Art Therapy: Journal of the American Art Therapy Association, 23(3) pp. 104-111 © AATA, Inc. 2006

Articles

The Utility of the Formal Elements Art Therapy Scale in Assessment for Substance Use Disorder

Pam Rockwell and Mardis Dunham, Murray, KY

Abstract

This study explored the use of the Formal Elements Art Therapy Scale (FEATS) with a population of persons with a DSM-IV diagnosis of Substance Use Disorder who were court ordered for treatment. Two groups of adults (N = 40) were closely matched on age, gender, race, socioeconomic status and education level, and were administered the Person Picking an Apple From a Tree (PPAT) art assessment. The drawings were scored using the Formal Elements Art Therapy Scale (Gantt & Tabone, 1998). Acceptable inter-rater reliability was established on 9 of the 12 FEATS elements. Statistical analyses indicated that three FEATS elements in the PPAT accurately predicted group membership and that together, the 12 FEATS elements successfully distinguished the two groups with 85% accuracy. Implications of these findings for art therapy research and practice are discussed.

Introduction

Assessment and diagnosis have become cornerstones of counseling and mental health services in the last few decades. While the degree of differentiation in diagnosis made possible with the assimilation of accumulated knowledge in the Diagnostic and Statistical Manual-Fourth Edition (DSM-IV-TR; American Psychiatric Association, 2000) is a relatively new phenomenon, society has utilized some manner of assessment for diagnosis of psychological dysfunction for many centuries. For example, court records in Europe indicate the use of mental status exams for differential diagnosis as far back as the thirteenth century when judges were called upon to distinguish between individuals with cognitive disorders and those with mental illness (Neugebauer, 1979).

Assessment is a means by which to gather information about the mental status of a client for the purpose of enabling a clinician to effectively design and direct treatment. According to Feder and Feder (1998), good evaluation procedures serve several important functions. They enable practitioners to identify a client's problems or needs so that effective treatment procedures can be established, and they provide a means by which to monitor progress so that clinicians know when to stop treatment and how to improve upon their treatment protocols.

In addition to assessing progress or effectiveness during and after treatment, often clinicians need to measure behaviors or mental states against a standard. Or they may need to measure frequency or duration of a behavior or treatment. In those instances, quantitative evaluation procedures are helpful. Through testing procedures, quantitative evaluation generates data usually expressed in numerical form. Carolan (2001) points out that many types of information in art therapy, from personal experience to imagery, can be effectively communicated in this manner. With the use of fixed standards and with little opportunity for personal judgment to enter into the data-gathering process, quantitative evaluation procedures are generally considered to render objective information (Feder & Feder, 1998).

There are a number of formal and informal assessment techniques available to assist mental health practitioners in making diagnostic decisions. Among these are interviews, developmental and personal history taking, observations, and norm-referenced psychological tests. Additionally, practitioners often employ projective personality assessment techniques. Drawing tests like the House-Tree-Person, or interpretive tests like the Rorschach Inkblot and Thematic Apperception Tests, are well-known among these. Objective personality measures (e.g., Minnesota Multiphasic Personality Inventory [MMPI], NEO Personality Inventory) are also widely used to gather important diagnostic and treatment information. Combined with developmental history and mental status exams, assessment instruments often provide the cornerstone of the diagnostic process. Which ones of these procedures a clinician employs depends upon training and treatment philosophy. For example, clinical psychologists, because of their training and experience, may rely upon the Rorschach or the MMPI while social workers may rely more heavily upon interviews and observations. Art therapists, with their

Editor's note: Pam Rockwell, EdS, LPCA, is a professional counselor in Kentucky. Mardis Dunham, PhD, is Coordinator of School Psychology at Murray State University in Murray, Kentucky. Correspondence concerning this article may be sent to pmrockw@aol.com or marty.dunham@coe.murraystate.edu.

training and expertise in the use of art media, often integrate drawing, painting and sculpture into their data gathering and treatment processes.

Historically, the use of drawing techniques and their interpretation has been impressionistic, largely psychoanalytically symbolic, and intuitive, and generally lacking in empirical validation (Feder & Feder, 1998; Gantt, 1992). Recognizing these limitations within the field of art assessment, Ulman and Levy (1968) as early as the 1960s were insisting that relating elements within paintings or drawings to psychopathological diagnoses was an assumption open to serious debate unless it could be empirically proven that even the crudest diagnosis—patient or normal—could be made with consistent precision on the basis of the artwork. Gantt (1998) similarly pointed out that while art therapists like to make claims about such things as how changes in art are related to changes in psychological state, or how certain features are indicative of neuroses or mental adjustment, those assertions can no longer be made without solid, scientific proof of their validity if arts therapists wish to maintain their reputability among other counseling professionals. That kind of proof cannot be established without empirical research techniques that render quantifiable data to support the claims.

One promising attempt to establish empirical support for drawing techniques comes from Gantt and Tabone (1998) who have utilized a standardized drawing instrument, the Draw a Person Picking an Apple From a Tree (PPAT), in conjunction with the Formal Elements Art Therapy Scales (FEATS) rating system. The instrument was developed to address the issues of validity, reliability and standardization within the field of art therapy assessment.

Rosal (1998) defined the FEATS as a measurement system that enables researchers to quantify global variables in two-dimensional works of art. The authors describe the instrument as a single-picture assessment (PPAT) and a method for rating that picture (FEATS; Gantt & Tabone, 1998). The instrument is used to quantify the formal art elements (line, shape, color, space, etc.) within the drawing.

Validity and reliability studies (Gantt & Tabone, 1998; Gantt, 2001; Munley, 2002) have demonstrated the FEATS to be a reliable instrument that measures variables (formal art elements) which the authors have equated with psychiatric symptoms from the DSM-IV-TR (APA, 2000). Using an objective rating scale, the FEATS enables the systematic study of how people draw (form), eliminating the confounds to quantitative analysis that the literature has demonstrated to be inherent in attempts at studying what they draw (content).

Gantt's research (Williams, Agell, Gantt, & Goodman, 1996; Gantt & Tabone, 1998) using the PPAT in conjunction with the FEATS rendered statistical data supporting her hypothesis that there are indicators within the quality of the lines, the size of the renderings, the presence or absence of colors, and so forth, that could be correlated with symptoms of mental illness. The pictures alone, in the absence of contact with or interpretation by the artists, provided enough information to enable judges to accurately assign them to not only patient or non-patient categories, but to one of five DSM diagnostic categories: major depression, schizophrenia, bipolar disorder (mania), organic mental disorder, cognitive disability and non-patient.

Apart from the studies conducted by the FEATS author (see Gantt, 2001), a review of literature produced no other studies with adults. Additional validity and replication studies are needed to verify the original findings specifying the statistical properties and utility of the FEATS. A further review of the literature produced one study utilizing the PPAT with the FEATS with children. Munley (2002) determined the association between PPATs drawn by boys diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD) and those of controls (age-matched boys with no known behavioral problem or learning disability). Analyzing FEATS variables individually using logistic regression analysis, Munley demonstrated the overall reliability of each FEATS element to accurately place subjects into the ADHD group. Reliability ranged from 48.63% to 88.25% on 15 scales overall ($p = < .05$). Predictive reliability did not fall below 70% on over half of the scales. Munley's study supports two hypotheses: PPATs as measured by the FEATS (1) reliably demonstrate differences on several scales between children with and without ADHD and (2) drawings by children with ADHD exhibit predictable similarities among themselves on several scales.

Research thus far conducted with the FEATS suggests that Gantt and Tabone have designed a valid, reliable instrument for gleaning diagnostic information from art. In her pilot study, Gantt (1990) determined that the patterns of symptoms that distinguish four major categories of psychiatric illness (major depression, mania, schizophrenia and organic mental disorder) have graphic equivalents which can be reliably measured through this standardized art assessment instrument. There have been no published or peer reviewed studies demonstrating the use of the FEATS with a substance abuse population. However, one other study was found that employed a drawing technique to study the personality characteristics of persons with substance use disorders. Francis, Kaiser, and Deaver (2003), using the Bird's Nest Drawing (BND; Kaiser, 1996), found that persons with substance use disorders differed from controls in the use of color. Specifically, those with substance use disorders used less color, in particular the color green. The authors, studying graphic representations of attachment security in the drawings, interpreted this finding as perhaps related to such symptoms as a lack of energy, a lack of motivation or engagement with the task, or as depressive symptoms.

The purpose of the present research was threefold: (1) to explore the inter-rater reliability of the FEATS; (2) to evaluate the utility of the FEATS in a mental health setting by determining if it could successfully distinguish between a group of individuals with substance use disorders and those without; and (3) to provide additional normative data for the instrument with an adult population to assist future researchers.

Method

Participants

Participants in this study included 40 adults (*mean age* = 33 years) from several counties in western Kentucky. Excluding the local state university, the ethnic make-up of the population is predominately Caucasian. Participants in the experimental group (n = 20) were comprised of 7 Caucasian females, 9 Caucasian males and 4 African American males. The mean age for the experimental group was 31.5 (*standard deviation* = 9.75). The control group (n = 20) was comprised of 10 Caucasian females, 1 African American female, 1 Asian female, 6 Caucasian males and 2 African American males. The mean age for the control group was 34.6 (*standard deviation* = 13.5).

The experimental and control groups were matched as closely as possible on socioeconomic and demographic variables. Specifically, both groups included participants from each of the following categories: factory workers, construction workers, restaurant employees, state employees, homemakers, farmers and business professionals. Both groups included participants with high school diplomas and both included participants with some college. The experimental group included three participants who had not completed high school while the control group included two participants without high school diplomas.

The twenty participants in the experimental group were individuals who were either on probation or parole for substance related charges. They had all been court-ordered to receive substance abuse counseling as a condition of their probation or parole. They were participating in individual counseling at a state-licensed substance abuse counseling center.

Instrument

The instrument used in this research was the Draw a Person Picking an Apple From a Tree (PPAT) art assessment as measured by the Formal Elements Art Therapy Scales (FEATS) (Gantt & Tabone, 1998). The PPAT is a controlled content drawing task that is rated on the FEATS. The FEATS consists of fourteen Likert Scales. Each scale assigns a numerical value between one and five to each of fourteen formal art elements observable in drawings. The purpose of the FEATS is to measure variability within formal art elements in PPAT drawings in order to determine quantifiable comparisons between groups, or to establish within-group similarities.

Research on the FEATS, though limited, indicates good inter-rater reliability on 12 of the 14 scales. Inter-rater reliability among those 12 scales ranges from .70 to .95 as computed using an intra-class correlation (Gantt & Tabone, 1998). Scales thirteen and fourteen of the FEATS measure Perseveration and Rotation, both of which are features associated with the drawings of either very young children or very old individuals. Since participants in the study belonged to neither of those age groups, those two scales were omitted from the study. Unlike previous studies, an aggregate score was computed for the current study by obtaining an average score of the 12 elements.

Procedure

For the experimental group, twenty drawings were collected as part of the intake process from twenty clients entering substance abuse counseling at a state-licensed, private counseling facility. For research purposes, the drawings were archived at the time of the study; there were no names or identifying information associated with any of the drawings. For the control group, twenty drawings were collected from twenty individuals from the community chosen by the researcher in such a way as to match the two groups as closely as possible with regard to demographic variables.

Participants in both groups were administered the PPAT in accordance with the instructions described in the testing protocol (Gantt & Tabone, 1998). Each participant was given one 12" x 18" sheet of white drawing paper that was handed to them in such a way as not to suggest either a vertical or horizontal orientation by the administrator. Each participant was given twelve felt tipped markers, one each of the following colors: red, orange, blue, turquoise, green, dark green, hot pink, magenta, purple, brown, yellow, and black. Each participant was given the instructions: "Draw a person picking an apple from a tree." The instruction could be repeated if necessary; however, no further instruction or clarification could be given. No time limit was set on the drawing task.

Raters and Rating Procedure

First, the forty drawings were coded and mixed so that group identity could not be ascertained during the rating process. The drawings were then rated by the researcher on 12 FEATS scales using the scoring guidelines in the FEATS manual (Gantt & Tabone, 1998). To avoid skewed results from rater fatigue, the forty drawings were divided into three equal groups and rated one group per session over three sessions that occurred within two consecutive days.

Next, to establish inter-rater reliability, an independent rater was recruited to rate twenty randomly selected drawings—10 from the experimental group and 10 from the control group. The drawings were randomly mixed so that the rater was unaware as to which group a drawing belonged during the rating process. The independent rater was an individual with a masters degree in painting who was given instruction by the researcher on how to rate the drawings in accordance with the FEATS rating manual.

Analyses

Once gathered, several statistical analyses were applied to the data. First, inter-rater reliability was established by correlating the ratings from the two raters. Second, significant differences between the two groups on the twelve elements and the aggregate score were evaluated using a Mann-Whitney U test. The Mann-Whitney is a nonparametric test that evaluates the hypothesis that two independent groups are from the same population. This test is

appropriate when distributions of scores are not uniform or when the two variances are not homogeneous (Glass & Hopkins, 1996). Third, to further assess the differences between the two groups on the aggregate score, an analysis of variance (ANOVA) was computed, followed by a logistic regression to discern which FEATS elements could most reliably predict group membership.

Results

Results of the inter-rater reliability study are presented in Table 1 and were consistent with other FEATS research (Williams, Agell, Gantt & Goodman, 1996; Munley, 2002) in that the Pearson correlations revealed a very high degree of consistency between the two raters. Specifically, nine of the twelve scales, as well as the aggregate score, were statistically significant, indicating that the FEATS Rating Manual (Gantt & Tabone, 1998) enables different raters to reliably score the formal art elements present in PPAT drawings.

Of the three FEATS scales that were not statistically significant, two (Integration and Line Quality) were positively correlated while the third scale (Problem Solving) revealed a very small negative correlation. In comparing the criteria set forth in the FEATS rating manual for scoring PPATs on the Problem Solving scale with the scores assigned the drawings by the two raters, it appeared that the negative correlation resulted from differing interpretations of the scoring criteria in the rating manual. In order to maintain consistency with the other studies employing the FEATS, all 12 scales were included in subsequent analyses.

The scores from the two groups on the FEATS as well as the results of the Mann-Whitney U Test are provided in Table 2. Consistent with previous research (Williams, Agell, Gantt & Goodman, 1996; Munley, 2002) an alpha level of .05 was set to determine the existence of statistically significant differences.

Table 1
Pearson Product Moment Coefficients of Inter-Rater Reliability

FEATS Scale	Pearson Coefficient
Prominence of Color	.901**
Color Fit	.804**
Implied Energy	.834**
Space	.817**
Integration	.251
Logic	.444*
Realism	.818**
Problem Solving	-.057
Developmental Level	.504*
Details of Objects and Environment	.788**
Line Quality	.416
Person	.937**
Aggregate	.886**

Note: ** p = <.01; * p = <.05

The Mann-Whitney test results indicate that the group means for the control group were significantly higher than for the experimental group on the aggregate score and on three scales—Developmental Level, Realism, and Person. Essentially, these three scales contributed to the significant difference on the aggregate score. The probability that the Mann-Whitney U test results could have occurred by chance alone were less than three out of one hundred; so small as to indicate that group means were derived from qualities that are observable within the art variables.

The experimental group produced drawings that indicated a Developmental Level comparable to that of latency-

Table 2
Experimental and Control Group Comparisons on the FEATS

	Experimental Group		Control Group			
FEATS Element	Mean	SD	Mean	SD	U Test	p
Color Prominence	2.56	.99	2.96	.74	1.228	.219
Color Fit	4.25	1.21	4.79	.57	1.80	.073
Implied Energy	3.00	.85	3.31	.59	1.02	.306
Space	3.44	.89	3.66	.93	1.04	.299
Integration	3.86	.60	4.16	.56	1.90	.058
Logic	4.90	.26	4.90	.31	.38	.706
Realism	3.25	.75	3.74	.55	2.21	.027*
Problem Solving	3.91	1.18	4.45	.89	1.74	.082
Dev. Level	3.51	.65	4.09	.43	2.59	.010**
Object/Envir. Detail	2.89	1.01	3.44	.91	1.71	.088
Line Quality	3.86	.43	3.99	.29	.91	.365
Person	4.18	.69	4.70	.31	3.16	.002**
Aggregate	3.62	.55	3.99	.34	2.29	.022*

Note: ** denotes .01 level of significance; * denotes .05 level of significance

Table 3
Analysis of Variance Using Aggregate Scores by Group

Source	Sum of Squares	*df*	Mean Square	*F*	*p*
Between Groups	1.38	1	1.38	6.666	.014
Within Groups	7.247	38	.207		
Total	**9.247**	**39**	**.207**		

N = 40

age children (Lowenfeld, 1957, cited in Gantt & Tabone, 1998), while the control group's drawings were comparable to those done by adolescents. The complexity of the drawings, as rated on the Realism scale, indicated that drawings produced by the experimental group were more simply rendered with fewer of the art elements (i.e., texture, shading and details within objects) that create the illusion of three-dimensionality in two-dimensional art work. Finally, the Person in the experimental group's drawings tended to be drawn as a stick figure, often lacking details such as hands, feet, fingers, or the articulated body parts (waist, neck, knees, elbows) that typify drawings that rate higher on the scale, such as those produced by the control group.

To further investigate group differences and the validity of the initial findings, an analysis of variance (ANOVA) was applied to the aggregate score (Table 3), followed by a logistic regression analysis, an approach used by Munley (2002). The significance of applying logistic regression analysis was in ascertaining, for purposes of practical application in mental health settings, the fewest number of FEATS elements that, together, could accurately predict group membership. The logistic regression analysis detailed which FEATS elements could classify the groups and at what level of accuracy.

Not surprisingly, the ANOVA was significant (F (1.39) = 6.67, p = .014), and the logistic regression indicated that together the twelve FEATS elements were able to correctly categorize, on average, 85% of the group members—80% for the experimental group and 90% for the control group (X^2 (12, N = 40) = 28.534, p = .005). Individually, three of the twelve FEATS elements were statistically correlated (p = < .05) with the model in predicting group membership. Individually, group membership could be predicted from scores on the Realism scale in 67.5% of instances, and on the Developmental Level scale in 62.5% of instances. In 77.5% of instances the Person scale, alone, could predict group membership. When combined, these three scales were able to successfully predict 72.5% of the cases.

The remaining nine elements did not explain any additional variance (individually) and were therefore not included in the equation. The logistic regression analysis data in Table 4 detail the individual predictive value for the twelve FEATS elements.

Discussion

Current art therapy literature indicates that sound empirical research is needed in the development of valid and reliable instruments for assessment and diagnosis (Feder & Feder, 1998; Gantt, 1998; Rosal, 1998; Ulman & Levy, 1968). This study employed empirical research methodologies and applied several statistical measures to data obtained from two groups, one (experimental) with a DSM-IV diagnosis of Substance Use Disorder and one (control) with no diagnosis. The PPAT assessment was used to collect drawings which were then numerically scored using the FEATS rating protocol to determine whether the experimental group responded differently to the drawing task than controls.

Previous studies conducted using the FEATS with PPATs have demonstrated that the instrument reliably measures formal art elements that have been equated with DSM symptoms, and that it is a valid tool for differentiating between groups (Munley, 2002; Williams et al., 1996). This research corroborates those findings.

The first noteworthy finding is the strong inter-rater reliability, with statistically significant correlations found on 9 of the 12 scales. Scores of such magnitude support the findings of previous studies (Gantt, 1986; Munley, 2002) which indicate that the FEATS, when used by different practitioners, produce similar results. These findings are important—if two raters cannot consistently obtain the same results from the same data, then the reliability of the scoring system is lacking and the validity of the constructs they attempt to measure cannot be determined.

Findings from a Mann-Whitney U analysis indicated that the aggregate score and scores on three of the twelve FEATS scales (Realism, Developmental Level and Person) were significantly different between groups, with individuals in the substance use group obtaining significantly lower scores as compared to controls. Two out of these three elements—Realism and Person—have been identified as prominent features of the PPATs drawn by individuals presenting with depression, schizophrenia, and dementia from among some 5,000 patients who rendered drawings collected at a psychiatric hospital in West Virginia (Gantt & Tabone, 1998). By contrast, the FEATS elements that define the cluster of symptoms for Bipolar Disorder discussed in the FEATS manual share no common elements with those identified in this study.

The FEATS authors indicate that the instrument's usefulness in assessment is in identifying clusters of scores that may be indicative of mental states in the same way that the DSM uses clusters of symptoms to support a particular diagnosis. The significance level of the scores on the Realism, Developmental Level and Person scales in this study indicate that, for purposes of assessment in substance

Table 4
Logistic Regression Analysis of FEATS Elements Indicating the Percentage of Correct Predictions for the Two Groups

	Group		
FEATS Element	**Experimental**	**Control**	**Overall**
Prominence of Color	40	75	57.5
Color Fit	45	80	62.5
Implied Energy	50	65	57.5
Space	55	60	57.5
Integration	90	45	67.5
Logic	0	100	50.0
Realism	65	70	67.5
Problem Solving	60	65	62.5
Developmental Level	45	80	62.5
Details of Objects and Environment	60	55	57.5
Line Quality	25	90	57.5
Person	80	75	77.5

Note: Values are percentage of correct classification.

use disorders, clinicians should pay particular attention to how their clientele score on these three scales. These findings support the use of the FEATS as part of the diagnostic process—whereas Gantt (1990) and Munley (2002) found differences in groups with depression, schizophrenia, bipolar disorder, and ADHD, the current findings suggest that persons with substance use disorders also differ reliably from controls on the FEATS. There appears to be no apparent similarity between the Bird's Nest Drawing study (Francis, Kaiser, & Deaver, 2003) and the current FEATS study. In the former study, persons in the substance abuse group used fewer colors than the control group, whereas in the current study, the groups differed primarily on elements not involving color.

Logistic Regression analysis was useful in specifying to what degree each of the elements represented by the four significant scores on the Mann-Whitney test was able to predict group membership. Individually, the Person scale was able to predict membership in the experimental group at the highest rate of accuracy (77.5%). The other two scales that achieved statistically significant correlations individually (Realism and Developmental Level) were less valuable in predicting group membership at 67.5% predictive accuracy and 62.5% respectively. Taken together, the three scales were able to successfully predict 72.5% of the cases. It is the aggregate score from all twelve scales, however, that would be most valuable to clinicians utilizing the PPAT to support a diagnosis of substance use disorders. Together, the twelve FEATS elements used in this research were able to correctly categorize 85% of the group members—80% for the experimental group and 90% for the control group. In contrast, the regression analysis in Munley's (2002) study of the PPATs of children diagnosed with ADHD identified significant scores on a completely different set of art variables: Prominence of Color, Details, and Line Quality.

Of the twelve individual FEATS scales utilized in this study, the Person scale had the highest predictive value. Taken alone, this scale predicted membership in the substance use group 77.5% of the time. The FEATS rating manual interprets Person as the graphic equivalent of several mental health symptoms: diminished interest, diminished ability to think or concentrate, delusions, hallucinations, incoherence, loosening of associations, impairment in abstract memory and thinking.

For illustration purposes, Figures 1 and 2 are typical of those collected from the experimental group (substance use group) in this research. Compared with those typically drawn by the control group (Figures 3 and 4), the drawings are relatively simply rendered. They have few details and lack three-dimensionality in the objects and in the overall composition. Also, artists in the experimental group often had difficulty rendering the person. Compared with Figures 1 and 2, the Control group's "person" has well-proportioned body parts, hands with fingers, arms and legs with joints, and clothing details.

Implications

This study employed sound research methodology using a valid and reliable, controlled-content art assessment instrument. As such, the findings should be valuable for clinicians interested in the use of art assessment in substance abuse. As part of an intake process, the PPAT could produce supportive criteria for establishing a diagnosis of Substance Related Disorders if similar results were found in replications of this study. Replication studies could establish a reliable pattern of significant scores around a cluster of FEATS elements that have a high predictive value for the diagnosis of Substance Related Disorders. Similarly, correlational research employing both the FEATS and the BND with the same substance abuse population would be helpful in determining if and how these two techniques are measuring the same construct. It is conceivable that they could be used in conjunction for diagnosis and treatment.

The FEATS rating manual interprets the Person and Realism scales as representing the graphic equivalents of several mental health symptoms related to cognitive functioning such as concentration, coherent thinking, and abstract memory. If used for both pre- and post-testing with substance abuse clients, as well as in longitudinal studies, the PPAT could provide data regarding the degree to which these capacities are recovered in clients who remain drug-free.

As noted earlier, art therapy literature indicates the need for correlational studies that evaluate client artwork against controls for the purpose of establishing norms for existing art assessment instruments. The current research provides a small but usable normative sample for future research. Without normative data for a broad range of clin-

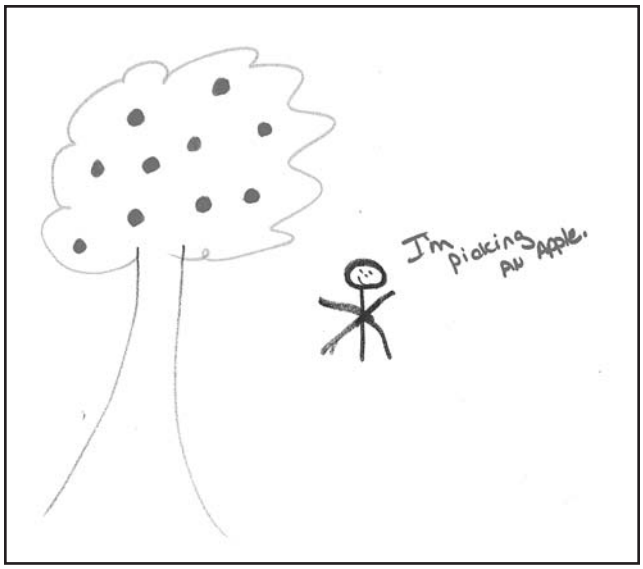

Figure 1

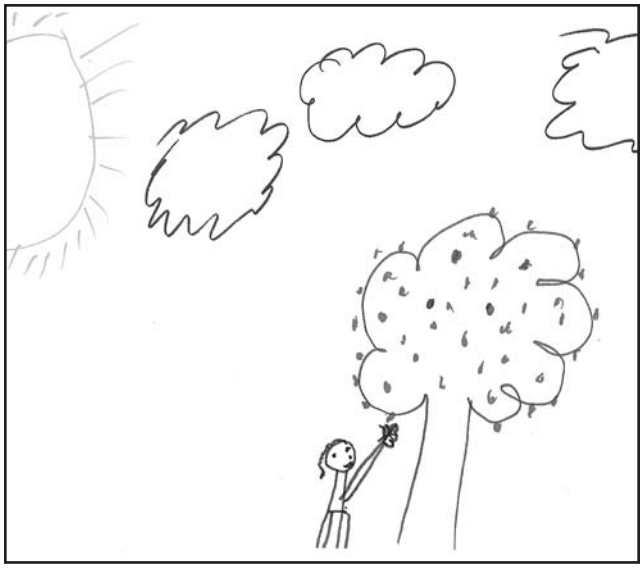

Figure 2

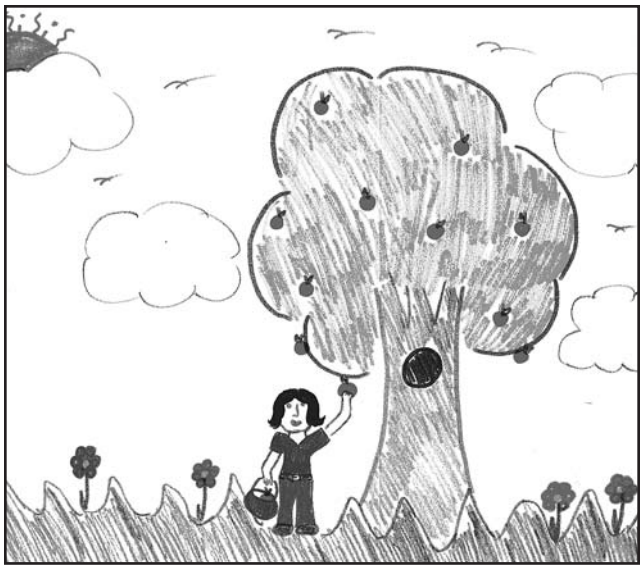

Figure 3

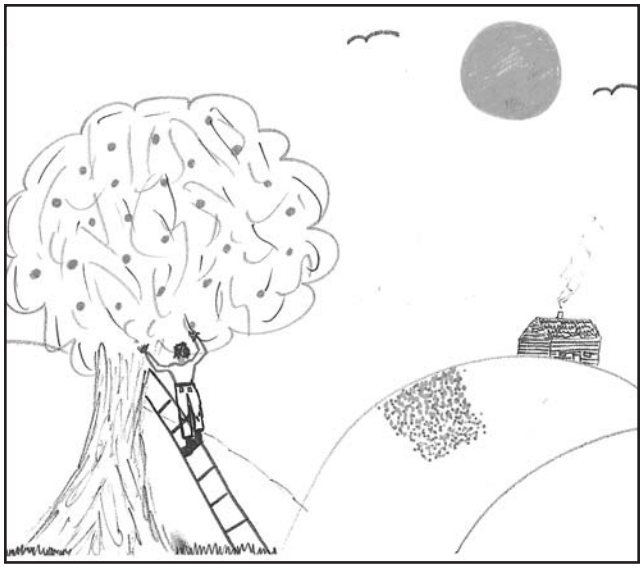

Figure 4

ical syndromes, along with validated cut scores to aid in differential diagnosis (clinical versus non-clinical populations) much like that employed by the MMPI-2 (Hathaway & McKinley, 1992) and the Rorschach Exner scoring system (Exner, 1991), the practical utility of the FEATS is limited.

Limitations

There are some limitations to this study that should be considered in interpreting and generalizing the findings. First, the control group for this study was a convenience sample of individuals recruited from within the researcher's sphere of acquaintances. In spite of the fact that the researcher made every effort to create a control group population that was matched as closely as possible in gender, socioeconomic status, race and education to that of the experimental group, the study would have been strengthened by collecting drawings from a much larger, anonymous sample population and then screening the drawings to create the matched pair sample. While the attempt at matching controlled for some confounding variables (e.g., cohort effect), it may have created other confounds that are presently unknown (e.g. unknown past history of substance abuse).

Second, the experimental group drawings were collected from a general pool of drawings created by clients court-ordered to counseling for Substance Related Disorders. As such, there were no subgroup distinctions made. For example, some of the clients had alcohol addictions, many had methamphetamine addictions, and some had polysubstance addictions. Research that distinguished between the substance of abuse could be helpful to determine whether subgroup differences are present and valuable for differential diagnosis when using the PPAT with this heterogeneous population.

Third, none of the participants in the experimental group were voluntary clients in substance abuse counseling. All had been arrested for substance related charges and had been court-ordered to counseling as a condition of their probation or parole. FEATS research with voluntary consumers may or may not produce different scores than those obtained from involuntary participants. Despite the fact that it was explained to individuals in the experimental group that they were under no obligation to create a drawing, and there would be no consequence of refusal to participate, certainly it is plausible that issues of "faking good" and/or defensiveness could be operating in a population of court-ordered mental health consumers.

Conclusions

This study found that the FEATS manual is sufficiently illustrative to yield adequate inter-rater reliability for most scales. As for all standardized instruments, inter-rater reliability is a prerequisite to establishing validity and other forms of reliability. These findings also suggest that the FEATS can be used to reliably distinguish the drawings of persons with substance abuse disorders from those without. Specifically, those persons treated for substance use disorders had lower ratings on the Person, Realism, and Developmental Level elements than those without, although utilizing all scales together provides the most utility in distinguishing the two groups. In fact, 11 of the 12 elements (Logic was rated equally for both groups), as well as the aggregate score were lower for drawings produced by clients in treatment for substance use.

In conclusion, this study demonstrated two important points. One, elements in drawings can be readily quantified, standardized, and communicated to art therapy practitioners, therefore providing empirical credence for the use of art in mental health service delivery. Two, drawings can be useful when combined with other data collection procedures to make differential diagnoses and to inform treatment decisions. The results of this study support the case for replicating studies that empirically quantify the connection between art and mental health assessment.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders,* (4th ed., text revision). Washington, DC: Author.

Carolan, R. (2001). Models and paradigms of art therapy research. *Art Therapy: Journal of the American Art Therapy Association, 18*(4), 190-206.

Exner, J. (1991). *The Rorschach: A comprehensive system* (2nd edition). New York: Wiley-Interscience.

Feder, B., & Feder, E. (1998). *The art and science of evaluation in the arts therapies: How do you know what's working?* Springfield, IL: Charles C Thomas.

Francis, D., Kaiser, D., & Deaver, S. (2003). Representations of attachment security in the Bird's Nest Drawings of clients with substance abuse disorders. *Art Therapy: Journal of the American Art Therapy Association, 20*(3), 125-137.

Gantt, L. (1986). Systematic investigations of art works: Some research models drawn from neighboring fields. *American Journal of Art Therapy, 24*(4), 111-118.

Gantt, L. (1990). *A validity study of the Formal Elements Art Therapy Scale (FEATS) for diagnostic information in patients' drawings.* Unpublished doctoral dissertation. University of Pittsburgh, Pittsburgh, PA.

Gantt, L. (1992). A description and history of art therapy assessment in research. In H. Wadeson (Ed.), *A guide to conducting art therapy assessment in research* (pp. 119-135). Mundelein, IL: American Art Therapy Association.

Gantt, L. (1998). A discussion of art therapy as a science. *Art Therapy: Journal of the American Art Therapy Association, 15*(1), 3-12.

Gantt, L. (2000). Assessments in the creative arts therapies: Learning from each other. *Music Therapy Perspectives, 18,* 41-46.

Gantt, L. (2001). The Formal Elements Art Therapy Scale: A measurement system for global variables in art. *Art Therapy: Journal of the American Art Therapy Association, 18*(1), 50-55.

Gantt, L., & Tabone, C. (1998). *The Formal Elements Art Therapy Scale: Rating manual.* Morgantown, WV: Gargoyle Press

Glass, G., & Hopkins, K. (1996). *Statistical methods in education and psychology.* Boston: Allyn & Bacon.

Hathaway, S., & McKinley, J. (1992). *Minnesota Multiphasic Personality Inventory* (2nd ed.). Minneapolis: The Regents of the University of Minnesota, National Computer Systems.

Kaiser, D. (1996). Indications of attachment security in drawing tasks. *Arts in Psychotherapy, 23*(4), 333-340.

Munley, M. (2002). Comparing the PPAT drawings of boys with AD/HD and age-matched controls using the Formal Elements Art Therapy Scale. *Art Therapy: Journal of the American Art Therapy Association, 19*(2), 69-76.

Neugebauer, R. (1979). Medieval and early modern theories of mental illness. *Archives of General Psychiatry, 36*(4), 477-483.

Rosal, M. (1998). Research thoughts: Learning from the literature and from experience. *Art Therapy: Journal of the American Art Therapy Association, 15*(1), 47-50.

Ulman, E., & Levy, B. (1968). An experimental approach to the judgment of psychopathology from paintings. *Bulletin of Art Therapy, 8,* 3-12.

Williams, K., Agell, G., Gantt, L., & Goodman, R. (1996). Art-based diagnosis: Fact or fantasy? *American Journal of Art Therapy, 35,* 9-13.