Utility of Drawings as a Screen for Emotional and Behavioral Concerns in Adolescents: An Exploration of Formal Elements in Drawings and the BASC-2.

BY

Selby M. Conrad

Submitted to the graduate degree program in Psychology and Research in Education and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Thomas S. Krieshok, Chairperson

Vicki Peyton

Tamara Coder Mikinski

Barbara A. Kerr

Philip C. McKnight

Date defended: May 7, 2010

The Dissertation Committee for Selby M. Conrad certified that this is the approved version of the following dissertation:

Utility of Drawings as a Screen for Emotional and Behavioral Concerns in Adolescents: An Exploration of Formal Elements in Drawings and the BASC-2.

Committee:

Thomas S. Krieshok, Chairperson

Vicki Peyton

Tamara Coder Mikinski

Barbara A. Kerr

Philip C. McKnight

Date approved: May 7, 2010

DEDICATION

This work is dedicated to my grandfather, Capehart Harney, who taught me to see the beauty in both art and science.

ACKNOWLEDGEMENTS

As with all things, this dissertation would not have been possible without others.

Thank you to my advisor, Tom Krieshok, for his unending support and guidance during the past six years. You encouraged me to trust my intuition and provided a solid foundation from which I could explore and experience the world of psychology to find my current path.

Many thanks to my committee, Vicki Peyton, Tammy Mikinski, Barbara Kerr, and Phil McKnight, for your feedback and thoughtful consideration of my research.

Much gratitude is extended to Vicki Peyton for seemingly endless hours of discussion related to measurement development and statistics, which resulted not only in this work but in developing my sense of mastery over statistical decision making.

Thanks to Heather Hunter, for her steady support. Your friendship, guidance, and support have added richness not only to this study but to life in general.

Thanks to Melissa Hill for her perspective on this project and friendship. You continue to inspire me with your creativity and inner strength.

Thank you to Sarah Brown, Maggie Syme, and Kerri Kim. Your friendships have challenged me to work harder, ask better questions and, now and then, have some fun.

Much gratitude is expressed to Marina Tolou-Shams, who appeared in my life just at the right time and continues to provide cherished support and mentorship. To future endeavors.

Many thanks to my family who has encouraged and guided me though life and academics:

To my brother, Dan Conrad. Your legal mind continues to impress me and challenge me to think of things in new ways. You have made me work for my opinions and support my findings. Thank you!

To my father, Robert Conrad, who has served as an inspiration to me in endless ways. Your support of my education has been unyielding, and your drive and commitment to education, both in and out of the classroom, your sense of social justice, and your love of family are present in all that I do.

Finally, to my mother, Helen Conrad, without whom this dissertation, and most of my academic work, would not have been possible. You have been a constant source of support and strength. Endless edits and drafts of papers have come your way and not once were they turned away. You have challenged me to be a better writer and a better person, with unconditional support and love. You have been, and continue to be, a sounding board, an artistic inspiration, a shining example of fun and whimsy, and my dear friend.

Abstract

The research related to the use of art in assessment of emotional and behavioral concerns in adolescent populations has met with inconsistent results. Much of the previous research in this area has focused on analyzing the content (i.e., what is drawn) of participants' drawings. This study explores the use of formal elements (i.e., how it is drawn) as a screening tool with a general sample of high school students (n=193). The aim of the project is two fold, a) explore the psychometrics and underlying factor structure of the FEATS an existing measure of formal elements, and b) assess the ability of the FEATS to accurately screen for emotional and behavioral concerns in high school students. The resulting data was analyzed for inter-rater reliability, distribution of item responses, and underlying factors. This resulted in the creation of 3, internally consistent, composite scales formed from 9 dichotomous items which had achieved acceptable levels of inter-rater reliability. A series of MANCOVAs between these composites and participant's scores on the BASC-2 suggest that the FEATS composites are not an acceptable screening tool for emotional and behavioral concerns. Results do indicate the FEATS composite may provide a means of assessing interpersonal relationships; however future research related to measurement design, factor structure, and ability to detect between group differences is needed before applied use.

Table of Contents

Appendices

Chapter I

Introduction

Humans have been creating art for thousands of years to communicate. In fact, images pre-date written language as a means to record significant events and important information. As humans, we observe the world around us and often are drawn to create a visual record of that which has impressed, awed, inspired, and terrified. Graphic communication has facilitated warnings, instruction, and creation of historical record. Neolithic men painted images of bison and deer on the walls of caves to illustrate the hunt; ancient Egyptians created a symbolic language; and modern artists continue this tradition of visual communication. Artistic media have mushroomed as humans continue to create visual records of thoughts and feelings related to self, interpersonal relationships, and societal experience.

Given humanity's use of symbolic imagery, it is not surprising that as the field of psychology began to develop early theories related to inner thoughts and feelings, the use of imagery and artwork was included. The belief that art provides a means of accessing emotion, self-perception, and experience provides the theoretical underpinning for the use of art in the assessment and treatment of patients of varying ages, diagnoses, and backgrounds. Art has been used for over a hundred years as an means of accessing patient information and remains one of the most widely used techniques for assessing children and adolescents (Archer & Newsom, 2000).

Falling under the umbrella of projective assessments, the use of drawings has changed as theoretical perspectives in psychology have developed. However, despite their widespread use, research focused on what items are present in patient drawings (i.e., content) have met much criticism due to ambiguous research support. For every study supporting the use of drawings in

this manner there appears to be another suggesting that drawing content is not a reliable or valid means of assessing personality, cognitive abilities or level of functioning (Anastasi, 1982; Dawes, 1994; Gittleman Klein, 1986; Lillenfeld, Wood, & Garb, 2000; Watkins, Campbell, Neiberding, & Hallmark, 1995).

Psychological research has largely focused on the content in drawings, such as inclusion or exclusion of body parts. Recent research focused on art-based formal elements - how the drawing was created - has provided a new direction for art-based assessment of emotional and psychological wellbeing (Cohen, 1985; Gantt& Tabone, 1998; Kramer, 1983; Rubin, 1984; Ulman, 1992). Artwork such as painting, drawing and sculpture, may provide insights into a person that other means cannot (Bruscia, 1988). It has been theorized that how a person draws (i.e., formal elements) is independent from the drawing's thematic content and could provide a means of screening clinically significant emotional and behavioral concerns in a variety of populations (Arnheim, 1986; Gantt, 2000; Gantt & Tabone, 1998; Hinz, 2008). If this theory is accurate, art-based assessment may provide a unique and non-threatening means of screening typically resistant populations, such as adolescents, for a variety of psychological concerns.

Researchers have been challenged to develop a standardized, reliable and valid art-based assessment of formal elements that allows for objective research without stifling the artistic expression needed to determine the absence or presence of formal elements. While several art-based assessments have been created to screen for emotional and behavioral concerns, only two provide objective rating scales. The Formal Elements Art Therapy Scale (FEATS; Gantt & Tabone, 1998) and the Diagnostic Drawing Series (DDS; Cohen, 1985) outline standardized instructions, materials and objective rating guides for the assessment of artwork. The objective rating measures have not been tested with two-dimensional artwork of varying thematic content,

but research conducted with a standardized drawing or series of drawings has yielded promising results (Anschel, Dolce, Schwartzman, & Fisher, 2005; Cheyne-King, 1990; Cohen, Cox, Mills, & Sobol, 1990; Cohen, Hammer, & Singer, 1988; Couch, 1992; Des Marais, 1996; Ferber, 1996; Fowler, 2002; Gussak, 2007; Heitmajer & Cohen, 1993; Kessler, 1992; Leavitt, 1988; McHugh, 1997; Mills, 1988; Morris, 1995; Munley, 2002; Neale, 1994; Rankin, 1994; Ricca, 1992; Rockwell & Dunham, 2006; Shlagman, 1996; Wallace, et al., 2004; White, Wallace, & Huffman, 2004; Woodward, 1998; Yahnke, 2000).

The current study explores the use of the FEATS measure of formal elements (Gantt & Tabone, 1998) to assess drawings of varying thematic content in an adolescent population. This project has two general purposes. The first is to determine the psychometric properties of the FEATS measure by assessing the inter-rater reliability of formal elements ratings, by analyzing the item response patterns and underlying factor structure of the measure, and by creating composite scales. The second purpose is to explore the use of the FEATS composites as screening tools for emotional and behavioral concerns in a general sample of adolescents, as measured by the Behavior Assessment System for Children-2 (BASC-2; Reynolds & Kamphaus, 2004).

Housed within the context of these two general research questions are five specific hypotheses:

I. Training raters in the use of the FEATS scale will result in acceptable levels of inter-rater reliability on all items.

II. The FEATS will contain an underlying factor structure allowing for the creation of composite scales.

III. FEATS composite scales will be stable across a 24 hour period, supporting that they are a measure of functioning independent of drawing content and themes.

IV. Formal elements, as measured by the FEATS composite scores, will predict participants' *emotional functioning* as measured by the related BASC-2 primary and composite scales while controlling for age and gender.

V. Formal elements, as measured by the FEATS composite scores, will predict participants' *behavioral functioning* as measured by the related BASC-2 primary and composite scales while controlling for age and gender.

There have been several studies that have explored the use of the FEATS, with the standardized Person Picking an Apple from a Tree (PPAT; Gantt & Tabone, 1998) drawing with clinical samples of adolescents (Ferber, 1996; Munley, 2002; Wallace, et al., 2004; White, Wallace, and Huffman, 2004); however, none has explored the FEATS with drawings other than the PPAT or in a general adolescent sample.

Adolescence is typically a time of great physical and psychological change, and it can be difficult to screen teens for emotional and behavioral problems. Teens are often guarded with adults and resist asking for help. Thus, a fast, cost effective, non-threatening and accurate means of screening for emotional concerns through artistic expression is appealing. With a sound measure of formal elements and a clear connection between formal elements and emotional and behavioral concerns, adolescents' drawings could be routinely screened, in a variety of settings, and intervention could occur as needed. This study represents an exploration of the psychometric properties of the FEATS and an initial attempt at understanding the connection, if any, between adolescents' emotional and behavioral concerns and the formal elements they include in their drawings.

Chapter II

Literature Review

For over a hundred years the field of psychology has been developing techniques to gain insight into human behavior and experience. As part of this search for understanding, psychologists have used a variety of tools to attempt to access the often elusive nuances of human behavior. Art production has often been used as an assessment of behavior, cognitive ability, developmental level and overall functioning, as it is viewed as means of accessing information that is parallel to, but distinctly different from, verbal communication. However, the research conducted on using art in this manner has yielded mixed results. This study has focused on a relatively new means of analyzing the drawings of adolescents in an attempt to understand their relationships to adolescent behavior.

Projective Techniques in Psychology

The use of art in psychological assessment is intertwined with the history and development of projective techniques. This is primarily due to a common foundation grounded in psychodynamic theory that posits internal thoughts, conflicts and feelings as the root of behavior. Thus, to change behavior, it was believed it necessary to understand this internal process by bringing thoughts and feelings to consciousness for analysis and reflection (Freud, 1911). Projective techniques are based on the principle that individuals unconsciously assign their personality traits and behaviors onto external stimuli including people, situations and images (Lillenfeld, Wood, & Garb, 2000). When patients are presented with an ambiguous stimulus, they are believed to project themselves onto this subject matter. Freud, who did not have much interest in the use of art production, clearly used this technique in his analysis of dreams and free association designed to elicit unconscious thoughts and feelings (Freud, 1911).

However, it was Jung (1967) who formally extended this technique to include both reflection on images and production of artwork as a means of accessing internal information through projection.

To those who subscribe to this theoretical perspective, the utility of the projective process may appear to be a perfectly reasonable means of accessing internal information otherwise unconsciously concealed. However, even a brief glance at the literature and research conducted over the last hundred years suggests that there is little consensus on the use of these techniques for assessment and treatment. If fact, much research indicates that projective assessment should no longer be used in any clinical setting due to a lack of reliability and validity which leads to inaccurate information for diagnosis and treatment (Anastasi, 1982; Dawes, 1994; Gittleman Klein, 1986; Lillenfeld, Wood, & Garb, 2000). That said, other research supports the notion that much can be gleaned from the clinical use of projective techniques (Watkins, Campbell, Neiberding, & Hallmark, 1995). With continued refinement to increase reliability and validity, they believe these tools can yield a wealth of rich information to help clinicians develop a comprehensive picture of their patients-- information that may be unavailable without this form of assessment.

What is clear is that despite this on-going debate, projective techniques remain some of the most frequently used means of gathering clinical information with children and adolescents. In their survey of psychological test use with adolescents, Archer and Newsom (2000) found that five of the top thirty assessments used by psychologists were projective techniques, and the Rorschach Inkblot Technique (Rorschach, 1921), the Thematic Apperception Test (TAT; Murray, 1943) and the House-Tree-Person Technique (H-T-P; Buck, 1948) were in the top ten. These results parallel Camara, Nathan, and Puente's (2000) study focused on psychologists who

work with adults indicating projective assessments are commonly used in clinical settings with all age ranges, despite concerns raised about their reliability and validity.

Projective techniques: Definition and subtypes. In light of the concerns raised about the accuracy of projective techniques, researchers have explored multiple questions related to this field. A brief literature search using the key word projective yielded approximately 13,500 peer reviewed articles. However, despite this prolific amount of literature and the long standing use of projective techniques in psychological assessment, it is perhaps surprising that the definition of projective techniques and a common system for classification of these tools remains ambiguous. In their 2000 review of projective techniques, Lillenfeld, Wood, and Garb point out this absence stating that the "definition of projective techniques is less clear cut than many authors have assumed" (p.28).

Projective techniques have been defined as "any test, device, or set of procedures designed to provide information about or insight into an individual's personality by allowing the individual the opportunity to respond in an unrestricted manner to art-based materials or visual constructs" (Reber & Reber, 2001, p.177). This definition supports the idea that the stimuli allow for varied response, but it does not indicate that the stimuli needs to be ambiguous. Other definitions do indicate that projective techniques are defined by ambiguous stimuli. For example, Garb indicates that the stimuli used in projective techniques tend to be more ambiguous than in other tests, and this leads to an increase in the variety and number of responses (Garb, Wood, Nezworski, Grove, & Stejskal, 2001).

The lack of agreement in the definition of projective techniques has compounded the problems in researching these techniques, and it has been suggested that concretely operationalizing the definition of projective techniques will aid in the development and

successful completion of methodologically sound research (Wagner, 2003). Further adding to the complexity of defining this category of assessment, Wagner suggests that any definition must highlight the distinction between *permissiveness* and *ambiguity*.

Permissiveness is described as the latitude in both task instructions and stimuli that fosters a wide range of participant responses. *Ambiguity* refers to the stimulus itself and the degree of ambiguity in a projective technique can range from clear (incomplete sentence blank stems) to obscure (ink blots) (Wagner, 2003). Thus, according to Wagner, all projective techniques would have a high degree of permissiveness but not all have ambiguous stimuli. While providing a simple definition of projective assessment, the combination of these two factors allows for a seemingly infinite number of variations in procedure, process, and rating. Thus, the Rorschach Ink Blot Test, a Sentence Completion Test, and Human-Figure Drawings fall into the same category of assessment.

To provide some structure to the broad topic of projective techniques, Lillenfeld, Wood, and Garb (2000) propose six categories of projective techniques including: Association (e.g. Rorschach Ink Blot Test; Rorschach, 1921); Completion (e.g. Washington University Sentence Completion Test; Loevinger, 1976); Arrangement/Selection (e.g. Szondi Test; Szondi, 1947); Expression (e.g. Projective Puppet Play; Woltmann, 1960); Handwriting Analysis (Beyerstein & Beyerstein, 1992); and Construction (e.g. Draw-A-Person Test; Machover, 1949). While all of these categories encompass projective assessments, it is the category of Construction that is most applicable to this study and will be explored further in this review. A projective technique in the Construction category asks the patient to create a response to a permissive stimulus or directive using a standardized set of materials. Standardized rating manuals can then be developed to

explore the variety of responses produced, allowing for drawings to be rated in a systematic fashion that may help decrease concerns related to reliability and validity.

Psychological projective drawing assessment. The production of drawings has been described as a standardized process though which feelings, behaviors, relationships and thoughts can be translated into visual representations and provide a record of thinking and problem solving abilities (Arnheim, 1969, 1986; Burgess & Hartman, 1993; Neal & Rosal, 1993; Taylor, Kymissis, & Pressman, 1998). When asked to draw, patients often tap into conscious and unconscious ideas, allowing for exploration of thoughts and feelings that might not have been accessible verbally (Taylor, Kymississ, & Pressman, 1998). Based on these assumptions, drawings of varying themes have been used as assessment tools in the evaluation and diagnosis of cognitive abilities, psychological states, and behavioral concerns (Abell, Heiberger, & Johnson, 1994; Amir & Lev-Wiesel, 2007; Hjorth & Harway, 1981; Joiner, Schmidt, & Barnett, 1996; Lev-Wiesel & Shvero, 2003; Lev-Wiesel & Hershkovitz, 2000).

Three of the most popular drawing assessments used are the House-Tree-Person (H-T-P; Buck, 1948), Kinetic Family Drawing (K-D-F; Burns & Kaufman, 1970) and Draw-A-Person (DAP; Machover, 1949), all of which provide a standardized set of materials and instructions for administration (Archer & Newsom, 2000; Camara, Nathan, & Puente, 2000). A review of the literature and research related to the use of human figure drawings in psychological assessment provides a context for the current study.

Drawings as a measure of intelligence. In her pioneering research on drawings, Goodenough (1926) explored the use of drawings as a measure of intelligence. She developed the first standardized drawing assessment, Draw-a-Man (D-A-M), to measure non-verbal intelligence. Goodenough collected the first large scale research on children's drawings,

collecting over 4,000 human figure drawings from children representing diverse ages, ethnic groups and geographic regions of the country to provide norms for rating. The final rating system for the D-A-M is divided into two main categories: recognizable as a human figure, and unrecognizable. If recognizable, drawings are then rated based on the inclusion of 51 specific items such as eyes, mouth, hands, and clothing. These items are dichotomous and the total number of points received is summed and compared with the age specific norms for intelligence. While not negating the possibility that human figure drawings could be used as indicators of emotional functioning, Goodenough avoided any discussion using children's drawings to this end (Hagood, 2003).

As the D-A-M increased in popularity, psychologists began to use this tool to subjectively assess psychopathology in children and adolescents. Harris (1963) criticized the use of drawings in this manner as diagnostic impressions based on intuitive impression rather than systematic scoring and research. Ironically, fifty years of literature continues to highlight this very concern and illustrates the complexity of using children's drawings in assessment. Continuing to promote the use of drawings as a measure of intelligence, rather than psychopathology, Harris expanded Goodenough's assessment of children's human figure drawings by adding two additional human figure drawings 1) a woman, 2) a self-portrait (Hagood, 2003). In each of the drawings, the patient is asked to draw the whole human figure which is then rated on 73 dichotomous items similar in content to Goodenough's original system. This resulted in the Goodenough-Harris Drawing Test (1963) and is the system that is used commonly today. The test was normed on 2,622 children and adolescents aged 5-17 from a variety of ethnic groups, and the original study reports acceptable inter-rater reliability on items ranging from .74 to .90.

However, other studies do not support the use of human figure drawings to assess intellectual ability and development. Abell, Horkheimer, & Nguyen (1998) explored the use of human figure drawings to assess intelligence in adolescent boys. The 200 drawings collected in their study, from boys ages 14-15, were rated using the Goodenough-Harris (1963) scoring system. Though acceptable inter-rater reliability was reported, the results of this study suggest that this tool is not sufficient for assessing intelligence. While a significant correlation between intelligence and human figure drawings was reported, researchers noted that drawing scores underestimated the IQ scores of adolescent participants. This suggests that while human figure drawings may be correlated to intelligence, current rating systems are insufficient in capturing this variable successfully and with the diagnostic accuracy needed.

Drawings as a measure of emotional functioning. Building on the Goodenough-Harris rating scale for intelligence, Koppitz (1968) developed a similar rating scale to be used with drawings of human figures that gauges socio-emotional functioning and assesses psychopathology, in addition to developmental level and visual-motor abilities. Much like the Goodenough-Harris, this scale focuses on the details in the drawings and includes items to note presence and absence in human figure drawings. Koppitz (1968) included many emotional indicators (EIs) on the rating scale which are believed to correspond to a variety of emotional concerns, including items such as legs pressed together, large hands, and presence of hands. The dichotomous rating scale can rated by someone who was not present at the collection of the drawing as items on the rating scale focus solely on the content of the image rather than including any verbalization about the images (Hagood, 2003). This work provided an objective means of rating drawings that could be utilized in research and expanded the use of drawings to include assessment of emotional functioning.

There has been much research exploring the use of human figure drawings and Koppitz's (1968) emotional indicators to distinguish among clinical populations. Much of this work is in the area of childhood sexual abuse. Wood, Howe, Burgess & McCormack (1987) found significant differences in the emotional indicators found in the human figure drawings of sexually abused adolescents compared to their non-abused counterparts. Such indicators included ambiguity of sexual identification of the figure, emphasis on facial features, faint sketchy lines, and limited range of color. Johnson (1989) found support for Koppitz's content rating of emotional indicators of anxiety related to sexual abuse. Broken lines, tiny heads, omission, and vacant eyes were found to be helpful in identifying children experiencing anxiety and stress related to traumatic events.

Sidun and Rosenthal (1987) also explored the use of the human figure drawing as a screening tool for sexual abuse with hospitalized adolescents. They noted several trends in drawings of sexually abused adolescent males such as inclusion of more phallic objects, poor body integration, omitted hands and fingers, and drawings of heads only, despite instructions to draw a whole human figure. Despite these trends, their hypothesis that the human figure drawings contained significant between group differences in self-esteem, anxiety, and sexuality was unsupported. Hibbard and Hartman (1990) found that if emotional indicators were analyzed in categories, rather than as individual items, human figure drawings could differentiate between groups of adolescent participants. This suggests that drawing items may comprise scales which are more sensitive to detecting between-group differences than individual items.

In addition to exploring between-group differences in drawings for survivors of sexual abuse, a variety of studies have examined the drawings of other groups, such as patients with eating disorders, substance abuse problems, and history of sexual abuse, to determine if drawings

can be used to accurately screen for a variety of mental and physical health concerns (Taylor, Kymissis, & Pressman, 1998). A few studies have used adolescent self-report scales of behavior and emotional experiences to compare to drawings.

In their 2005 study, Milne, Greenway, and Best rated the content of children's drawings to determine if these items were related to their scores on the Behavioral Assessment System for Children (BASC; Reynolds & Kamphaus, 1992) and the Child Behavior Check List (CBCL; Achenbach, 1991). Additionally, an objective scoring rubric was used to rate each child's drawings. The content of the human figure drawings of self and parents included details such as teeth, facial hair, eyes, lips, etc. Ratings on twenty two content variables highlighted similarities and differences in inclusion of these items across drawings.

These scores were then compared to the BASC and the CBCL and revealed significant gender differences. Boys' externalizing behaviors, as reported on the BASC and the CBCL, were significantly correlated with emotional indicators found in their self portraits and the drawings of their mothers. However, girls' BASC and CBCL scores were not predicted by any of the content items rated. Unfortunately, there is no mention of inter-rater reliability being conducted on the drawing ratings. In the absence of a reliable rating rubric, other findings are questionable. Additionally, the small number of participants (n=75), paired with the large number of variables being explored, limits the power of the statistical analysis. However, this study does support the feasibility of collecting adolescent self-report measures and drawings in a group context and rating them according to an objective rating scale.

In response to much criticism related to the reliability and validity of human figure drawings in psychological assessment, Naglieri (1991) further honed rating scales for human figure drawings for use with adolescent populations though the creation of the Draw-A-Person:

Screening Procedure for Emotional Disturbance (DAP:SPED; Naglieri, 1991). The DAP:SPED incorporates the more traditional content elements of human figure drawings described in Koppitz's (1968) emotional indicators with a global approach (e.g., space usage, line quality) to ratings aimed at assessing emotional functioning.

The DAP: SPED includes 55 items that are applied to three drawings: a man, a woman, and a child. Items are then summed for each drawing resulting in a total score which can be used in statistical analysis. This tool was standardized on a sample of 2,355 children and adolescents ages 6 to 17. Researchers report acceptable inter-rater reliability (above .91 on all scales) and the scales appear to be independent of level of intelligence, which is notable since DAP has traditionally been used to assess level of cognitive development (Naglieri & Pfeiffer, 1992).

The DAP:SPED appears to show promise in identifying adolescents with behavioral disturbance. Naglieri and Pfeiffer (1992) collected D-A-P drawings from 108 adolescents (54 control; 54 Oppositional Defiant Disorder). Drawings were collected in a group setting with other self-report measures including intelligence. This study's results suggest that adolescents diagnosed with ODD have lower total scores on the DAP: SPED than their non-clinical counterparts, and support the hypothesis that the DAP:SPED can be used to identify clinical and non-clinical samples.

Additionally, human figure drawings have been used to explore a variety of characteristics in child, adolescent and adult populations. Studies have researched perception of self (Berryman, 1959; Leibowitz, 1999); anxiety (Fox, Davidson, Lighthall, Waite, & Sarason, 1958; Johnson, 1989; Sturner & Rothbaum, 1980); externalizing behaviors (Koppitz, 1968, 1984; Milne, Greenway & Best, 2005); sexual abuse and neglect (Burgess & Hartman, 1993; Hibbard & Hartman, 1990); and relationships with others (Main, Kaplan, & Cassidy, 1985).

Most, if not all, of these studies utilize objective scoring rubrics that focus on the content (emotional indicators) of the human figure drawing by rating what is and what is not present in the drawing (e.g., hands, eyes, feet). While this represents only a small sample of the psychological rating scales developed for use with children's drawings, it is clear that human figure drawings have the potential to assess a variety of different aspects of child and adolescent functioning but continue to have equivocal results related in finding significant differences between groups.

Problems with projective drawing assessment. In their review of the literature on human figure drawings, Veltman and Browne (2002) attempted to ascertain if drawing assessments are valuable in the assessment of children and adolescents. Their findings were inconclusive due to the large amount of contradictory research using drawings as an indicator of emotional state, relationships or behavioral concerns. On the one hand, it appears that drawings may be useful in overall assessment of children and adolescents (Falk, 1981; Landgarten, 1987; Malchiodi, 1990; Veltman & Brown, 2002). However, the literature also continually echoes concerns over a variety of methological problems including ways of analyzing the drawings, low inter-rater reliability, and small sample size.

Currently, drawings produced by children and adolescents in assessment or therapeutic settings tend to be analyzed in two different ways. The first is the "global impression" method that uses assessors' experience and reactions to the drawing to provide information (Lally, 2001). This method of analysis, though frequently used, is obviously not objective, structured, or reliable. While not negating clinical experience and expertise, it is extremely difficult to reproduce results gleaned in this subjective manner, and thus it is not an acceptable means of systematically assessing patient's cognitive, behavioral, or emotional well being.

The second method of analyzing drawings is through the development of an objective rating guide. Even with a rating rubric that objectively guides raters through the process of scoring drawings, inter-rater reliability continues to be a problem in studies that explore the use of drawings as assessment tools. For example, in his review of the literature, Kahill (1984) found that a majority of the indicators scored in human figure drawings were reliable at .80, and about two-thirds were above .70, which falls within acceptable limits. He also reported that inter-rater reliability could range from .45 to .96 for shading, .54-.99 for head size, and -.13 to .60 for facial expression. Palmer, et. al (2000) and Vass (1998) reported similarly poor inter-rater reliabilities ranging from .27 to .75 on items rated in human figure drawings. This inconsistent inter-rater reliability raises the concern that raters are not seeing the same elements in drawings, which makes any statistical analysis resulting from these items questionable.

Additionally, validity continues to be a challenge to the use of drawings in an assessment context. For every study supporting the ability of emotional indicators (content items) in drawings to distinguish between different behavioral and emotional concerns (Hammer, 1954, 1969; Holtzman, 1993; Reithmiller & Handler, 1997a, 1997b), there is another asserting that drawings possess little to no validity in this context (Kahill, 1984; Klopfer & Taulbee, 1976; Lillenfeld, Wood & Garb, 2000; Motta, little, & Tobin, 1993; Roback, 1998; Suinn & Oskamp, 1969; Thomas & Jolly, 1998). Additionally, several studies have shown that the variability in human figure drawings can be accounted for by artistic ability rather than emotions or behavior, and this ability appears not to have a strong relationship to psychopathology of any kind (Adler, 1952; Cresssen, 1975; Feher, VanderCreek & Teglasi, 1983; Nicholas & Stumpfer, 1962).

While there is no doubt that the use of drawings as a means of accessing and measuring psychological states is an appealing means of collecting information, the current challenges with

the content rating of drawings suggest that different ways of quantifying drawings in an objective manner are necessary.

A Case for a Different Kind of Drawing Assessment

Projective drawing techniques can provide a wealth of information about diagnostic categories and personality traits, but they are limited in the information that can be assessed due to the restricted range of materials and no freedom to select the theme of the drawing (Hinz, 2008). In psychological projective drawing assessments there is little to no consideration paid to the way in which an individual approaches a variety of art material and how they use these materials to create art (Gantt, 2004). Thus, art based assessments that allow choice from a wider range of media along with the freedom to create images based on personal choice could provide a wealth of information about the internal state of the patient that may be limited in more traditional projective techniques (Hinz, 2008). Arnheim (1969, 1986) noted the difficulty in reducing artistic production to non-artistic processes and suggested that artwork would be best understood based on the structural and dynamic principles that underlie artistic creation.

Arts such as painting, drawing and sculpture, may provide a unique perspective and novel insights into a person that other disciplines cannot (Bruscia, 1988). The field of art therapy, which combines knowledge of art materials and appreciation for the richness of artistic expression, with supportive and specific counseling techniques found in psychology, provides an interesting theoretical foundation for research that combines these two divergent fields. As such, art therapy researchers have begun to explore developing art based assessments that allow for artistic expression as part of the information to be assessed. Researchers and clinicians have begun develop art rating scales based on Arnheim's suggestion focusing on both the structure and the artistic qualities of patients' drawings.

Kramer (1983) developed an art assessment that instructs patients to create three works of art using drawing materials, paint and clay. While the materials are suggested, they are by no means standardized, and the directives for this assessment are completely open-ended and encourage participants to create whatever they would like. There are no specific instructions on how to inquire about each piece of art, or standardized rating guides for this evaluation. This limits its diagnostic utility and research potential. However, this assessment marks a development in using the art-making process to assess patients' personality traits, cognitive development and level of emotional functioning rather than solely analyzing specific emotional indicators.

Others have also suggested the use of multi-media art-based assessment. Rubin (1984) suggested using a variety of materials (wood, paint, clay, and drawing material) and providing no specific directives related to subject matter to assess both cognitive and emotional functioning. Rather than a rating based on a standardized scale, the examiner observes the art-making process and qualitatively assesses the patient's selection and use of materials, approach to the activity, content of the artwork, and verbalizations during the process. While this technique may provide a wealth of qualitative data, it has not been systematically studied. The diagnostic impressions and clinical inferences are based on observation and intuition which has been shown inaccurate (Ulman & Levy, 1967, 1992). Additionally, the assessment is difficult to give because of the numerous materials provided. While this form of assessment provides a wonderful way to engage children and adolescents in a novel non-threatening activity, the lack of standardization and difficult administration makes its utility as an assessment tool questionable.

In an attempt to provide a more structured art based assessment, Ulman (1992) developed the Ulman Personality Assessment Procedure (UPAP) to aid in diagnostic assessment of adult

psychiatric patients. The UPAP consists of a standardized set of materials: gray construction paper and 12 hard pastels, and standardized instructions for administration of four drawings which allow for freedom of artistic process and expression: 1) free drawing, 2) have patients practice swinging arms up and down, and in circular motion then record this on paper, 3) a scribble drawing which is used as a starting point for the patient to develop a drawing, 4) choice between another scribble drawing or free drawing. The UPAP does not provide standardized rating instructions, and the examiner forms diagnostic impressions of the patient based on intuition, previous experience and supervision rather than on an objective research base.

These art-based assessments are based on a qualitative understanding of human expression, lack scientific support and echo concerns in the human figure drawing literature related to reliability and validity (Betts, 2006). Given the training focus on artistic creation and freedom of expression, many art therapists are resistant to standardizing art-based assessment, believing that it stifles the creative process (Gantt, 2000). However, several studies have raised concerns about the ability of art therapists and psychologists to intuitively assess patients' drawings. While Ulman and Levy (1967) report that blind judges are able to distinguish between drawings of patient and non-patient undirected drawings at a 90% success rate, this analysis does not include any diagnostic specificity. Rubin and Schachter's (1972) study suggests that when asked to blindly judge children's drawings with diagnostic specificity, the success rate drops significantly (20%-40% success). This highlights the necessity of objective rating scales and statistical support for diagnostic conclusions based on drawings.

In addition to the importance of a standardized objective rating scale, Eitel, Szkira, Pokorny, and Von Wietersheim (2008) illustrate the importance of training raters in scale use. Their study explored the ability of raters to agree on the presence and presentation of formal

elements (shape, lines, and colors) in drawings. Their study explored the use of a 5 point Likert scale with untrained raters. Researchers collected drawings from 156 participants and used a standardized rating manual with untrained raters, including psychologists, students, and art therapists. Results of this study suggest that without training in the specific rating scale and scoring system, raters, regardless of level of education and training, were unable to produce acceptable inter-rater reliability on seemingly objective rating scales. In fact, no group obtained an inter-rater reliability over .53 of any of the formal elements. This study indicates that proper training in the use of any art-based rating scale is essential to ensure reliable results.

Additional concerns that have been raised in the art therapy and psychology literature surrounding art based assessment research includes small sample size (Betts, 2006; Hagood, 2002), poor inter-rater reliability due to subjective rating scales (Hacking, 1999; McNiff, 1998; Ulman & Levy, 1992), methodological concerns around scale development (Betts, 2006; McNiff, 1998), and failure to build directly on previous research in the field (Hacking, 1999). Researchers are attempting to address some of these concerns and contribute meaningfully to the dialogue about the use of art as an assessment of cognitive ability, personality functioning and behavior. Building on this artistic foundation, several researchers have worked towards developing assessments focused on the art-based process that provide an objective and standardized process for rating that allows for meaningful statistical analysis.

Joiner, Schmidt, and Barnett (1996) explored the reliability and validity of three art based indicators traditionally associated with emotional distress in children's drawings: size, detail, and line pressure. Using data collected from an in-patient sample of 80 children and adolescent ages 6 to 16 (53 male), researchers collected a large battery of psychological tests including the Children's Depression Inventory (CDI; Kovacs, 1992), the Revised Children's Manifest Anxiety

Scale (RCMAS; Reynolds & Richmond, 1985), Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), Roberts Apperception Test for Children (RATC; McArthur & Roberts, 1990), and several drawings including the Kinetic-House-Tree-Person (KHTF; Burns, 1987), a Kinetic-Family-Drawing (KFD; Burns & Kaufman, 1970), and a Draw-a-Family (DAF; Hulse, 1951). The drawings were then rated on a 10 point Likert scale according to size, detail, and line pressure, and researchers reported acceptable inter-rater reliability (size=.95; detail=.91; pressure=.92). Despite findings in previous studies, researchers found that these drawing dimensions were not significantly related to depression and anxiety.

LaRoque and Obrzut (2006) used technology to explore the relationship between line pressure and trait anxiety in 50 children ages 6 to 11. Participants were asked to draw several images, including DAP, on a pressure sensitive palette that recorded the pressure exerted during drawing. Results from this study suggest that children with higher levels of trait anxiety use less pressure when drawing, which is consistent with previous research (Handler & Reyher, 1966). This study supports the notion that how children and adolescents draw may be related to how they are feeling, and the different findings between the LaRoque and Obrzut (2006) study and the Joiner, Schmidt, and Barnett (1996) study highlight the on-going challenges with researching drawings. However, the LaRoque and Obrzut (2006) study illustrates the degree of operational specificity that may be needed to reliably rate drawings and use them as a valid measure of mood.

Cohen (1985, 1986) developed the Diagnostic Drawing Series (DDS) in an attempt to address some of the concerns around the scientific imprecision in using art as part of a psychological battery of assessment (Mills, Cohen, & Meneses, 1993). Using standardized materials (three 18 by 24 inch pieces of white paper and a 12-color box of square, soft chalk

pastels) the patient is asked to complete three drawings that include: 1) “Make a picture using these materials”, 2) “Draw a picture of a tree”, and 3) “Make a picture of how you are feeling using lines, shapes and colors” (Cohen, 1985). Using a formal objective rating guide, these drawings are scored on 23 scales using objective and specific criteria that include both formal elements (such as color use, blending of colors, and space use) and content items (such as water and animals). Inter-rater reliability on these items is reported at above 95% but requires two months of training to reach this level (Mills, Cohen, & Meneses, 1992).

While providing a structured and objective means of rating the DDS, the rating scale is categorical and presents some challenges to statistical analysis. Several studies have been conducted on this instrument to establish norms in a variety of diagnostic categories including: non clinical control studies with adults (Cohen, Hammer, & Singer, 1988; Morris, 1995), seniors (Couch, 1992) and children and adolescents (Leavitt, 1988; Neale, 1994; Shlagman, 1996); and clinical samples of individuals with Schizophrenia (Cohen, Hammer, & Singer, 1988; Morris, 1995; Ricca, 1992), Mood Disorders (Cohen, Hammer, & Singer, 1988; Leavitt, 1998; McHugh, 1997), Dissociative Disorders (Fowler, 2002; Heitmajer & Cohen, 1993; Ricca, 1992), Eating Disorders (Kessler, 1992), Borderline Personality Disorder (Mills, 1988), Post Traumatic Stress Disorder (Des Marais, 1996), and Conduct Disorder (Neal, 1994).

In addition to collecting normative data from individuals with specific DSM-IV-TR diagnoses, researchers exploring the DDS have conducted studies with a variety of clinical groups including child witnesses to violence (Woodward, 1998; Yahnke, 2000), childhood sexual abuse survivors (Cohen, Cox, Mills, & Sobol, 1990; Rankin, 1994), individuals with traumatic brain injury (Cheyne-King, 1990), women with dysmorphic body image (Anderson, 1997), military personnel with Gulf War Syndrome (Des Marais, 1996), and men who have been

incarcerated for murder (Mitchell, 1995). These studies have attempted to provide a research base for the utility of the DDS as a diagnostic tool in a variety of clinical settings, and results indicate that formal elements related to artistic expression may be a fruitful direction for future research.

Formal elements art therapy scale. It has been suggested that focusing on the artistic structure of drawings may provide insight into the psychological functioning of patients that may unavailable if only the content of the drawing is explored (Durken, 1954; Kahn & Jones, 1965; Lehmann & Risquez, 1953; Ogdon, 1975; Ulman & Levy, 1992). Gantt and Tabone (1998) also observed that much like psychologists working with the same art materials, they were not able to achieve reliable diagnostic information from the content of patients' artwork (Gantt, 2001). Moving away from the content of patients' artwork and projective assessments, they began to develop The Formal Elements Art Therapy Scale (FEATS; Gantt & Tabone, 1998) that focused on *how* people draw (formal elements), rather than emphasizing *what* they draw (content/emotional indicators).

The FEATS was developed as an objective measurement system for rating the formal elements in any two-dimensional work of art on a 5-point Likert scale (Gantt, 2001) and was validated by rating the Person Picking an Apple from a Tree (PPAT) drawing of adults from a variety of different diagnostic groups (Gantt & Tabone, 1998). While research conducted on the FEATS has consistently used the PPAT drawings, Gantt (2001) clearly states that the rating scale focuses on the formal elements of drawings and thus can be utilized with drawings of different thematic content and with artists of all ages.

Gantt and Tabone (1998) clearly state that the FEATS is designed to look at drawings as art, and thus the scales reflect items that are important to artists, including items such as color,

line quality, space usage and developmental quality. The resulting scale is grounded in an artistic approach to the image and attempts to address some of the limitations with psychological drawing and projective assessments. Additionally, the FEATS is designed to explore clinical states rather than personality traits which is believed to increase its utility in the assessment and diagnosis of a variety of psychiatric disorders (Gantt, 2004).

After several years of development, Gantt and Tabone (1998) proposed the current FEATS which is comprised of 14 scales: Prominence of Color, Color Fit, Space, Integration, Implied Energy, Logic, Realism, Developmental Level, Details of Objects in the Environment, Line Quality, Person, Rotation and Perseveration. While some of these scales echo items used to rate artwork in the psychology literature, the ways in which these scales have been operationalized have a decidedly artistic slant (see Methods for further details on scale development). The objective and concrete manner in which these items are presented have made the FEATS a useful tool for research. Additionally, the premise that the FEATS may be sensitive enough to detect change over time makes it an interesting tool for outcome research and, in addition to studies related to diagnostic and between group differences, there have been several studies to that end (Gantt, 2001).

Rockwell and Dunham (2006) explored the use of the FEATS using PPAT drawings in identifying patients with a diagnosis of Substance Use Disorder. Using a small sample (N=40, 20 with substance use disorders) of participants matched for age, gender, race and socio-economic status, PPAT drawings were collected and blindly rated using 12 of the 14 FEATS scales (omitting Rotation and Perseveration). Rockwell and Dunham report acceptable inter-rater reliability on most of the items with the exception of Line Quality (*p=.416)*, Integration *(p=.251)*, and Problem Solving *(p=-.057)*. Despite these low reports, all 12 scales were included

in their analysis which found that the drawings of the Substance Use Disorders group significantly differed from the control group on Realism *(p=.027)*, Developmental Level *(p=.010)*, Person *(p=.002)*, and an aggregate score *(p=.022)*. This study concludes that the FEATS can be used in a reliable manner and may provide a means of distinguishing between groups.

Gussak (2007) used the FEATS in conjunction with PPAT drawings to explore the effectiveness of art therapy in the reduction of depression in a pilot therapy program with prisoners. Using the FEATS, a survey of inmate interactions and compliance with rules, and the Beck Depression Inventory, Gussak explored the utility of the FEATS in highlighting symptom change over time for those involved in a 4-week art therapy program. Despite his relatively small sample (N=48) and no control group, Gussak observed that inmates involved in the study had significant changes in their PPAT drawings on five scales (Prominence of Color, Color Fit, Energy, Details of Objects in the Environment, and Space) that were consistent with observations of symptom reduction. In a similar follow up study, Gussak was unable to replicate these results, finding no change on BDI scores or the FEATS for those participating in the art therapy program. Despite the ambiguous results for art therapy as an intervention, it appears in both studies that the FEATS yielded similar results to the more traditional means of assessing symptom reduction, and thus these studies provide valuable support for the use of the FEATS in the detection of change over time.

Anschel, Dolce, Schwartzman, and Fisher (2005) explored the use of the FEATS as a diagnostic tool for adults with epilepsy. Researchers blindly collected three drawings, a PPAT, a Free Drawing, and a body outline, from 60 adult participants who fell into one of four groups: seizures, partial seizures, complex partial seizures, or non-epileptic. The PPAT was rated using

the FEATS, and the Free Drawing was rated using a modified version of the FEATS which included all scales except Problem Solving, Rotation and Perseveration. Data were analyzed using a multivariate logistic regression, and results of this pilot suggest that the FEATS has the sensitivity to distinguish between the epileptic group and controls. Additionally, the Free Drawing task appeared to further predict membership in the complex partial seizure group. Researchers did not discuss inter-rater reliability obtained on the FEATS for this study, nor did they discuss the rater training procedures, but they confirmed the feasibility of using art assessment and the FEATS rating scale in this medical setting.

The FEATS was modified in Swan-Foster, Foster, and Dorsey's (2003) study of the use of human figure drawings with pregnant women. This study was aimed at determining the utility of the FEATS paired with a Human Figure Drawing to assess pregnant women for mental health concerns during pregnancy. Researchers used a modified FEATS scale that included prominence of color, space, detail, person, realism, developmental level, line quality, perseveration, and implied energy to score 30 drawings completed by women falling into three risk categories (high risk outpatient, high risk inpatient, and low risk) for mental health concerns during pregnancy. A multivariate analysis of variance highlighted statistically significant between-group differences on color, energy, space, details, and person on the modified FEATS. Differences on these scales were present across all three groups. The researchers report acceptable inter-rater reliability on the FEATS scales but do not discuss their training procedures. This study suggests that several of the FEATS scales are sensitive enough to distinguish between groups with varying mental health concerns using drawings other than the PPAT.

Most of the research using the FEATS has been with adults. However, a small number of studies has explored this rating scale with children and adolescents. Munley (2002) explored the ability of the FEATS to distinguish non-medicated ADHD in boys aged 5-12 from a medicated ADHD control group. While this matched pair study found differences in the drawings of the ADHD and control groups on three of the subscales: Color Prominence, Details of Objects and Environment, and Line Quality, the results are limited due to an extremely small sample size (n=13). While the authors reported acceptable general reliability between blind raters, they do not report specific inter-rater reliability on the subscales of the FEATS.

Ferber (1996) conducted a study with high school freshmen (n=32) that explored the correlation between the FEATS and the Coopersmith Self-Esteem Inventory-School Form (CSI-SF; Coopersmith, 1981). This study reported correlations between lower scores on the Details of Objects and Environment, Line Quality and Person subscales of the FEATS and lower reports of self-esteem among adolescents. However, even the strongest correlations in this study would be considered weak ($r>.4$). Nonetheless, this study illustrated the feasibility of drawing and self-report data collection in the high school setting. Additionally, researchers were able to maintain acceptable levels of inter-rater reliability ($a<.8$) when using the FEATS to code drawings.

In 2004, Wallace, et al., used the FEATS to assess depression and emotional trauma in pediatric transplant patients. A sample of 64 renal transplant patients, ages 6-21, were asked to complete the Childhood Depression Inventory (CDI; Kovacs, 1992) and a PPAT drawing which was then rated using the FEATS. In contrast to the CDI, which was not collected for everyone, all participants in this study were able to complete the drawing assessment. While the results of this study indicate that the FEATS is not sensitive enough to detect depression or experience of trauma in this population, they also suggest that using art in this pediatric population was not

only feasible, but perhaps a less threatening and more acceptable means of collecting data. White, Wallace, and Huffman (2004) extended the use of the FEATS to include children and adolescents by comparing the FEATS to the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), an established self-report measure of emotional and behavioral problems. This sample consisted of 53 adolescents ages 8-17 (*M*=12.2, *SD*=2.3) enrolled in a therapeutic day school with varying co-morbid psychiatric diagnoses. Researchers report inter-rater reliability coefficients of .73 to .93 on 10 of the 14 FEATS scales. Integration (.45), Line Quality (.47), Rotation (.13), and Perseveration (.26) were the scales with the lowest inter-rater reliability, but researchers chose to drop only Rotation and Perseveration from the analysis. Participants were divided into two groups: impaired and non-impaired thinking, based on their CBCL scores. ANOVAs yielded significant between-group differences on the Problem-Solving, Integration, Realism, Developmental Level, and Details of Objects and Environment subscales. A follow-up discriminate function analysis was significant in predicting group membership based on these variables, most of the variance being accounted for by Problem-Solving. The study's results indicate that the FEATS may be able to distinguish between groups based on thought process and encourage further research in this area.

Advantages of Art with Adolescents

Art-based assessments, such as the FEATS, have been used successfully with a variety of age groups. However, the current study targets using art-based assessment with an adolescent population and may prove to be an advantageous means of collecting information during this typically guarded developmental period.

Like other skills children learn as they get older, artistic expression follows a developmental course (Golomb, 1999; Lowenfeld & Brittain, 1987; Veltman & Browne, 2002).

In addition to the profound physical, sexual, cognitive, and psychological changes that mark the adolescent stage of development, this period is often a stage of great creativity and self-exploration (Cox & Price, 1990; Emunah, 1990). Adolescents are experiencing a major transition in emotional and cognitive development that has been described as a turning point in developing independence with the opportunity for tremendous growth (Craig & Dunn, 2007; Emunah, 1990; Erickson, 1968).

In the midst of this transition, adolescents can often draw on creative means to express themselves using art, music, and prose. Thus, these techniques provide an engaging, non threatening means of accessing information from an often guarded population (Veech & Gladding, 2007). As adolescents are developing new thinking and reasoning capabilities, including abstract thought, they often have a profound need to express and communicate their internal world but have not yet mastered the skills to do this verbally (Emunah, 1990; Erickson, 1968; Piaget, 1952). Creative expression can help process the new dilemmas, moral choices, and paradoxes that this new way of thinking presents (Blos, 1962; Emunah, 1990; Piaget, 1952).

May (1975) postulates that creativity is not a process of illness but rather a product of healthy expression and emotional health, and having adolescents embrace this form of self-expression can encourage appropriate expression of thoughts, emotions and experiences that can carry over into adulthood. To this end, the creative process, including art, provides containment for the mirage of new thoughts and feelings that adolescents are navigating for the first time (Emunah, 1990). By providing an appropriate means of expressing and processing much of what they are thinking and feeling, adolescents are able to expand their thought possibilities to include past, present and future orientations.

Researchers have postulated that traumatic experiences and significant life events are often coded in visual and sensory form and thus are more easily accessed and processed via imagery rather than words (Veltman & Browne, 2001). Children and adolescents are continuing to develop their verbal skills and may have difficulty completely expressing themselves though conversation. As such, some researchers have begun to explore the use of drawings to facilitate narrative discussion in children and adolescents. Lev-Wiesel and Liraz (2007) found that children of substance abusing fathers who were first asked to draw a picture of life in the shadow of their father's addiction and then describe this were more likely to engage in detailed and revealing conversations than those who were only asked to describe this experience. This suggests that the use of drawings in assessment may provide a window to feelings that remain inaccessible or guarded when asked verbally.

In addition to the developmental appropriateness of using art techniques with adolescents, art appears to be a means of engaging adolescents in a therapeutic or assessment process to which they may otherwise be defended. Several authors have noted an increased level of adolescent engagement in art based therapeutic and assessment activities (Cox & Price, 1990; Gross & Hayne, 1998). By allowing adolescents to think through actions and consequences (Emunah, 1990) and tapping into this stage of development in a non-threatening way (e.g., creative expression), it may be possible to screen adolescents for a variety of mental health concerns though a process that comes naturally and prompts healthy expression. Thus, the use of art in the assessment of adolescents continues to be highly appealing despite many of the methodological challenges.

In addition to the theoretical reasons that creative assessment, specifically art, is advantageous with adolescents, drawing assessments can be done in group settings, and can

often be more practical and economical than other individualized approaches to screening adolescents (Veech & Gladding, 2007). Additionally, there appear to be small differences in how males and females approach drawings tasks, and these differences are often found in the content items rather than the formal elements. This suggests that art could be used as an assessment with adolescents of both genders (Goodenough, 1926; Hagood, 2003; Kellogg, 1970; Machover, 1960; Naglieri, 1993; Rubin, Schacter, & Ragins, 1983). The development of art-based assessments and rating systems that can reliably and accurately screen adolescents for a variety of emotional or cognitive concerns is therefore very appealing.

As art-based assessments provide a unique and non-threatening means of assessing adolescents, the current study is aimed at exploring the use of the FEATS as a screening tool for social, behavioral and emotional disturbance in drawings of varying thematic content. The psychometrics of the FEATS was explored with an adolescent sample, drawings of varying thematic content were rated using the FEATS, and these scores were compared to the Behavioral Assessment System for Children -2 (BASC-2; Reynolds & Kamphaus, 2004) to determine the utility of the FEATS in predicting clinical scores. Chapter III outlines the methods used in this study.

CHAPTER III

Method

Art provides an engaging and developmentally appropriate means of screening adolescents for a variety of emotional concerns and general level of functioning. However, research in the use of drawings as an assessment tool has reported ambiguous results and suggests that alternative methods for analyzing drawings may be needed to increase the utility of this form of assessment. Art-based assessments provide a relatively new way of exploring art work in relation to psychological functioning, and the FEATS scale is a standardized means of operationalizing this information. The current study's purpose is to explore the use of the FEATS as a screening tool for emotional disturbance in a high school sample and hypothesizes that: 1) formal elements (purported to measure state emotions) in participants' drawings will remain consistent across two drawings of varying thematic content, and 2) the formal elements found in adolescents' drawings will relate to socio-emotional functioning as reported on the Behavioral Assessment System for Children (BASC-2; Reynolds & Kamphaus, 2004). This chapter details the methods used in this study and is divided into sections focused on describing the larger study from which data is drawn, participants included in this sample, measures, procedures specific to the current study, hypotheses and the analyses of data.

Study of Health Risk Behaviors in Adolescents

The current study is a sub-sample of a larger program of research designed to explore a variety of health risk behaviors in adolescents. To provide a context, the larger week-long study will be briefly described. Working in conjunction with a traditional high school and a therapeutic high school, researchers collected data from various measures focused on behaviors including general mental health, substance use, sexual behavior, eating behavior, dating

relationships and perception of parental conflict. In collaboration with the school's administration and classroom teachers, researchers also developed and provided psycho-educational groups and activities focused on health topics similar to data collected via self-report measures (e.g., substance use, sexual behavior, and interpersonal relationships). The topic of each presentation was an area identified by the school as an area in which students could benefit from further information and education.

On each day of the program, approximately one half of a class time was devoted to data collection. The remainder of the class was dedicated to the health-related presentation and activity. Two days included drawing activities. To prevent contamination of the results, students were asked to complete all study questionnaires relevant to the day's topic prior to the presentation. For example, the group discussion about substance use was conducted after all surveys containing information related to substance use were collected. All questionnaires were read aloud by a researcher to control for differences in reading level. At the end of the research program, feedback sessions were conducted with students regarding the program and data collection.

All procedures in this study were approved by the University of Kansas Human Subjects Committee Lawrence (HSCL) as well as the staff at the site (health educators and art therapists), the school principal, and the research coordinator for the school district (Appendix A). Two weeks prior to data collection, letters describing the project were sent home with students to their parents (Appendix B). Copies of the measures were also provided to the principals and instructors (at each site). A waiver of the requirement for parental permission for participation was granted in accordance with the HSCL policy 45 CFR 46.117 (b) (2) because data were

collected as part of a curriculum-based exercise, participation presented minimal risk to students, and no identifying information was obtained.

Students were recruited for this study from their semester-long health or art therapy class. All students enrolled in these classes were invited to participate in the study. There were no mental or physical characteristics unique to the sample, and no student was excluded based on race, sex, or age. At the beginning of each class period, researchers obtained students' verbal assent (Appendix C). Students were assured that their responses to items would remain confidential and that their participation was voluntary. Students who chose not to participate were provided with an alternative class activity by their teacher. Of the 243 students approached at the traditional high school, one chose to participate in the alternative educational activities. At the therapeutic school, four students indicated that they would prefer not to participate in research.

The two sites selected for the larger study were high schools located on the east coast of the United States. The traditional high school is a large public high school in the southern region of the country which provides mostly college preparatory courses. The second site is a small Level 4 (therapeutic) high school in the Mid-Atlantic region of the country. All students enrolled in this school meet the criteria for at least one DSM-IV-TR diagnosis and have a variety of emotional and behavioral challenges. Students are only selected to attend a Level 4 school if a traditional public education has been unsuccessful.

As part of the larger study, students at the public high school were recruited from their health class, which is required of all students at the school. Of the 243 students enrolled in a health class, 99% (n=242) chose to participate in the larger study. However, due to absences during the program and attrition through passively choosing not to complete the measures, only

80.54% (n=194) of the total sample completed all of the procedures for the larger study. Of those choosing to participate in the research component of the program, 51.51% were male, and the average age of participants was 16.02 years (SD=1.04, range 14-19). Adolescents in this sample identified as European American (56.8%), African American (13.8%), Asian or Pacific Islander (11.2%), Biracial (6.9%), Other (6.5%), and Multiracial (4.8%).

At the therapeutic high school, 51 students, 84% (n=47), recruited from their art therapy class assented to participate in the study. However, due to absences, behavioral problems, and attrition only 60.84% (n=28) completed the entire research protocol. Ninety-six percent of students choosing to participate in the research were male, which accurately reflects the general population of the school. Students responding to demographic questions reported an average age of 16.86 (SD=1.71, range=14-19). Adolescents in this sample identified as African American (42.9%), Multiracial (21.4%), Hispanic/Latino (10. 7%), Asian or Pacific Islander (10.7%), Other (10.7%) and Native American (3.6%).

Participants

Participants in the current study were selected from the larger sample from the two high school sites and were not required to complete the entire study protocol to be included in this sample. The criteria for inclusion in this study were that the participant was between the ages of 13 to 19, completed a psychological measure (BASC-2), and completed the two drawings as part of the psycho-educational program. Participants meeting these criteria were screened into the current study's sample. Table 1 provides a breakdown of participants approached for the larger study, those who provided assent for research, and those who were included in this sample.

Table 1

Recruitment statistics for larger study, completers and final sample

	Traditional school	Therapeutic school	Total sample
Total recruited	243	51	294
Total assent	242 (99%)	47 (92%)	289 (98%)
All measures*	194 (80%)	28 (60%)	222 (77%)
BASC & drawings*	194 (80%)	41 (87%)	235 (81%)
Final sample**	165 (85%)	28 (68%)	193 (82%)

*Total number of participants completing assessments.
** Final sample number excludes participants removed due to invalid BASC-2

The BASC-2 provides several scales to determine the validity of the profile for each participant (see Measures section). Of those meeting the initial entry requirements for this study, BASC-2 scores were analyzed for internal validity, and 43 participants were dropped from the sample due to invalid BASC-2 profiles. The BASC-2 validity scales are described in detail in the measures section of this chapter. Information related to the reduction of this sample due to invalid BASC-2 profiles is outlined in Table 2.

Table 2

Participants removed from study due to invalid BASC-2 profiles

	Removed Traditional school	Removed Therapeutic school	Total Removed
F-index	5	3	8
L-index	4	1	5
V-index	8	3	11
Consistency index	9	11	20
Response pattern	8	2	10
Total removed	30	14	43

Note: Total removed from sample is not equal to the number removed from each category as several participants scored within the caution range on multiple indexes.

After removing participants without complete drawing sets, BASC-2, or invalid BASC-2, the final sample for this study includes 193 participants from both sites. Fifty-eight percent of students choosing to participate in the research were male. Students responding to demographic questions reported an average age of 16.51 (SD=1.19, range=14-19). Adolescents in this sample identified as European American (49.2%), African American (15.7%), Multiracial (9.9%), Hispanic/Latino (6.8%), Asian or Pacific Islander (11%), Other (7.3%). Table 3 includes the breakdown of demographics variables by site.

Table 3

Demographics variables by site

	Traditional school (*n=165)*	Therapeutic school *(n=28)*	Total sample *(n=193)*
Gender	85 male (52%)	27 male (96%)	112 male (58%)
Age	16.01 (SD=1.04)	17 (SD=1.61)	16.15 (SD=1.19)
Ethnicity			
African-American	18 (10%)	12 (44%)	30 (15%)
Asian/Pacific Islander	18 (11%)	3 (11%)	21 (11%)
European-American	94 (57%)	0 (0%)	94 (49%)
Hispanic/Latino	10 (6%)	3 (11%)	13 (7%)
Multiracial	13 (8%)	6 (22%)	19 (10%)
Other	11 (7%)	3 (11%)	14 (7%)

Measures

Demographics questionnaire. A simple paper and pencil demographics questionnaire was used in this study. In addition to other demographics variables, the questionnaire asked students their age, year in school, and ethnicity (Appendix D). This form was read aloud to control for weaker reading levels.

Behavioral Assessment System for Children -2 (BASC-2). The Behavioral Assessment System for Children (BASC-2) is a multi-modal, multi-dimensional assessment that

is used to evaluate behavior and self-perception of children and adolescents ages 2 through 25 (Reynolds & Kamphaus, 2004). It was designed to aid in the differential diagnosis process and, if the entire battery is given, it provides information from multiple sources including: 1) Self-Report of Personality (SRP), 2) Teacher Rating Scale (TRS), 3) Parent Rating Scale (PRS), 4) Structured Developmental History (SDH), and 5) Student Observation System (SOS). These forms can be used independently or in combination to provide an increasingly comprehensive assessment of behavior and psychological functioning.

Since this study's focus was on adolescents' drawings as a reflection of self, and it was not feasible to collect multiple forms, only the SRP was collected. There are several BASC-2 SRP forms designed for specific age groups, and the computer scored paper and pencil version of the SRP (ages 12-21) was used for this study. All BASC-2 protocols were computer-scored using the appropriate computer scoring system.

The SRP is a self-report scale that consists of 176 items and takes 20 to 30 minutes to complete (Reynolds & Kamphaus, 2004). Items on the SRP are categorized by two types of response choices. One is True or False and the other is a five-point Likert scale ranging from Never to Almost Always. While the reading level is appropriate for students in this age range, items were read aloud to control for those with lower reading levels.

When scored, the SRP yields a variety of clinical scales, including 16 primary and 5 composite (Reynolds & Kamphaus, 2004). Scales with T-scores of over 60 fall into the at-risk category, while those over 70 are deemed clinically significant. The exception to this are the adaptive scales for which anything below 40 is considered at-risk and clinical significance is a T-score of 30 or below.

Composite scales were created using a confirmatory factor analysis and a principle-axis analysis (Reynolds & Kamphaus, 2004). Table 4 provides a list of the primary and composite scales and well as their definitions when respondents score in the at-risk or clinically significant range.

Table 4

*BASC-2 scale descriptions: SRP scales ages 12-21 (*Reynolds & Kamphaus, 2004)

Primary Scale	Definitions
Anxiety	Feelings of nervousness, worry and fear; the tendency to be overwhelmed by problems
Attention problems	The tendency to report being easily distracted and unable to concentrate more than momentarily
Attitude to school	Feelings of alienation, hostility and dissatisfaction regarding school
Attitude to teachers	Feelings of resentment and dislike of teachers; beliefs that teachers are unfair, uncaring, or overly demanding
Atypicality	The tendency toward bizarre thoughts or other thoughts and behaviors considered “odd”
Depression	Feelings of unhappiness, sadness, and dejection; a belief that nothing goes right
Hyperactivity	The tendency to report being overly active, rushing through work or activities and acting without thinking
Interpersonal relationships	The perception of having good social relationships and friendships with peers
Locus of control	The belief that rewards and punishments are controlled by external events or people
Relationship with parents	A positive regard towards parents and a feeling of being esteemed by them
Self-esteem	Feelings of self-esteem, self-respect, and self-Acceptance

Table 4 continued

*BASC-2 scale descriptions: SRP scales ages 12-21 (*Reynolds & Kamphaus, 2004)

Primary Scale	Definitions
Self-reliance	Confidence in one's ability to solve problems; a belief in one's personal dependability and decisiveness
Sensation seeking	The tendency to take risks and seek excitement
Sense of inadequacy	Perceptions of being unsuccessful in school, unable to achieve one's goals, and generally inadequate
Social stress	Feelings of stress and tension in personal relationships; a feeling of being excluded from social activities
Somatization	The tendency to be overly sensitive to, to experience, or to complain about relatively minor physical problems and discomforts
Composite scales	*Primary scales used*
Emotional symptoms index	Social stress, anxiety, depression, sense of inadequacy, self-esteem (inverted), self-reliance (inverted)
Inattention/hyperactivity	attention problems, hyperactivity
Internalizing problems	Atypicality, locus of control, social stress, anxiety, depression, sense of inadequacy, somatization
Personal adjustment	Relationship with parents, interpersonal relations, Self-esteem, Self-reliance
School problems	Attitude to school, Attitude to teachers, sensation seeking

In addition to the primary and composite scales, the BASC-2 has 3 indexes which are designed to determine the validity of the respondent's scores. These include indicators of: exaggerated symptoms (F-index), minimized symptoms (L-index), or misunderstanding the assessment (V-index) (Reynolds & Kamphaus, 2004). The measure also contains a Response Pattern Index and a Consistency Index, two other measures of validity.

Caution scores on the Response Pattern Index indicate that a participant selected the same response repeatedly or rotated through item choices in a pattern (1, 2, 3, 1, 2, 3). The Consistency Index pairs like-items to determine if a participant is answering like questions in a similar manner. As indicated above, participants whose F-Index, L-Index, V-Index, Response Pattern Index, or Consistency Index score fell within the "extreme caution" range were not included in this study as their BASC-2 profiles were considered invalid (see Table 2).

Reynolds and Kamphaus (2004) also report acceptable external validity of the BASC-2 SPR used in this study. Several studies were conducted to explore the concurrent validity of the primary and composite scales of the adolescent SPR paired with other well researched behavioral measures and profiles of adolescents with specific DSM-IV-TR diagnosis. These studies suggest that the SPR form of the BASC-2 has acceptable external validity.

The internal reliability of the SRP is also reported to be generally high across primary and composite scales range from .68 (Somatization) to .95 (Emotional Symptoms Index) with most scales achieving above .8. Additionally, test-retest reliability is also reported as acceptable from a sample of 107 adolescent participants.

The BASC-2 was chosen for use in this study because it is a well researched, reliable, and valid self-report measure for adolescents that screens for a number of psychological, relational, and behavioral concerns. It provides a comprehensive assessment of a participant's level of functioning in 20 minutes and the computer scoring and objective nature of the measure limit error. As such, it was deemed appropriate standard upon which to measure the FEATS.

Standardized Drawings. Two standardized drawings were collected from the adolescent participants at both sites. While this study's primary concern was the inclusion of formal elements, focus topics were given to participants that were complementary to the larger project.

The students at both schools were asked to draw: 1) their perfect romantic date, and 2) a place where they might be exposed to alcohol or drugs, and who would be there (Appendix E).

Participants were provided with standardized materials including: 1) an 11 x 18 inch piece of 80lb. white paper, and 2) a 12 pack of Mr. Sketch markers. These materials were selected because they are typically used with the FEATS rating scale (Gantt, 1998), are inexpensive and easily accessible, and non-toxic. Each piece of paper contained a student ID number and a label containing a statement of permission to photograph the image for research and educational purposes. After completing their drawing, participants were asked to check a Yes or No statement giving permission to photograph their drawing. Approximately 93% of participants gave consent for their drawings to be photographed. Students failing to check either box were not assumed to give consent and thus their drawings were not photographed but were included in the study as they had provided assent for use in research. Additionally, upon completion of the drawing, students were prompted to write a brief description of the drawing on the back. The drawings in this study were collected within 24 hours of each other to control for any significant temporal changes.

Formal Elements Art Therapy Scale (FEATS). A review of the literature yielded two primary measures of formal elements in art-based assessment: 1) the Formal Elements Art Therapy Scale (FEATS; Gantt & Tabone 1998) and 2) the Diagnostic Drawing Series (DDS; Cohen, 1988). The FEATS was selected for use in this study because it is reported to be appropriate for use with any two dimensional drawing, while the DDS was designed for use with a specific series of drawings, and the non-task specific formal elements captured by the DDS rating scale are common to the FEATS.

The FEATS includes 14 scales described in Table 5. Gantt and Tabone (1998) report these art-based scales were created from three sources: 1) symptoms from the DSM that could have graphic equivalents, 2) art therapy and psychology literature about spontaneous and directive drawings, and 3) clinical observations (Gantt & Tabone, 1998). All of the scales, with the exception of Scale 9, Developmental Level, are believed to be graphic representations of common symptoms in the DSM-IV-TR. For example, Prominence of Color is reported to be related to affective expression (Gantt & Tabone, 1998). Patients with depressed mood will score low on this item, while those with manic presentation will score high. Appendix F provides the permission letter to use the FEATS in this research and Appendix G includes the FEATS rating scale and examples of FEATS rating guide.

Table 5

Summary of FEATS scales (Gantt, 2001; Gantt & Tabone, 1998)

Scale	Description of scale
Prominence of color	The way in which color is applied to areas of the drawing and objects in the drawing. Is color only used to outline objects or is the whole paper covered in color?
Color fit	The way in which color is used. Is color used conventionally, or in an idiosyncratic manner?
Implied energy	The amount of energy expended to create the drawing.
Space	The amount of space on the paper that contains drawing. This is a true ratio scale focused on the proportion of space use.
Integration	The way in which the objects in the drawing are balanced and related to each other.
Logic	The number of elements or objects in drawing that are bizarre or illogical.

Table 5 continued

Summary of FEATS scales (Gantt, 2001; Gantt & Tabone, 1998)

Scale	Description of scale
Realism	The extent to which the items in the drawing are recognizable and realistically drawn.
Problem-solving	The degree to which the artist shows the person getting the apple out of the tree. It is specific for use with the PPAT drawing.
Developmental level	Lowenfeld's (1978) levels of artistic development which are analogous to chronological developmental levels.
Details of objects environment	The degree to which the individual concretely followed & the directions of the task versus embellishment of the drawing. Do they use extra details and embellishments in the drawing?
Line Quality	The overall quality and degree of control that the artist has used in drawing the lines.
Person	The development of the human figure in the drawing. Does it look like a person? How is the figure proportioned?
Rotation	The extent to which the drawing is rotated.
Perseveration	The repetitive act of making a short line over and over.

Scales are rated on a 5-point Likert scale and higher scores indicate a greater presence of the element being rated. Gantt and Tabone (1998) selected a 5 point Likert scale believing it would provide a more sensitive measure of patient change. However, there is little evidence to suggest that a continuous measure of art's formal elements provides more diagnostic information than one that is dichotomous. Once these scales were formalized, a detailed objective rating rubric was established, and the FEATS was tested for validity and reliability.

Using the Person Picking an Apple from a Tree Assessment (PPAT) which includes standardized directions and materials, researchers tested the FEATS for reliability and validity (Gantt & Tabone, 1998). The FEATS was validated using over 5,000 drawings from both control and clinical samples (Gantt & Tabone, 1998). The sample included predominantly European American adults who were affiliated with a state university hospital. FEATS developers encourage continued research with increased variability in race and ethnicity, age, and socio-economic status.

Collecting PPAT drawings from different diagnostic groups including Major Depressive Disorder, Bipolar Disorder, Schizophrenia, and Cognitive Disorders, items rated on the FEATS were correlated with diagnostic groups. Correlations between scales and diagnostic groups provide support for a relationship between formal art-based items measured by the FEATS and patients in different diagnostic categories. In addition to the initial validation studies, several other studies have explored the use of the FEATS to distinguish diagnostic categories from non-clinical samples, including substance users (Rockwell & Dunham, 2006), epileptics (Anschel, Dolce, Schwartzman, & Fisher, 2005), general mental health concerns (Swan-Foster, Foster, & Dorsey, 2003; White, Wallace, & Huffman 2004), and ADHD (Munley, 2002). All of these analyzed the FEATS item by item rather than exploring an underlying factor structure.

Inter-rater reliability on the FEATS is reported as acceptable. In a reliability study, rating manuals were provided to three groups of raters (art therapists, social work students, and recreation therapy students) and, with limited training, they achieved interclass correlations greater than or equal to .90 between and within groups of raters on all scales except Perseveration and Rotation (Gantt & Tabone, 1998). These finding were replicated in several additional studies (Anschel, Dolce, Schwartzman, & Fisher, 2005; Munley, 2002; Rockwell &

Dunham, 2006; Swan-Foster, Foster, & Dorsey, 2003; White, Wallace, & Huffman 2004) and suggest that the FEATS can be scored reliably by trained raters which increases its utility in research.

Current research on the FEATS suggests that measure development has been limited and there is no report of item analysis or exploration of underlying factors analysis. The FEATS manual does not discuss item analysis or factor analysis of the measure. While Gantt and Tabone (1998) note that there may be composite scales present in the FEATS, this analysis has not been conducted. Additionally, while authors report correlations between clusters of formal elements and diagnostic categories, no cluster analysis or model has been presented to support these assumed relationships. Thus there is much to be done to support the use of this scale for research and clinical purposes.

Procedures

As the general study is described above, the procedures described in this section relate to the collection of the BASC-2 and the two drawings specific to this study. On the second day of research, participants were guided though the assent process and those choosing to participate were asked to complete the BASC-2. Each participant received a pencil and a BASC-2 protocol that included a study ID number. Participants were instructed not to put their names on the protocol to ensure anonymity. A researcher provided directions and then read items aloud to control for reading level and ensure comprehension of the questions. During this process, an additional researcher was available to respond to participant questions and facilitate the general data collection process.

Upon completion of the BASC-2, participants were transitioned to the psycho-educational portion of the project which included their first drawing: their perfect romantic date.

This topic was related to the previous day's measures on relationships. Participants were provided with a sheet of paper that included their study ID and a small label asking for consent to photograph their image for research and educational purposes. Additionally, each participant was provided with a sandwich-sized plastic bag containing twelve Mr. Sketch markers. They were then provided with the instructions for the drawing and asked to use the materials provided to create their drawing. After drawings were completed, participants were encouraged to write a brief description of their perfect date on the back of their drawing and indicate consent to photograph. Participants who did not expressly give consent to photograph were assumed not to give consent, and thus their drawings were not photographed (Table 6). Perfect date drawings were then used to facilitate a discussion related to dating and safe sexual behaviors. All drawings were collected at the end of the class.

On the third day of data collection, participants were asked to provide assent to participate in research, and those willing completed a variety of measures related to their substance use behaviors and beliefs. Participants were next asked to complete the second drawing, which focused thematically on substance use. Those choosing to draw were provided with the same standardized materials (white paper and Mr. Sketch markers) and asked to draw a place where they may be exposed to drugs or alcohol. Upon completion of this drawing, participants were asked to indicate consent to photograph (Table 6) and write a brief description on the back of their drawing. These drawings were then used to facilitate a discussion related to substance use behaviors and beliefs. A sample of drawings from both sites can be found in Appendix H.

Table 6

Consent to photograph drawings

	Traditional School (n=165)	Therapeutic School (n=28)	Total Sample (n=193)
Drawing 1: Date	137 (83%)	19 (67%)	156 (80%)
Drawing 2: Drugs	131 (80%)	19 (67%)	150 (78%)

Hypotheses

This research project had two overall goals. As indicated by Gantt and Tabone (1998), formal elements are reported to be reflections of mood and behavioral symptoms and thus should be consistent across drawings, regardless of content, when controlling for temporal changes. The first goal was, therefore, to establish the stability of formal elements across drawings. The second goal was to explore the use of the FEATS as a screening measure for emotional and behavioral disturbances with adolescents, providing support for the scale's construct validity in an adolescent population. To meet these goals, the four following hypotheses were explored:

I. Training raters in the use of the FEATS scale will result in acceptable levels of inter-rater reliability on all items.

II. The FEATS will contain an underlying factor structure allowing for the creation of composite scales.

III. FEATS composite scales will be stable across a 24 hour period, supporting that they are a measure of functioning independent of the drawings content and theme.

IV. Formal elements, as measured by the FEATS composite scores, will predict participants' *emotional functioning* as measured by the related BASC-2 primary and composite scales, while controlling for age and gender.

V. Formal elements, as measured by the FEATS composite scores, will predict participants' *behavioral functioning* as measured by the related BASC-2 primary and composite scales, while controlling for age and gender.

Analysis

Drawing ratings and inter-rater reliability. Given the repeated concerns raised in the literature related to inter-rater reliability on drawing rating scales, the training and reliability of raters in this study was viewed as essential. The primary two raters for this study were undergraduate psychology students at a large mid-western university who expressed interest in gaining research experience and class credit though helping in graduate student research. They had no art background or special training in psychology. A third rater was a psychology graduate student, and co-investigator on the larger project who had some familiarity with the use of art in therapy but no training in the use of art-based assessment.

Three raters were trained in the use of the FEATS by the Principle Investigator of this study, who has a Master's Degree in Art Therapy and training in the use of art-based assessment. Training included a four hour session in which scales were described and drawings were rated collaboratively as a group to increase understanding of the different dimensions and uniformity in rating. Upon completion of this training, raters were asked to complete independent training on additional drawings.

Twenty training drawings were collected from a graduate level statistics course which all raters were asked to independently score. Upon completion of the initial ten drawings, an additional training booster session was held. Ratings were compared and discussed to ensure understanding of the FEATS. Raters independently rated the remaining ten training drawings and

achieved high inter-rater reliability on all FEATS scales ($k \geq .79$). This was determined to be acceptable training, and raters were then asked to begin rating the studies drawings.

All drawings were rated by these three raters on 13 of the 14 FEATS scales. Scale 8 (Problem-solving) was not included because it is specific to the PPAT which was not used in this study. To ensure that inter-rater reliability was maintained, the Principle Investigator rated a randomly selected 20% (n=77) of drawings from each site in this sample. Any discrepancies were analyzed using a collaborative team approach, the correct rating was determined for each discrepant scale, and the correct coding was entered into SPSS. Qualitatively, disagreements on drawing ratings often appeared to be due to rater fatigue or error rather than true inability to see the same formal element in the drawing.

There appear to be many commonly accepted ways to compute and report inter-rater reliability, including Pearson Correlation Coefficient and Cohen's Kappa Statistic (Cohen, 1960; Hayes & Krippendorff, 2007; Saal, Downey & Lahey 1980; Shrout & Fleiss, 1979; Uebersax, 1987). Because demonstrating acceptable levels of inter-rater reliability was paramount to this study, it was computed using both correlations and Kappa. Inter-rater reliability appeared to be low when rating the 5-point Likert scale FEATS and was recomputed using the dichotomized FEATS items. These analyses are presented in Chapter IV.

Data entry and descriptive statistics. All data were entered into SPSS 16.0 by undergraduate research assistants. Data were double entered, compared and cleaned to reduce any entry error. Demographic information was analyzed and collapsed into groups. Age and ethnicity were collapsed into smaller categories to provide enough power in each group for analysis. Age was recoded into three groups: 1) 14 and 15 year olds, 2) 16 and 17 year olds, and

3) 18 and 19 year olds; and ethnicity was also collapsed into three groups: 1) European-American, 2) African-American/Black, 3) Other.

BASC-2 primary and composite scores, as well as the F, V, and L indexes, were entered as well as FEATS scores on each scale. Before the primary analysis could be conducted, several modifications were made to the data. The BASC-2 scores were recoded into two variables: 1) Non-clinical (below 60), 2) At-risk or Clinically Significant (60 or above). As it was hypothesized that the FEATS would be able to function as a screening tool for at-risk students, given the low number of students scoring above 70 (Clinically Significant), it was determined that these two general groups should be analyzed. Adaptive scales on the BASC-2 were reverse coded appropriately to reflect the difference in scoring.

As limited information related to the measure development and analysis of the FEATS is published, this measure required a more substantial item analysis and measure exploration and modification, described below.

Analysis and modification to the FEATS. As no current item analysis was available on the FEATS scales, this was conducted as the first step in exploring this measure. The first thing revealed in this analysis of the FEATS scales was that three of the items had no variability within this study's population. Perseveration, Rotation, and Line Quality scales were rated almost uniformly across the drawings. These scales were removed from the FEATS for this study as they did not contribute any information. Frequency distribution tables for the FEATS are presented in Chapter IV.

The other nine scales on the FEATS appeared to capture different levels of each construct. However, rather than showing the expected normal distribution across each construct, the FEATS scales emerged as bi-modal. It was apparent in this study that the FEATS items were

largely dichotomous, indicating the presence or absence of an item in the drawing. Thus the remaining 9 FEATS scales were recoded to dichotomous (0, 1) variables.

Due to the dichotomous nature of the FEATS scales, an Exploratory Factor Analysis was determined to be an inappropriate model for this structure. Items were analyzed using a Phi Coefficient to explore potential factors. This analysis yielded three factors. To provide external face validity and assure that the factors suggested by the Phi Coefficients held together theoretically, the 9 FEATS items were provided to an art therapist familiar with the scoring system. The Master's level Art therapist created factors from the individual items and was not provided with any statistical information. She independently produced three factors identical to those found using Phi Coefficients which suggests that the FEATS is comprised of underlying factors which consist of 1) Specific Drawing Factors (scales: Integration, Realism, Developmental Level, Details of Objects, and Person), 2) Logic (scales: Color Fit and Logic), 3) Artistic Energy (scales: Prominence of Color and Implied Energy). Phi Coefficients and frequencies for each of these scales is further described in Chapter IV.

These three factors were analyzed for skew and kurtosis, and the Artistic Energy composite scale was transformed using a log transformation to approach normal distribution. Specific Drawing Factors and Logic composite scales were normally distributed and did not require transformation. All composite scales were analyzed for internal consistency. Specific Drawing Factors produced an acceptable internal consistency of .73. However, due to few items in Logic and Artistic Energy internal consistency was not acceptable. The Spearman-Brown Prophecy Formula was used to determine how many items would be required to have these achieve acceptable alpha levels. Originally used to compute the number of items that need to be added to a scale to achieve acceptable levels of internal consistency, the Spearman-Brown

Prophecy formula was used to gage internal reliability on scales with a low number of items. While additional items would not be added to the FEATS composite scales (which had only two items), the formula served as a measure of internal consistency. It is computed using the following formula: $r_{SB1} = (k^* r_{ij})/[1 + (k-1)^* r_{ij})]$, where r_{SB1} = the Spearman-Brown split-half reliability, r_{ij} = the Pearson correlation between forms i and j, k = total sample size divided by sample size per form (k is usually 2).

Stability of formal elements across thematically unrelated drawings. Gantt and Tabone (1998) suggest that the formal elements captured by the FEATS that are used in assessment and diagnosis should be independent from the thematic content of the drawing. If formal elements are truly independent from content, they should be consistent across multiple drawings, despite different drawing topics, assuming they are collected across a short time frame to prevent significant changes in mood.

The literature supports several methods for computing test-retest reliability which explores temporal relationships, including 1) Pearson's Correlation Coefficient, 2) Paired Sample *t*-test, and 3) Confirmatory Factor Analysis (Beekhuizen, Davis, Kolber & Cheng, 2009; Clarke, Murnen, & Smolak, 2010; Kielhofner, Dobria, Forsyth, & Kramer, 2010; McGraw & Wong, 1996; Reb & Greguras, 2010; Shrout & Fleiss, 1970). For this study Pearson's Correlation Coefficient methods were employed to explore the stability of formal elements across drawings that are thematically unrelated, which is essentially test-retest reliability.

Analysis of the BASC-2 and modified FEATS. Once the demographic variables and the BASC-2 scales were recoded into categorical variables, and the FEATS measure was analyzed to develop composite scales, the primary analysis for this study could be conducted. As the two FEATS composite scales represent multiple dependent variables, it was determined that a

Multiple Analysis of Covariance (MANCOVA) was the most appropriate statistical procedure to test the remaining hypothesis. A MANCOVA is a statistical test that is used to explore between group differences on more than one dependent variable, while controlling for confounding factors (Gall, Borg, & Gall, 1996). This type of multivariate analysis is preferred to an independent sample test, as it reduces Type I error and captures the interrelated nature of the FEATS composite scales (Tabachnick & Fidell, 2007).

Pearson Correlation Coefficients were explored between demographic variables, BASC-2 scales, and FEATS composites to determine relatedness of these variables. These correlations indicated that the MANCOVA needed to control for gender and school, but showed that age and ethnicity/race were not significantly related to the dependent or independent variables in this study. This is consistent with the wealth of literature on gender difference in the production of artwork (Buck, 1948; Burns & Kaufman, 1970; Goodenough, 1926; Harris, 1963; Kaufman, Niu, Sexton & Cole, 2010) and the expected difference in populations between the traditional and therapeutic high schools. Due to the categorical nature of the ethnicity/race variable, an ANOVA was conducted between the FEATS composites, BASC-2 scales and ethnicity/race to explore any interrelatedness that would require control in the final analyses. It was determined that race/ethnicity was not related to any of the BASC-2 scales or FEATS composites, and thus race/ethnicity was not included as a co-variate in the MANCOVA.

A series of 21 MANCOVA tests were conducted using the three FEATS composite scales as dependent variables, gender and school as covariates, and the dichotomous BASC-2 primary and composite scales as independent variables. Omnibus tests were explored using Pillai's Trace. Like Wilk's lambda, Pillai's Trace is a multi-variate test that indicates the proportion of the variance in the outcomes that is not explained by an effect and is used to detect

between group differences in MANOVA and MANCOVA models. To calculate Pillai's trace, divide each eigenvalue by 1 + the characteristic root, and then sum these ratios. Pillai's trace is a more conservative test of between group differences and is frequently used when group sizes are unequal to reduce Type I error (Tabachnick & Fidell. (2007). Pillai's Trace was the appropriate MANCOVA test for this sample as the independent variable, BASC-2 scores, consists of unequally sized groups. In addition to the omnibus MANCOVA, the between-subjects tests were explored for both significant and non-significant differences.

Using the procedures outlined above, the hypotheses related to the development of the FEATS, stability of FEATS scales over time, and utility of FEATS composite scales to screen for social, emotional and behavioral concerns identified by the BASC-2 were tested, and results of these tests are presented in Chapter IV.

CHAPTER IV

Results

As detailed in Chapter III, this study has three general purposes, a) determining the potential for accurate inter-rater reliability rating of the FEATS, b) analysis and reduction of the FEATS including the creation of composite scales, and c) establishing the utility of the FEATS composite scales, on drawings of varying themes, as a screen for social, emotional and behavioral concerns in adolescents to be measured by the BASC-2. The results presented in this chapter are thus divided into two main categories: 1) FEATS development and 2) Analysis of the FEATS by means of the BASC-2.

FEATS Development

Inter-rater reliability. As reported in Chapter III, inter-rater reliability was deemed to be of paramount importance to the success of this study. Before any analysis could be conducted on the FEATS scales or the relationship between the FEATS and the BASC-2, it was important to first determine levels of inter-rater reliability and ascertain the ability of raters to distinguish between common elements of drawings in a reliable manner. Cohen's Kappa and Pearson Correlation Coefficient's were conducted on approximately 20% of randomly selected drawings from each site (n=77; 63 traditional high school). Results are shown in Table 7.

Table 7

Inter-rater reliability: Kappa and Pearson Coefficient for 5-point Likert FEATS

Scale	*K*	*r*
Prominence of color	.76*	.83**
Color fit	.40*	.48**
Implied energy	.27*	.46**
Space	.65*	.76**
Integration	.21*	.43**
Logic	.31*	.35**
Realism	.22*	.35**
Developmental level	.56*	.43**
Details of objects	.32*	.31**
Line quality	.20*	.44**
Person	.33*	.44**
Rotation	.85*	.68**
Perseveration	.51*	.59**

* Cohen's Kappa significant at .01 level
** Pearson Correlation Coefficient significant at .01 level

Kappa results ranged from .20 (Line Quality) to .85 (Rotation). Despite all scales achieving significance at the .01 level, suggesting that agreement was not due to chance, the Kappa levels reported on all scales except Prominence of Color, Space Use, Developmental Level and Rotation were unacceptable levels of inter-rater reliability. Pearson Coefficients produced moderate to high correlations ranging from .31 (Details of Objects) to. 83 (Prominence of Color) all of which are significant at the .01 level and indicate raters produced similar results across scales but were also not high enough to suggest acceptable levels of inter-rater reliability.

The section entitled Exploration of FEATS Scales below details the reduction of the FEATS items from continuous to dichotomous. While further discussed below, this change was made primarily due to lack of variability on the Likert scales which accounts for the poor inter-rater reliability seen in Table 7. Once items were collapsed, inter-rater reliability was re-run and

produced highly acceptable inter-rater reliability on all scales except Line Quality. These results are reported in Table 8.

Table 8

Inter-rater reliability: Kappa and Pearson Coefficient for dichotomous FEATS items

Scale	*K*	*r*
Prominence of color	1.0*	1.0**
Color fit	.55*	.70**
Implied energy	.66*	.71**
Space	1.0*	1.0**
Integration	1.0*	1.0**
Logic	.45*	.63**
Realism	1.0*	1.0**
Developmental level	1.0*	1.0**
Details of objects	1.0*	1.0**
Line quality	-.02	-.03
Person	1.0*	1.0**
Rotation	1.0*	1.0**
Perseveration	1.0*	1.0**

* Cohen's Kappa significant at .01 level
** Pearson Correlation Coefficient Significant at .01 level

Inter-rater reliability was acceptably achieved on all dichotomously coded items except for Line Quality which was removed from further analysis of the FEATS.

Exploration of FEATS scales. The next steps in the analysis of the FEATS included item analysis and collapse of items into composite scales. The distribution of scores on each scale was explored to determine variability and to confirm normal distribution. Table 9 presents the frequency of responses, mean and standard deviation for each of the 13 FEATS scales.

While the scales Prominence of Color and Developmental Level appeared to have close to normal distribution, most scales were heavily concentrated on one or two of the Likert response choices. Additionally, Space Use (*M=3.91, SD=.21)*, Line Quality (*M=4.81, SD=.33),*

Rotation (*M=1.04, SD=.06)*, and Perseveration (*M=1.20, SD=.54)* presented with extremely limited variability and were removed from further analysis. The remaining nine FEATS items were dichotomized to reflect a more normal distribution of responses and more appropriately capture the drawing elements. This collapse of scales is presented in Table 10.

Table 9

Percentages: FEATS response distribution by scale (n=338)

	Likert scale response (%)					
Scale	1	2	3	4	5	*M* (*SD)*
Prominence of color	15.7	14.0	60.5	7.00	2.90	2.75 (0.93)
Color fit	0.60	11.0	10.5	13.4	64.0	54.2 (1.23)
Implied Energy	0.00	6.40	40.1	48.8	4.70	3.43 (0.70)
Space	1.20	1.20	2.30	94.4	0.80	3.99 (0.21)
Integration	4.10	3.50	3.50	30.2	58.7	4.28 (1.04)
Logic	4.10	3.50	4.70	12.8	75.0	4.60 (0.92)
Realism	2.40	2.90	31.2	57.1	6.50	3.59 (0.78)
Developmental level	0.60	10.5	54.1	30.8	4.10	3.23 (0.73)
Details of objects	6.40	5.80	44.8	33.7	9.30	3.26 (0.93)
Line quality	0.60	0.00	1.40	1.40	96.6	4.87(0.33)
Person	21.5	3.50	22.7	42.6	8.70	3.07 (1.26)
Rotation	82.6	17.4	0.00	0.00	0.00	1.04 (0.06)
Perseveration	67.4	11.5	9.10	4.20	7.80	1.20 (0.54)

Table 10

Percentages: Dichotomous distribution of FEATS scales (n=338)

	Dichotomized response (%)		
Scale	0	1	*M* (*SD)*
Prominence of color	15.1	74.9	.65 (.47)
Color fit	35.5	64.5	.63 (.48)
Implied energy	46.5	53.5	.48 (.50)
Integration	41.3	58.7	.53 (.50)
Logic	25.0	75.0	.78 (.41)
Realism	36.5	63.5	.60 (.49)
Developmental level	65.1	34.9	.30 (.46)
Details of objects	57.0	43.0	.40 (.49)
Person	47.7	52.3	.49 (.50)

While the scales were coded 0 or 1, this did not necessarily indicate an absence or presence of each item. Items were recoded based on the assumption that they were normally distributed. Additionally, an attempt was made to code responses that were, in theory, more developmentally appropriate responses for adolescents' developmental level as a 1 and responses that were considered more aberrant were coded as 0. A description of the categorization of each scale is provided.

For Scale 1: Prominence of Color, a rating of 1 or 2 indicated that color was used only to outline objects or fill in 1 object in the drawing. These were coded as a 0. Likert response choices 3-5 indicated that the participant used color to fill in more than 2 objects in the drawing and were coded as 1. The recoded Prominence of Color scale had a mean equal to .65 and a standard deviation of .47.

Scale 2: Color Fit captures the use of idiosyncratic color choices in each drawing. In this case, a score of 0 includes Likert responses 1-4 and indicates the presence of any idiosyncratic color. A rating of 1 reports no atypical color choices are present. After recoding, the Color Fit scale had a mean of .63 and a standard deviation of .48.

For Scale 3: Implied Energy Likert responses 1-3 were collapsed into a 0. Responses indicating a lower level of perceived energy used to create the image were coded as 0 and those with increased energy were coded as 1 (4 or 5). Recoding this scale produced a mean score of .48 with a standard deviation of .50.

Scale 5: Integration indicates how well the objects in the drawing work together to form a composed image, rather than random pieces. Likert scores 1-4 were recoded as 0 and indicate the drawing is not completely integrated and contains at least one object that doesn't fit with the rest of the image. Those drawings receiving a Likert score of 5 were recoded as 1 and reflect the

presence of several well integrated images in the drawing. This dichotomized Integration scale has a mean of .53 and a standard deviation of .50.

Logic, Scale 6, signifies the presence of bizarre or incongruent objects (not colors) in the drawing. A Likert score of 1-4 was coded as a 0 and indicated the presence of at least one illogical object or aspect of the image. Likert scores of 5 reflect that all objects or elements in the drawing appear to make sense. The dichotomized Logic scale has a means of .78 and a standard deviation of .41.

Scale 7: Realism gauges the extent to which items are recognizable and realistically drawn. Likert scores of 1-3 indicate a less realistic rendering of aspects in the drawing and were recoded as 0. Scores of 4 and 5, recoded as 1, signify items are not only recognizable but realistically drawn. This scale obtained a mean of .60 and standard deviation of .49 when dichotomized.

Developmental Level, Scale 9, assesses the elements of the drawing according to Lowenfeld's (1978) levels of artistic development which are analogous to chronological age. Likert scores of 1-3 signify a lower level of artistic development and were recoded as 0. Scores of 4 and 5 indicated a higher level of development more congruent with chronological age and were scored as 1. When recoded, the Developmental Level scale yielded mean of .30 and standard deviation of .46.

Scale 10: Details of Objects in the Environment reflects the number of details the participant included in their drawings. Participants whose drawings contained fewer details were recoded as 0 and included Likert responses 1-3. A drawing that included more details received a 4 or 5 and was recoded to 1. The dichotomous Details of Objects scale had mean of .40 and standard deviation of .49.

Finally, Scale 12: Person gauges how well the person in the picture is drawn. People in the drawings who were unrecognizable or poorly drawn were scored 1-3 and recoded to 0. Those images that included more correctly proportioned people with correct details received a score of 4 or 5 which was recoded to 1. The Person scale produced a mean of .49 and a standard deviation of .50.

Reduction of FEATS to composite scales. After creating dichotomous scales for nine of the FEATS scales, underlying relationships between these scales were explored to aid in the development of composite scales. Phi Coefficients were computed for each pair of scales and examined for the strongest correlation which led to the creation of three composite scales: Specific Formal Elements, Logic, and Energy. As Phi is not typically used in data reduction, but needed in this case due to binary data. Exploratory Factor Analysis would have been another option in exploring the underlying factor structure, but the statistical software needed to conduct this analysis on dichotomous variables was unavailable. As such, Phi was determined to be an appropriate solution provided the statistical outcome made theoretical sense.

To this end, an art therapist with no knowledge of the statistical analysis was asked to create composite scales from the 9 FEATS scales based on theoretical relationships. She produced the same three composite scales adding to their face validity. Phi Coefficients and *p*-values are presented in Tables 11, 12 and 13.

The first composite scale, Specific Formal Elements, includes 5 items: Integration, Realism, Developmental Level, Details of Objects, and Person. These items gage the quality of specific formal elements a participant included in their images. This scale was slightly positively skewed but had an unacceptable level of kurtosis (*M=2.32, SD=1.61, Skewness=.03, Kurtosis=-1.18).* A Log10 transformation (X=Log10 *(6-specific formal elements)) was conducted on this

scale to create a less kurtotic distribution (*M=1.86, SD=.44, Skewness= -.37, Kurtosis= -.88).*

This scale produced an acceptable level of internal consistency (*a*=.73).

Table 11

Phi Coeffiecients and reliability for FEATS composite: Specific Formal Elements

All Drawings (n=338) (α = .73)	Realism	Dev. level	Details	Person
Integration	.32*	.32*	.40*	.17*
Realism		.32*	.43*	.44*
Developmental level			.28*	.15
Details				.18*
Drawing 1 (n=172) (α = .76)	**Realism**	**Dev. level**	**Details**	**Person**
Integration	.37*	.30*	.32*	.22*
Realism		.33*	.45*	.49*
Developmental level			.25*	.17
Details				.27*
Drawing 2 (n=166) (α = .70)	**Realism**	**Dev. level**	**Details**	**Person**
Integration	.30*	.32*	.42*	.16*
Realism		.30*	.41*	.38*
Developmental level			.28*	.15
Details				.16*

* Significant at the .01 level
** Cronbach's Alpha computed using the Spearman-Brown Prophecy Formula

Table 12

Phi Coeffiecients and reliability for FEATS composite: Logic

All Drawings (n=338) (α = .37, if 4 items added α = .70**)	
	Color fit
Logic	.27*
Drawing 1: (n=172) Logic (α = .36, if 4 items added α = .68**)	
	Color fit
Logic	.25*
Drawing 2 (n=166) (α = .40, if 4 items added α = .71*)	
	Color fit
Logic	.29*

* Significant at the .01 level
** Cronbach's Alpha computed using the Spearman-Brown Prophecy Formula

Table 13

Phi Coeffiecients and reliability for FEATS composit: Energy

All Drawings (n=338) (α = .53, if 4 items added α = .82**)	
	Prominence of color
Implied energy	.47*
Drawing 1 (n=172) (α = .43, if 4 items added α = .71**)	
	Prominence of color
Implied energy	.32*
Drawing 2 (n=166) (α = .56, if 4 items added α = .81**)	
	Prominence of color
Implied energy	.46*

* Significant at the .01 level
** Cronbach's Alpha computed using the Spearman-Brown Prophecy Formula

The second composite scale, Logic, includes 2 FEATS scales: Color Fit and Logic. These two items indicate idiosyncratic color and bizarre objects being included in the drawing. This scale appeared close to a normal distribution and produced acceptable levels (below 1) of skewness and kurtosis (*M=1.4, SD=.72, Skewness= -.80, Kurtosis= -.65)*. Due to the low number of items included in the Logic scale, the initial level of internal consistency was not acceptable (*a*=.37). The Spearman-Brown Prophecy Formula determined that 4 items would need to be added to obtain acceptable internal consistency (*a*=.70).

The final composite scale produced by the reduction of the FEATS was Energy. This composite, that includes Prominence of Color and Implied Energy, appears to be related to the amount of perceived energy used to create the image. This scale also appeared close to being normally distributed and produced acceptable levels (below 1) skewness and kurtosis (*M=.55,*

SD=.60, Skewness=.59, Kurtosis= -.58). Similar to the Logic composite scale, the two items present in this scale did not reach an acceptable level of internal consistency (*a*=.53). Again, the Spearman-Brown Prophecy Formula was used and determined that 2 items would need to be added to obtain acceptable internal consistency (*a*=.82).

Pearson Correlation Coefficients were computed among each of the composite scales. The Specific Formal Elements composite scale yielded low to moderate correlations with both the Logic composite (*r=-.30*) and the Energy composite (*r=-.13*). Additionally, Logic and Energy also produced a low correlation (*r=.20*) suggesting that the composite scales are measuring different aspects of the drawing.

Stability of formal elements composite scales. After FEATS scales were reduced to three composite scales, test-retest reliability was conducted to determine the stability of these scales on subsequent drawings across a discrete period of time (24 hours). Test-retest reliability was determined to be the most appropriate statistic to test stability because Drawing One and Drawing Two are hypothesized to be versions of the same test. In other words, the formal elements in each drawing, captured by the FEATS at time one and time two, are what were tested across time.

As such, composite scale scores were computed on each drawing and Pearson Correlation Coefficients were conducted between the Specific Formal Elements Composite Scale for each of the two drawings. Results of this indicate that the composite scales of Specific Elements (r=.30, p>.01*)* is significantly correlated across drawings. However, the composite scale Artistic Energy had a weak and non-significant correlation across the two drawings (r=.06, p<.01*)*. Results are presented in Table 14.

Table 14

Test-retest reliability: Drawings 1 and 2 on FEATS composite scales (n= 172)

	Specific elements (D1)	Logic (D1)	Energy (D1)
Specific elements (D2)	.30**	.09	.38**
Logic (D2)	.12	.28**	.13
Energy (D2)	.05	.02	.06

** Correlation significant at the .01 level

The results of the internal consistency tests and the test-retest reliability suggest that the two composite scales Logic and Energy have low reliability and need further development and exploration before routinely being used in research. However, as this study's hypotheses are focused on the utility of all composite scales, they are included in the final analyses in this study. That said, results of the MANCOVA should be interpreted cautiously.

Primary Analysis: FEATS and BASC-2

Descriptive statistics and collapse of demographic variables. Demographic variables of gender, age, and race/ethnicity were explored through descriptive statistics and are provided in Table 15. The total sample was approximately half male (58%). However, the therapeutic high school was overwhelmingly male (96.4%). Both genders were included in the MANCOVA analyses and gender was included as a co-variate to control for any between group differences. Most students participating in the study were about 16 years of age (*M=16.5, SD=1.1),* and approximately half the total sample identified as European-American (49%). However, the therapeutic high school's participants predominantly identified as African-America (44%), and none of the participants identified as European-American.

Table 15

Frequencies and percentages: Demographic variables (n=193)

	Traditional (n=165)	Therapeutic (n=28)	Total sample (n=193)
Gender	85 male (51.8%) 80 female (48.2%)	27 male (96.4%) 1 female (3.6%)	112 male (58%) 81 female (42%)
Age			
14 years	3 (1.8%)	2 (7.1%)	5 (2.6%)
15 years	57 (34.8%)	4 (14.3%)	61 (31.8%)
16 years	59 (36%)	4 (14.3%)	63 (32.8%)
17 years	29 (17.6%)	7 (25%)	36 (18.8%)
18 years	13 (7.9%)	4 (14.3%)	17 (8.9%)
19 years	1 (1.8%)	7 (25%)	10 (5.2%)
Race/Ethncity			
African American	18 (11%)	12 (44.4%)	30 (15%)
Asian American	18 (11%)	-	21 (11%)
European American	94 (57.3%)	-	94 (49.2%)
Hispanic/Latino	10 (6.1%)	3 (11.1%)	13 (6.7%)
Multiracial	13 (7.9%)	6 (22.2%)	19 (9.9%)
Other	11 (6.7%)	3 (11.1%)	14 (7.3%)

As there were large number of age and race/ethnicity responses that were not evenly distrabuted and several contained a relatively small number of respondents, these two variables were recoded into groups. Age was collapsed into three groups: 1) 14 and 15 year olds, 2) 16 and 17 year olds, and 3) 18 and 19 year olds. Race/Ethnicity was recoded into 3 groups as well and included: 1) European-American, 2) African-American, and 3) Other. The descriptive statistics for these recoded demographics variables are presented in Table 16.

Table 16

Descriptive statistics: Recoded demographics variables (n=193)

	Traditional (n=165)	Therapeutic (n=28)	Total sample (n=193)
Age			
14 and 15 years	60 (36.6%)	6 (21.4%)	66 (34.4%)
16 and 17 years	88 (53.7%)	11(39.3%)	99 (51.3%)
18 and 19 years	16 (9.8%)	11 (39.3%)	27 (14%)
Race/Ethncity			
African American	18 (11.9%)	12 (57.1%)	30 (17.4%)
European American	94 (62.3%)	-	94 (54.7%)
Other	39 (25.8%)	9 (42.9%)	48 (27.9%)

BASC-2 data reduction and descriptive statistics. The BASC-2 was also explored for variability and distribution. This analysis demonstrated that exploring primary and composite scales by three levels (Clinically Significant, At-risk, Non-Clinical) produced significantly unequal groups. As such, these scales were recoded into two groups: Clinical/At-Risk and Non-Clinical. Numbers of participants in these groups is presented in Table 17.

Table 17

Frequencies: BASC-2 clinically significant, at-risk, and non-clinical original and recoded distributions (n=193)

BASC-2 scales	Traditional (n=165)			Therapeutic (n=28)			Total sample (n=193)		
	sig.	risk	non	sig.	risk	non	sig.	risk	non
Anxiety	4	25	136	-	1	27	4	26	163
Anx recode	-	29	136	-	1	27	-	30	163
Attention problems	11	39	115	-	2	26	11	41	141
Attention recode	-	50	115	-	2	26	-	52	141
Attitude to school	10	30	125	2	7	19	12	37	144
Att to sch recode	-	40	125	-	9	19	-	49	144
Attitude to teachers	12	24	129	1	7	20	13	31	149
Att to teach recode	-	36	129	-	8	20	-	44	149
Atypicality	5	15	145	-	3	25	5	18	170
Atypicality recoded	-	20	145	-	2	25	-	23	170
Depression	15	12	138	2	5	21	17	17	159
Depression recode	-	27	138	-	7	21	-	34	159
Hyperactivity	9	18	138	1	3	25	10	20	163
Hyper recode	-	27	138	-	4	25	-	30	163
Interpersonal*	-	15	150	-	1	27	-	16	177
Locus of control	11	27	127	4	4	20	15	31	147
Locus recode	-	38	127	-	8	20	-	46	147
Rel with parents*	-	18	127	-	-	28	-	18	175
Self-esteem*	-	9	156	-	7	21	-	16	177
Self-reliance*	-	27	138	-	3	25	-	30	163
Sensation seeking	6	17	142	1	7	20	7	24	162
Sensation recode	-	23	142	-	8	20	-	31	162
Sense of inadequacy	13	26	126	1	3	24	14	29	150
Sense of inad recode	-	39	126	-	4	24	-	43	150
Social stress	10	12	143	1	2	25	11	14	168
Social recode	-	22	143	-	3	25	-	25	168

Table 17 continued

Frequencies: BASC-2 clinically significant, at-risk, and non-clinical original and recoded distributions (n=193)

BASC-2 scales	Traditional (n=165)			Therapeutic (n=28)			Total sample (n=193)		
	sig.	risk	non	sig.	risk	non	sig.	risk	non
Somatization	7	14	144	-	3	25	7	17	169
Somat recode	-	21	144	-	3	25	-	24	169
Composite scales									
Emotional symptoms	8	21	136	1	3	24	9	24	160
Emotional recode	-	29	136	-	4	24	-	33	160
Inattention/hyperactivity	14	28	123	-	2	26	14	30	149
Inatt/hyper recode	-	42	123	-	2	26	-	44	149
Internalizing problems	9	25	131	2	2	24	11	27	155
Interng problems recode	-	34	131	-	4	24	-	38	155
Personal adjustment*	-	16	149	-	1	27	-	17	176
School problems	7	34	124	2	5	21	9	39	145
School problems recode	-	41	124	-	7	21	-	48	145

*Items were not recoded as they were already dichotomous due to no respondents in the clinically significant group.

Interpersonal Skills, Relationship with Parents, Self-Esteem, Self-Reliance, and Personal Adjustment did not have any participants score within the Clinically Significant range and thus these variables were not recoded. Once recoded, the At-risk group continued to be much smaller per variable than the non-clinical group with number of at-risk respondents ranging from 4 (Anxiety) to 17 (Depression).

The BASC-2 primary and composite scales were also explored using descriptive statistics and revealed that more than half of students participating in this study scored in one or more at-risk groups (80.8%). The traditional high school had fewer at-risk scores (78.8%) than the therapeutic high school, which as expected found most (92.9%) students included in some at-risk

category. A detailed breakdown of participants scoring in the at-risk/clinical group is provided in Table 18.

Table 18

Percentages: BASC-2 at-risk/clinically significant respondents

	Traditional (n=165)	Therapeutic (n=28)	Total sample (n=193)
Any at-risk/clinical	130 (78.8%)	26(92.9%)	156 (80.8%)
Anxiety	29 (17.6%)	1 (3.6%)	30 (15.5%)
Attention problems	50 (30.3%)	2 (7.1%)	52 (26.9%)
Attitude to school	40 (24.2%)	9 (32.1%)	49 (25.4%)
Attitude to teachers	36 (21.8%)	8 (28.6%)	44 (22.8%)
Atypicality	20 (12.1%)	3 (10.7%)	23 (11.9%)
Depression	27 (16.4%)	7 (25%)	34 (17.6%)
Hyperactivity	27 (16.4%)	3 (10.7%)	30 (15.5%)
Interpersonal relationship	15 (9.1%)	1 (3.6%)	16 (8.3%)
Locus of control	38 (23%)	8 (28%)	46 (23.8%)
Relationship with parents	18 (10.9%)	-	18 (9.3%)
Self-esteem	9 (5.5%)	7 (25%)	16 (8.3%)
Self-reliance	27 (16.4%)	3 (10.7%)	30 (15.5%)
Sensation seeking	23 (13.9%)	8 (28.6%)	31 (16.1%)
Sense of inadequacy	39 (23.6%)	4 (14.3%)	43 (22.3%)
Social stress	22 (13.3%)	3 (10.7%)	25 (13%)
Somatization	21 (12.7%)	3 (10.7%)	24 (12.4%)
Composite scales			
Emotional symptoms	29 (17.6%)	4 (14.3%)	33 (17.1%)
Inattention/hyperactivity	42 (25.5%)	2 (7.1%)	44 (22.8%)
Internalizing problems	34 (20.6%)	4 (14.3%)	38 (19.7%)
Personal adjustment	16 (9.7%)	1 (3.6%)	17 (8.8%)
School problems	41 (24.8%)	7 (25%)	48 (24.9%)

Attention problems were the most commonly self-reported concern among the entire sample (26.9%) followed by school problems. Most of those students reporting attention problems were at the traditional high school (30.3%) rather than the therapeutic high school (7.1%). At both schools, roughly a quarter of the students reported school problems. Thirty-two

percent of the therapeutic school sample reported problems with attitudes towards their teachers, making this the most commonly reported at-risk scale for this school.

The least common at-risk scales for the total sample included self-esteem (8.3%) and interpersonal relationships (8.3%). The responses of those participants at the traditional high school echoed these total percentages (Self-Esteem =5.59%, Interpersonal Relationships= 9.7%). While the therapeutic high school yielded a similarly low percent reporting interpersonal concerns (3.6%), the report of self-esteem concerns was quite different than reported for the whole sample and the traditional high school (25%).

Descriptive statistics and bi-variate correlations of demographics, FEATS composites, and BASC-2 primary and content scales. Once all data were analyzed descriptively and collapsed into appropriate categories, descriptive statistics were computed for BASC-2 primary and composite scales and FEATS composite scale Specific Formal Elements on each of the two drawings. Table 19 provides descriptive statistics for the scales that were used.

Table 19

Descriptive statistics: BASC-2 primary and composite scales and FEATS composite

Scale (n=193)	*M*	*SD*	Range
Primary scales: BASC-2			
Anxiety	.16	.36	0-1
Attention problems	.27	.45	0-1
Attitude to school	.25	.44	0-1
Attitude to teachers	.23	.42	0-1
Atypicality	.12	.33	0-1
Depression	.18	.38	0-1
Hyperactivity	.16	.36	0-1
Interpersonal relationship	.10	.28	0-1
Locus of control	.24	.43	0-1
Relationship with parents	.09	.30	0-1
Self-esteem	.08	.28	0-1
Self-reliance	.16	.36	0-1
Sensation seeking	.16	.37	0-1
Sense of inadequacy	.22	.42	0-1
Social stress	.13	.34	0-1
Somatization	.12	.33	0-1
Composite scales: BASC-2			
Emotional symptoms	.17	.38	0-1
Inattention/hyperactivity	.23	.42	0-1
Internalizing problems	.20	.40	0-1
Personal adjustment	.09	.30	0-1
School problems	.25	.43	0-1
FEATS composites by drawing			
Specific formal elements (D1)	1.93	.46	1-2
Specific formal elements (D2)	1.93	.42	1-2

Pearson Correlation Coefficients were conducted across variables to determine independence of variables that were used in the MANOVA (Table 20). All FEATS composite scales were significantly correlated with school and gender. However, while the correlations were statistically significant, the correlation coefficients produced were relatively low ($r \leq .27$). As a statistically significant correlation exists between the dependant variable in this study

(FEATS composite scale: Specific Formal Elements) and demographic variables of school and gender, these will be controlled for in the primary analysis.

Table 20

Bi-variate correlations and eta squared: demographics and FEATS composites

FEATS composite	School (r)	Gender (r)	Age (r)	Ethnicity/race (η^2)
Specific formal elements	-.27*	-.25*	.05	.19
Logic	.17*	.19*	.01	.16
Energy	.21*	.19*	-.04	.17

*Significant at the .05 level

**Coding: School (0=traditional high school); Gender (0=male)

Pearson correlations were also analyzed between demographic variables and BASC-2 scales. The strongest correlation that was produced between the demographics and BASC-2 scales was between site and self esteem ($r=.25$, $p<.01$). All other combinations produced Pearson Coefficients of less than or equal to .25 suggesting independence in variables. Details of this analysis are provided in Table 21.

Table 21

Pearson Coefficients and eta squared: Demographics and BASC-2 scales (n=193)

Scales	School (*r*)	Gender (*r*)	Age (*r*)	Race (η^2)
Primary scales				
Anxiety	.13	.15*	.06	.08
Attention problems	.18*	-.09	-.10	.07
Attitude to school	-.06	-.13	.05	.09
Attitude to teachers	-.05	-.16*	.09	.18
Atypicality	.02	-.08	-.03	.15
Depression	-.08	-.03	.16*	.15
Hyperactivity	.06	.07	-.09	.12
Interpersonal relationship	.07	.04	-.05	.19
Locus of control	-.04	-.03	.01	.11
Relationship with parents	.13	.19*	-.11	.09
Self-esteem	-.25*	-.18*	.09	.16
Self-reliance	.05	-.04	-.04	.06
Sensation seeking	-.14	-.15	.01	.11
Sense of inadequacy	.08	.05	.08	.11
Social stress	.02	-.14	.04	.11
Somatization	.02	.03	-.02	.15
Composite scales				
Emotional symptoms	.03	-.08	.09	.16
Inattention/hyperactivity	.15*	.01	-.08	.06
Internalizing problems	.05	-.03	.09	.15
Personal adjustment	.07	-.11	-.07	.15
School problems	-.01	-.12	.08	.10

*Pearson Coefficient significant at the .05 level
** Eta significant at the .05 level

Finally, Pearson Correlations were computed between BASC-2 primary and composite scales and the FEATS composite scales to determine the necessary independence to conduct a MANCOVA between these variables. Only one combination yielded a statistically significant correlation: Anxiety and Specific Formal Elements (*r=.20, p=.02)*. Despite the statistical significance of this correlation, the coefficients are low enough to consider the variables minimally interdependent. These results are detailed in Table 22.

Table 22

Bi-variate correlations: BASC-2 scales and FEATS composite scales

BASC-2 scales	Specific elements	Logic	Energy
Primary scales			
Anxiety	.20*	.10	-.01
Attention problems	-.11	-.04	-.01
Attitude to school	-.09	-.01	.02
Attitude to teachers	-.14	-.03	-.01
Atypicality	-.02	-.08	.03
Depression	-.11	-.05	-.05
Hyperactivity	.08	-.01	.10
Interpersonal relationship	.12	-.01	.06
Locus of control	-.13	-.02	-.04
Relationship with parents	.08	-.02	.14
Self-esteem	-.04	-.23*	-.13
Self-reliance	-.03	-.04	-.04
Sensation seeking	-.04	-.07	-.06
Sense of inadequacy	.05	-.03	.06
Social stress	-.06	-.05	.01
Somatization	-.06	-.09	.06
Composite scales			
Emotional symptoms	-.02	-.07	-.08
Inattention/hyperactivity	.08	.06	.05
Internalizing problems	-.05	-.06	.01
Personal adjustment	.04	-.10	.02
School problems	-.04	.02	-.02

*Significant at the .05 level

Pearson Correlations were computed between BASC-2 primary and composite scales, and the two scales with the highest correlations were Depression and Interpersonal Relationships (*r=.63)*, and the Emotional Symptoms Index and Internalizing Problems Index (*r=.73)*. The constructs with the highest correlations have much in common, and thus a possible relationship between the two is not unexpected. The remaining scales produced correlation coefficients of less than or equal to .50 which suggests that scales are measuring independent constructs. The full correlation matrix is provided in Appendix I.

Multivariate analysis of covariance. A series of twenty-one 2x2 between-subjects multivariate analyses of covariance (MANCOVA) were conducted to determine the relationships among the 16 primary BASC-2 scales and 5 BASC-2 composite scales and the three dependant variables: FEATS composite scales (Specific Formal Elements, Logic and Energy) computed for each of the two drawings. Due to the significant correlations, school and gender differences were considered covariates and controlled for in each model. Significant results are presented here and non-significant MANCOVA results can be found in Appendix I.

Drawing 1. 2x2 between-subjects MANCOVAs were performed on the dependant variables computed for Drawing 1: FEATS composite scales of Specific Formal Elements, Logic and Energy. Adjustment was made for two covariates: school and gender. Independent variables were scores on the BASC-2 scale of Self-Esteem and Locus of Control (Clinical/At-risk and Non-Clinical).

SPSS MANCOVA was used for the analyses with an eventual sample of 172. This was reduced from the total sample of 193 due to the deletion of those participants who did not complete Drawing 1. There were no univariate or multivariate within-cell outliers at a=.01. Results of the evaluation of assumptions of normality, linearity, and multicollinearity were satisfactory. A non-significant Box's M on both BASC-2 scales ($p>.16$), indicated a lack of evidence that the homogeneity of variance-covariance matrix assumption was violated and the MANCOVA was considered to be the appropriate statistical test.

With the use of Pillai's Trace, due to substantial differences in group sizes on the independent variable, the combined dependant variables were significantly related to scores on the BASC-2 scale of self-esteem. Significant differences were found among Clinical/At-Risk levels of Self-Esteem and non-clinical levels while controlling for two co-variates (school and

gender), Pillai's Trace=.05, F (3, 171) =2.63, p=.05. However, there were no significant omnibus results for the Locus of Control BASC-2 scale. Descriptive statistics for this test are presented in Table 23 and results from this MANCOVA are presented in Table 24.

Table 23

Descriptive statistics: Drawing 1 FEATS composite scores and BASC-2 (n=172)

Variables	*M*	*SD*	range
Self-Esteem	.08	.28	0-1
Locus of Control	.24	.43	0-1
Specific Formal Elements Composite	.48	.25	0-1
Logic Composite	1.40	.73	0-2
Energy Composite	.64	.63	0-2

Table 24

Multivariate effects: Drawing 1and BASC-2 scales: Locus of control (n=172)

Variables	Pillai's trace	*F*	*df*	Error *df*	P-value
Self-Esteem	.05	2.63	3	166	.05*
Locus of control	.03	1.54	3	166	.21

*Significant at the .05 level

Univariate analyses of covariance (ANCOVAs) for each dependent variable were conducted as follow-up tests to the MANCOVA. Using the LSD method for controlling Type I error rates for two comparisons, each ANOVA was tested at the .025 level. These follow-up tests found significant differences between FEATS composite scales and the BASC-2 scales of Self-Esteem and Locus of Control.

The ANOCOVA conducted indicated the Energy FEATS composite scale was a significant predictor of Self-Esteem scores on the BASC-2, F (1, 171) = 5.92, p=*.01*. While resulting in non-significant results for the Specific Formal Elements FEATS composite scale, F (1, 171) = .02, p=.88, and the Logic FEATS composite scale, F (1, 171) = .65, p = .42. This indicates the Energy composite scale was the only indicator of Self-Esteem on Drawing 1.

While not producing significant main effects, the BASC-2 scale of Locus of Control yielded significant between-subjects effects. Using the LSD method for controlling Type I error rates for two comparisons, each ANOVA was tested at the .025 level. The ANOCOVA showed the Specific Formal Elements FEATS composite scale was a significant predictor of Locus of Control scores on the BASC-2, F(1, 171) = 4.44, p=*.01*. The Logic FEATS composite scale, F (1, 171) = .35, p=.55, and the Energy FEATS composite scale, F (1, 171) = .10, p = .76, were not significant predictors of Locus of Control on Drawing 1. Results for follow-up tests for Drawing 1 are presented in Table 25.

Table 25

Between-subject effects: Drawing 1and BASC-2 scales: Locus of control(n=172)

Variables	Sum squares	*df*	Mean square	*F*	P-value
Self-Esteem*Specific	.01	1	.01	.02	.88
Self-Esteem*Logic	.31	1	.31	.65	.42
Self-Esteem*Energy	2.19	1	2.19	5.92	.01*
Locus of Cont*Specific	.24	1	.24	4.44	.03*
Locus of Cont*Logic	.17	1	.17	.35	.55
Locus of Cont*Energy	.04	1	.04	.10	.76

*Significant at the .05 level

Drawing 2. 2x2 between-subjects MANCOVAs were performed on three dependant variables computed for Drawing 1: FEATS composite scales of Specific Formal Elements, Logic, and Energy. Adjustment was made for two covariates: school and gender. Independent

variables were scores on the BASC-2 scale of Self-Reliance, Attitude towards Parents, and Attitude towards Teachers (Clinical/At-risk and Non-Clinical).

SPSS MANCOVA was used for the analyses with an eventual sample of 166. This was reduced from the total sample of 193 due to the deletion of those participants who did not complete Drawing 2. There were no univariate or multivariate within-cell outliers at a=.01. Results of the evaluation of assumptions of normality, linearity, and multicolinearity were satisfactory. A non-significant Box's M (p>.43) on the three independent variables, indicated a lack of evidence that the homogeneity of variance-covariance matrix assumption was violated and the MANCOVA was considered to be the appropriate statistical test.

With the use of Pillai's Trace, due to substantial differences in group sizes on the independent variable, the combined dependent variables were significantly related to scores on the BASC-2 scale of Self-Reliance. Significant differences were found among Clinical/At-Risk levels of Self-Reliance and non-clinical levels while controlling for two co-variates (school and gender), Pillai's Trace=.06, F (3, 165) =3.16, p=.03. Descriptive statistics for this test are presented in Table 26 and the results from this MANCOVA are presented in Table 27.

Table 26

Descriptive statistics: Drawing 2 FEATS composite scores and BASC-2 (n=166)

Variables	*M*	*SD*	range
Self-Reliance	.15	.36	0-1
Relationship with Parents	.10	.31	0-1
Attitude to Teachers	.26	.44	0-1
Specific Formal Elements Composite	.54	.22	0-1
Logic Composite	1.43	.71	0-2
Energy Composite	.46	.56	0-2

Table 27

Multivariate effects: Drawing 2 and BASC-2 scales: Self-reliance, relationship w parents, and attitude to teachers (n=166)

Variables	Pillai's trace	*F*	*df*	Error *df*	P-value
Self-reliance	.06	3.16	3	160	.03*
Relationship w/parents	.05	2.54	3	160	.06
Attitude to teachers	.03	1.60	1	160	.19

*Significant at the .05 level

Using an LSD test to control for Type I error, ANCOVAs for each dependent variable were conducted as follow-up tests to the MANCOVA and tested at the .025 level. These follow-up tests found significant differences between FEATS composite scales for Drawing 2 and the BASC-2 scales of Self-Reliance, Attitude towards Parents, and Attitude towards Teachers.

The Energy FEATS composite scale was a significant predictor of Self-Reliance scores on the BASC-2, F (1, 165) = 7.57, p=.01. While the Specific Formal Elements FEATS composite scale, F (1, 165) = 2.31, p=.13, and the Logic FEATS composite scale, F (1, 165) = 1.51, p = .22, were not significant predictors of Self-Reliance.

While not producing significant omnibus tests, the BASC-2 scales of Attitude towards Parents and Attitude towards Teachers produced significant between-subjects effects. Using the LSD method for controlling Type I error rates for two comparisons, each ANCOVA was tested at the .025 level. The Specific Formal Elements FEATS composite scale was a significant predictor of Attitude towards Parents on the BASC-2, F (1, 165) = 6.70, p=*.01*. While the Logic FEATS composite scale, F (1, 165) = .60, p=.44, and the Energy FEATS composite scale, F (1, 165) = .06, p = .82, were not significant predictors of student's attitudes towards parents. Additionally, the Logic composite scale was a predictor of Attitude toward Teachers, F (1, 165) = 4.30, p=.02, but Specific Formal Elements, F (1, 165) = 0.0, p=.99, and Energy, F (1, 165) = .66, p=.42, were not significant predictors of attitude towards teachers. Results for follow-up tests for Drawing 2 are presented in Table 28.

Table 28

Between-subject effects: Drawing 2and BASC-2 scales: Self-reliance, relationship w parents, and attitude to teachers (n=166)

Variables	Sum squares	*df*	Mean square	*F*	P-value
Self-Reliance*Specific	.10	1	.10	2.31	.13
Self-Reliance*Logic	.73	1	.73	1.51	.22
Self-Reliance*Energy	2.15	1	2.15	7.57	.01*
Rel Parent*Specific	.28	1	.28	6.70	.01*
Rel Parent *Logic	.29	1	.29	.60	.44
Rel Parent *Energy	.02	1	.02	.06	.82
Att to Teacher*Specific	9.28	1	9.28	0.00	.99
Att to Teacher*Logic	2.05	1	2.05	4.30	.04*
Att to Teacher*Energy	.20	1	.20	.66	.42

*Significant at the .05 level

The present chapter was divided into two sections which corresponded to the two larger research questions asked in this study. The first section detailed the analysis of the FEATS and outlined the development of composite scales for this measure of the formal elements in art work. This section included analyses of item distribution, creation of composite scales using Phi

Coefficients, reliability analyses of composite scales and test-retest relaibility of FEATS composites to explore stability over a descrete period of time.

The second section was devoted to the analysis of the relationship between the FEATS composite scales and the primary and composite scales of the BASC-2. The distribution of demographic variables and BASC-2 scores were analyzed and reduced for the primary analysis and Pearson Correltions between variables were explored to assess covariates. Several Multivariate Analyses of Covariance (MANCOVA) were conducted to explore the FEATS composites as predictors of scores on scales of the BASC-2, and follow-up ANCOVA were run to determine specific FEATS composites that were related to the BASC-2 scales. Discussion of these results is presented in Chapter V.

CHAPTER V

Discussion

In the wake of over a hundred years of ambiguous research on the utility of projective drawings for assessment of personality and level of functioning, some research has begun to explore alternative means of using artwork as a screening tool. Changing the focus away from rating the content of drawings, research using formal art-based elements as a means of assessing level of distress, social interactions, behavioral concerns and psychological functioning has yielded encouraging results (Durken, 1954; Kahn & Jones, 1965; Lehmann & Risquez, 1953; Ogdon, 1975; Ulman & Levy, 1992). A few rating scales have been developed to capture art-based formal elements which allows for quantative research into their efficacy in screening for psychological concerns in varying populations (Cohen, Hammer, & Singer, 1988; Gantt & Tabone, 1998).

The FEATS is one such measure that provides standardized definitions and objective rating of the formal elements in two-dimensional art work. Based on the belief that how one draws (e.g., formal elements) may provide a novel approach to assessing social, emotional and behavioral concerns in the artwork of children, adolescents and adults, the FEATS provides a simple means of assessing drawings using fourteen 5-point Likert scales (Gantt & Tabone, 1998). Previous research conducted using the FEATS has suggested that it may be a useful tool in assessing for a variety of psychological concerns in many populations across age groups (Anschel, Dolce, Schwartzman, & Fisher, 2005; Ferber, 1996; Gantt & Tabone, 1998; Gussak, 2007; Munley, 2002; Rockwell & Dunham, 2006; Swan-Foster, Foster, & Dorsey, 2003; Wallace, et al., 2004; White, Wallace, & Huffman, 2004). These studies have typically paired a previously identified clinical sample with a non-clinical control, rated PPAT drawings using the

FEATS and used varying statistical analyses to detect between-group differences. However, the FEATS, has yet to be explored as a general screening tool for a variety of emotional disturbances in adolescent populations with drawings other than the PPAT.

The purpose of this study was threefold: 1) to establish the inter-rater reliability and explore the psychometric properties of the FEATS when used in conjunction with adolescents' drawings, 2) to explore any underlying factor structure present and assess the stability of the FEATS scales over a discrete period of time, and 3) to test the ability of the FEATS composite scales to accurately identify adolescents at-risk for emotional and behavioral disturbance as indicated by scores on the self-report BASC-2. At the conclusion of this study, the original 14 item FEATS had 5 items removed due to poor inter-rater reliability and low variability of rating response. The remaining 9 items were collapsed into three composite scales that appear to have acceptable internal reliability and face validity.

Using these composite scales, several MANCOVA tests were conducted to explore the between-group differences in the formal elements found in two drawings of differing thematic content by adolescent participants. Results of this study suggest the currently proposed FEATS composite scales are not an appropriate screening tool for emotional, behavioral and social problems in adolescents and need continued research before being used in applied settings. Additional findings suggest that the formal elements captured by the FEATS are not independent from the thematic content of an adolescent's drawing and may change when teens are presented with differing drawing topics.

This chapter is broken into three main sections. The first section is a discussion of Hypotheses 1 and 2 related to inter-rater reliability and stability of FEATS composite scales. This section also includes a discussion of the modifications made to the FEATS measure. The

second focuses on the use of the FEATS composite scales as a screening tool for adolescents. Within each of these sections, specific limitations are presented related to each topic. Finally, the last section of this chapter is a discussion of the implications of this study and directions for future research.

Formal Elements Art Therapy Scale (FEATS)

Inter-rater reliability. Without a reliable measure of formal elements, the foundation for the additional analyses in this study would be weak. It was important to determine that raters were indeed indentifying the same formal elements in each participant's drawings. As such, inter-rater reliability was the focus of the first hypothesis in this study, and it was postulated that trained raters could reliably rate drawings using the 5-point Likert FEATS.

FEATS 5-point Likert Scale. Despite multiple reports of acceptable inter-rater reliability using the FEATS 5-point Likert scale, this study did not replicate these results. Kappa values and correlations were poor on all scales except Prominence of Color, Space Use, Developmental Level, Rotation, and Perseveration.

This was an unexpected finding and suggests that the FEATS may not be an appropriate measure to rate formal elements in artwork with varying thematic content in adolescent populations. The FEATS measure had not been previously explored within the current study's context which may explain the findings related to inter-rater reliability. Previous studies reporting acceptable levels of inter-rater reliability have used the PPAT drawing and targeted specific clinical populations (Ferber, 1996; Gantt & Tabone, 1998; Rockwell & Dunham, 2006; Swan-Foster, Foster, & Dorsey, 2003; White, Wallace, & Huffman, 2004). The few studies that explored the FEATS measure with adolescent samples did so outside of a general school environment (Ferber, 1996; Munley, 2002; Wallace, et al., 2004; White, Wallace, & Huffman,

2004), and two out of the four studies did not discuss inter-rater reliability (Munley, 2002; Wallace, et al., 2004).

It is possible that the change in standardized drawing instructions (i.e., not the PPAT) makes it difficult to rate the FEATS reliably. The implications of the findings may suggest that a change is needed in the conceptualization of formal elements as measured by the FEATS. Originally theorized to capture formal elements independent from the content of drawings, this study suggests that the operational definitions of the FEATS are specific to the PPAT and not generalizable to other drawings.

Additionally, it is possible that the sample used in this study is not appropriate for use with the FEATS, which would limit its utility. Perhaps there is a confound in the non-clinical high school sample or group data collection that compromises the inter-rater reliability of the FEATS and implies limited utility as a screening tool for socio-emotional and behavioral disturbances in a general high school population.

Finally, it is possible that the conclusions of previous studies supporting the reliable use of the FEATS with adolescents were inaccurate. The two studies which reported acceptable inter-rater reliability in adolescent populations had sample sizes of less than fifty-five participants (Ferber, 1996; White, Wallace, & Huffman, 2004). Neither study reports the percentage of drawings collected that were double coded for reliability, the means by which raters were trained, or how differences in ratings were reconciled. Even if all drawings in these studies were used to compute inter-rater reliability, the current study has a 20% larger sample size and provides more power when computing inter-rater reliability than previous studies. It is possible that previous support for inter-rater reliability was based on small samples with unique characteristics, which do not exist in larger more diverse samples.

Limitations. While this study clearly does not provide support for reliably rating the 5-point Likert scale version of the FEATS, there are several limitations which could affect this finding including rater related issues such as training, bias and fatigue, and sample concerns.

Raters were trained to acceptable levels of agreement by an art therapist, who had familiarity with the specific FEATS scales and the general concept of formal elements, using the established objective rating guide. That said, the trainer did not develop the FEATS, nor was she trained by its developers. It is possible that some particulars of the scales were not described accurately in training, and that nuances of rating specific scales were lost, increasing ambiguity between raters. If inadequate training is contributing to poor inter-rater reliability, the FEATS manual may need to be further refined with increasing specificity to increase its utility. Additionally, training workshops or certification in the use of the FEATS system may be required to adequately train raters.

Another factor that may be contributing to low levels of inter-rater reliability is the variation in thematic content of the drawings collected in this study. The FEATS was initially designed to be used with the Person Picking an Apple from a Tree Drawing (PPAT; Gantt & Tabone, 1998). While theorized that it could be used with a variety of two dimensional arts, this hypotheses has been previously untested. As acceptable levels of inter-rater reliability have consistently been reported using the 5-point Likert FEATS with the PPAT, the change in drawing theme could be affecting the outcome in this study.

Compounding this problem is that inter-rater reliability was computed between trained raters and the PI, who trained them in the use of the scale. Thus if the PI had misinterpreted the scale or had an unintentional bias towards dichotomous variables, it is possible that this was transferred to the raters. Ideally, the inter-rater reliability would have been computed using the

trained raters and an independent control rater who was not involved in the training of the primary raters. However, this was not feasible in this study.

It is also possible that raters became fatigued during the rating process and began to rate drawings similarity. A total of 338 drawings were rated for this study, and the two primary raters in this study each coded approximately 160 drawings each. While not all these drawings were rated at the same time, rating sessions were often several hours at a time and could have resulted in similar ratings across drawings by individual raters. This, in turn, could have contributed to lower levels of inter-rater reliability, as the fatigue may have resulted in inaccurate coding.

Additionally, drawings selected for inter-rater reliability represented 20% of the total sample. While care was taken to select random drawings from each site, it is possible that 76 drawings had some unknown common characteristics or challenges that contributed to rating discrepancies. Ideally, most, if not all, drawings would have been double rated and discrepancies analyzed to understand variation in inter-rater reliability. However, this was not feasible in the current study.

FEATS dichotomous scales. As discussed in Chapter IV, the FEATS items were dichotomized based on analyses of item response patterns for each scale. Not only did this produce more equally distributed FEATS scales, it allowed for inter-rater reliability to be recomputed using the dichotomous variables. The result of these analyses suggest that when dichotomized, the FEATS items are able to be rated with acceptable levels of inter-rater reliability. In fact, nine out of the thirteen items analyzed had no discrepancies between raters. Of the four with discrepancies, only Line Quality failed to produce acceptable Kappa levels.

Other studies have reported inter-rater reliability concerns on the Line Quality scale but have not chosen to remove it from further analyses (Rockwell & Dunham, 2006; White, Wallace, & Huffman, 2004). It is interesting that raters consistently have a difficult time accurately determining Line Quality which is defined as the overall quality and degree of control that the artist has used in drawing the lines (Gantt & Tabone, 1998). While it is possible that the definition needs further specificity to improve accurate rating, it is also likely that the drawing materials provided (markers) did not provide the opportunity for participants to exert varying degrees of control when drawing lines. As the materials provide limited occasion for differences in control, raters may have been asked to create artificial distinctions in the line quality of drawings which led to inaccurate rating.

Excluding Line Quality, the results of inter-rater reliability on the remaining FEATS scales suggests that raters are able to indicate the presence, or absence, of a formal element in a drawing. This is different from the inter-rater reliability computed on the 5-point Likert version of the FEATS. Previous research related to the FEATS appears not to have explored the item response patterns of the scales in different populations, and no other findings related to inter-rater reliability on dichotomous items has been reported. Thus, there is little with which to compare this finding.

It is of interest that the other standardized measure of formal elements in art work, the Diagnostic Drawing Series (DDS: Cohen, 1985) is comprised of categorical items similar to those produced when the FEATS was dichotomized in this study. The DDS is reported to have acceptable levels of inter-rater reliability using categorical items (Mills, Cohen, & Meneses, 1992) and several studies conducted using the DDS have supported its use in detecting between-group differences (Cohen, Hammer, & Singer, 1988; Kessler, 1992; Leavitt, 1988; Morris, 1995;

McHugh, 1997; Neale, 1994; Shlagman, 1996). Paired with the current study's findings, this suggests that formal elements may actually be dichotomous in nature.

Limitations. There are several limitations to the findings which yielded acceptable inter-rater reliability using the FEATS dichotomized items. Much like the limitations to the initial findings related to the 5-point Likert FEATS, these include rater related issues including training, bias and fatigue, and sampling error. As these were previously discussed in detail, they will not be repeated here.

However, there is a significant limitation to the interpretation of inter-rater reliability with the dichotomized FEATS scales that was not present with the 5-point Likert results. The order in which statistical analysis were conducted confounds this finding. The response pattern analyses of each FEATS scale and subsequent dichotomization of each scale were conducted using the 5-point Likert scales that had not achieved acceptable levels of inter-rater reliability. In other words, the dichotomous scales were created using unreliable scales. While there is little that could be done to control for this confound, it is possible that FEATS items were recoded based on inaccurate ratings making any levels of acceptable inter-rater reliability on the dichotomous scales incorrect. As such, it is important that any results of this study be interpreted with caution due to this methodological concern.

Future research related to inter-rater reliability. There are several future research directions that would clarify the results of this study related to inter-rater reliability. A study that collected, in a random order, both PPAT and drawings of varying content and blindly rated each using the 5-point Likert FEATS would allow questions related to any effect of thematic content on the FEATS to be explored. Such a study would also allow for exploration of response patterns and provide further information related to the categorical or continuous nature of these

formal elements. Additionally, a study that compared the inter-rater reliability of raters trained by an individual independent of the study and experienced in the FEATS would help determine the effect of training bias on inter-rater reliability. Finally, studies related to inter-rater reliability in a general population of adolescents, using the FEATS and the PPAT, should be conducted to determine any effect the more general sample in the study may have had on the inter-rater reliability.

FEATS measure development. Many of the conclusions in this study are built on the modifications made to the FEATS scale including response pattern analysis, dichotomizing the FEATS scales and the creation of the FEATS composite scales. As such, they will be discussed and limitations related to this area of the study will be provided.

FEATS item response patterns explored. The analyses of response patterns for each item on the FEATS indicated that 3 of the 13 scales used in this study needed to be removed due to limited response variability. In addition to Line Quality, which had been previously removed due to unacceptable inter-rater reliability, Space Use, Perseveration, and Rotation were removed because most participants scored the same on each scale.

While no other studies using the FEATS have reported concerns related to item variability (or reported that they conducted analyses of item distribution), other concerns have been reported related to several of these items. Perseveration and Rotation have been consistently reported to have low inter-rater reliability, and often these two scales are dropped from final study analyses (Anschel, Dolce, Schwartzman, & Fisher, 2005; Rockwell & Dunham, 2006, Swan-Foster, Foster, & Dorsey, 2003; White, Wallace, & Huffman, 2004).

Though the concern in this study was not related to inter-rater reliability, it was not unexpected for this study to echo previous problems with the Perseveration and Rotation scales.

Rotation is theorized to screen for organic concerns, and this study's sample did not include participants identified with organic problems. Thus, it makes sense that there would be no variability on this item. Like Rotation, Perseveration is theorized to screen for organic concerns. However, it is also purported to increase with traumatic dissociation. This study did include participants with traumatic backgrounds, but few had identified dissociative symptoms which may have limited the variability of this item. It is possible that findings supporting other studies' conclusions related to inter-rater reliability were not present because there was no variability in the items to allow for potential disagreement.

This study's finding related to item variability for Space Use was surprising. Traditionally, it has been argued that the amount of space used and the size of the drawing on the paper are related to symptoms of depression (Buck, 1948; Burns & Kaufman, 1970; Koppitz, 1968). This was the theoretical underpinning for the scale's inclusion on the FEATS (Gantt & Tabone, 1998). However, the results of this study support the conclusions of Joiner, Schmidt, and Barnett (1996) which found that space and line quality were not significant indicators of mood state in adolescents. If accurate, this finding has significant implications for the way that drawings are assessed in clinical and non-clinical settings.

Several factors may be confounding this result, including materials used, the manner in which space use was operationally defined, and the power available for analysis of this variable. Many of the early findings supporting space use as an indicator of depression were conducted using pencil and paper (8x12), and the materials in this study included colorful markers and larger paper (12x18). While Gantt & Tabone (1998) also found that space use was an indicator that a participant endorsed symptoms of depression, their studies were not conducted with adolescent samples.

The FEATS objectively defines Space Use by dividing the paper into quadrants. Raters are then asked to determine the number of quadrants that include any mark of color. Drawings with any color in all 4 quadrants are rated as full space use. This is different than methods used in traditional projective drawing assessments which allow a more qualitative assessment of size of drawing and space used. It is possible that the objective criterion proposed by the FEATS does not effectively capture the nuances of space use, and thus the item was rated consistently as full space use (n=330 of total n=338). Conversely, it is also possible that adolescents tend to draw on most of the piece of paper, regardless of their socio-emotional functioning.

Finally, it is possible that the finding related to the variability of Space Use was due to limited statistical power. This study was conducted using a general high school sample which did include students endorsing symptoms of depression on the BASC-2. However, it is possible that there were not enough students in this category to provide the necessary power to detect significant differences in space use.

FEATS items dichotomized. Once the item response patterns were explored and items with low variability were removed, it was determined that the scales should be dichotomized into categorical variables rather than conceptualize them as continuous. This created more equally distributed scales which aided in further statistical analyses.

One of the reasons that the FEATS was chosen for this study was because it objectively defined formal elements in a continuous way. It was believed that this would allow for increased specificity in screening adolescents when paired with the BASC-2 scores. However, once the response patterns were explored, it was very clear that items were being rated categorically, as present or absent, rather than with any specificity. While there have been no published findings related to response patterns on FEATS items, it is assumed this information

has not been reported because the patterns of response have not been explored. As such, comparison with other studies is not possible, but it is feasible that the FEATS measure is artificially imposing a continuous structure on naturally dichotomous variables.

There are several factors that may have contributed to this finding, including poor rater training, problems with the continuous definitions on the FEATS, and use of a drawing other than the PPAT. As mentioned above, raters were trained by an Art Therapist with training in formal elements and significant familiarity with the FEATS measure. However, as none of the training came directly from the developers of the FEATS, it is possible that some aspects of the continuous scales were not properly understood. If this was the case, as mentioned earlier, the rating manual for the FEATS needs to be further refined or advanced training made available to increase understanding and rating accuracy. Also, the FEATS was developed using the PPAT drawing, and it is possible the continuous item ratings are appropriate for use with this drawing but are unable to be generalized effectively to drawings with differing subject matters.

FEATS composite scales developed. After FEATS items were dichotomized, relationships were explored to determine an underlying factor structure. Phi Coefficients were used to detect probable relationships between items, and the statistical analyses were supported by the independent art therapist's development of scales based on theoretical relationships. An independent art therapist with training in formal elements and no knowledge of the statistical analysis grouped FEATS items together into composite scales based on theorized relationships. Her theorized composite scales were identical to the statistically determined composites, adding to their face validity.

This analysis yielded three composite scales that appear to tap different aspects of the way in which a drawing was created. The first scale, Specific Formal Elements, which is

comprised of Integration, Realism, Developmental Level, Details of Objects, and Person, reflects the specific details of formal elements that are included in the drawing. Logic, the second scale, includes the Logic and Color Fit items. This composite scale determines if the objects and colors included are idiosyncratic to the drawing. The final scale, Artistic Energy, includes Prominence of Color and Implied Energy items, and reflects the perceived energy used to create the drawing. It is encouraging that correlations between the composite scales were low and suggests that they are measuring independent concepts.

In this study, these composite scales demonstrated appropriate internal consistency and were included in the primary analysis. However, future research needs to explore these underlying factors in greater detail as there are several factors that limit these results. Ideally, should the continuous scale be supported, factor analysis would confirm any underlying scales. The modifications made to the FEATs, and the tenuous findings related to inter-rater reliability suggest that any further analyses be interpreted conservatively.

Stability of Formal Elements

The second hypothesis in this study was that the FEATS composite scales would be stable across a 24 hour period on two different drawings. Formal elements have been theorized to reflect specific mood states independent of drawing theme or content. As such, drawings of differing thematic content were thought to be the equivalent of different forms of a test designed to capture the same factors at different periods of time. Thus, FEATS composites should be stable across the 24 hour period, achieving acceptable levels of test-retest reliability, if reflecting clinically significant levels of mood and behavioral concerns.

This hypothesis was not supported and indicated that FEATS composite scales were not stable across drawings. Test-retest reliability computed between drawings using Pearson Coefficients yielded weak correlations on all composite scales.

This finding was unexpected and suggests that formal elements may not be direct representations of participants' chronic mood, behavioral and social concerns. Rather, it appears that formal elements may be influenced by mild changes in mood as a result of feelings related to drawing directive. There are no reports of test-retest reliability in the FEATS literature. Additionally, the FEATS manual does not report test-retest reliability but does state that it expects formal elements to change as clinically significant symptoms increase or decrease, but it does not discuss stability when symptoms do not change (Gantt & Tabone, 1998). Gussak (2006) found that participants' FEATS scores did change with their level of depression but was unable to replicate these findings in his 2007 study. However, given the results of this study, any changes in the FEATS scores could be the result of unstable formal elements rather than a reflection of clinically significant changes.

There are several reasons the hypothesis regarding the stability of the FEATS composite scales could have resulted in non-significant findings. A connection between formal elements and drawing themes may be present. Additionally, the FEATS composite scales may be sensitive to non-clinical moods and behaviors. Finally, measurement error could be contributing to perceived instability of formal elements over time.

Formal elements measured by the FEATS have been suggested to be independent of drawing content (Gantt & Tabone, 1998). However, most research focused on the FEATS measure used the PPAT as the drawing collected from study participants and thus Gantt and Tabone's (1998) hypothesis has been untested. The results of this study indicate that formal

elements are not independent of drawing thematic content. Rather, the FEATS composite scales appear to change with differing drawing content. This does not necessarily negate their utility as measures of psychological function, but it would limit the generalizability of the FEATS to a variety of drawings that had been well researched in specific populations.

Another conclusion that could be drawn based on this finding is that the FEATS measure is not detecting changes in clinical levels of functioning. FEATS composite scales appear be sensitive to non-clinical changes in mood, behavior, and interpersonal interactions which may a reflect response to the drawing directive or could be reflective of daily changes in mood.

Limitations. There are several limitations to the findings related to this hypothesis including data collection and analysis of the FEATS measure, and development of composite scales. The drawings in this study were not collected in a random order. Due to the larger study's data collection and psycho-educational presentations, all drawings reflecting adolescents' ideal date were collected first, and drawings about exposure to drugs or alcohol were collected 24 hours later. Ideally, drawings would have been collected in a random order to control for any temporal variations that may have occurred at the school and any changes in perception of the research across time.

While the time point at which drawings were collected may contribute to limitations around the understanding of formal elements' stability over time, this result may also be attributed to measurement error. To conduct this analysis, the FEATS measure had to be modified in several ways, and it is possible that this affected the outcome. It is very clear that more extensive research and development of the FEATS is needed prior to additional exploration of the measure as a screening tool for between-group differences. The modifications to the FEATS items, paired with the findings related to inter-rater reliability, make conclusions related

to the stability of formal elements across a discrete period of time and the utility of the FEATS composite scales tenuous. With that caveat, results of the primary analysis of this study will be discussed.

Future research related to stability of the FEATS. As no studies have explored the stability of the FEATS across a discrete period of time, such studies are needed to assess the validity of these results. Studies in clinical and non-clinical populations in which drawings are collected at specific intervals should be conducted to determine the stability of these concepts over time. Additionally, studies exploring any differences in stability between the PPAT drawing and those of varying content would add to the discussion related to test re-test reliability of the FEATs. Finally, it would be of interest to explore if the setting contributed to changes in the formal elements. In this study, drawings were collected in a group and it is possible that viewing other's artwork changed the way that adolescents in this study drew. Research exploring individual and group administration would add to the understanding of the FEATS.

FEATS Composite Scales as Screening for Socio-Emotional Concerns

The final two hypotheses in this study were focused on the ability of the FEATS composite scales, computed on two drawings of varying thematic content, to function as a screening for adolescents' emotional and behavioral functioning as measured by the BASC-2 primary and composite scales. This section will include two main categories. The first will discuss the FEATS composites as a general screening tool, and the second will focus on analyses that yielded interesting significant results.

FEATS as a general screen and BASC-2 scores. Results of the MANCOVA did not support the use of FEATS composite scales to screen for behavioral disturbance in adolescents. Nor did MANCOVA results support the use of the FEATS composites in screening for a variety

of emotional concerns. There were some significant results that will be further described below, but the FEATS composite scales do not appear to be an appropriate general screening tool for adolescents in a general high school setting. While studies have supported the use of content items to screen for adolescent emotional and behavioral concerns (Naglieri, 1991), this study's results do not support the use of formal elements in this capacity. Studies exploring the FEATS in adolescent populations are limited and have yielded mixed results in the ability of the FEATS to distinguish between-group differences (Ferber, 1996; Munley, 2002; Wallace, et al., 2004; White, Wallace, & Huffman, 2004).

There are several reasons why this study's results may differ from those of previous studies with adolescents. It is possible that despite previous research findings, formal elements are not indicators of clinically significant emotional and behavioral concerns in adolescents. Additionally, this study used different drawing themes than those used in previous research. It is also feasible that formal elements are drawing specific, and that the FEATS measure should only use the PPAT drawing to screen for possible mental health concerns. Either of these explanations would undermine the current theoretical conceptualization of formal elements, as they suggest that formal elements are not reflecting clinically significant mood state, nor are they independent from drawing content.

Another explanation for the divergent findings related to the FEATS composites, as a screen with adolescents, is the setting in which drawings were collected. This is one of the first large scale studies focused on the formal elements in general samples of adolescents where data collection occurred within a group classroom setting. Adolescence is a period of development marked by peer interaction and influence (Berndt, 1979; Costanzo & Shaw, 1966; Csikszentmihalyi, Larson, & Prescott, 1977; Harris, 1995). As such, it is probable that

participants' drawings were influenced by their peers. The pressure or peer influence could contribute to profound changes in the way each participant drew, affecting not only drawing content but formal elements. In turn, this would suggest that the information rated by the FEATS composites may have been inaccurate.

While all measures in this study were administered in a group setting, it is likely that the drawings were the most affected by peer dynamics. Given the administration form and multiple choice responses, it is less likely that group pressure influenced the scores. In fact, one reason the BASC-2 was selected for this study was that group data collection in both clinical and research samples had been previously explored (Reynolds & Kamphaus, 2004). Group administration was selected for this study because it appeared to be the most convenient means of accessing information and most closely resembled situations in which a drawing screening might be used. However, this study's non-significant results may be related to the venue, and it suggests future research is needed to determine the most efficacious environment in which to collect drawings.

A final reason hypotheses III and IV may have been unsupported could be the use of composite scales, rather than individual scales, to rate formal elements. As this study appears to be the first to explore an underlying factor structure of the FEATS and the development of composite scales, there is no other data with which to compare these results. However, it is possible that the modifications to the FEATS measure have changed the way in which formal elements are captured. It could be more appropriate to screen adolescents' drawings without using the composite scales. This could increase the number of available patterns of scoring. By providing more options for FEATS scoring patterns, a greater number of diagnostic concerns could possibly be screened. However to explore this, a study with a very large sample would need to be conducted to provide enough power to run the statistical analyses.

Limitations. There are several limitations to this study's findings related to the efficacy of the FEATS composite scales as a screen for adolescent emotional and behavioral concerns, including some methodological concerns and measurement problems. Many of these limitations have been outlined in the above sections as they relate to the inter-rater reliability and measurement development. An additional limitation of these findings is related to the study's design.

This study was a convenience sample and not an experimental design. While great effort was made to collect data from sites reflecting different adolescent populations, participants were not randomly assigned to groups, nor were any variables manipulated to detect cause and effect. While this design was necessary for this study,, it is possible that the design limits the findings.

FEATS composites and BASC-2 significant results. While the hypotheses related to using the FEATS as a general screen for emotional and behavioral concerns in adolescents was unsupported, the analyses did yield some surprising results. These suggested that the thematic topic of each drawing elicited specific FEATS composites that might distinguish between-group differences for a few specific emotional concerns.

Drawing 1 asked participants to draw an image of their perfect romantic date. This drawing appeared to elicit participants' level of self-esteem and locus of control. Participants with lower levels of self-esteem appeared to put less energy into their drawings of the perfect romantic date. Additionally, those with lower levels of internal locus of control included less detail in this image. Both of these findings have face validity. It makes sense that adolescents with low self-esteem and external locus of control would respond using fewer details and less energy on a drawing focused on interpersonal relationships.

The second drawing asked participants to focus on drawing an image of a place in which they might be exposed to drugs and/or alcohol. Interestingly, this drawing appears to tap feelings of self-reliance, relationships with parents and attitudes towards teachers. These BASC-2 scales all focus on interpersonal relationships with adults. Participants with higher levels of self-reliance were more likely to draw with more energy. Those with poor relationships with parents used fewer specific formal elements, and those with poor attitudes towards teachers drew with more idiosyncratic items and colors in these images.

These results suggest that FEATS composite scales are not unrelated to drawing content. Rather it appears that the topic upon which participants were asked to reflect and draw was related to the formal elements that they included in their drawings. This finding has profound effects for the way in which formal elements are conceptualized, providing further support for their connection with drawing content. However, no other studies have explored the use of the FEATS with drawings other than the PPAT so these findings must be interpreted cautiously and reproduced in other studies before conclusions are drawn.

Limitations. While these findings are very interesting, there are limitations to these findings. The limitations to these findings are largely those previously discussed including the data collection setting, variation in drawings, sample self-selection, modification to the FEATS and measurement concerns, and statistical limitations and all must be taken into account as findings are interpreted. The most significant limitation to these findings is the FEATS measure, as the psychometrics of the scale are tenuous and need to be further explored before future studies, focusing on its utility as a screening measure are conducted.

General Implications

At the conclusion of this study, it was determined that the FEATS composites, used with a variety of two dimensional art work, are not an appropriate screening tool for emotional and behavioral concerns in adolescents. However, the FEATS composites do appear to have the potential to screen for emotional concerns when used with specific drawings. More research is needed to determine the accuracy of these results.

These findings have theoretical, research and applied implications. The results suggest that the formal elements captured by this scale are not independent from drawing content. As such, the way in which formal elements are conceptualized may need to be readdressed, which may mean reworking the theoretical understanding of art-based formal elements. In addition to theoretical concerns, this study highlights many problems with the current research base using the FEATS. It suggests that additional measure development is needed to reduce measurement error. Finally, this study's results indicate that formal elements may have some ability to detect adolescents' emotional functioning but warn that additional research is needed before being used in an applied setting.

Limitations. While many of the limitations related to specific hypotheses of this study have already been addressed, there are a few general limitations worthy of discussion. As previously mentioned, this was not an experimental study and there were many variables that could provide possible confounds. There was no control for artistic background, artistic ability or interest in art activities. The participants' ability levels and exposure to artistic media could have affected the outcome of this study and should be controlled in future research.

Additionally, there could have been some differences present in the data collection. While standardization was followed throughout the study, modifications in presentation and data

collection could have occurred across classroom environments and sites. This, paired with data collection occurring in a specific order at each site, rather than randomly collecting drawings and the BASC-2, could have affected internal validity, as participants may have lost interest in the general study. A final threat to internal validity is rater bias. As previously discussed in detail, steps were taken to reduce rater bias, but it is possible that rater bias was present in this study.

The sample in this study is another limitation which could effect external validity. As the sample included a general high school population, it is possible that there were some unique characteristics of those enrolled in the health and art therapy courses at the each respective site. Future studies may wish to target specific adolescent populations to reduce the sample size needed for appropriate statistical analyses. While great thought was given to the statistical analyses in this study and in general the study's power was appropriate, the number of participants that fell within the at-risk/clinical range on the specific BASC-2 scales was small and limited the power of the statistical analyses to detect between-group differences.

Finally, as discussed throughout, there are significant measurement concerns which limit the findings of this study. In addition to the concerns related to the psychometrics of the FEATS and collection of the drawings in a group setting, all measures in this study were self-report in nature. While not feasible, had additional outside information been collected, the results of this study may have been different.

Future Directions

Results of this study suggest two main directions for future research: 1)further psychometric development and testing of the FEATS measure, and 2) assessment of the utility of formal elements as a screening tool for emotional and behavioral disturbance in adolescents.

FEATS measure development. The FEATS measure would benefit from continued refinement and research in several areas, including inter-rater reliability, item analyses and factor exploration, and use with drawings other than the PPAT. As inter-rater reliability in this study was questionable, future studies exploring inter-rater reliability of the FEATS using the PPAT and other drawings would be beneficial. Provision of a solid research base with acceptable levels of inter-rater reliability is the first step in assuring the psychometrics of the FEATS measure. Researchers using the FEATS in future studies are encouraged to rigorously explore inter-rater reliability, report their findings in the literature, and remove items that are not reliable for analyses.

Additionally, the FEATS measure would be well served by further psychometric development. The current study's findings suggest that the response patterns for items on the FEATS are not normally distributed across a continuous scale, and that they may be more appropriate for dichotomous categorical coding. However, given that there is no other report of response pattern analysis on the FEAT, this finding needs additional research to determine the most appropriate scale with which to rate formal elements. Once inter-rater reliability is well established and the appropriate scale is determined, the underlying relationships between scales on the FEATS should be thoroughly explored using exploratory factor analyses or SEM, and items reduced to scales accordingly. This would increase the utility of the FEATS both in research and clinical settings.

Finally, continued research related to the use of the FEATS with drawings other than the PPAT needs to be explored. This study's results indicate that the FEATS may not be a stable indicator of formal elements across drawings of varying content. This would limit its use to the PPAT drawing. Research exploring the FEATS scores on PPAT and other drawings, controlling

for artistic interest and experience, would aid in determining the stability and generalizability of formal elements.

Formal elements as an indicator of mood state. Results of this study do not support the current use of the FEATS composite scales to screen for emotional and behavioral concerns. It does appear that the FEATS composite scales may be able to distinguish broadly between a clinical and non-clinical sample as there are strong differences in drawing response by school. However, due to the concerns related to the FEATS measure, it is not clear if these findings are due to measurement error or that formal element are not appropriate screening tools. Given that art-based assessment provides a non-threatening means of screening adolescents in a variety of contexts for mental health concerns, it would be beneficial to continue research in this area. Once a reliable and valid measure of formal elements is established, future studies with adolescent populations may yield different results.

Research focused on using the PPAT in conjunction with other drawings would help determine if screening requires a specific drawing directive. Additionally, using objective measures to capture mood and behavioral concerns, from a variety of sources, would control for any problems related to self-report data collection. Future studies may wish to control for artistic ability, skills and training, as well as other variables such as intelligence and openness to experience which may influence the way in which drawings are approached.

While the results of this study did not support the proposed hypotheses, they do provide evidence of feasibility of conducting large scale drawing data collection with adolescent populations. Additionally, the findings suggest concrete areas for future research which could solidify the theory and psychometrics related to measuring formal elements. This is an exciting new area of research with much possibility, and further carefully constructed and controlled

research would add valuable dimensions to the discussion about the utility of drawings in assessing adolescents.

REFERENCES

Abell, S., Heigberger, A.M., & Johnson, J.E. (1994). Cognitive evaluations of young adults by means of human figure drawings: An empirical investigation of two methods. *Journal of clinical psychology, 50,* 900-905.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist (4-18) and 1991 profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Adler, P.Y. (1952). Evaluation of the figure drawing technique: Reliability, factorial structure and diagnostic usefulness. *Journal of consulting and clinical psychology, 33,* 52-57.

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

Amir, G., & Lev-Wiesel, R. (2007). Dissociation as depicted in the traumatic event drawings of child sexual abuse survivors: A preliminary study. *The arts in psychotherapy, 34,*114-123.

Anastasi, A. (1982). *Psychological testing.* New York: MacMillian.

Anderson, J (1997). Tree drawings as isomorphic representations of disturbance in female body image. Poster presented at the American Art Therapy Association, Milwaukee, WI.

Anschel, D.J, Dolce, S., Schwartzman, A., & Fisher, R.S. (2005). A blinded pilot study of artwork in a comprehensive epilepsy center population. *Epilepsy and behavior, 6,* 196-202.

Archer, R.P., & Newsom, C.R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment, 7,* 227-235.

Arnheim, R. (1969). *Visual thinking.* Berkeley, CA: University of California Press.

Arnheim, R. (1986). *New essays on the psychology of art.* Berkeley, CA: University of California Press.

Anschel, D.J., Dolce, S., Schwartzman, A., & Fisher, R.S. (2005). A blinded pilot study of artwork in a comprehensive epilepsy center population. *Epilepsy & Behavior, 6,* 196-202.

Beekhuizen, K.S., Davis, M.D., Kolber, M.J., & Cheng, M.S. (2009). Test-retest reliability and minimal detectable change of the Hexagon Agility Test. *Journal of strength and conditioning research, 27,* 2167-2171.

Berndt, T (1979). Developmental changes in conformity to peers and parents. *Developmental psychology, 15,* 608-616.

Berryman, E. (1959). The self-portrait: A suggested extension of the HTP. *Perceptual and motor skills, 9,* 411-414.

Betts, D. J. (2006). Art therapy assessments and rating instruments: Do they measure up? *The arts in psychotherapy, 33,* 422-434.

Beyerstein, B.L. & Beyerstein, D.F. (1992). *The write stuff: Evaluations of graphology – the study of handwriting analysis.* Buffalo, NY: Prometheus Books.

Blos, P. (1962). *On adolescence: A psychoanalytic interpretation.* New York: The Free Press.

Brooke, S.L. (1996). *A therapist's guide to art therapy assessments: Tools of the trade.* Springfield, Illinois: Charles C Thomas Publisher, LTD.

Bruscia, K. (1988). Standards for clinical assessment in the arts therapies. *The arts in psychotherapy, 15,* 5-10.

Buck, J.N. (1948). The H-T-P technique: A qualitative and quantitative scoring manual. *Journal of clinical psychology, 4,* 317-397.

Burgess, A.W., & Hartman, C.R. (1993). Children's drawings. *Child abuse and neglect, 17,* 161-168.

Burns, R.C., & Kaufman, S.H. (1970). *Kinetic family drawings (KFD): An introduction to understanding children though kinetic drawing.* New York: Brunner/Mazel.

Burns, R.C. (1987). *Kinetic-House-Tree-Person Drawings (K-H-T-P): An interpretive manual.* New York, NY: Bruner-Mazel.

Camara, W.J., Nathan, J.S., & Puente, A.E. (2000). Psychology test usage: Implications in professional psychology. *Professional psychology: Research and practice, 31,* 141-154.

Cheyne-King, S.E. (1990). Effects of brain injury on visual perception and art production. *The arts in psychotherapy, 17,* 69-74.

Clarke, P.M, Murnen, S.K., & Smolak, L. (2010). Development and psychometric evaluation of a quantitative measure of "fat talk". *Body image, 7,* 1-7.

Cohen, B.M. (Ed.) (1985). *The Diagnostic Drawing Series Handbook.* (Available from Barry M. Cohen, P.O. Box 9853, Alexandria, VA 22304).

Cohen, B.M. (Ed.) (1986). *The Diagnostic Drawing Series Rating Guide.* (Available from Barry M. Cohen, P.O. Box 9853, Alexandria, VA 22304).

Cohen, B.M., Cox, C.T., Mills, A., & Sobol, B. (Speakers). (1990). *Art by abuse*

survivors: A lifecycle. (Cassette No. 569-05-90). Alexandria, VA: Audio Transcripts.

Cohen, B.M., Hammer, J.S., & Singer, S (1988). The Diagnostic Drawing Series: A systematic approach to art therapy evaluation and research. *The arts in psychotherapy, 15,* 11-21.

Cohen, F.W., & Phelps, R.E. (1985). Incest markers in children's artwork. *The arts in psychotherapy, 12,* 265-283.

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Education and Psychological Measurement, 20,* 37–46.

Coopersmith, S. (1981). *The antecedents of self-esteem.* Palo Alto, CA: Consulting Psychologists Press.

Couch, J.B. (1994). Diagnostic Drawing Series: Research with older people diagnosed with organic mental syndrome and disorders. *Art therapy, 11,* 111-115.

Costanzo, P., & Shaw, M. (1966). Conformity as a function of age level. *Child Development, 37,* 967-975.

Cox, K.L., & Price, K. (1990). Breaking through: Incident drawings with adolescent substance abusers. *The arts in psychotherapy,* 17, 333-337.

Craig, G.J. & Dunn, W.L. (2007). *Understanding human development.* Upper Saddle River, NJ: Prentice Hall.

Cressen, R. (1975). Artistic quality of drawings and judges' evaluations of the DAP. *Journal of personality assessment, 39,* 132-137.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of youth and adolescence, 6,* 281-294.

Dawes, R.M. (1994). *House of cards: Psychology and psychotherapy built on myth.* Mahwah, New Jersey: Erlbaum.

Des Marais, S. (Speaker). (1996). *Working towards closure: Art therapy with soldiers of the Gulf War.* (Cassette Recording No. 123). Denver, CO: National Audio Video, Inc.

Durken, H. (1954). The reliability and validity of spontaneous finger paintings. *Journal of projective techniques, 18,* 169-182.

Eitel, K., Szkura, L., Pokorny, D., & Von Wietersheim, J. (2008). Do you see what I see? A study about the interrater reliability in art therapy. *Psychotherapy research, 18,* 216-224.

Emunah, R. (1990). Expression and expansion in adolescence: The significance of creative arts therapy. *The arts in psychotherapy, 17,* 101-107.

Erikson, E. (1968). *Identity, youth and crisis.* New York: Norton.

Falk, J.D. (1981) Understanding children's art: An analysis of the literature. *Journal of personality assessment, 45,* 465-472.

Feber, R. (1996) *Measuring the validity of the FEATS using the apple tree projective drawing to assess self-esteem in adolescents.* Unpublished master's thesis, College of Notre Dame, Belmont, CA.

Feher, E., VanderCreek, L., & Teglasi, H. (1983). The problem of art quality in the use of human figure tests. *Journal of clinical psychology, 39,* 268-275.

Fowler, J.P., & Ardon, A.M. (2002). Diagnostic Drawing Series and dissociative disorders: A Dutch study. *The arts in psychotherapy, 29,* 221-230.

Fox, C., Davidson, K., Lighthall, F., Waite, R., & Satason, S.B. (1958). Human figure drawings of high and low anxious children. *Child development, 29,* 297-330.

Freud, S. (1911). *General psychological theory: Papers on metapsychology.* New York: Collier Books.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (7th edition). White Plains, NY: Longman.

Gantt, L. (2004). A case for formal art therapy assessments. *Art therapy, 21,* 18-29.

Gantt, L. (2001). The Formal Elements Art Therapy Scale: A measurement system for global variables in art. *Art therapy, 18,* 50-55.

Gannt, L. (2000). Assessment in the creative arts therapies: Learning from each other. *Music therapy perspectives, 18,* 41-46.

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

Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Toward are solution of the Rorschach controversy. *Psychological Assessment, 13*, 423-448.

Gittelman Klein, R. (1996). Questioning the clinical usefulness of projective psychological tests for children.

Golomb, C. (1999). Art and the young: The many faces of representation. *Visual arts research, 25,* 27-50.

Goodenough, F.L. (1926). *Measurement of intelligence by drawings.* New York: Harcourt, Brace, and World.

Gross, J., & Hayne, H. (1998). Drawing facilitates children's verbal reports of emotionally laden events. *Journal of experimental psychology: Applied, 2,* 163-179.

Groth-Marnat, G. (1990). *Handbook of psychological assessment,* (2nd ed.). New York, NY: John Wiley & Sons.

Gussak, D. (2006). Effects of art therapy with prison inmates: A follow-up study. *The arts in psychotherapy, 33,* 3, 188-198.

Gussak, D. (2007). The effectiveness of art therapy in reducing depression in prison populations. *International journal of offender therapy and comparative criminology, 51,* 4, 444-460.

Gwet, K. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British journal of mathematical & statistical psychology,61,* 29-48.

Hacking, S. (1999). *The psychology of everyday art: A quantative study.* Dissertation, University of Keele, Sheffield, UK. Published online at http://musictherapyworld.de/modules/archive/stuff/papers/hacking.pdf.

Hagood, M.M. (2002). A correlational study of art-based measures of cognitive developmental: Clinical and research implications for art therapists working with children. *Art therapy: Journal of the American Art Therapy Association, 19,* 63-68.

Hagood, M.M. (2003). The use of the Naglieri Draw-a-Person Test of Cognitive Development: A study with clinical and research implications for art therapists working with children. *Art therapy, 20,* 2, 67-76.

Hammer, E. (Ed.). (1958). *The clinical application of projective drawings.* Springfield, IL: Charles C Thomas.

Hammer, E.F. (1969). DAP: Back against the wall? *Journal of consulting and clinical psychology, 33*, 151-156.

Handler, L., & Reyher, J. (1966). Relationship between GSR and anxiety indexes in projective drawings. *Journal of consulting psychology, 30,* 60-67.

Harris, D.B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough Draw-a-Man Test.* New York, NY: Harcourt Brace Jovanovich.

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological review, 102,* 458-489.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77-89.

Hays, R.E., & Lyons, S.J. (1981). The bridge drawing: A projective technique for assessment in art therapy. *The arts in psychotherapy, 8*, 207-217.

Heitmajer, O.A., & Cohen, B.M. (Speakers). (1993). MPD and the DDS: A Dutch replication study. *In B.G. Braun (Ed.), Proceedings of the 10th International Conference on Multiple Personality/Dissociative States.* Chicago: Rush Hospital. Audiotape: Audio Transcriptions, Alexandria, VA.

Hibbard, R.A., & Hartman, G.L. (1990). Emotional indicators in human figure drawings of sexually victimized and nonabused children. *Journal of clinical psychology, 46,* 211-219.

Hinz, L.D. (2008). Walking the line between passion and caution in art therapy: Using the Expressive Therapies Continuum to avoid therapist errors. *Journal of the American Art Therapy Association, 25,* 38-40.

Hjorth, C.W., & Harway, M. (1981). The body image of physically abused and normal adolescents. *Journal of clinical psychology, 37,* 863-866.

Holtzman, W.H. (1993). An unjustified, sweeping indictment by Motta, et al. of human figure drawings for assessing psychological functioning. *School psychology quarterly, 8,* 189-190.

Hulse, W.C. (1951). The emotionally disturbed child draws his family. *Quarterly journal of child behavior, 3,* 152-174.

Johnson, G.S. (1989). Emotional indicators in the human figure drawings of hearing impaired children: A small sample validation study. *American annals of the deaf, 134,* 205-208.

Joiner, T.E., Jr., Schmidt, K.L., & Barnett, J. (1996). Size, detail, and line heaviness in children's drawings as correlates of emotional distress; (more) Negative evidence. *Journal of personality assessment, 67,* 127-141.

Jung, C.J. (1967). *The spirit of man, art, and literature.* Princeton, New Jersey: Princeton University Press Bollingen Foundation.

Kahill, S. (1984). Human figure drawings in adults: An update of the empirical evidence, 1967-1982. *Canadian psychology, 25,*269-292.

Kahn, B. (1999). Art therapy with adolescents: Making it work for school counselors. *Professional school counseling, 2,* 291-299.

Kahn, M., & Jones, N. (1965). Human figure drawings as predictors of admission to a psychiatric hospital. *Journal of projective techniques and personality assessment, 29,* 319-322.

Kaufman, J.C., Niu, W., Sexton, J.D., & Cole, J.C. (2010). In the eye of the beholder: Differences across ethnicity and gender in evaluating creative work. *Journal of applied social psychology, 40,* 496-511.

Kellogg, R. (1970). *Analyzing children's art.* Palo Alto, CA: National Press Books.

Klopher, W.F., & Taulbee, E. (1976). Projective tests. *Annual review of psychology, 27,* 543-567.

Kielhofner, G., Dobria, L., Forsyth, K., & Kramer, J. (2010). The Occupational Self Assessment: Stability and the ability to detect change over time. *Occupational therapy journal review: Occupation, participation and health, 30, 11-19.*

Koppitz, E.M. (1968). *Psychological evaluation of children's human figure drawings.* New York, NY: Grune & Stratton.

Koppitz, E.M. (1984). *Psychological evaluation of human figure drawings made by middle school pupils.* New York, NY: Grune & Stratton.

Kovacs, M. (1992). *Children's Depression Inventory manual.* Los Angeles, CA: Western Psychological Services.

Kramer, E., & Schehr, J. (1983). An art therapy evaluation session for children. *Art therapy, 23,* 3-12.

Kressler, J. (1994). A study of the Diagnostic Drawing Series with eating disordered patients. *Art therapy, 11,* 116-118.

Kymissis, P. (1992). The prospective kinetic family drawing. *American journal of art therapy, 31,* 17-21.

Lally, S. (2001). Should human figure drawings be admitted to court? *Journal of personality assessment, 76,* 135-149.

Landgarten, H. (1987). *Family art therapy.* New York: NY: M. Evans.

Landis, J.R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

LaRoque, S.D., & Obrzut, J.E. (2006). Pencil pressure and anxiety in drawings: A techno-projective approach. *Journal of psychoeducational assessment,24,* 381-393.

Leavitt, C.G. (1988). A validity study f the Diagnostic Drawing Series used for assessing depression in children and adolescents (Doctoral dissertation, California School of Professional Psychology, Los Angeles).

Lehmann, H. & Risquez, E. (1953). The use of finger paintings in the clinical evaluation of psychotic conditions: A quantative and qualitative approach. *Journal of mental science, 99,* 763-777.

Leibowitz, M. (1999). *Interpreting projective drawings.* New York, NY: Brunner/Mazel.

Lev-Wiesel, R. & Liraz, R. (2007). Drawings versus narratives: Drawings as a tool to encourage verbalizations in children whose fathers are drug abusers. *Clinical child psychology and psychiatry, 12,* 65-75.

Lev-Wiesel, R., & Hershkovitz. (2000). Detecting violent aggressive behavior among male prisoners through the Machover Draw-A-Person test. *Arts in psychotherapy, 27,* 171-177.

Lev-Wiesel, R., & Shvero, T. (2003). An exploratory study of self-figure drawings of individuals diagnosed with schizophrenia. *Arts in psychotherapy, 30,* 13-16.

Levy, B.I., & Ulman, E. (1967). Judging psychopathology from painting. *Journal of abnormal psychology, 72,* 182-187.

Lilienfeld, S.O., Wood, J.M., & Garb, H.N. (2000). The scientific status of projective techniques. *Psychological science in the public interest, 1,* 27-66.

Loevinger, J. (1976). *Ego development: Conceptions and theories.* San Francisco, CA: Jossey-Bass.

Lowenfeld, V., & Brittain, W. (1987). *Creative and mental growth* (8th ed.). New York: Macmillian.

Machover, K. (1949). *Personality projection in the drawing of the human figure.* Springfield, Il: Charles C. Thomas.

Machover, K. (1960). Sex differences in the developmental patterns of children as seen in Human Figure Drawings. In A.I. Rabin & M.R. Haworth (Eds.), *Projective techniques in children.* New York, NY: Grune & Stratton.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood and adulthood: A move to the level of representation. *Monograph for the Society for Research in Child Development, 50,* 66-104.

Malchiodi, C. (1990). *Breaking the silence: Art therapy with children from violent homes.* New York, NY: Brunner/Mazel.

May, R. (1975). *The courage to create.* New York, NY: Norton.

McArthur, D.S., & Roberts, G.E. (1990). *Roberts Apperception Test for Children manual.* Los Angeles, CA: Western Psychological Services.

McGraw, K.O., & Wong, S.P. (1996) Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*, 30-46.

McHugh, C. (1997). *A comparative study of structural aspects of drawings between individuals diagnosed with major depressive disorder and bipolar disorder in the manic phase.* Unpublished master's thesis, Eastern Virginia Medical School, Norfolk, Virginia.

McNiff, S. (1998). *Art based research.* Philadelphia, PA: Jessica Kingsley.

Mills, A (Speaker). (1988). *The Diagnostic Drawing Series: The borderline personality disorder sample.* (Cassette Recording No. W327-39). Garden Grove, CA: Infomedix.

Mills, A., Cohen, B.M., Meneses, J.Z. (1993). Reliability and validity tests of the Diagnostic Drawing Series. *Art in psychotherapy, 20*, 83-88.

Milne, L.C., Greenway, P., & Best, F. (2005). Children's behavior and their graphic representation of parents and self. *Arts in psychotherapy, 32*, 107-119.

Mitchell, C. (1995). *The Diagnostic Drawing Series: Research with male inmates incarcerated for murder.* Unpublished Master's thesis. College of Notre Dame, Belmont, CA.

Morris, M.B. (1995). The Diagnostic Drawing Series and the Tree Rating Scale: An isomorphic representation of multiple personality disorder, major depression, and schizophrenia populations. *Art therapy, 12*, 118-128.

Mostkoff, D.L., & Lazarus, P.J. (1983). The kinetic family drawing: The reliability of an objective scoring system. *Psychology in the schools, 20,* 16-20.

Motta, R.W., Little, S.G., & Tobin, M.I. (1993). The use and abuse of human figure drawings. *School psychology quarterly, 8,* 162-169.

Munley, M. (2002). Comparing the PPAT drawings of boys with ADHD and aged-matched controls using the Formal Elements Art Therapy Scale. *Art therapy: Journal of the American art therapy association, 19,* 69-76.

Murray, H. (1943). *Thematic Apperception Test.* Cambridge, MA: Harvard University Press.

Naglieri, J.A. (1991). *Draw-a-Person: Screening Procedure for Emotional Disturbance.* Austin, TX: ProEd.

Naglieri, J.A. (1993). *Draw-a-Person: A quantative scoring system.* San Antonio, TX: The Psychological Corporation.

Naglieri, J.A., & Pfeiffer, S.I. (1992). Performance of behavior disordered and normal samples on the Draw-a-Person: Screening Procedure for Emotional Disturbance. *Psychological assessment, 4,* 156-159.

Neal, E.L., & Rosal, M.L. (1993). What can art therapists learn from the research on projective drawing techniques for children? A review of the literature. *The arts in psychotherapy, 20,* 37-49.

Neal, E.L. (1994). The Children's Diagnostic Drawing Series. *Art therapy, 11,* 119-126.

Nichols, R.C., & Stumpfer, D.J. (1962). A factor analysis of Draw-A-Person test scores. *Journal of consulting psychology, 26,* 156-161.

Ogdon, D. (1975). *Psychodiagnositcs and personality assessment: A handbook (2nd ed.)*. Los Angeles: Western Psychological Services.

Palmer, L., Farrar, A.R., Valle, M., Ghahary, N., Panella, M., & DeGraw, D. (2000). An investigation of the clinical use of the house-tree-person projective drawings in the psychological evaluation of child sexual abuse. *Child maltreatment, 5*, 169-175.

Peterson, C.A., & Schilling, K.M. (1983). Card pull in projective testing. *Journal of personality assessment, 47,* 265-275.

Piaget, J. (1952). *The origins of intelligence in children.* New York, NY: International Universities Press.

Rankin, A. (1994). The tree drawings and trauma indicators: A comparison of past research with current findings from the Diagnostic Drawing Series. *Art therapy, 11*, 127-130.

Reb, KJ., & Greguras, G. J. (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. *Journal of applied psychology,95,* 213-220.

Reber, A.S. & Reber, E.S (2001). *Dictionary of psychology*. London, Penguin/Viking.

Reithmiller, R.J., & Handler, L. (1997). Problematic methods and unwarranted conclusions in DAP research: Suggestions for improved research procedures. *Journal of personality assessment, 69*, 459-473.

Reynolds, C.R., & Kamphaus, R.W. (1994). *Behavior Assessment System for Children manual (BASC).* Circle Pines, MN: American Guidance Service.

Reynolds, C. R. & Kamphaus, R. W. (2004). *Behavior assessment system for children, 2nd edition* (*BASC-2*). Circle Pines, MN: American Guidance Service.

Reynolds, C.R., & Richmond, B.O. (1985). *Revised Children's Manifest Anxiety Scale (RCMS) manual.* Los Angeles, CA: Western Psychological Services.

Ricca, D. (1992). *Utilizing the Diagnostic Drawing Series as a tool in differentiating a diagnosis between multiple personality disorder and schizophrenia.* Unpublished master's thesis, Hahnemann University, Philadelphia.

Roback, H. (1968). Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. *Psychological bulletin, 70,* 1-19.

Rockwell, P. & Dunham, M. (2006). The utility of the Formal Elements Art Therapy Scale in assessment for Substance Use Disorder. *Art therapy, 23,* 3, 104-111.

Rorschach, H. (1921) *Psycodiagnostik.* Bern, Switzerland: Bicher.

Rubin, J.A. (1984). Imagery in art therapy: The source, the setting and the significance. *Journal of mental energy, 9,* 71-81.

Rubin, J.A. (2005). *Child Art Therapy.* Hoboken, NJ: John Wiley and Sons, Inc

Rubin, J.A. & Shacter, J. (1972). Judgment of psychology from art productions of children. *Confinia psychiatric, 15*, 237-250.

Rubin, J.A., Shacter, J., & Ragins, N. (1983). Intra-individual variability in H-D-Fs: A developmental study. *American journal of orthopsychiatry, 53,* 654-667.

Saal, F.E., Downey, R.G., & Lahey, M.A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88,* 413–428.

Shlagman, H. (1996). *The Diagnostic Drawing Series: A comparison of psychiatric inpatient adolescents in crisis with non-hospitalized youth.* Unpublished master's thesis, Hahnemann University, Philadelphia.

Shrout P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological bulletin, 86,* 420-428.

Sidun, N.M., & Rosenthal, R.H. (1987). Graphic indicators of sexual abuse in draw-a-person tests of psychiatrically hospitalized adolescents. *The arts in psychotherapy, 14,* 25-34.

Stein, M. (1997). The use of family drawings by children in pediatric practice. *Developmental and behavioral pediatrics, 18,* 334-339.

Sturner, R.A., & Rothbaum, F. (1980). The effects of stress on children's human figure drawings. *Journal of clinical psychology, 36,* 324-331.

Suinn, R.M., & Oskamp, S. (1969). *The predictive validity of projective measures: A fifteen year evaluative review of research.* Springfield, IL: Charles C. Thomas.

Swan-Foster, N., Foster, S., & Dorsey, A. (2003). The use of human figure drawing with pregnant women. *Journal of reproductive and infant psychology, 21,* 293-307.

Szondi, L. (1947). *Experimentelle tribsdiagnostik.* Bern: Verlag Has Huber.

Taylor, S.A., Kymisssis, P., & Pressman, M. (1998). Prospective Kinetic Family Drawing and adolescent mentally ill chemical abusers. *Arts in psychotherapy, 25,* 115-124.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th edition). Boston: Allyn and Bacon.

Thomas, G.V., & Jolley, R.P. (1998). Drawing conclusions: A re-examination of empirical and conceptual bases for psychological evaluations of children from their drawings. *British journal of clinical psychology, 37,* 127-139.

Uebersax, J.S. (1987). Diversity of decision making models and the measurement of interrater agreement. *Psychological Bulletin, 101*, 140–146.

Ulman, E. (1992). A new use of art in psychiatric diagnosis. *American journal of art therapy, 30,* 78-88.

Ulman, E., & Levy, B. (1967). Judging psychopathology from paintings. *Journal of abnormal psychology, 72,* 182-187.

Ulman, E., & Levy, B. (1992). An experimental approach to the judgment of psychopathology from paintings. *American journal of art therapy, 30*, 107-112.

Vass, Z. (1998). The inner formal structure of the H-T-P drawings: An exploratory study. *Journal of clinical psychology, 54,* 611-619.

Veach, L.J., & Gladding, S.T. (2007). Using creative group techniques in high schools. *The journal for specialists in group work, 32,* 71-81.

Veltman, M.W.M, & Browne, K.D. (2001). Identifying childhood sexual abuse through favorite kind of day and kinetic family drawings. *The arts in psychotherapy, 28,* 251-259.

Veltman, M.W.M, & Browne, K.D. (2002). The assessment of drawings from children who have been maltreated: A systematic review. *Child abuse review, 11*, 19-37.

Wagner, E.E. (2003). Defining projective techniques: Finding common ground for discourse. *Psychological reports, 92*, 951-956.

Wallace, J., Yorgin, P.D., Carolan, R., Moore, H., Sanchez, J., Belson, A., Yorgin, L., Major, C., Granucci, L., Alexander, S. R., & Arrington, D. (2004). The use of art therapy to detect depression and post traumatic stress disorder in pediatric and young adult renal transplant recipients. *Pediatric transplantation, 8*, 52-59.

Watkins, C.E., Campbell, V.L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologist. *Professional psychology: Research and practice, 26,* 54-60.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of Positive and Negative Affect: PANAS scales. *Journal of personality and social psychology, 54,* 1063-1070.

Wechsler, D. (1991). *Manual of the Wechsler Intelligence Scale for Children-Third Edition.* San Antonio, TX: The Psychological Corps.

White, C.R., Wallace, J., & Huffman, L.C. (2004). Use of drawings to identify thought impairment among students with emotional and behavioral disorders: An exploratory study. *Art therapy: Journal of the American art therapy association, 21,* 210-218.

Williams, K.J., Agell, G., Gantt, L., & Goodman, R.F. (1996). Art-based diagnosis: Fact or fantasy? *American journal of art therapy, 35,* 9-30.

Woltmann, A.G. (1960). Spontaneous puppetry by children as projective method. In A.I. Rabin & M.R. Haworth (Eds.), *Projective techniques with children.* (pp. 305-312). New York, NY: Grune & Stratton.

Wood Howe, J., Burgess, A.W., & McCormack, A. (1987). Adolescent runaways and their drawings. *Arts in psychotherapy, 14,* 35-40.

Woodward, S. (1998). Understanding the Child Diagnostic Drawing Series within the child witness to domestic violence population. *Canadian art therapy journal, 12,* 11-33.

Yahnke, L. (2000). *The Diagnostic Drawing Series as a assessment for children who have witnessed marital violence.* Doctoral dissertation, Minnesota School of Professional Psychology, Minneapolis, MN.

Appendix A

KU HUMAN SUBJECTS COMMITTEE APPROVAL LETTER

KU RESEARCH & GRADUATE STUDIES
The University of Kansas

2/5/2008
HSCL #15714

Selby Conrad
PRE
621 J R Pearson

The Human Subjects Committee Lawrence reviewed your research update application for project

15714 Conrad/Beasley/Hunter/Kim/O'Byrne (PRE) The Impact of Heterosocial Competence on Adolescent Adjustment, Family Relationships, Pee Relationships, and Health Behavior

and approved this project update through an expedited review process according to 45 CFR 46.110 (b)(2) minor changes (or no changes) in a previously approved project. Your project has continued approval to 2/24/2009. Approximately one month prior to 2/24/2009, HSCL will send to you a Status Report request, which will be necessary for you to complete in order to obtain continued approval for the next twelve months. Please note that you must stop data gathering if you do not receive continued HSCL approval.

HSCL approves the SSBQ instrument which asks about sexual behaviors and sexual risk taking. HSCL also approves your revised parent letter and assent which describe the new measure.

Please use the HSCL "approval stamp" on your consent forms. Just cut and paste. You may resize and reshape the text to fit your documents.

Approved by the Human Subjects Committee Lawrence (HSCL) on 2/5/2008. Approval expires one year from 2/24/2008.

If you complete your project before the renewal date, please notify HSCL. Thank you for providing us with this update information.

Sincerely,

David Hann

David Hann
HSCL Coordinator

Human Subjects Committee Lawrence
Youngberg Hall | 2385 Irving Hill Road | Lawrence, KS 66045 | (785) 864-7429 | Fax (785) 864-5049 | www.research.ku.edu

Appendix B

INFORMATION FOR PARENTS

Dear Parent or Guardian:

Later this month, visitors from the University of Kansas will be coming to _____ High School to work with students in health class over a three day period. During this time, the visitors will be providing information about health behaviors such as substance abuse and relationships. They will also be asking students to answer questions about their health behaviors, peer relationships and parents' relationships as part of a study about adolescent health. These survey questions are designed to provide information about the connection between relationships and health behaviors. In addition to answering questions about these topics, students will be asked some background questions and will be involved in an art activity.

The visitors have been working with health teachers at our school to select topics for these presentations. While the presentations will be part of regularly scheduled health classes, answering the visitors' study questions about health behavior and relationships is completely voluntary. Your child may choose not to provide this information for any reason. If your child would not like to participate, he or she may choose to complete a regularly scheduled health activity provided by his or her teacher. Regardless of what your child chooses, his or her grade will not be affected. If your child does answer questions about his or her health behavior and relationships, all information will remain completely confidential. In fact, materials with information provided by your child will not contain his or her name.

Information collected will be used to learn more about students at your child's school and to help plan future activities within the health classes. Additionally, the information will help professionals better understand adolescents in general and help develop better intervention programs in the future. If you have any questions, would like additional information, or would like to view the surveys your child will be invited to complete, you may contact me, in the health department at your child's school by calling _____. In addition, you may feel free to contact the presenters by calling (785) 864-4416 or the University of Kansas Human Subjects Protection Office at (785) 864-7429.

We are looking forward to these activities and believe that students will benefit from participation.

Appendix C

ASSENT FOR RESEARCH

Hi, our names are (Selby Conrad, Heather Hunter). We will be in your health class for the next three days presenting some information about health topics like substance abuse and relationships. We are also interested in finding out what you think about your relationships with your family, friends, classmates, etc. In addition, we are interested in learning more about your feelings and your behaviors, so we would like to ask you some questions today in order to help us better understand the views of students your age. Your answers to these questions will help us to learn more about students at your school and to help plan future activities for health classes. The information will also help doctors better understand adolescents in general and help them to develop better intervention programs.

The questions we are going to ask you are from these forms (will show the study measures). We will read you the questions and you will choose the answers you think are most appropriate for yourself. We will not look at what you are writing or the answer you are circling. It should take us about 20 to 25 minutes to go through these questions. So that other people don't know your answers or anything about you, we will not put your name or other personal information on these forms. Your answers will not be shared with your parents, teachers, or anyone else. If you have any questions you can stop and ask me at any time. After you finish the questions on these forms, we will complete an activity for the remainder of class time.

You can decide not to talk with us today and that will be okay with everyone, including your health teacher. Even if you decide to participate, you can stop at any time and that will be okay too. Also, if you feel sad or upset while answering the questions or doing the activity, you can talk to someone about how you are feeling, like teachers, counselors, or one of us. Also, if you have questions about this after we leave, you can reach us by calling (785) 864-4416. You can also call the University of Kansas Human Subjects Protection Office at (785) 864-7429 if you have any concerns about the forms or activities. If you choose not to participate, your teacher has another, regularly scheduled, health activity for you to work on.

Do you want to participate?

Appendix D

DEMOGRAPHICS QUESTIONNARIE

1. How old are you? _____

2. When is your date of birth? _____ (MM/DD/YYYY)

3. What grade are you in? _____

4. What is your gender? (Please select 1)

    a. Male
    b. Female
    c. Other

5. What is your race or ethnic background? (Please select all that apply)

    a. Asian/ Pacific Islander
    b. Black/ African-American, Non-Hispanic
    c. Hispanic/ Latino
    d. Native American/ Alaskan Native
    e. White, Non-Hispanic
    f. Other: _____

6. How tall are you in feet and inches? _____

7. How much do you weight in pounds? _____

Appendix E

STANDARDIZED DRAWING INSTRUCTIONS

Materials provided for both drawings included:

- 12 x 18 piece of 80 pound white paper
- 12 pack of Mr. Sketch makers

Drawing 1:

There are lots of different things that teens do on romantic dates. Using the materials provided, please draw a picture of your idea of the perfect date. There is no right or wrong way to do this and it is not about making the drawing look perfect. It is about the process of art making rather than the final product. Any questions?

Drawing 2:

People see drugs and alcohol in lots of different places. Using the materials provided, please draw a place where you might see drugs or alcohol. There is no right or wrong way to do this and it is not about making the drawing look perfect. It is about the process of art making rather than the final product. Any questions?

Appendix F

PERMISSION FOR USE OF THE FEATS RATING SCALES

314 Scott Avenue
Morgantown, WV 26508

PERMISSION TO USE COPYRIGHTED MATERIAL

Selby Conrad is hereby granted permission to use the FEATS Rating Scales, the FEATS Rating Sheet, and the Content Tally Sheet in her doctoral dissertation. She is also granted permission to use selected quotations (with proper citations) from the *Formal Elements Art Therapy Scale Rating Manual* by Linda Gantt and Carmello Tabone, published by Gargoyle Press.

If Ms. Conrad wishes, she may include a copy of the FEATS Rating Sheet and/or Content Tally Sheet in her dissertation provided she retains the credit line at the bottom of the form.

11 December 2008 Linda Gantt, President
Gargoyle Press

Appendix G

FEATS Rating Scale and Sample of FEATS Rating Instructions

SCALE #1 - PROMINENCE OF COLOR

How much color is used? Is the color only used to define an item or shape or is it used to color in the item or shape?

CRITERIA	RATING
This variable cannot be rated. The person did not do the drawing or the person did not use the required materials.	**0**
Color is used only to outline the forms or objects in the picture, or to make lines; none of the forms are colored in.	**1**
Color is used for outlining most of the forms or objects, but only one form or object is filled in (such as a tree trunk or a person's body). [For small elements such as an apple, it may be difficult to decide whether it is just outlined or filled in as well as outlined.]	**2**
Two or more (but not all) forms or objects are colored in.	**3**
Color is used for both outlining the forms and objects **and** filling them in.	**4**
Color is used to outline the forms and objects, to color them in, and to fill in the space around the forms (for example, a completely colored sky).	**5**

Gantt & Tabone, *FEATS Rating Manual*

#7 - Realism

Not realistic (cannot tell what was drawn)	0	1	2	3	4	5	Quite realistic

#8 - Problem-solving

No evidence of problem-solving	0	1	2	3	4	5	Reasonable solution to picking apple

#9 - Developmental Level

Two-year-old level	0	1	2	3	4	5	Adult level

#10 - Details of Objects and Environment

No details or environment	0	1	2	3	4	5	Full environment, abundant details

#11 - Line Quality

Broken, "damaged" lines	0	1	2	3	4	5	Fluid, flowing lines

#12 - Person

No person depicted	0	1	2	3	4	5	Realistic person

#13 - Rotation

Pronounced rotation	0	1	2	3	4	5	Trees & people, upright, no rotation

#14 - Perseveration

Severe	0	1	2	3	4	5	None

SCALE #1 - PROMINENCE OF COLOR

How much color is used? Is the color only used to define an item or shape or is it used to color in the item or shape?

CRITERIA	RATING
This variable cannot be rated. The person did not do the drawing or the person did not use the required materials.	**0**
Color is used only to outline the forms or objects in the picture, or to make lines; none of the forms are colored in.	**1**
Color is used for outlining most of the forms or objects, but only one form or object is filled in (such as a tree trunk or a person's body). [For small elements such as an apple, it may be difficult to decide whether it is just outlined or filled in as well as outlined.]	**2**
Two or more (but not all) forms or objects are colored in.	**3**
Color is used for both outlining the forms and objects **and** filling them in.	**4**
Color is used to outline the forms and objects, to color them in, and to fill in the space around the forms (for example, a completely colored sky).	**5**

Appendix H

TRADITIONAL HIGH SCHOOL DRAWING SAMPLES

Figure 1.1. Photographs of Participant Drawings

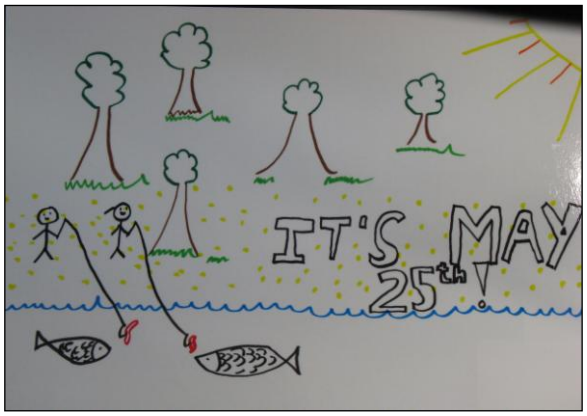
a.

b.

Figure 1.1. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 15 year old European-American female. BASC-2 screened at-risk/clinical for no scales. FEATS composite ratings included: (a) Specific Formal Elements: 5, Logic: 1, Energy: 1; (b) Specific Formal Elements: 1, Logic: 2, Energy: 1.

Figure 1.2. Photographs of Participant Drawings

a.



b.

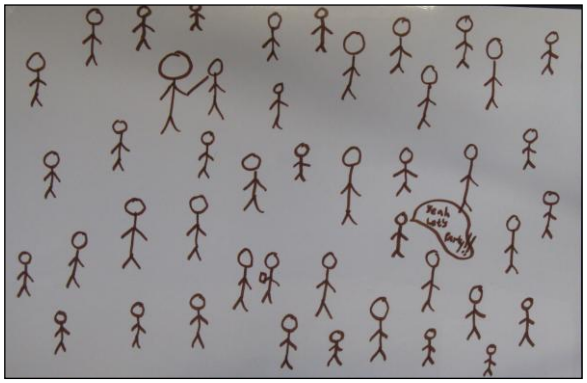


Figure 1.2. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 16 year old Asain-American male. BASC-2 screened at-risk/clinical for: Hyperactivity, Inattention, Locus of Control. FEATS composite ratings included: (a) Specific Formal Elements: 5, Logic: 1, Energy: 1; (b) Specific Formal Elements: 1, Logic: 2, Energy: 1.

Figure 1.3. Photographs of Participant Drawings

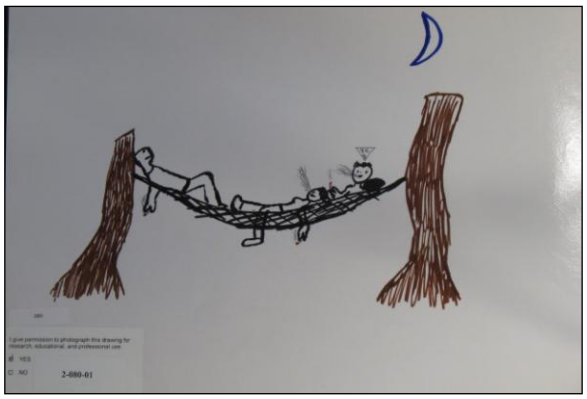
a.

b.

Figure 1.3. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 16 year old European-America female. BASC-2 screened at-risk/clinical for: Atypicality, ESI, Hyperactivity, Nattenion, Locus of Control. FEATS composite ratings included: (a) Specific Formal Elements: 3, Logic: 1, Energy: 1; (b) Specific Formal Elements: 3, Logic: 2, Energy: 1.

Figure 1.4. Photographs of Participant Drawings

a.



b.

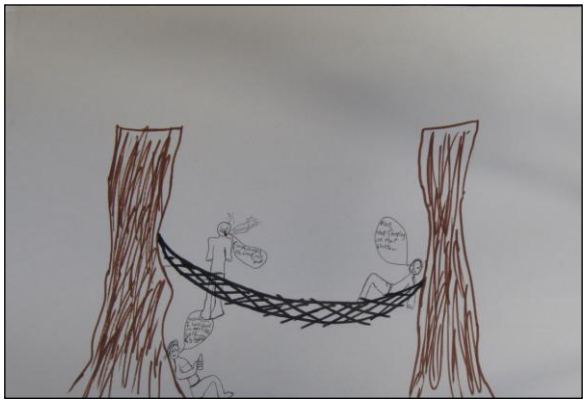


Figure 1.4. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 16 year old European-American male. BASC-2 screened at-risk/clinical for: Attention Problems, and Interpersonal Problesm. FEATS composite ratings included: (a) Specific Formal Elements: 2, Logic: 2, Energy: 0; (b) Specific Formal Elements: 3, Logic: 2, Energy: 0.

Figure 1.5. Photographs of Participant Drawings

a.

b.

Figure 1.5. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 15 year old Eurpean-American female. BASC-2 screened at-risk/clinical for: Attenion Problems, Depression, Hyperactivity, Inattention, Interalizing Problems, Locus of Control, and Somatizaiton. FEATS composite ratings included: (a) Specific Formal Elements: 0, Logic: 1, Energy: 0; (b) Specific Formal Elements: 0, Logic: 2, Energy: 0.

Figure 1.6. Photographs of Participant Drawings

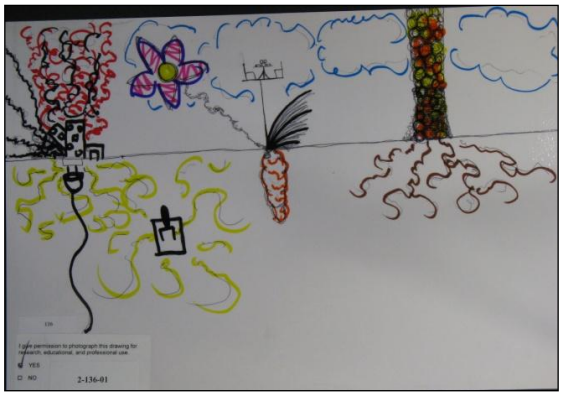
a.

b.

Figure 1.6. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 16 year old European-American male. BASC-2 screened at-risk/clinical for: Anxiety, Attenion Problems, Hyperactivity, Sense of Inadaquacy, Locus of Control, and School Problems.FEATS composite ratings included: (a) Specific Formal Elements: 5, Logic: 1, Energy: 1; (b) Specific Formal Elements: 3, Logic: 1, Energy: 1.

Figure 1.7. Photographs of Participant Drawings

a.



b.

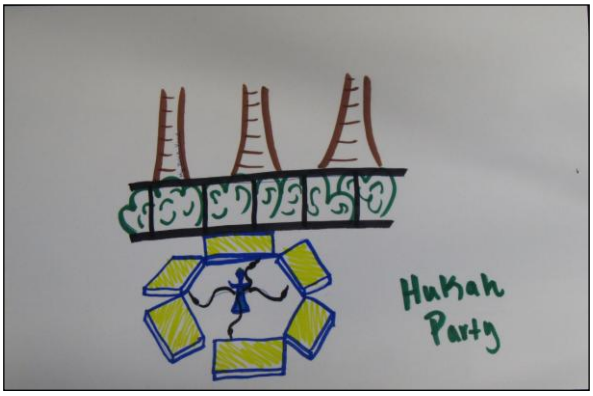


Figure 1.7. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 17 year old European-American female. BASC-2 screened at-risk/clinical for: Atypicality, Internalizing Problems, Somatization. FEATS composite ratings included: (a) Specific Formal Elements: 1, Logic: 0, Energy: 1; (b) Specific Formal Elements: 4, Logic: 2, Energy: 1.

Figure 1.9. Photographs of Participant Drawings

a.

b.

Figure 1.9. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 17 year old Eurpoean-American male. BASC-2 screened at-risk/clinical for: Hyperactivity, Inattention, Attitude to School, School Problems, Self-Reliance, and Sensation Seeking. FEATS composite ratings included: (a) Specific Formal Elements: 5, Logic: 2, Energy: 1; (b) Specific Formal Elements: 4, Logic: 1, Energy: 1.

THERAPEUTIC HIGH SCHOOL DRAWING SAMPLES

Figure 2.1. Photographs of Participant Drawings

a.

b.

Figure 2.1. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 17 year old Muli-racial male. BASC-2 did not have any positive screens for at-risk/clinical behaviors. FEATS composite ratings included: (a) Specific Formal Elements: 1, Logic: 0, Energy: 0; (b) Specific Formal Elements: 3, Logic: 2, Energy: 0.

Figure 2.2. Photographs of Participant Drawings

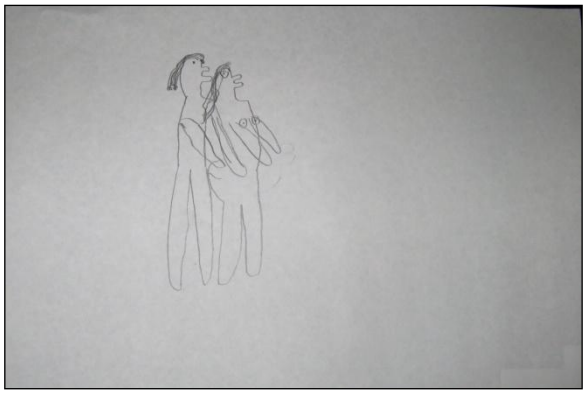
a.

b.

Figure 2.2. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 16 year old African-American male. BASC-2 screened at-risk/clinical for: School Problems, Attitude to School, Attitude to Teachers. FEATS composite ratings included: (a) Specific Formal Elements: 4, Logic: 2, Energy: 2; (b) Specific Formal Elements: 2, Logic: 2, Energy: 1.

Figure 2.3. Photographs of Participant Drawings

a.

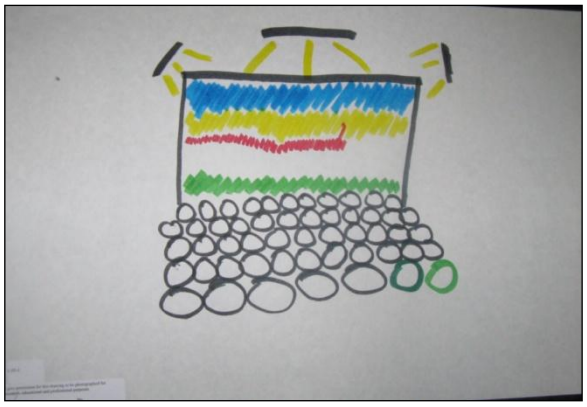


b.

Figure 2.3. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 16 year old Multi-racial male. BASC-2 screened at-risk/clinical for: Hyperactivity, Attitude to School, Self-Esteem and Self-Reliance. FEATS composite ratings included: (a) Specific Formal Elements: 2, Logic: 1, Energy: 0; (b) Specific Formal Elements: 3, Logic: 1, Energy: 1.

Figure 2.5. Photographs of Participant Drawings

a.



b.

Figure 2.5. Perfect romatic date drawing (a) and place exposed to drugs or alcohol drawing (b) from a 17 year old African-American male. BASC-2 screened at-risk/clinical for: Atypicality, Sense of Inadaquacy, Locus of Control, Self-Reliance, and Sensation Seeking. FEATS composite ratings included: (a) Specific Formal Elements: 1, Logic: 1, Energy: 0; (b) Specific Formal Elements: 4, Logic: 2, Energy: 1.

Appendix I

PEARSON CORRELATIONS BETWEEN BASC-2 SCALES

Table for Appendix I: *Correlation Coefficients between BASC-2 Scales*

Primary scales	Anx.	AttProb	Att. School	Att. Teacher	Atyp	Depression	Hyper
Anxiety	--	.13	.14*	.11	.20**	.25**	.05
Att. problems	--	--	.18*	.28**	-.04	.24**	.35**
Att to school	--	--	--	.22**	.15**	.29**	.35**
Atti to teachers	--	--	--	--	.07	.20**	. 21**
Atypicality	--	--	--	--	--	.08	.11
Depression	--	--	--	--	--	--	.06

Primary scales	Interpersonal	Locus of Control	Rel. Parent	Self-Esteem	Self-Reliance
Anxiety	.13	.26**	-.04	-.13	-.07
Att problems	.07	.24**	-.19**	-.14	-.01
Att to school	.31**	.17*	-.18*	.04	.18*
Att to teachers	.25**	.21**	-.09	.02	-.13
Atypicality	.34**	.17*	-.12	-.05	.15
Depression	.63**	.28**	-.12	-.04	-.08
Hyperactivity	.18*	.23**	-.14	.03	.13
Inter relationship	--	.45**	-.16*	-.15*	.04
Locus of control	--	--	-.18*	-.12	-.04
Rel with parents	--	--	--	.44**	.42**
Self-esteem	--	--	--	--	.23**

Primary scales	Sensation Seeking	Sense of Inad	Social Stress	Somatization
Anxiety	.01	.35**	.39**	.19**
Att problems	-.04	.32**	.32**	.19**
Att to school	.20**	.37**	.31**	.11*
Att to teachers	.01	.24**	.19**	.21**
Atypicality	.10	.34**	.29**	.20**
Depression	-.01	.47**	.51**	.19**
Hyperactivity	.20**	.11	.13	.18**
Interpersonal	.07	-.07	-.06	-.05
Locus of control	.05	.37**	.30**	.16*
Rel. w. Parents	.01	-.13	-.07	-.07
Self-esteem	.12	-.16*	-.11	-.11
Self-reliance	.32**	-.02	.05	-.08
Sen seeking	--	-.03	.04	-.04
Sense of inad	--	--	.42**	.18*
Social stress	--	--	--	.23**

Table for Appendix I continued: *Correlation Coefficients between BASC-2 Scales*

Composite scales	Inattention	Internalizing	Personal Adj	School Probs
Emotional symptoms	.28**	.74**	-.14	.18*
Inattention/hyperactivity	--	.29**	-.12	.34**
Internalizing problems	--	--	-.15*	.26**
Personal adjustment	--	--	--	-.18*

* Pearson Correlation significant at the .05 level
** Pearson Correlation significant at the .01 level

Appendix J

NON-SIGNIFICANT MANCOVA RESULTS DRAWING 1

Table 1: Appendix J

Multivariate effects: Drawing 1and BASC-2 scales: (n=170)

Variables	Pillai's trace	*F*	*df*	Error *df*	P-value
Age	.03	.86	6	328	.53
Ethnicity	.04	1.05	6	290	.40
Primary Scales					
Anxiety	.03	1.80	3	164	.15
Attention problems	.02	1.09	3	164	.35
Attitude to school	.01	.78	3	164	.51
Attitude to teachers	.01	.62	3	164	.60
Atypicality	.01	.44	3	164	.72
Depression	.01	.48	3	164	.69
Hyperactivity	.01	.60	3	164	.62
Interpersonal	.01	.65	3	164	.59
Rel. w. Parents	.02	.98	3	164	.41
Self-reliance	.01	.30	3	164	.88
Sensation seeking	.01	.08	3	164	.97
Sense of inadequacy	.01	.21	3	164	.89
Social stress	.01	.16	3	164	.92
Composite Scales					
Emotional symptoms	.01	.63	3	164	.60
Inattention/hyperactivity	.01	.26	3	164	.85
Internalizing problems	.01	.44	3	164	.72
Personal adjustment	.02	.98	3	164	.40
School Problems	.01	.37	3	164	.78

Table 2 Appendix J
Between-subject effects: Drawing 1and BASC-2 scales (n=170)

Variables	Sum squares	*df*	Mean square	*F*	P-value
Age*Specific	.09	2	.04	.80	.45
Age*Logic	1.29	2	.64	1.35	.26
Age*Energy	.56	2	.28	.73	.49
Ethn*Specific	.13	2	.07	1.20	.31
Ethn*Logic	.77	2	.40	.83	.44
Ethn*Energy	1.18	2	.60	1.47	.23
Anx*Specific	.23	1	.23	3.41	.06
Anx *Logic	.20	1	.20	.42	.52
Anx*Energy	.48	1	.48	1.26	.26
Att*Specific	.18	1	.18	3.29	.07
Att *Logic	.01	1	.01	.01	.95
Att *Energy	.15	1	.15	.40	.53
Att.Sch*Specific	.18	1	.18	3.29	.07
Att.Sch *Logic	.01	1	.01	.01	.95
Att.Sch *Energy	.15	1	.15	.40	.53
Att.Tea*Specific	.18	1	.18	3.29	.07
Att.Tea *Logic	.01	1	.01	.01	.95
Att.Tea *Energy	.15	1	.15	.40	.53
Atyp*Specific	.01	1	.01	.09	.76
Atyp *Logic	.23	1	.23	.47	.49
Atyp *Energy	.35	1	.35	.92	.34
Dep*Specific	.07	1	.07	1.37	.24
Dep *Logic	.06	1	.06	.16	.69
Dep *Energy	.03	1	.03	.08	.78
Hyper*Specific	.02	1	.02	.40	.53
Hyper*Logic	.72	1	.72	1.51	.22
Hyper *Energy	.01	1	.01	.02	.88
Inter*Specific	.08	1	.08	1.44	.23
Inter *Logic	.29	1	.29	.60	.44
Inter*Energy	2.90	1	2.90	.00	.99
Inad*Specific	.00	1	.00	.01	.94
Inad*Logic	.19	1	.19	.40	.54
Inad*Energy	.09	1	.09	.26	.61
Rel.Par*Specific	.03	1	.03	.56	.46
Rel.Par *Logic	.54	1	.54	1.13	.30
Rel.Par *Energy	.38	1	.38	.98	.32

Table 2 Appendix J continued
Between-subject effects: Drawing 1and BASC-2 scales (n=170)

SelfRel*Specific	6.07	1	6.07	.00	.99
SelfRel*Logic	.11	1	.11	.22	.64
SelfRel*Energy	.14	1	.14	.37	.54
Sen*Specific	.01	1	.01	.03	.86
Sen *Logic	.04	1	.04	.09	.77
Sen *Energy	.05	1	.05	.12	.73
SocialSt*Specific	.01	1	.01	.01	.92
SocialSt *Logic	.09	1	.09	.18	.68
SocialSt *Energy	.13	1	.13	.34	.56
Composite Scales					
ESI*Specific	.00	1	.00	.01	.95
ESI*Logic	.44	1	.44	.92	.34
ESI *Energy	.26	1	.26	.67	.42
Inatt*Specific	.03	1	.03	.60	.44
Inatt*Logic	.07	1	.07	.17	.71
Inatt *Energy	.09	1	.09	.26	.61
Intern*Specific	.06	1	.06	1.07	.30
Intern *Logic	.01	1	.01	.03	.87
Intern*Energy	.22	1	.22	.58	.45
PerAdj*Specific	.03	1	.03	.55	.46
PerAdj *Logic	.12	1	.12	.25	.62
PerAdj*Energy	.59	1	.59	1.54	.22
SchProb*Specific	2.55	1	2.55	.00	.98
SchProb *Logic	.07	1	.07	.15	.70
SchProb *Energy	.31	1	.31	.82	.37

NON-SIGNIFICANT MANCOVA RESULTS DRAWING 2

Table 3 Appendix J

Multivariate effects: Drawing 2and BASC-2 scales: (n=166)

Variables	Pillai's trace	*F*	*df*	Error *df*	P-value
Age	.02	.50	6	318	.81
Ethnicity	.08	1.93	6	286	.08
Primary Scales					
Anxiety	.02	1.26	3	160	.29
Att problems	.01	.33	3	160	.80
Att to school	.01	.78	3	160	.51
Atypicality	.01	.52	3	160	.66
Depression	.01	.42	3	160	.74
Hyperactivity	.01	.58	3	160	.63
Interpersonal	.01	.79	3	160	.50
Locus of Control	.01	.71	3	160	.55
Self-Esteem	.02	1.07	3	160	.36
Sensation seeking	.01	.27	3	160	.85
Sense of inadequacy	.01	.41	3	160	.74
Social stress	.01	.25	3	160	.88
Composite Scales					
Emotional symptoms	.01	.19	3	160	.91
Inattention/hyperactivity	.01	.64	3	160	.60
Internalizing problems	.01	.60	3	160	.62
Personal adjustment	.02	1.21	3	160	.30
School Problems	.02	1.45	3	160	.22

Table 4 Appendix J

Between-subject effects: Drawing 2and BASC-2 scales (n=166)

Variables	Sum squares	*df*	Mean square	*F*	P-value
Age*Specific	.03	2	.01	.30	.74
Age*Logic	.86	2	.43	.88	.41
Age*Energy	.14	2	.07	.23	.80
Eth*Specific	.07	2	.04	.86	.42
Eth*Logic	1.70	2	.85	1.88	.16
Eth*Energy	.64	2	.32	1.03	.36
Anx*Specific	.20	1	.20	.45	.50
Anx *Logic	.38	1	.38	.78	.38
Anx*Energy	.92	1	.92	3.15	.08
Att*Specific	.03	1	.03	.74	.39
Att *Logic	.12	1	.12	.25	.62
Att*Energy	.12	1	.12	.40	.53
Att.Sch*Specific	.18	1	.18	3.29	.07
Att.Sch *Logic	.01	1	.01	.01	.95
Att.Sch*Energy	.15	1	.15	.40	.53
Atyp*Specific	.02	1	.02	.35	.55
Atyp *Logic	.40	1	.40	.83	.36
Atyp *Energy	.07	1	.07	.25	.62
Dep*Specific	.01	1	.01	.19	.66
Dep*Logic	.24	1	.24	.48	.50
Dep*Energy	.11	1	.11	.38	.54
Hyper*Specific	.01	1	.01	.18	.67
Hyper *Logic	.29	1	.29	.60	.44
Hyper*Energy	.01	1	.01	.01	.91
Inter*Specific	.02	1	.02	.45	.51
Inter *Logic	.79	1	.79	1.63	.21
Inter*Energy	.04	1	.04	.15	.70
Inad*Specific	.05	1	.05	1.22	.27
Inad*Logic	7.06	1	7.06	.00	.99
Inad*Energy	.04	1	.04	.13	.77
Locus*Specific	.08	1	.08	1.72	.19
Locus *Logic	.07	1	.07	.14	.71
Locus *Energy	.03	1	.03	.10	.76
SelfEst*Specific	.09	1	.09	2.19	.14
SelfEst *Logic	1.04	1	1.04	2.13	.15
SelfEst *Energy	.14	1	.14	.46	.50

Table 4 Appendix J

Between-subject effects: Drawing 2and BASC-2 scales (n=166)

Variables	Sum squares	*df*	Mean square	*F*	P-value
Sen*Specific	.02	1	.02	.51	.48
Sen*Logic	.12	1	.12	.24	.63
Sen*Energy	.01	1	.01	.01	.93
SocialSt*Specific	.05	1	.05	.43	.52
SocialStr*Logic	.08	1	.08	.16	.69
SocialStr *Energy	.02	1	.02	.05	.82
Composite Scales					
ESI*Specific	.00	1	.00	.01	.95
ESI*Logic	.01	1	.01	.02	.89
ESI *Energy	.14	1	.14	.54	.50
Inatt*Specific	.07	1	.07	1.54	.22
Inatt *Logic	1.02	1	1.02	2.01	.15
Inatt*Energy	.14	1	.14	.48	.49
Intern*Specific	.06	1	.06	1.07	.30
Intern *Logic	.01	1	.01	.03	.87
Intern *Energy	.22	1	.22	.58	.45
PerAdj*Specific	.10	1	.10	2.24	.09
PerAdj *Logic	.52	1	.52	1.07	.30
PerAdj*Energy	.15	1	.15	.51	.48
SchProb*Specific	.09	1	.09	2.09	.15
SchProb *Logic	.49	1	.49	.99	.32
SchProb *Energy	.12	1	.12	.42	.52