Development of the Children’s Apperceptive Story-Telling Test

Mary F. Schneider and Jan Perney
National-Louis University

This study reports the development of the Children’s Apperceptive Story-Telling Test (CAST), which uses story telling to evaluate the emotional functioning of children ages 6–13. The CAST yields profile scores on 4 major factors (Adaptive, Nonadaptive, Immature, and Uninvested) and 15 adaptive, nonadaptive, and problem-solving scales. The CAST is an attempt to address the psychometric issues present with traditional apperceptive techniques in that it offers a nationally representative standardization sample (N = 876); the development of an objective Likert-type scoring system based on an Adlerian theoretical model; colored picture stimuli reflective of contemporary family, peer, and school contexts; separate but equivalent stimulus pictures for boys and girls that have been subjected to studies on ethnic sensitivity; and demonstrated viable validity and reliability with school-based behavior-disordered children (N = 322).

Story telling has a long tradition as a psychological technique for diagnosis as well as for therapeutic intervention (Exner, 1976; Mills & Crowley, 1986; Vane, 1981). Apperceptive techniques, which have included the Thematic Apperception Test (Murray, 1971) and the Children’s Apperception Test (Bellak & Bellak, 1975) use story telling as a diagnostic tool. A client creates stories from picture stimuli, and these stories help the examiner to draw psychological inferences regarding the client’s beliefs and potential behavior. Although research regarding the use of apperceptive techniques supports their popularity as diagnostically and interview tools in school and clinical settings (Brown & McGuire, 1976; Goh, Teslow, & Fuller, 1981), the technique has been criticized for a lack of psychometric rigor (Anastasi, 1982; Batsche & Peterson, 1983; Gittelman, 1980) and diagnostic usefulness (Gittelmann-Klein, 1978).

The need exists for a valid, reliable, theory-based, objectively scored, apperceptive instrument for children that is standardized on a representative sample of well-adjusted and behavior-disordered individuals. A new instrument, the Children’s Apperceptive Story-Telling Test (CAST; Schneider, 1989), designed for children ages 6 through 13 years, addresses this need for psychometric rigor. The nationally standardized instrument consists of 17 colored stimulus picture cards (parallel sets for boys and girls are provided) designed to elicit relevant and contemporary family, school, and peer issues. Apperceptive stories generated by children in response to the picture cards are objectively scored on the basis of an Adlerian theoretical model. The purpose of this study was to examine the construct validity and interrater, intrarater, test–retest, coefficient alpha, and split-half reliability of the CAST with samples of well-adjusted and behavior-disordered children.

Method

The stimulus picture cards of the CAST were constructed to reflect the Adlerian life task areas of family, peer relations, and school (Adler, 1963). To maximize identification, the stimulus picture cards reflect typical contemporary childhood experiences. Because a basic tenet of individual psychology is that the family is the first and primary social unit—the training ground for the individual’s beliefs—the pictures were designed to elicit substantial thematic material relative to the family life task. An initial set of 25 experimental pictures were pilot-tested in a series of studies designed to select the final set of stimulus picture cards as well as refine the scoring system. The final set of 17 cards (two parallel sets of 14 cards for boys and 14 cards for girls and 3 cards common to both sets) were selected for their ability to (a) reflect diverse thematic categories and content populars (minimizing stereotypic content), (b) evidence the least amount of response delays or refusals, and (c) elicit self-validation of thematic content (i.e., subject volunteers direct information that the content of story parallels his or her own life experience).

The CAST scoring system consists of four adaptive thematic scales (Instrumentality, Interpersonal Cooperation, Affiliation, and Positive Affect), five nonadaptive thematic scales (Inadequacy, Alienation, Interpersonal Conflict, Limits, and Negative Affect); and six problem-solving scales (Positive Preoperational, Positive Operational, Refusal, Unresolved, Negative Preoperational, and Negative Operational). In addition, six thematic indicators (Sexual Abuse, Substance Abuse, Divorce, Hypothetical Thought, Emotionality, and Self-Validation) are provided not as scales but to highlight apperceptive material essential to interpreting the child’s responses to the CAST.

The CAST also addresses sensitivity to ethnicity in assessment (Malgady, Costantino, & Rogers, 1984; Padilla, 1979) by presenting stimulus figures reflective of different races as well as by including Black, Hispanic, and Asian children in the standardization sample. The experimental stimulus cards were subjected to ethnic evaluation by 40 children (20 boys and 20 girls) ages 6 through 13 years. The 40 raters included 10 White, 10 Black, 10 Hispanic, and 10 Asian children. The children were asked to tell what race they perceived the figures in the cards to be. These initial ratings were used in the drafting of the final revised color pictures. The final set was again rated by 48 children (24 boys and 24 girls) ages 6 through 13 years. The 48 raters included 12 White, 12 Black, 12 Hispanic, and 12 Asian children. The final ratings...
The standardization sample of 876 children was representative of the population of elementary school students in the United States. The children ranged in age from 6 through 13 years, with 87% attending public schools and 13% attending private schools. The sample was selected from 19 states located in diverse regions of the United States and, as Table 1 indicates, reflected the composition of the U.S. population in terms of race, gender, parental educational level, community size, and geographic location (U.S. Bureau of the Census, 1983). School districts and special education cooperatives throughout the United States with populations reflective of the targeted standardization sample plan were requested to participate in the standardization process. Approximately 8,000 permission forms were distributed to the children in the schools participating in the standardization study. From the set of approximately 5,000 completed and returned permission forms, subjects were selected who reflected the demographic variables for the standardization sample. When several subjects met the standardization sample plan criterion, the final subjects were randomly selected from the set of eligible individuals.

In addition to the nationally representative sample of 876, another 322 students identified as behavior disordered (BD) were tested. The behavior-disordered students provided a comparison group used to establish the construct validity of the CAST. Table 1 also shows the representation of age, gender, race, and parental educational level of the behavior-disordered students selected for the sample. The smaller number of 6-year-olds appeared to be the result of a tendency not to classify first graders as behavior disordered. The ratio of boys to girls in the sample is consistent with the 4:1 ratio of boys to girls present in behavior-disordered populations in general (Kauffman, 1985).

### Administration and Scoring of the CAST

All subjects were tested individually by the principal researchers, trained psychologists, or trained researchers within their school environment. After developing rapport with the subject, the examiner administered the CAST in accordance with the standard instructions. The examiner asked each child to make up a story for each of the 17 cards of the CAST. The child was asked to tell what was happening in the story, how the people were feeling, and what they were saying. The child was also asked to let the story have a beginning, a middle, and an end.

Subjects responded to the administration directions with stories for the 17 picture cards. These stories were tape-recorded to ensure accuracy and to maintain natural linguistic flow relative to the stories' contents. The examiner, freed from the transcription task, could maintain rapport and make additional behavioral observations. Administration time ranged from 20 to 45 min. The tapes were transcribed by a trained transcription team. Transcriptions were checked for accuracy during the training period and spot-checked after initial training to ensure transcription quality.

Scoring for the standardization sample and the behavior-disordered sample was conducted by a trained scoring team consisting of Mary F. Schneider and three trained research students. Scorers were trained on sample profiles until interrater reliability for 25 profiles reached the .90s. After initial training, the scoring team worked at one location, with all questionable scoring issues flagged for team discussion. Throughout the scoring process, each profile was randomly checked to ensure that the quality of scoring was maintained. The scoring team was blind to the subjects' identities. Although detailed scoring directions and examples are provided in the manual, the following is an example of the scoring for one of the problem-solving scales, Positive Operational Problem-Solving.

Positive Operational Problem-Solving has three levels and can be worth 1, 2, or 3 points. Positive Operational Problem-Solving represents a solution to a problem that is positive and reflects the presence of cognitive operations—operations that are either concretely temporal (1 point), functional (2 points), or abstract (3 points). The concrete temporal (1 point) type of Positive Operational Problem-Solving involves positive endings to the apperceptive story that are not directly tied to the dilemma presented in the story line; that is, the ending does not resolve the central dilemma present in the story but rather reflects what would happen in the next temporal time frame of the story sequence. An example for Card 4, which is a stimulus picture of a mother, two children, and a disheveled living room ("...and the mother told them to clean up the messy room." How will the story end?), is "He will scratch his nose."

Functional (2 points) Positive Operational Problem-Solving is scored for constructive solutions to the apperceptive story—solutions that re-

---

1 The CAST manual describes the development of the CAST instrument and scoring system in detail.
solved the story's dilemma. An example for Card 4 would be "The children will clean up the mess in the living room."

Abstract (3 points) positive operational problem-solving is scored for endings that reflect the positive and realistic cause-and-effect solutions to the central dilemma presented in the apperceptive story and which in addition provide a governing cognitive rule or abstract generalization for the solution. An example for Card 4 would be:

The children will clean up the mess in the living room because they know that they were the ones who made the mess in the first place and that means that they are the ones who need to clean the mess up. What is fair is fair, you make the mess, you clean it up.

Results

Validity

This section demonstrates the results of the factor analysis of the scales as well as construct validity. Construct validity was established by calculating t-tests to examine the differences between known groups and conducting a discriminant analysis to determine which set of scales would discriminate between well-adjusted and behavior-disordered subjects.

Factor analysis. To analyze the relationship among the scales of the CAST, a principal-components factor analysis with orthogonal varimax rotation and Kaiser normalization was conducted for all subjects in the standardization sample (N = 876). With a cutoff criterion of 1 or greater for the eigenvalues, four factors emerged (Table 2). After examining the scales that loaded on the respective factors, we named the four factors Nonadaptive, Adaptive, Immature, and Uninvested. The Nonadaptive factor, which consists of the Inadequacy, Alienation, Interpersonal Conflict, Limits, Negative Affect, and the Negative Operational Problem-Solving scales, is again reflective of the CAST theoretical model. The Immature factor, which consists of the Positive Preoperational Problem-Solving and the Negative Preoperational Problem-Solving scales indicates a tendency to solve problems using a simplistic nonrational approach. The Uninvested factor, which consists of the Unresolved and Refusal scales, measures the extent to which a child is cognitively or emotionally uninvested, or both, in engaging in the story-telling task or in solving the dilemmas evidenced in their story-telling. Although the Immature and Uninvested factors were underdetermined, because only two scales loaded on each factor, valuable clinical information would be lost if these two factors were not considered as part of the total profile of the individual child.

In addition, factor analyses were conducted for contrasting groups on the basis of gender, race, and adjustment (well-adjusted vs. behavior-disordered subjects). The results showed that the factor structure for the different gender, racial, and adjustment groups was substantially similar.

When separate factor analyses were conducted by race, gender, and age, the factor structures differed from the structure for the standardization sample in the following ways:

By race: White (N = 655)—the factor structure was the same as for the entire standardization sample; Black (n = 126)—the Nonadaptive factor was reduced to the Negative Operational and Alienation scales, with the remaining scales on the Nonadaptive factor loading with the scales on the Adaptive factor; Hispanic (n = 68)—a fifth factor emerged, which was defined by the Conflict Scale, and the Alienation Scale loaded on the Adaptive factor; Asian/Other (n = 27)—there were too few students to conduct a separate factor analysis.

By gender: Girls (n = 438)—the factor structure was the same as for the entire standardization sample; Boys (n = 438)—three factors emerged, with the Adaptive and Uninvested factors loading as polar opposites on the same factor.

By age: 6- and 7-year-olds (n = 215)—three factors emerged, with the

<table>
<thead>
<tr>
<th>CAST scale</th>
<th>Nonadaptive</th>
<th>Adaptive</th>
<th>Immature</th>
<th>Uninvested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumentality</td>
<td>.09</td>
<td>.74</td>
<td>-.11</td>
<td>.04</td>
</tr>
<tr>
<td>Interpersonal Cooperation</td>
<td>.37</td>
<td>.71</td>
<td>-.21</td>
<td>.14</td>
</tr>
<tr>
<td>Affiliation</td>
<td>.32</td>
<td>.57</td>
<td>-.25</td>
<td>.30</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>-.40</td>
<td>.44</td>
<td>.51</td>
<td>-.01</td>
</tr>
<tr>
<td>Inadequacy</td>
<td>.76</td>
<td>.27</td>
<td>-.07</td>
<td>.14</td>
</tr>
<tr>
<td>Alienation</td>
<td>.62</td>
<td>.30</td>
<td>-.13</td>
<td>.23</td>
</tr>
<tr>
<td>Interpersonal Conflict</td>
<td>.61</td>
<td>.10</td>
<td>-.03</td>
<td>.11</td>
</tr>
<tr>
<td>Limits</td>
<td>.77</td>
<td>.30</td>
<td>-.06</td>
<td>.01</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>.45</td>
<td>-.12</td>
<td>-.02</td>
<td>.56</td>
</tr>
<tr>
<td>Positive Preoperational</td>
<td>-.17</td>
<td>-.08</td>
<td>.85</td>
<td>.00</td>
</tr>
<tr>
<td>Positive Operational</td>
<td>-.10</td>
<td>.52</td>
<td>-.66</td>
<td>.45</td>
</tr>
<tr>
<td>Refusal</td>
<td>-.05</td>
<td>-.06</td>
<td>.03</td>
<td>-.71</td>
</tr>
<tr>
<td>Unresolved</td>
<td>-.10</td>
<td>-.23</td>
<td>-.01</td>
<td>-.73</td>
</tr>
<tr>
<td>Negative Preoperational</td>
<td>.03</td>
<td>-.22</td>
<td>.67</td>
<td>.10</td>
</tr>
<tr>
<td>Negative Operational</td>
<td>.72</td>
<td>-.33</td>
<td>.04</td>
<td>.01</td>
</tr>
</tbody>
</table>

Eigenvalues: 4.49, 2.05, 1.54, 1.12

Note. Variance accounted for with four factors = 61%. CAST = Children's Apperceptive Story-Telling Test.
Table 3
Comparison Between the Behavior-Disordered Sample and a Random Selection From the Standardization Sample

<table>
<thead>
<tr>
<th>Scale/factor</th>
<th>Random selection from the standardization sample (n = 322)</th>
<th>Total behavior-disordered sample (n = 322)</th>
<th>F(1, 641)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Adjusted M</td>
</tr>
</tbody>
</table>
| Instrumentality             | 24.70       | 12.29       | 23.30       | 16.00       | 9.83        | 17.40       | 159.39*
| Interpersonal Cooperation   | 6.15        | 5.26        | 5.44        | 3.43        | 3.71        | 4.14        | 142.23*
| Affiliation                 | 11.14       | 5.74        | 10.32       | 8.56        | 4.90        | 9.38        | 87.68*
| Positive Affect             | 7.98        | 3.68        | 7.93        | 7.48        | 4.04        | 7.53        | 2.71
| Inadequacy                 | 14.23       | 9.61        | 12.79       | 16.24       | 11.47       | 17.69       | 10.54*
| Alienation                  | 4.18        | 4.19        | 3.62        | 3.98        | 4.03        | 4.54        | 6.9
| Interpersonal Conflict      | 3.11        | 2.81        | 2.74        | 4.73        | 6.19        | 5.10        | 21.40*
| Limits                      | 8.12        | 6.00        | 7.14        | 9.68        | 8.50        | 10.66       | 12.60*
| Positive Preoperational     | 1.08        | 2.31        | 1.19        | 2.06        | 3.45        | 1.95        | 18.73*
| Positive Operational        | 23.57       | 8.89        | 22.55       | 14.25       | 9.68        | 15.27       | 228.44*
| Refusal                     | .14         | .75         | .28         | .78         | 2.35        | .64         | 25.16*
| Unresolved                  | .82         | 2.39        | 1.10        | 2.30        | 3.81        | 2.02        | 42.69*
| Negative Preoperational     | .38         | 1.36        | .43         | .95         | 2.14        | .91         | 16.33*
| Negative Operational        | 4.45        | 4.75        | 4.15        | 5.18        | 5.72        | 5.49        | 3.32
| Adaptive factor             | 187.02      | 73.39       | 176.43      | 128.09      | 62.60       | 138.62      | 285.27*
| Nonadaptive factor          | 109.68      | 61.71       | 99.02       | 125.22      | 80.94       | 135.88      | 16.06*
| Immature factor             | 1.46        | 3.06        | 1.62        | 3.01        | 4.69        | 2.86        | 25.67*
| Uninvested factor           | .96         | 2.69        | 1.38        | 3.07        | 4.76        | 2.66        | 67.10*

* p < .001.

Comparison of well-adjusted and behavior-disordered samples. To determine whether the scales of the CAST would discriminate between the well-adjusted and behavior-disordered groups, an overall multivariate test (Hotelling's $T^2$), univariate $F$ tests, and a discriminant analysis were used. The results from the multivariate test showed that the 15 scales as a group did discriminate between the behavior-disordered sample and a randomly selected group of 322 well-adjusted subjects from the standardization sample (Hotelling's $T^2 = .45, F = 32.66, p < .001$). Because the number of themes present in the children's stories correlated moderately with the score on the scales and factors, the number of themes was used as a covariate when comparing the behavior-disordered and well-adjusted students on the scales and factors of the CAST. The issue of verbal fluency (i.e., the number of themes generated by a child) is discussed in the following section.

Means, standard deviations, means adjusted for the number of themes for each child, and analysis of covariance $F$ tests for the CAST scoring scales are presented for the behavior-disordered sample and the well-adjusted sample in Table 3. Independent $F$ tests demonstrate that significant differences were found for the two groups for all scales and factors, with the exception of four scales: Positive Affect, Negative Affect, Alienation, and Negative Operational Problem-Solving. The results were in the expected direction for each of the scales. Not only were the differences between the groups statistically significant, but differences between the means expressed as proportions of the standard deviation values are fairly substantial for a number of the scales and factors. The results indicate that affect is not a strong predictor and that alienation themes are equally present in the apperceptive stories of well-adjusted children. Indeed, it was a thematically popular response for subjects to fear parental divorce (scored as formal alienation) when stories centered on a parental fight (Card 11) or to feel left out of peer interaction (scored as informal alienation) in a peer situation (Card 5). As the manual discusses in detail, Negative Operational Problem-Solving was associated with conduct-disordered and oppositional-disordered subjects, whereas the Preoperational and Unresolved Problem-Solving solutions were associated with attention-deficit-disordered subjects.

Table 3 shows that the well-adjusted subjects evidenced significantly higher scores on the Instrumentality, Interpersonal Cooperation, Positive Preoperational Problem-Solving, and Positive Operational Problem-Solving scales. The behavior-disordered subjects evidenced significantly higher scores on the Inadequacy, Interpersonal Conflict, Limits, Refusal, Unresolved, and Negative Preoperational Problem-Solving scales.

To determine which scales would most effectively discriminate between the total behavior-disordered group and the total standardization sample, a discriminant analysis was conducted. The standardized discriminant function coefficients are given in the following discriminant equation:

\[ Y = .30 \text{ (Instrumentality)} + .27 \text{ (Interpersonal Cooperation)} + .01 \text{ (Affiliation)} - .06 \text{ (Positive Affect)} - .37 \text{ (Inadequacy)} - .11 \text{ (Alienation)} - .21 \text{ (Interpersonal Conflict)} - .41 \text{ (Limits)} - .03 \text{ (Negative Affect)} + .40 \text{ (Positive Preoperational Problem-Solving)} + 1.08 \]
(Positive Operational Problem-Solving) — .10 (Refusal) + .15 (Unresolved) + .21 (Negative Preoperational Problem-Solving) + .69 (Negative Operational Problem-Solving). On the basis of the one statistically significant discriminant function, $R = .55$, $\lambda = .70$, $x^2 (15, N = 1,196) = 423.42$, $p < .0001$, 73% of the behavior-disordered sample and 80% of the well-adjusted sample were correctly classified. This represents an overall correct classification rate of 78%. The three most highly discriminating scales in order of their discriminating power were Positive Operational Problem-Solving, Instrumentality, and Interpersonal Cooperation.

**Verbal Fluency**

Because verbal fluency is an issue related to the validity of projective techniques, verbal fluency on the CAST is accounted for through a measure of thematic frequency (thematic verbosity). The CAST provides the examiner with comparison information regarding the thematic verbosity levels of behavior-disordered and well-adjusted children. The CAST has addressed verbosity not as word count but as thematic count, because the presence of theme, not the number of words, is used to determine a child’s score on the scales. The number of discrete themes per story is counted and summed for the entire 17 stories. The resulting value is compared to the high, average, and low standardization values for children at the subject’s age level. The high, average, and low categories for verbosity were determined by setting cutoff points one standard deviation above and below the mean of the number of themes generated by the standardization sample. Thus, low thematic verbosity represents one or more standard deviations below the mean of the standardization group in terms of number of themes generated. High thematic verbosity is one or more standard deviations above the standardization mean, and average thematic verbosity is less than one standard deviation from the mean number of themes generated by the standardization sample. It is important for the examiner to take into account the client’s level of verbosity when interpreting a child’s performance on the CAST. For example, when a child generates very few themes and therefore attains a score in the low thematic verbosity category, the examiner should consider the issue of low generation in the process of interpretation. In addition to the possibility of cognitive immaturity, the examiner should be alert to signs of other potential hypotheses such as shyness, lack of rapport, or speech and language problems.

Table 4 shows the comparison of CAST themes generated by age and sex for the behavior-disordered validity sample and the standardization sample. These results indicate that, with the exception of 6-year-old and 13-year-old subjects, the trend was for the subjects in the standardization sample to express significantly more themes in their stories than the subjects in the behavior-disordered sample.

**Reliability**

Reliability on the CAST was determined through five methods: interrater, intrarater, test—retest, coefficient alpha, and split-half.

**Interrater and intrarater reliability.** Both interrater and intrarater reliability of the CAST were investigated. These studies were conducted independent of the scoring team’s work on the standardization sample scoring. The interrater approach involved two studies. In the first study, 6 raters conducted a one-time scoring of six protocols. In the second study, 12 raters conducted a one-time scoring of four typical protocols with each rater scoring four to six scales of the four protocols. The intrarater approach required 2 of the original 6 raters to rescore all 6 of the children’s protocols after a 7-week time interval.

The six interrater reliability raters consisted of one clinical psychologist, three school psychologists, one clinical social worker, and one school social worker. The six protocols were selected to reflect three different levels of scoring difficulty: easy, moderate, and difficult, as judged by the four members of the standardization scoring team. The raters were given a set of stimulus picture cards, the administration and scoring sections of the manual, one sample case illustrating the scoring, and the scoring guidelines that were refined during the scoring of the standardization protocols. Raters were unmonitored and worked independently of each other and without time limits. Pearson correlations were calculated for all possible pairs of raters. Age was partialed out of the correlations reported. Three of the interrater reliabilities were in the .80s, whereas the remaining 11 reliabilities were in the .90s. The median of the reliability coefficients for all 15 scales was .95.

The second interrater reliability study involved 12 graduate students enrolled in a master’s degree program in school psychology and counseling psychology. The raters were given a set of stimulus picture cards, the administration and scoring sections of the manual, one sample case illustrating the scoring, and the scoring guidelines that were refined during the scoring of the standardization protocols. Using a matrix sampling approach (Lord, 1962), we asked raters to score four to six scales of the CAST (four raters scoring each scale) for four entire protocols. Scoring fatigue was addressed through a counterbalanced ordering of protocols.

Raters were monitored during scoring, and scoring time ranged from 25 min to 2 hr, with a median of 60 min for scoring all four protocols. Pearson correlations were calculated for all possible pairs of raters. Age was partialed out of the correlations reported. One of the interrater reliabilities was in the .70s, 4 were in the .80s, and the remaining 10 were in the .90s. The median of the reliability coefficients for all 15 scales was .93. Comments by the raters participating in this study indicated that ease of scoring may be attributed to the detailed scoring examples found in the manual.

In the intrarater reliability approach, the two raters were practicing school psychologists who rescored the six protocols after a 7-week period. The raters were provided with the same materials used in the original scoring exercise. The 7-week period was considered long enough for the raters to forget the original scores given to the subjects. Pearson correlations were calculated between the two raters, with age partialed out of the correlations. One of the 26 interrater reliabilities was .33, one was .64, one was .71, three were in the .80s, and the remaining 20 fell in the .90s. For a number of the scales, the scores for each rater decreased from the first to the second rating. For two of the scales, Negative Operational Problem-Solving and Refusal, it was impossible to compute interrater reliabilities because of
The Negative Preoperational Problem-Solving Scale for children ages 12 and 13 years was .76. The correlations are consistent across age levels, with the primary exception being the Negative Preoperational Problem-Solving Scale for children ages 12 and 13 years. Because children in the older age group exhibited little Negative Preoperational Problem-Solving, this lower reliability coefficient is likely due to the low scores and limited variability of this scale within this age group. Coefficient alpha was calculated for the four factors of the CAST based on the scores from the total standardization sample. With the exception of the Immature factor at the highest age level (r = .76), the reliabilities ranged from .90 to .93. The lower reliability for the Immature factor was likely due to the reduced presence of Preoperational Problem-Solving at the highest age level.

**Split-half reliability.** Following the traditional procedure of split-half reliability, the 17 stimulus picture cards were separated into two groups based on the content of the stimulus pictures. Scores on the 15 scales of the CAST were then calculated for the two halves of the test for the subjects in the standardization sample and for the behavior-disordered sample. Pearson correlations between the half scales were calculated, and the Spearman-Brown formula was applied (Nunnally, 1978). The split-half reliabilities were highest for the problem-solving scales and somewhat lower for the nonadaptive scales. For the standardization sample the median value of the split-half reliabilities for all scales was .74; the median value for the behavior-disordered sample was .79.

**Discussion**

Although apperceptive story-telling techniques continue to be of value for clinicians who work with young children (Piotrowski, 1984), the psychometric rigor necessary to establish confidence in the information provided by such instruments has been lacking. The Children's Apperceptive Story-Telling Test (CAST) joins the tradition begun by the Roberts Apperceptive Test for Children (McArthur & Roberts, 1982) in attempting to address psychometric issues such as validity, reliability, and an objective scoring system. The CAST builds on this tradition by using a theory-based scoring model, racially sensitive and colored stimulus pictures, a representative standardization sample of well-adjusted children, and a large comparison sample of behavior-disordered children.
Although arguments exist against the use of traditional psychometric measures with apperceptive instruments (LaVoie, 1984), many scales and factors of the CAST achieve reliabilities that allow confidence, especially with the use of the scores based on the four factors. The reliabilities for the factors are sufficient to allow the examiner to construct working hypotheses regarding the adjustment (adaptive or nonadaptive) and problem-solving styles of the children taking the CAST. The apperceptive content is a sample of the cognitive constructs and problem-solving capacity of the subject. This content requires hypothesis formation and direct confirmation with the subject and often with significant others (e.g., parents, teachers, and peers) before an intervention plan is instituted. Because several sources of data should always be considered in making diagnostic decisions concerning intervention for an individual child, the CAST's working hypotheses may be confirmed in a clinical interview (a process delineated in depth in the CAST manual) or through the use of additional measures.

Although the CAST evidences substantial attention to the psychometric concerns raised regarding apperceptive instruments, limitations exist. For example, the procedure used in securing a well-adjusted sample of subjects for the standardization may have produced an extremely well-adjusted group. Perhaps the use of categories from the Diagnostic and Statistical Manual of Mental Disorders (3rd ed.; American Psychiatric Association, 1980) to classify behavior-disordered subjects introduced limitations due to the potential subjectivity of multidisciplinary team decision making. Future research might include further investigation of the relationship between language development and productivity on apperceptive instruments for behavior-disordered children. Future studies could focus on the establishment of normative data for children with a primary language other than English, pretest-posttest group studies designed to assess the effects of treatment, the use of the CAST to predict therapeutic outcomes of individuals, and reliability and validity studies with larger groups and groups with specific characteristics (e.g., gifted or Native American children). Although research within the area of diagnostic and therapeutic story telling requires further development, the validity and reliability studies reported here offer evidence that the Children's Apperceptive Story-Telling Test can provide the clinician and researcher with a potentially viable diagnostic tool designed specifically for use with children who display behavioral and emotional difficulties.

References

Received July 7, 1989
Revision received October 25, 1989
Accepted November 13, 1989