anxiety. In the latter case, the clinicians would not be using the scale in the prescribed way but would probably add objectivity to their observations by reference to it.

In short, behavior observation measures of anxiety seem to have clear advantages over other types of anxiety measures and relatively minor drawbacks. But despite the potential usefulness of such measures, we have found little systematic research on behavior indicators of anxiety in children: In fact, there appears to be no carefully validated behavior rating scale for anxiety in children. Perhaps the closest approach is that of Grossman (1968), who used observer ratings of anxiety with 6-year-olds. However, his scale consisted of a mixture of specific objectively observable behaviors (e.g., nail-biting) and general indicators requiring subjective inference (e.g., "reactions suggesting that the child was frightened"); as described by the author, the scale appeared to contain only six items, thus showing limited sensitivity to the range of behaviors through which children may display anxiety. Validity data consisted of only two nonsignificant correlation coefficients relating the behavior scale to the General Anxiety Scale for Children (Sarason, Davidson, Lighthall, & Waite, 1958), and to our knowledge, no independent studies have been done to support the use of the scale as an anxiety measure. Melamed and Siegel (1975) used their Observer Rating Scale of Anxiety in a study on the reduction of anxiety through filmed modeling in 4- to 12-year-old children facing surgery. Again, only very limited validity data on this scale are available. Also, only four items of the scale were published, and these represent an exceedingly modest sampling of the possible behavioral manifestations of anxiety.

In the present study, a much more detailed list of behavior items was assembled by means of a systematic search of the anxiety literature. This list was used to form the Preschool Observation Scale of Anxiety (POSA). The interobserver reliability of the POSA was assessed by two independent judges. The validity of the scale was determined (a) through assessment of its relation to three independent inventory measures of anxiety and (b) through an experimental manipulation of stressors. Significant positive relations between the POSA scores and the inventory scores were anticipated, but these relations were expected to be modest, since (a) the inventory measures and the POSA were designed to assess somewhat different aspects of anxiety, and (b) the inventory measures relied on parents' and teachers' judgments, whereas the POSA involved relatively objective observations by trained observers. The experimental manipulation involved a contrast between an initial mother-absent condition and a subsequent mother-present condition. Because the format and experimental procedure would be more familiar to the children in the second session, and because mothers would be present, it was predicted that children would show lower levels of anxiety in Session 2 than in Session 1, and thus score lower on the POSA.

Method

Subjects

The 36 Cornell University Nursery School children who formed the sample ranged in age from 32 to 59 months ($M = 47$ months; $SD = 7.7$ months). Mean Hollingshead (Note 1) socioeconomic status (SES) was 1.47 (1 = highest; 7 = lowest; $SD = .94$). There were 21 girls and 15 boys.

Procedure

The Preschool Observation Scale of Anxiety. The first step in devising the scale was a systematic search of the *Education Index* (1929–1976), *Psychological Abstracts* (1927–1976), *Resources in Education/Re-*

search in Education [1966 (year or origin) to 1976], and *Current Index to Journals in Education* [1969 (year of origin) to 1976], for studies using or mentioning behavioral indicators of anxiety in children or adults. Next, three child-clinical psychologists examined a list of items based on the literature search to suggest additions, to make minor modifications, and to give their final approval to the scale. Some behavior indicators suggested by the literature or the clinicians (e.g., hand perspiration) were not included in the final scale because pilot testing indicated that they were too difficult to observe accurately and reliably. Table 1 includes a description of each behavioral indicator of the final POSA, along with a reference to the article(s) supporting the use of each indicator in the scale.

Parent and teacher questionnaires: Independent anxiety measures. Several independent measures of anxiety were obtained for the purpose of validating the POSA. One was a questionnaire, the Parent Anxiety Rating Scale (Doris, McIntyre, Kelsey, & Lehman, 1971), completed by each child's parents and comprised of six questions about the child's separation anxiety (PARSEP) and 19 questions about the child's general anxiety (PARGEN). Two weeks before school began, this questionnaire was sent to the children's parents, who completed it by the beginning of the nursery school year. A second independent measure of anxiety was the Teachers' Separation Anxiety Scale (TSAS; Doris et al., 1971), comprised of 11 items about the child's reaction to separation from his/her mother or father when left at the nursery school at the beginning of the school day. One of two teachers (one for each of two nursery school sessions) rated the child on each of the first 10 consecutive days of the child's attendance at nursery school for the year. Ratings were made at the end of each school day and pertained to the period from the child's arrival at the nursery school with parent to the parent's departure. If the child was not delivered to the nursery school by his/her parent, the child was not rated that day, and his/her TSAS score was prorated.

Experimental manipulation of anxiety. In addition to obtaining questionnaire scores of the children's anxiety, we used a manipulation designed to create differing levels of anxiety in two experimental sessions, both involving cognitive tests. The first session was expected to be more anxiety arousing, since it involved an unfamiliar adult who individually tested each child in an unfamiliar setting during one of the first few days of the nursery school year. In contrast, the second session occurred approximately 2 weeks following the first session, after the children had had a chance to settle in to their new surroundings; furthermore, the child's mother was present at the second session, along with the same examiner who had tested the child at the first session. Finally, the children were given the same three tests in the second session that they had already taken in the first session.

Tests were given because it was felt that an evaluative atmosphere would make the experimental situation more anxiety arousing, especially in the first session when the materials would be unfamiliar. Also, performance on the cognitive tests by the high-anxious children compared to the low-anxious children was of interest, since existing evidence is in conflict over this question. Some evidence suggests that high anxiety should interfere with test performance, especially when the tests are fairly difficult ones for the subjects (e.g., Feldhusen & Klausmeier, 1962; Tamaroff, 1976; Young & Brown, 1973). However, other evidence (e.g., Denny, 1966; Hodges & Durham, 1972; Katahn, 1966; Spielberger, 1966) suggests that anxiety may have facilitating effects on performance for children of higher socioeconomic and intellectual levels. (See other conflicting evidence in Fischer & Awrey, 1973; Mazzei & Goulet, 1969.) Since, according to the evidence, anxiety might lead to performance enhancement in some children and performance decrements in other children, no prediction was advanced regarding the relation between anxiety and task performance in the present study. However, the relation between task performance and anxiety scores was reported in an effort to shed light on the controversy just described.

Global anxiety self-ratings and ratings by the examiner. In the Introduction to this article, we emphasized the disadvantage of both self-reports by children and global (and thus subjective) ratings by adults. To determine whether our negative assessment was correct, we included two such measures in the present study, so that their effectiveness might be contrasted with that of the more specific and presumably more objective POSA. At the end of each of the two testing sessions, the examiner rated the child's general anxiety level during the session on a scale of 1 to 6. Also, a teacher asked each child to choose one of six pictures depicting progressively more fearful facial expressions to show how anxious the child felt during the testing session. The teacher rather than the examiner asked this question, since it was felt that the children would be more candid about their feelings with their teacher. The wording of the question was the following:

> These pictures show a picture of a child who is more and more scared as you go from this top picture to the bottom picture here [point]. You see, up here the child is not scared at all [point]; here [second picture from top] the child is a little more scared but still pretty happy; here [third from top], he's getting more scared; and here [bottom], very, very scared. Which picture would show how scared you felt when you went downstairs to the testing room with that lady? One of these up here where the child isn't scared at all, or one of the bottom ones showing a child who is more and more scared?

The observation periods. At the beginning of the first session, the examiner, a 26-year-old experienced female teacher, approached each child individually in the nursery school and told the child

Table 1
Items of the Preschool Observation Scale of Anxiety and Interrater Agreement During the First 10 Minutes of the First Session

Item	Range of frequency	Interrater reliability[a]	Concordance 1	Concordance 2
1. Physical complaint: Child says he or she has a headache, stomachache, or has to go to the bathroom (B,C,D,E,G,K,Q)	0–0	nc[b]	nc	nc
2. Desire to leave: Child says he or she wants to leave the testing room or makes excuses about why he or she must leave; desire or "need" to leave must be explicit (D,E,P,Q).	0–5	.99*	.99	.46
3. Expression of fear or worry: Child complains about being afraid of or worried about something; must use the word "afraid," "scared," "worried," or a synonym (F).	0–0	nc	nc	nc
4. Cry: Tears should be visible (G,I,K,P,Q).	0–2	nc	nc	nc
5. Scream (P).	0–0	nc	nc	nc
6. Whine or whimper (G,P).	0–2	.69*	.97	.17
7. Trembling voice (F,G,I,M).	0–1	nc	nc	nc
8. Stutter (A,F,G,H,M,O).	0–0	nc	nc	nc
9. Whisper: Child speaks softly, without vocal cords; should not be a playful whisper (E,G).	0–11	.67*	.87	.39
10. Silence to one question in the interval (E).	0–3	.74*	.98	.50
11. Silence to more than one question in the interval (E).	0–5	.82*	.95	.22
12. Nail-biting: Child actually bites his or her nails in the testing room (F,G,I).	0–3	.83*	.98	.46
13. Lip-licking: Tongue should be visible (G).	0–13	.94*	.89	.63
14. Fingers touching mouth area: not counted if bites nails while touching mouth.	0–17	.96*	.91	.79
15. Sucking or chewing object: not fingernails (G,P).	0–1	.47*	.99	.33
16. Lip contortions.	0–13	.67*	.76	.44
17. Trembling lip (B).	0–0	nc	nc	nc
18. Gratuitous hand movement at ear area (G,I,J,N,P).	0–3	.56*	.96	.39
19. Gratuitous hand movement at top of head (G,I,J,N,P).	0–3	.62*	.95	.30
20. Gratuitous hand movement at an object separate from body or at a part of clothing separate from body (G,I,J,N,P).	0–18	.81*	.76	.61
21. Gratuitous hand movement at some part of body (not ear, hair, mouth, or genitals) (G,I,J,N,P).	0–12	.82*	.75	.56
22. Gratuitous hand movement (N).	0–9	.52*	.80	.35
23. Gratuitous leg movement (M, N).	0–20	.89*	.81	.77
24. Gratuitous foot movement: below ankles, distinguish from foot merely moving along with leg (M,N).	0–20	.49*	.66	.56
25. Trunk contortions (e.g., arching back) (N).	0–20	.89*	.81	.68

Table 1 (*continued*)

Item	Range of frequency	Interrater reliability[a]	Concordance 1	Concordance 2
26. Rigid posture: Part of body is held unusually stiff or motionless for the entire 30-sec interval (B,G,N).	0–11	.77*	.94	.16
27. Masturbation: touches genital area (K).	0–3	.81*	.99	.50
28. Fearful facial expression (E,I).	0–4	.92*	.98	.26
29. Distraction: Must be indicated by a verbal reminder by the examiner to the child to pay attention (B,I,L,P).	0–7	.86*	.94	.38
30. Avoidance of eye contact: Examiner should be having clear trouble making eye contact with child (G, M).	0–1	−.04[c]	.98	.00
Total	26–107	.78*	.92	.58

Note. Letters in parentheses following each item description refer to studies suggesting that item as an anxiety indicator. The following code was used: A: Boland (1953); B: Buss, Wiener, Durkee, and Baer (1955); C: Cowen, Zax, Klein, Izzo, and Trost (1965); D: Endler, Hunt, and Rosenstein (1962); E: Fink (1956); F: Grossman (1968); G: Insel and Spencer (1972); H: Kasl and Mahl (1965); I: McReynolds (1965); J: Melamed and Siegel (1975); K: Miller, Barrett, Hampe, and Noble (1971); L: Nottelman (1975); M: Paul (1966); N: Raskin (1962); O: Santostefano (1960); P: Tamaroff (1976); Q; Wolff (1969). Items 14 and 16 were suggested by our clinician consultants. The range of frequency column gives the lowest score obtained by any subject and the highest score obtained by any subject. The interrater reliability column gives the correlation coefficients for the Pearson correlations between the scores of the two observers. The Concordance 2 column gives the interrater agreements on the occurrence of indicators in pairs of adjoining intervals (the number of agreements divided by the sum of agreements plus disagreements). The Concordance 1 column gives the interrater agreement on both the occurrence and nonoccurrence of indicators in pairs of adjoining intervals (again, the number of agreements by the sum of agreements plus disagreements).

[a] n = 33 for interrater reliability and concordance calculations.

[b] nc = not calculated due to infrequent occurrence (i.e., no children or only one child had nonzero scores on the indicator).

[c] Based on only two children with nonzero scores.

* $p < .001$.

that she wanted him or her to go with her to another room "to do some tests."[1]

The testing room was 5.7 m × 3.5 m, with one-way mirrors on top of its four walls and a microphone about 30 cm above the child's head for transmitting sounds from the testing room to the observation areas. A male and a female observer, both in their early 30s, sat behind the one-way mirrors at a distance of about 1.2 m from the child. The observers sat around a corner from each other, and were separated by about 1.5 m and a partition. Each observer spoke softly into a tape recorder each time he or she observed the child emit one of the 30 behavioral indicators. The two observers could not hear each other speaking. Also, the observers were blind to the independent anxiety ratings of the children whom they observed and to the fact that lower anxiety ratings were expected for the second session in comparison to the first session.

A standard time-sampling procedure involving 30-sec intervals was used. A Davis Scientific Instruments General Purpose Time Interval Generator (Model B501) emitted to each observer a beep and a red-light flash at the 30-sec intervals. Aside from preventing observers from habituating to the behavioral indicators, this procedure was used to aid data analysis: Behaviors were given a score of 1 for each interval in which they occurred.

During the testing session, the examiner gave each child the three tests (Digits, Blocks, and Sentences) in standard form and in as neutral a manner as possible. The examiner was instructed to try to keep the children in the testing room for at least 10 minutes. The Digits test was taken from the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1968) and involved repeating a series of orally presented digits from memory. Sentences were taken from Tamaroff's (1976) adaptation of the Sentences subtest of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967). This test requires the child to repeat increasingly complex orally presented sentences from memory. Finally, Blocks was taken from Tamaroff's (1976)

[1] Two children refused to do any tests in the first session, and one refused to do any in the second session.

adaptation of the Block Construction Test of the Yale Scale of Child Development. It requires the child to copy visible block constructions with separate blocks, and it is timed.

In the second testing session, the same procedure was followed, except for two variations. First, the child was accompanied to the testing room by his or her mother as well as by the examiner, and the mother sat unobtrusively reading a magazine in the corner of the testing room during the session. Second, only the female observer was behind the one-way mirrors, since the second observer was used just during the first session to obtain interrater reliabilities. Of the 36 children in Session 1, 32 participated in Session 2. One child refused, and the parents of 3 children were unable to participate. Technical difficulties with recording equipment further reduced the sample (for whom complete data on both sessions were available) to 29.

Training for the observers. Both observers spent 2–3 hours memorizing the behavioral indicators and 6 hours in observation-training sessions. In training, the observers watched videotapes of three children being tested as other children would be tested in the actual experiment. Training also involved pilot observations of three children in the actual setting and conditions of the experiment. Finally, the observers spent approximately 2 hours discussing disagreements in their observations and ways to achieve better consensus. The detailed description of the items of the POSA given in Table 1 includes all of the details that the observers devised in order to maximize agreement.

Results

Scores for all subjects on the POSA were calculated by using the number of 30-sec intervals in which a given indicator occurred during the first 10 min (20 intervals) of the sessions. The examiner had been instructed to try to keep each subject in the testing room and working for a full 10 min; thus, a 10-min interval was chosen as the target time, in part, because of concern that the experimenter's behavior might have changed significantly after the 10-min period had expired.[2]

Interrater Reliabilities

To assess interrater reliability, Pearson correlations between the ratings of the two observers were calculated using the Session 1 scores for each indicator separately and for the sum of all 30 indicators (see Table 1 third column from the right). The most important correlation coefficient, that for the 30 indicators together, was .78 ($p < .001$). The intraclass correlation was .77 ($p < .001$).

Even though the preceding analysis answered the central question about reliability of overall POSA and individual indicator scores, we also sought to learn the level of specificity at which observer agreement took place. Toward this end we used a demanding procedure designed to gauge the degree of concordance between observers *within* observation intervals. The procedure involved grouping every 2 adjoining intervals (resulting in 19 interval groups) and counting how often the observers agreed or disagreed as to whether an indicator occurred within each interval group. Adjoining rather than single intervals were used for this agreement measure, since the observers sometimes reported the same behavior at slightly different times so that the interval cutoff occurred between their reports. Using this method of measurement, a quotient of concordance was calculated by dividing the number of agreements between the observers by the number of their "agreements" plus "disagreements" for each indicator separately and for the total of all 30 indicators. Table 1 (last two columns) shows the results of these calculations (a) when agreements between observers that a given indicator did *not* occur were included in the "agreement" score and (b) when such negative agreements were excluded from the agreement score. Considering the rigorous nature of the procedure (particularly b above), the percentages shown in Table 1 represent rather substantial agreement between observers at the level of brief observation intervals.

---

[2] The alternate scoring method of dividing the frequency of intervals in which each indicator occurred during the total session by the number of intervals in the total session would have posed an additional problem, in that fatigue and familiarity could have affected the scores of children having longer sessions. Unfortunately, using the 10-minute cutoff, all subjects were not engaged in the same activities during the target time, but this seemed to be a less significant consideration than those noted above.

Table 2
Correlations Among Anxiety Measures and Test Scores

	2	3	4	5	6	7	8	Sentences
1. POSA	.37*	.30*	.47**	−.12	−.02	−.04	.38*	−.16
2. PARSEP		.41**	.36*	.01	.06	.05	.28	.24
3. PARGEN			.22	−.18	.07	.03	.05	.05
4. TSAS				−.05	.19	−.07	.07	−.18
5. Self-rating					.14	−.45**	−.35*	−.24
6. Examiner rating						−.24*	−.10	−.13
7. Blocks							.42**	.62***
8. Digits								.72***

Note. POSA = Preschool Observation Scale of Anxiety; PARSEP = questions about the child's separation anxiety; PARGEN = questions about the child's general anxiety; TSAS = Teachers' Separation Anxiety Scale.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Correlations With Independent Measures of Anxiety

As an initial step in assessing the validity of the POSA, the correlations of the POSA with the PARSEP, PARGEN, and TSAS were calculated using data from the first session (since correlations with data of the second session would have been less meaningful due to the mothers' presence). As Table 2 indicates, all three correlations were significant. Other correlations noted in the table are also of interest. Note that the self-rating and examiner ratings were not correlated with one another or with POSA, PARSEP, PARGEN, or TSAS scores. Thus, even though the POSA did meet the first set of validity criteria that we established (i.e., significant correlations with the three inventory measures), the self-ratings and examiner ratings did not. The significant negative correlations between self-ratings and two of the cognitive tests suggests that self-ratings may have been influenced by the children's awareness of the quality of their test performance.

To assess the contributions of individual POSA indicators to the correlations between the scale and the three inventory measures, the correlations of each of the 30 items with the three inventory scores were calculated. Most of these individual correlations were nonsignificant, suggesting that the predictive power of the indicators lies mainly in their combination with one another. Those correlations that did attain statistical significance were lip contortions with PARSEP (.39, $p < .01$) and TSAS (.43, $p < .01$), gratuitous hand movement in the ear area with PARGEN (.32, $p < .05$), gratuitous hand movement at the top of the head with PARSEP (−.38, $p < .01$), gratuitous hand movement toward object with TSAS (.53, $p < .001$), gratuitous hand movement at other body part (−.31, $p < .05$), gratuitous leg movement with PARSEP (.34, $p < .05$) and TSAS (.35, $p < .05$), gratuitous foot movement with PARSEP (.35, $p < .05$) and TSAS (.46, $p < .01$), trunk contortions with PARSEP (.29, $p < .05$) and TSAS (.42, $p < .01$), and masturbation with PARGEN (.39, $p < .01$) and TSAS (.58, $p < .001$).

Anxiety Scores in the First and Second Sessions

As a second way of assessing the validity of the POSA, the mean POSA score for all children in the first session was compared with the mean score for all children in the second session (designed to be less anxiety producing than the first). As predicted, the children obtained significantly higher POSA scores in the first than in the second session, $t(28) = 2.53$, $p < .01$ (one-tailed). Of the 29 children, 22 had higher scores in the first than in the second session, $\chi^2(1) = 6.76$, $p < .01$. Further comparisons between the first and second sessions were made for children

Table 3
Comparisons of POSA Scores Between Sessions

Group	Session 1		Session 2			
	M	*SD*	*M*	*SD*	*t*	*df*
All children	61.4	21.3	51.3	14.0	2.53*	28
High PARSEP scorers	70.0	19.6	49.7	15.5	4.35**	14
High PARGEN scorers	64.3	19.2	49.0	14.3	3.17*	15
High TSAS scorers	61.4	22.4	47.6	14.0	2.66*	17

Note. POSA = Preschool Observation Scale of Anxiety; PARSEP = questions about the child's separation anxiety; PARGEN = questions about the child's general anxiety; TSAS = Teachers' Separation Anxiety Scale.
* $p < .01$, one-tailed.
** $p < .001$, one-tailed.

who scored above the median on the PARSEP, PARGEN, and TSAS, since it was thought that these children might be especially sensitive to the situational manipulations of stressors. Again, the high scorers on the three questionnaire scales showed significantly more behavioral indicators of anxiety in the first than in the second session. The results for these analyses are shown in Table 3. Note that in each of the three groups, the magnitude of the Session 1 – Session 2 difference is greater than for the entire sample. In fact, of the four groups, children rated by their parents as high in *separation* anxiety showed the highest mean POSA scores during Session 1 with mothers absent and the largest Session 1 – Session 2 difference—more than doubling the difference shown by the full sample.

Session 1 – Session 2 differences were also calculated for the children's self-ratings and the examiner ratings. The change in the children's self-ratings from Session 1 to Session 2 indicates a nonsignificant increase in anxiety, whereas the change in the examiners' ratings indicates a highly significant decrease. Anxiety ratings on the children's 6-point picture scale averaged 2.22 in Session 1 and 2.59 in Session 2 ($p = .28$). Ratings on the examiner's 6-point scale averaged 3.07 in Session 1 and 2.00 in Session 2, $t(29) = 3.71$, $p < .001$. Thus, both the POSA and examiner ratings of anxiety met the second validity criterion, that is, significantly higher scores in Session 1 than in Session 2.

Finally, Session 1 – Session 2 differences for each of the 30 POSA indicators were calculated in an effort to gauge the contribution of the individual indicators to the overall sessions difference in total POSA scores. As was true in the correlational analysis reported earlier, most individual item effects were nonsignificant, suggesting that the discriminative power of the indicators lies principally in their combination with one another. However, there were five indicators that had significantly different frequencies in Sessions 1 and 2 (all in the predicted direction): silence to one question ($p < .05$), touching mouth area ($p < .01$), gratuitous arm movement ($p < .001$), trunk contortions ($p < .001$), and rigid posture ($p < .05$).

Discussion

The results of the present study support the use of the POSA as a measure of anxiety in young children. Independent judges achieved strong agreement on both total POSA scores and total scores for most of the 30 individual indicators. The large number of indicators used seemed to interfere with interobserver concordance at the micro level of 1-min observation blocks (at least by the most rigorous method of analysis); and this suggests that the two observers may have differed frequently in the particular behavioral incidents that they observed. Yet, this is a relatively trivial limitation considering the high interobserver correlations obtained for total POSA and individual indicator scores. In sum, the difficulties introduced by requiring observers to watch for

30 indicators appear to be outweighed by the value of including a variety of potential anxiety manifestations in order to capture indications of anxiety in individuals with differing expressive styles.

Two types of evidence support the view that the POSA yields a valid index of anxiety. As predicted, the scale was significantly correlated with all three inventory measures of anxiety. Also as predicted, the POSA yielded significantly higher scores during a presumably high-anxiety test session than during a session designed to provoke less anxiety. The data generally supported our original belief that broad-based measures such as the POSA would outperform simpler approaches such as self-ratings by young children and global judgments by examiners. Children's self-ratings and examiner ratings did not correlate significantly with any of the three inventory measures, with POSA scores, or with each other. Self-ratings showed very slight differences (and in the wrong direction) between Sessions 1 and 2; however, Session 2 ratings by the examiner were significantly lower than Session 1 ratings. It is uncertain whether this latter finding derived from an expectation by the examiner that children's anxiety would be lower in the second session with mothers present. However, whatever the basis for the finding, it constitutes the only bit of evidence supporting the use of either self- or examiner ratings. This pattern of findings seems to indicate the superiority of structured observations of carefully delineated behaviors over global, unstructured, and thus subjective ratings, though this conclusion must be qualified by the fact that the structured observations were made by trained observers, whereas the global ratings were not.

As indicated in the Introduction, the relation between anxiety and problem-solving performance appears to be quite complex. In the present study none of the inventory measures was significantly related to test performance, and the POSA was significantly related only to Digits performance ($r = .38$). So, anxiety, as measured by the POSA and inventory scores, did not appear to have a debilitating effect on test performance for children in the present study, and may have had an enhancing effect on one test, perhaps due to factors related to the high socioeconomic status of the present subjects.

Further research with the POSA should include investigations of the scale's capacity to reflect the effects of stressors other than those devised in the present study. Children varying more widely in age and other demographic characteristics than those of the present sample should be included. And, more importantly, there is a need to assess the scale's usefulness with clinical populations of children who (unlike those of the present sample) suffer from pronounced behavior problems. In addition, research using longer observation periods than the 10-minute-plus sessions of the present study could be useful; some behaviors that did not occur in the relatively brief sessions of the present investigation might prove to be useful indices of anxiety if children were given a lengthier opportunity to display them. Even the behaviors that occur infrequently might well prove to be potent anxiety indicators when they do appear. Finally, the degree of anxiety is apt to be reflected not only by the *frequency* of anxiety behaviors but also by their *intensity*. Although intensity may be difficult to quantify, its potential for increasing the precision of anxiety measurement would seem to justify efforts in this direction.

For the present, however, the POSA represents a potentially useful approach to the assessment of anxiety in children. Given the sensitivity of the POSA to situational variations in stressors, the scale itself should be particularly useful to reseachers concerned with the interplay between specific situational conditions and affective states (e.g., Spielberger et al., 1972). In addition, the POSA could contribute usefully to our understanding of sex differences in anxiety manifestations at various ages (see Maccoby & Jacklin, 1974, pp. 182–190), and to our capacity to evaluate therapeutic techniques for children (see Achenbach, 1974, pp. 606–650). Research on these topics, focusing on anxiety states in children, clearly must circumvent the problems of projective techniques, self-reports, global and subjective

observer reports, and physiological measures, all outlined in the Introduction. The evidence presented in the present study suggests that instruments such as the POSA may provide logical, valid, and reliable alternatives to these more traditional approaches.

Reference Note

1 Hollingshead, A. B. *Two-factor index of social position.* Unpublished manuscript, Yale University, 1957.

References

Achenbach, T. M. *Developmental psychopathology.* New York: Ronald Press, 1974.

Boland, J. L. A comparison of stutterers and nonstutterers on several measures of anxiety. *Speech Monographs,* 1953, *20,* 144.

Buss, A., Wiener, M., Durkee, A., & Baer, M. The measurement of anxiety in clinical situations. *Journal of Consulting Psychology,* 1955, *19,* 125–129.

Cowen, E. L., Zax, M., Klein, R., Izzo, L. D., & Trost, M. A. The relation of anxiety in school children to school record, achievement, and behavioral measures. *Child Development,* 1965, *36,* 685–695.

Denny, J. P. Effects of anxiety and intelligence on concept formation. *Journal of Experimental Psychology,* 1966, *72,* 596–602.

Doris, J., McIntyre, A., Kelsey, C., & Lehman, E. Separation anxiety in nursery school children. *Proceedings of the 79th Annual Convention of the American Psychological Association,* 1971, *79,* 145–146. (Summary)

Endler, N. S., Hunt, J. M., & Rosenstein, A. J. An S-R Inventory of Anxiousness. *Psychological Monographs,* 1962, *76*(17, Whole No. 536).

Feldhusen, J. F., & Klausmeier, H. J. Anxiety, intelligence, and achievement in children of low, average, and high intelligence. *Child Development,* 1962, *33,* 403–409.

Fink, R. M. How children tell us they are afraid. *Childhood Education,* 1956, *32,* 264–266.

Fischer, D. G., & Awrey, A. Manifest anxiety, test anxiety, and intelligence in concept formation. *Journal of Social Psychology,* 1973, *89,* 153–154.

Grossman, B. D. Anxiety as a factor in the child's responsiveness to social reinforcement. *Journal of Genetic Psychology,* 1968, *112,* 165–174.

Hodges, W. F., & Durham, R. L. Anxiety, ability, and digit span performance. *Journal of Personality and Social Psychology,* 1972, *24,* 401–406.

Insel, S. A., & Spencer, T. D. *Emotional concomitants of reading readiness thinking in kindergarten.* Washington, D. C.: National Center for Educational Research and Development, December 1972. (ERIC Document Reproduction Service No. ED 072 871)

Kasl, S. V., & Mahl, G. F. The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology,* 1965, *1,* 425–433.

Katahn, M. Interaction of anxiety and ability in complex learning situations. *Journal of Personality and Social Psychology,* 1966, *3,* 475–479.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. *Illinois Test of Psycholinguistic Abilities.* Urbana: University of Illinois Press, 1968.

Lazarus, R. S. *Psychological stress and the coping process.* New York: McGraw-Hill, 1966.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences.* Stanford, Calif.: Stanford University Press, 1974.

Mazzei, J., & Goulet, L. R. Test anxiety and intelligence in concept formation. *Psychological Reports,* 1969, *24,* 842.

McReynolds, P. On the assessment of anxiety: I. by a behavior checklist. *Psychological Reports,* 1965, *16,* 805–808.

McReynolds, P. The assessment of anxiety: A survey of available techniques. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 1). Palo Alto, Calif.: Science and Behavior Books, 1968.

Melamed, B. G., & Siegel, L. J. Reduction of anxiety in children facing hospitalization and surgery by use of filmed modeling. *Journal of Consulting and Clinical Pychology,* 1975, *43,* 511–521.

Miller, L. C., Barrett, C. L., Hampe, E., & Noble, H. Revised anxiety scales for the Louisville Behavior Check List. *Psychological Reports,* 1971, *29,* 503–511.

Nottelman, E. *Test anxiety and off-task behavior in evaluative situations.* Washington, D.C.: National Institute of Education, April 1975. (ERIC Document Reproduction Service No. ED 113 022)

Paul, G. L. *Insight vs. desensitization in psychotherapy.* Stanford, Calif.: Stanford University Press, 1966.

Phillips, B. N., Martin, R., & Meyers, J. Interventions in relation to anxiety in school. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 2). New York: Academic Press, 1972.

Raskin, A. Observable signs of anxiety or distress during psychotherapy. *Journal of Consulting Psychology,* 1962, *26,* 389.

Santostefano, S. Anxiety and hostility in stuttering. *Journal of Speech and Hearing Research,* 1960, *3,* 337–347.

Sarason, S. B. The Measurement of anxiety in children: Some questions and problems. In C. D. Spielberger (Ed.), *Anxiety and behavior.* New York: Academic Press, 1966.

Sarason, S. B., Davidson, K. S., Lighthall, F. F., & Waite, R. R. Rorschach behavior and the performance of high and low anxious children. *Child Development,* 1958, *29,* 277–285.

Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. *Anxiety in elementary school children.* New York: Wiley, 1960.

Spielberger, C. D. The effects of anxiety on complex learning and academic achievement. In C. D. Spielberger (Ed.), *Anxiety and behavior.* New York: Academic Press, 1966.

Spielberger, C. D. Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1). New York: Academic Press, 1972.

Spielberger, C. D., et al. *Needed research on stress and anxiety. A special report of the OSOE-sponsored grant study: Critical appraisal of research in the personality-emotions-motivation domain* (IBR Report No. 72-10). Washington, D.C.: Office of Education, July 1972. (ERIC Document Reproduction Service No. ED 113 649)

Tamaroff, M. H. Separation anxiety and cognitive functioning in nursery school children. (Doctoral dissertation, Cornell University, 1975) *Dissertation Abstracts International,* 1976, *36,* 5820-B. (University Microfilms No. 76-9569)

Wechsler, D. *Wechsler Preschool and Primary Scale of Intelligence.* New York: Psychological Corporation, 1967.

Weisz, J. R. Transcontextual validity in developmental research. *Child Development,* 1978, *49,* 1–12.

Wolff, S. *Children under stress.* London: Penguin Press, 1969.

Young, F. A., & Brown, M. Effects of test anxiety and testing conditions on intelligence test scores of elementary school boys and girls. *Psychological Reports,* 1973, *32,* 643–649.

Received August 17, 1977 ■