Art Therapy: Journal of the American Art Therapy Association, 26(1) pp. 4-11 © AATA, Inc. 2009

Articles

A Normative Study of Children's Drawings: Preliminary Research Findings

Sarah P. Deaver, Norfolk, VA

Abstract

This paper describes methodology, data analysis, and initial results of a research study with the long-term goal of establishing contemporary normative data on drawings from children living in the United States. The pool of participants was composed of 316 fourth graders (mean age 9.69 years) and 151 second graders (mean age 7.56 years) who each created a Human Figure Drawing (HFD) that was scored on five modified Formal Elements Art Therapy Scales (FEATS) (Gantt & Tabone, 1998). Data were analyzed along several dimensions: age, gender, ethnic group, and mean scores on each of the five scales. Second graders included more details and used significantly more color and space than the fourth graders. Fourth graders scored significantly higher on the scale measuring congruence with Lowenfeld's stages of drawing development. There was a significant difference between boys' and girls' mean scores on one scale only, with girls using color more realistically than boys. There was no significant main effect for ethnicity (all p *values > .01).*

Introduction

In art therapy, children's drawings of people are of interest not only as a focus of the therapy process, but also because they are often used in assessment and diagnosis. In both assessment and diagnosis, art therapists rely primarily upon Victor Lowenfeld's 1947 scheme for categorizing children's artwork into developmental stages and for understanding what is "normal" or expected in children's drawings at specific ages. Lowenfeld, an art educator, conceived a stage theory of children's drawing development that corresponds approximately with Piaget's theories of child cognitive development (Lowenfeld, 1947; Malchiodi, 1998; Piaget & Inhelder, 1971). Lowenfeld's theory embodied his conviction that artwork produced by children manifested all aspects of their growth including psychological, cognitive, social, and physical development (Lowenfeld, 1947; Lowenfeld & Brittain, 1987).

In terms of specific diagnoses, Human Figure Drawings (HFDs) have been studied either through the widely used "global impression" method (Lally, 2001, p. 137) or though the matching method (Groth-Marnat, 1999). In the global impression method, assessors use their "phenomenological experience of the drawing, affective or visceral reactions to it, and relatively loosely reined impressions and associations" to interpret the meaning of an HFD (Scribner & Handler, 1987, p. 112). Using this approach, art therapists call upon their knowledge of human psychological development, psychopathology, and children's drawing development; their own in-depth art-making experience; and knowledge of art-based projective assessment techniques to arrive at an understanding of a child based not only on the child's artwork but also upon the therapists' clinical "sense" of the child. In contrast, in the matching method, assessors match specific drawing details (or combinations of details) with particular diagnoses or personality characteristics. Guides such as those written by Buck (1948), Jolles (1986), Machover (1949), and Ogden (1996) provide lists that might be used with the matching method; however, these guides are largely compendia of case studies and small researched-based studies that were often conducted from a "deficit" perspective, that is, specific drawing characteristics are seen as evidence of pathology rather than of health.

Large Scale Studies of Children's Art

The purpose of most large scale empirical studies of children's artwork has been to devise assessment methods for the detection of psychological distress resulting from emotional, physical, or sexual abuse (Peterson & Hardin, 1995), or to identify cognitive problems (Groth-Marnat, 1999). In other words, most research in this area focuses upon deviations from the norm—aspects of drawings assumed to represent maladjustment, impairment, or dis-

Editor's note: Sarah Deaver, MS, MS Ed, ATR-BC is Associate Professor and Research Director, Graduate Art Therapy Program, Eastern Virginia Medical School, Norfolk, VA. Funded by the Norfolk Foundation, the results of this study were presented in a paper at the 2005 Annual Conference of the American Art Therapy Association. Correspondence concerning this article should be sent to deaversp@evms.edu The author thanks David Elkins, MS, and Richard Handel, PhD, for their expertise with statistical analysis, and Matthew Bernier, ATR-BC, and Kay Stovall, ATR-BC, for their scholarly contributions.

turbance rather than upon what "the norm" actually is. For example, although Koppitz (1968) attempted to establish norms for both drawing development and healthy personality development, a primary aim of her work was to develop an objective assessment to identify characteristics of children's HFDs that were correlated with symptoms of emotional or behavioral disturbance. Another example is Naglieri (1988), who aimed to modernize and improve on the work of Harris (1963). Naglieri developed a system for rating three drawings (man, woman, and self) to yield a score representing the child artist's cognitive maturity; this assessment was designed as an initial screen for intelligence and achievement levels. Expanding upon his original 1988 work, Naglieri subsequently developed a variation of the Draw a Person test that screened for emotional disturbance in children (Naglieri, McNeish, & Bardos, 1991). Koppitz and Naglieri, as well as many other researchers, focused solely on measuring the presence or absence of specific aspects of the drawn human figure, such as arms, hair, nose, and so forth (Groth-Marnat, 1999), and did not attempt to measure formal artistic elements of the drawings such as color or the amount of space the drawing occupies on the paper.

Koppitz's (1968) normative study attempted to construct a "developmental test of mental maturity" (p. ix) by examining the HFDs of 1,856 children aged 5–12 to yield data about HFD characteristics associated with different age groups and genders. Through extensive test development, Koppitz was able to identity 30 "quality signs, special features, and omissions" (pp. 35–36) that she called *emotional indicators* (the elements of the drawings theorized to relate to disturbance) and a scoring system to rate their presence or absence in HFDs. Despite being labeled as "the standard for quantitative interpretation" of drawings (Peterson & Hardin, 1995, p. 24), Koppitz's findings have limited relevance today. The drawings were created over 40 years ago by mostly White middle- to high-income children in only one Midwest state and one eastern state, limiting generalizability to today's diverse population of U.S. children. Furthermore, only Koppitz herself scored the sample of normative drawings, introducing the possibility of scoring bias, and the children were a maximum of 12 years of age, which limited the scoring system's use with adolescents.

Naglieri's 1988 work created a large normative sample for developing his method for scoring children's drawings of people. Naglieri was more successful than Koppitz in establishing a diverse normative sample; his 1984 collection of drawings were from 4,468 children 5–17 years of age, across diverse geographical areas in the United States. From this group, 2,622 drawings that reflected the demographic makeup of the U.S. population were selected for the normative sample. However, since Naglieri's sample was collected, the population of the United States has changed dramatically, particularly in terms of the Hispanic population. It is likely that Naglieri's 25-year-old sample does not represent today's U.S. children.

Rationale and Purpose

A problem exists for art therapists engaged in treatment, assessment, and diagnosis who use either the global impression or matching method, or a combination of both, for understanding children's drawings: It appears that there is no large scale contemporary research that quantifies what constitutes children's "normal" or expected development as reflected in their drawings. Without such information, mental health professionals, school counselors, pediatricians and nurses, art educators, and art therapists cannot make valid inferences about children's drawings and the children who drew them.

This study addresses the lack of current normative data about children's drawings. Although the long-term goals of this research include collecting a variety of children's drawings, the study has begun with Human Figure Drawings. HFDs were chosen because of children's natural proclivity to draw people and the existence of extensive literature about HFDs (Golomb, 1974, 1992; Koppitz, 1968; Malchiodi, 1998; Naglieri, 1988). Investigators at the Eastern Virginia Medical School (EVMS) Graduate Art Therapy Program propose to collect, organize, and analyze over several years at least 5,000 American children's drawings, creating a drawing archive that will be a resource for clinicians and researchers nationwide. Such an archive will provide extensive normative data against which researchers can compare experimental samples. For example, HFDs of 7-year-olds undergoing dialysis might be compared to a random sample of HFDs in the database created by 7-year-olds to discover whether there are significant differences in the drawings along specific dimensions. In another example, comparisons could be made between drawings by sexually abused adolescent girls, drawings by physically abused adolescent girls, and drawings by a normative sample of adolescent girls from the database. Results of such comparisons would inform clinicians and others who regularly use children's drawings regarding deviations from the norm, indications for further testing, and implications for diagnosis and/or treatment.

We hypothesized that in a sample of 5,000 U.S. children's HFDs, clusters of drawing elements will characterize particular age groups. This study will address the following main research question and three sub-questions: What are the characteristics of a normative sample of American children's Human Figure Drawings, as measured by five modified Formal Elements Art Therapy Scales (FEATS) (Gantt & Tabone, 1998)? What formal elements (e.g. color, space, detail) characterize drawings from different age groups? What is the impact of age, gender, ethnicity, and geographic location on the drawings? Does the sample of drawings support Lowenfeld's 1947 stage theory of children's drawing development?

This initial report describes our research methods, data gathering procedures, preliminary findings based upon our sample of 467 HFDs, and plans for enlarging the study to multiple sites.

Method

Approval and Consent Procedures

Approval to conduct the study was granted by the EVMS Institutional Review Board. We reasoned that children in public schools would be appropriate study participants, so several local school systems were approached regarding their willingness to participate in the study, and four agreed to review the proposal. Each school system has its own research review procedures to evaluate risks and benefits of proposed studies; eventually our study was approved by the research review committees of these four school systems. Upon approval, schools and teachers were identified via emails sent from the school systems' research directors. In every case, the teachers interested in cooperating with the study were art teachers.

Consent was obtained from parents via our EVMS IRB-approved information letters and permission forms sent from the cooperating teachers. Interested parents gave consent and completed a brief demographic questionnaire regarding the age, gender, and ethnicity of the children, and sent the forms back to the teachers.

Data Collection

The EVMS Graduate Art Therapy Program contacted the interested art teachers to arrange times for data collection. In each participating classroom, data collectors (art therapists) distributed an envelope of art materials to each student. The drawing materials included a pack of 8 markers, a box of 12 oil pastels, a pencil with eraser, and gray 9" x 12" 80 lb. paper. Unlike many projective tests used in psychology in which the assessee uses a pencil and a piece of white paper, we included a variety of materials to increase the possibilities for creative expression. For example, the markers and pastels allow for pure and blended color applications; the pencil with eraser allows for changes, elaboration, and detail (Silver, 2002); and the gray paper gives visibility to the white oil pastel.

Students were told, "Use any of the drawing supplies to draw a person from head to toe. Try to draw a whole person, not a cartoon or stick figure. You will have up to 15 minutes to complete this drawing." Directions were derived from Koppitz (1968), and were standardized and read verbatim in each classroom. The 15 minute time limit was decided upon based on the investigators' collective clinical experience; 15 minutes seemed adequate for children to fully respond to the drawing directive.

After the children finished drawing, they were instructed, "Please turn the paper over and write a title." Children then placed their drawings and the supplies inside the envelopes and wrote their names on the outside. Envelopes were collected, and later, names were matched with signed permission forms. Drawings for which there was no parental permission were destroyed and not included in the study.

Using information supplied by the parents on the permission forms, the following data about each child participant were entered into the database: age, gender, ethnicity, and location where the drawing was collected.

Participants

We sought a convenience sample from local public schools. Although the ultimate goal of the study is to collect drawings from children aged 4–17 years, we began our study arbitrarily with second and fourth graders. Two different groups that were fairly close together in age were chosen in part to establish construct validity; that is, to determine if our instrument could distinguish between the two age groups. Furthermore, because we are looking for a normative sample of drawings and want to avoid a clinical sample, we did not collect from classrooms consisting solely of students receiving special education services. Other than that, there were no exclusion criteria for participation.

To date, we have 467 HFDs drawn by local public school students. There were 316 fourth grade participants (mean age 9.69 years) and 151 second grade participants (mean age 7.56 years). Among fourth graders, there were 118 boys and 198 girls. Among second graders, there were 67 boys and 84 girls. Ethnicity was distributed as follows: 297 Caucasians, 119 African Americans, 2 Native Americans, 10 Asians, 10 Hispanics, and 29 Other (usually described as "Bi-racial").

Instrumentation

The Formal Elements Art Therapy Scale (FEATS) (Gantt & Tabone, 1998) is becoming a widely used and researched art therapy assessment rating method. It is used to evaluate formal elements in the art therapy assessment task "Draw a Person Picking an Apple from a Tree" (PPAT). Formal elements in artwork include such attributes as integration of composition, realism, and line quality. The FEATS was chosen for use in the current study because it quantifies formal elements in drawings without assigning positive or negative psychological or diagnostic value to those elements. In consultation with Linda Gantt, PhD, EVMS Graduate Art Therapy Program faculty modified 5 of the 14 FEATS scales for use with Human Figure Drawings. The five scales were selected by various criteria such as ease of adaptability for use with an HFD. FEATS scales related to mental illness or organic brain disorders, such as the Rotation, Logic, and Perseveration scales, were eliminated. Wording of the five FEATS scales was changed to pertain to a Human Figure Drawing rather than to a PPAT. The resulting "FEATS/ HFD" scales are as follows: Scale I (Prominence of Color) measures how much color is used in the drawing; Scale II (Color Fit) measures how realistically color is used; Scale III (Space) measures how much of the paper space is occupied by the drawing; Scale IV (Developmental Level) measures the presence or absence of indicators of Lowenfeld's theorized stage theory of artistic development; and Scale V (Details of Objects and Environment) measures the amount of detail in the drawing. As with Gantt and Tabone (1998) in their use of the PPAT, we were not interested in the symbolism that may or

CRITERIA	RATING
This variable cannot be rated. The person did not do the drawing or the person did not use the required materials.	0
The drawing materials are used only to outline the forms or objects in the picture, or to make lines; none of the forms are colored in.	1
Drawing materials are used for outlining most of the forms or objects but only one form or object is filled in. An object that is made with just a dot (such as an eye) does not qualify as being "filled in."	2
Two or more (but not all) forms or objects are colored in.	3
Drawing materials are used for both outlining the forms and objects and filling them in.	4
Drawing materials are used to outline the forms and objects, to color them in, and to fill in the space around the forms (for example, the background is completely colored in).	5

Figure 1 Modified FEATS: Prominence of Color

CRITERIA	RATING
This variable cannot be rated. The person did not use the specified materials, or the colors are difficult or impossible to distinguish from each other.	0
The entire figure is drawn in only one color, and that color is blue, green, gray, or purple.	1
The entire figure is drawn in only one color, and that color is white, yellow, black, brown, red, orange, or pink.	2
Some colors (but not all) are used appropriately.	3
Most of the colors are used appropriately.	4
All of the colors are appropriate to the specific objects in the picture.	5

Figure 2 Modified FEATS: Color Fit

may not be inherent in the HFD, or even whether the HFD represents the self, as is theorized by many. Instead, we were interested in describing HFDs drawn by a normative sample of U.S. children by quantifying the drawings' formal elements.

Scoring

Each drawing is rated on each of the five scales, and for each scale a drawing is assigned a value from 0 to 5. Half values (.5, 1.5, 2.5, etc.) may be used in scoring. Rather than reflecting positive or negative connotations about

CRITERIA	RATING
This variable cannot be rated; or, the person did not do the drawing.	0
Less than 25% of the space on the paper is used.	1
Approximately 25% of the space is used.	2
Approximately 50% of the space is used.	3
Approximately 75% of the space is used.	4
100% of the space is used.	5

Figure 3 Modified FEATS: Space

CRITERIA	RATING
This variable cannot be rated because the individual elements cannot be identified.	0
The drawing consists solely of scribbles or masses of prefigural circles, lines, loops, and swirls.	1
The drawing has no baseline; the person's arms appear to come from the head or neck. Parts are distorted or omitted. Clothes, hair, and other details may be included.	2
There is a baseline and/or skyline. Objects may be lined up on the baseline. The body is composed of geometric shapes. Arms and/or legs show volume, and are correctly placed.	3
Objects are overlapping, and each object is a realistic size in relation to other objects. Figures appear stiff. Details such as belts and hair bows are present.	4
The drawing reflects an awareness of joints and body actions, facial expressions, and sexual characteristics. Special clothing details such as a pattern on a shirt or hats with ribbons or headbands are included.	5

Figure 4 Modified FEATS: Developmental Level

each drawing, the scores simply reflect the amount of each measured variable in each drawing. For example, a drawing with a score of 1 on the Space scale would not be considered "worse" than a drawing with a score of 4; the "1" simply indicates that less space was used, and the "4" indicates that more space was used. In other words, the numbers used in the rating scales are used to quantify and describe similarities and differences, not to assign value. Thus, the system we developed for quantifying the contents of the drawings is simple, objective, and atheoretical. See Figures 1–5 for the modified FEATS/HFD scales used in the study. All scales were adapted from those in Gantt and Tabone's FEATS rating manual (1998).

CRITERIA	RATING
This variable cannot be rated because individual items cannot be identified.	0
There is nothing but a person.	1
In addition to the person, there is a horizon line or baseline.	2
In addition to the person, there is a horizon line or baseline and/or one or two additional details such as flowers or sun, or the suggestion of interior space.	3
In addition to the person, there are a number of details such as clouds, birds, a tree, or furniture in a room.	4
In addition to the person, there are abundant and inventive details such as fences, houses with shutters, rooms with furniture, and decorative elements.	5

Figure 5
Modified FEATS: Details of Objects and Environment

Interrater Reliability

To establish interrater reliability (IRR) on the operational definitions of each scale, a faculty member not involved with the study selected 10 children's HFDs from an existing group of drawings unrelated to the research study described in this article. Using the modified FEATS scales, three faculty investigators then rated each of these 10 test HFDs. Interrater reliability was calculated using an intraclass correlation coefficient, and the results may be viewed in Table 1.

Interrater reliability values were considered acceptable, although the Scale IV value was not as strong as we would have liked. We developed a scoring manual similar to the FEATS manual (Gantt & Tabone, 1998), containing sample drawings and examples of operational definitions of each scale score. In an effort to avoid bias due to preconceived notions about children's artwork, we deliberately chose not to have an art therapist as a rater; instead, we engaged a local college student majoring in sociology. Using the manual, the rater (who was blind to the nature of the study) was trained in the use of the rating scales and then scored the same 10 test HFDs that had been scored by the faculty. The rater's scores were combined with the scores attained by the faculty, IRR was recalculated, and the results may be seen in Table 2.

The independent trained rater's scores elevated the Scale IV IRR to a more robust level. Thus, the IRR seen in Table 2 was determined to be the cut off point below which IRR would be considered insufficient. In training additional raters and others using the modified FEATS scales, the IRR seen in Table 2 must be obtained to be considered reliable. The trained rater then scored all 467 HFDs.

Results Based Upon Initial Sample of 467 HFDs

Data were analyzed along several dimensions: age, gender, ethnic group, and mean scores on each of the five scales. Regarding age, there were significant differences between second and fourth graders' mean scores on all scales except Scale II (Color Fit). On Scales I, III, and V, second graders' mean score was higher than the fourth graders'. On Scale IV, fourth graders' mean score was higher than the second graders'. Regarding gender, there was a significant difference between boys' and girls' mean scores on Scale II only, with girls scoring higher than boys. Univariate ANOVAs conducted on individual scales revealed no significant main effect for ethnicity (all p values > .01). Figures 6–9 illustrate some of these findings.

Scale I, Prominence of Color, measures how much color is used in the drawing. On this scale, second graders (M = 3.81, SD = 0.83) scored significantly higher than fourth graders (M = 3.53, SD = 0.93) (t = 3.3, p < .01). For example, in Figure 6, the fourth grader used color (grey) only for outlining forms, whereas the second grader outlined forms with color, and colored in the forms and the background.

Scale II, Color Fit, measures how realistically color is used. There were no significant differences by age on this scale (p > .01), but girls scored significantly higher than boys (t = 3.8, p < .01).

Scale III, Space, measures how much of the paper space is filled by the drawing. Second graders (M = 3.69, SD = 1.09) scored significantly higher than fourth graders (M = 3.08, SD = 1.15) on this scale (t = 5.5, p < .01). For Scale III, we use transparencies gridded with black lines; by overlapping and placing them on a drawing, the rater is able to measure how much paper space is used by the HFD. For

Table 1
Faculty Interrater Reliability

Scale I	Prominence of Color	.98
Scale II	Color Fit	.97
Scale III	Space	.99
Scale IV	Developmental Level	.81
Scale V	Details of Objects & Environment	.99

Table 2
Revised Interrater Reliability, Including Rater's Scores

Scale I	Prominence of Color	.99
Scale II	Color Fit	.98
Scale III	Space	.98
Scale IV	Developmental Level	.86
Scale V	Details of Objects & Environment	.99

Figure 6
Prominence of Color: Comparison of drawings by fourth grader (left) (score of 1) and second grader (right) (score of 5)

Figure 8
Developmental Level: Comparison of drawings by second grader (left) (score of 2.5) and fourth grader (right) (score of 4.5)

Figure 7
Space: Comparison of drawings by fourth grader (left) (score of 2) and second grader (right) (score of 4.5)

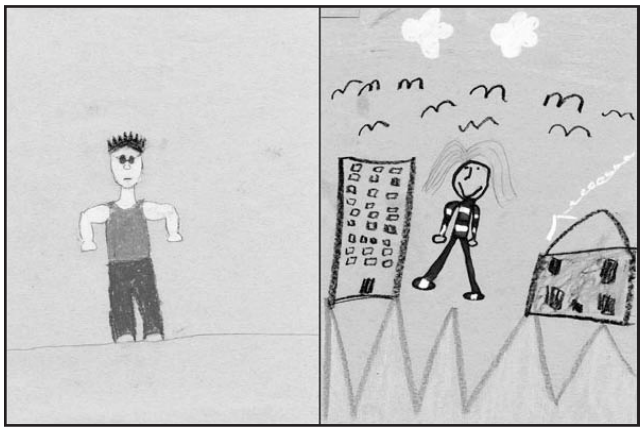

Figure 9
Details of Objects and Environment: Comparison of drawings by fourth grader (left) (score of 2) and second grader (right) (score of 4.5)

example, in Figure 7, the fourth grader's HFD filled about 25% of the paper space, whereas the second grader's HFD filled between 75% and 100% of the paper space.

Scale IV, Developmental Level, measures the presence or absence of indicators of Lowenfeld's theorized stage theory of artistic development. Fourth graders (M = 3.25, SD = 0.66) scored significantly higher than second graders (M = 3.00, SD = 0.59) on this scale (t = 4.0, p < .01). For example, in Figure 8, the second grader's drawing contains a skyline. The figure's arms appear to be coming out of the neck; interesting clothing and hair details are present. In contrast, the fourth grader's drawing depicts overlapping drawing elements; the figure is either standing in front of a walkway or is wearing a long cape. Although the figure appears stiff, there are special details such as ruffled sleeves and eyelashes.

Scale V, Details of Objects and Environment, measures the amount of detail in the drawing. Second graders (M = 3.08, SD = 1.24) scored significantly higher than fourth graders (M = 2.37, SD = 1.32) on this scale (t = 5.6, p < .01). To illustrate this result, in Figure 9, the fourth grader has drawn a person positioned on a baseline, earning a score of 2. In contrast, the second grader has included not only these elements but also a number of details such as buildings and birds, earning a score of 4.5.

Discussion

Because of our very limited sample, few conclusions can be drawn at this point. However, it is clear that, given free rein, the children in our study tended to place their human figures in an environment, despite having been asked to simply draw a person. Thus, our sample provides new information about drawing characteristics beyond earlier normative databases containing only pencil drawings of people.

The second graders' significantly higher mean scores on Scale I (Prominence of Color) and Scale III (Space) seem to reflect the younger children's less refined motor skills and characteristic use of bold color (Gardner, 1980; Malchiodi, 1998). However, the younger children's significantly higher mean score on Scale V (Details of Objects

and Environment) is contrary to widely held assumptions that as children grow older and more aware of their environment, more details appear in their artwork (Lowenfeld & Brittain, 1987; Naglieri, 1988).

The significantly higher scores that fourth graders attained on Scale IV (Developmental Level) appear to support, at least in part, Lowenfeld's stage theory of artistic development in children. These preliminary findings are congruent with those of Alter-Muri (2002), who studied 156 drawings created by children aged 3 to 11 years from five European countries and compared markers in the pictures to Lowenfeld's (1947) stages. Problems in sampling methods and the small number of subjects prevent generalization of results, but Alter-Muri concluded that there was some indication that Lowenfeld's theories may be applicable to European demographics.

Boyatzis and Albertini (2000) discussed the impact of social factors upon gender differences seen in children's drawings. They cited differences in the socialization of boys versus that of girls, and the impact of cultural traditions upon children's artwork. In our study, the girls' significantly higher mean score on Scale II (Color Fit) may reflect the contemporary Western cultural norm of socializing girls to attend to details of clothing and appearance, or it may be related to the impact of the media upon girls' self-awareness and self-appraisal (Malchiodi, 1998; Pipher, 1994). Regarding the amount of color used (Scale I, Prominence of Color), our results contrasted with those found by Milne and Greenway (1999). These two researchers studied the use of color in a variety of drawings created by a clinical sample of 61 boys and girls aged 4 to 14 years, and discovered that the boys' use of color decreased with age, depending upon the subject of the drawing. In our study there was no significant difference by gender on the Prominence of Color scale, but it is likely that such a difference may occur with a larger, more diverse sample.

The fact that ethnic group identity was not a significant variable in this sample is congruent with Naglieri's (1988) findings, and lends some support to the generalizibility of artistic developmental stages across ethnic groups. In addition, although we have entered all of the titles given to the drawings into the database, and have occasionally used them to shed light on the content of the drawings, we have not at this point considered titles in the analysis process. They constitute a rich source of data for future study.

Limitations

Threats to Internal Validity

Although the modified FEATS/HFD scales used in this study appear reliable and valid to a certain extent, the measure is relatively untested with children and may have weaknesses unknown at this time that would compromise the study's internal validity. For example, in our initial sample, the measure did not find significant differences between second and fourth graders' drawings regarding the realistic use of color (Scale II, Color Fit). Because it is unknown whether a difference actually exists, the validity of Scale II with 7- to 9-year-olds is not yet clearly established. We believe that analysis of a larger sample in the future will establish construct validity.

One way to strengthen the instrumentation would be through establishing criterion validity of the modified FEATS/HFD scales. To do this, correlations would be calculated between two sets of scores on the five scales used in this study (color fit, prominence of color, space, developmental level, and details of objects and environment): scores on the 10 test HFDs rated with the FEATS/HFD, and scores on 10 PPAT drawings rated with the original FEATS. Correlations of .80 and above would be considered evidence of acceptable criterion validity of the FEATS/HFD scales.

In addition, the time of data collection during the academic year may constitute a threat to internal validity. For example, collecting data from classrooms on the eve of winter break likely would have an impact upon the content of the drawings. This might be addressed either through limiting data collection to specific periods of time, or by ensuring that data collection is ongoing throughout the academic year.

Threats to External Validity

Generalization of results to the larger population of children is compromised by our small, non-random volunteer sample from one geographic area. Furthermore, despite the exclusion of classrooms dedicated to special education students, some students with special needs probably participated. Thus it is possible (although unlikely, due to the small percentage of public school students who receive special education services) that the sample was skewed. Because we did not collect drawings from special education classrooms, we assumed that drawings by any child participants receiving special education services would either be statistical outliers or would score similarly to those by children who were not receiving special education services. Only when our database is large enough to represent a diverse sample of children who do not have special needs will we be able to use it to compare to populations of children with special needs.

In addition, an ecological threat exists in that the sample does not contain sufficient participants who reflect the ethnic and geographic diversity of the U.S. population. To address these limitations, we plan to gather drawings created by children in kindergarten through 12th grades in school systems across the country, with the goal of 5,000 drawings.

Regarding socioeconomic status, we plan to add zip codes to the data entered for each drawing in our database. In collecting zip codes of participants' schools, we may be able to discern whether the socioeconomic statuses of children in our sample reflect that of the U.S. population as a whole. We plan to collaborate with other graduate art therapy programs on this study and hope that the resulting geographical diversity will improve the geographic, socioeconomic, and ethnic diversity of the database.

Lack of control over data collectors working with collaborating institutions constitutes a threat to the integrity of the study. This might be addressed through annual

training sessions and work groups with representatives from collaborating institutions at the American Art Therapy Association conferences, as well as though periodic email and telephone contact.

A final concern is that we trained only one rater who rated all of our drawings to date. Clearly, to reach our goal of 5,000 drawings, we will need to train additional raters. Furthermore, we will need to employ a method to ensure the accuracy of the scores on the drawings. Thus, as we progress through the study, for every 100 drawings, we will need to have a random sample of 10 of the drawings double-rated by two trained raters, and interrater reliability recalculated. As long as the IRR continues to meet the levels seen in Table 2, those two raters would then each rate half of the remaining 90 drawings.

Conclusion

Our study suggests that the FEATS/HFD scoring system that we developed has promise as a tool for developing a large scale normative database of children's Human Figure Drawings. The approach we have taken has value because our scoring system is atheoretical and objective. Furthermore, the procedures and art materials we chose to include allowed the children freedom to freely express themselves using color, and thus we have added new knowledge about children's drawings to the field of art therapy. However, our plans for expansion to multiple sites are complex and involve recruiting collaborators at various universities, approval of the research protocol by multiple Institutional Review Boards, training multiple raters, and setting up a more extensive database than we now use. Our hope is that despite these challenges, we will eventually be able to realize our goal of establishing a database of children's drawings that will be a valuable resource for art therapy educators, researchers, and clinicians.

References

Alter-Muri, S. (2002). Viktor Lowenfeld revisited: A review of Lowenfeld's preschematic, schematic, and gang age stages. *American Journal of Art Therapy, 40,* 170–192.

Boyatzis, C., & Albertini, G. (2000, Winter). A naturalistic observation of children drawing: Peer collaboration processes and influences in children's art. In C. Boyatzis & M. Watson (Eds.), Symbolic and social constraints on the development of children's artistic style. *New Directions for Child and Adolescent Development, 90,* 31–48.

Buck, J. N. (1948). The H-T-P Test. *Journal of Clinical Psychology, 4,* 151–159.

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

Gardner, H. (1980). *Artful scribbles: The significance of children's drawings.* New York: Basic Books.

Golomb, C. (1974). *Young children's sculpture and drawing.* Cambridge, MA: Harvard University Press.

Golomb, C. (1992). *The child's creation of a pictorial world.* Berkeley, CA: University of California Press.

Groth-Marnat, G. (1999). *Handbook of psychological assessment.* (3rd ed.). New York: Wiley.

Harris, D. (1963). *Children's drawings as measures of intellectual maturity.* New York: Harcourt, Brace, and World.

Jolles, I. (1986). *A catalog for the qualitative interpretation of the House-Tree-Person (HTP).* Los Angeles: Western Psychological Services.

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

Lally, S. (2001). Should Human Figure Drawings be admitted to court? *Journal of Personality Assessment, 76*(1), 135–149.

Lowenfeld, V. (1947). *Creative and mental growth.* New York: Macmillan.

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

Machover, K. (1949). *Personality projection in the drawing of the human figure: A method of personality investigation.* Springfield, IL: Charles C Thomas.

Malchiodi, C. (1998). *Understanding children's drawings.* New York: Guilford Press.

Milne, L., & Greenway, P. (1999). Color in children's drawings: The influence of age and gender. *The Arts in Psychotherapy, 26*(4), 261–261.

Naglieri, J. (1988). *Draw A Person: A quantitative scoring system.* San Antonio, TX: The Psychological Corporation.

Naglieri, J., McNeish, T., & Bardos, A. (1991). *Draw A Person: Screening Procedure for Emotional Disturbance examiner's manual.* Austin, TX: PRO-ED.

Ogden, D. (1996). *Psychodiagnostics and personality assessment: A handbook.* (2nd ed.). Los Angeles: Western Psychological Services.

Peterson, L. W., & Hardin, M. (1995). *Children in distress: A guide for screening children's art.* New York: Norton.

Piaget, J., & Inhelder, B. (1971). *Mental imagery in the child.* New York: Basic Books.

Pipher, M. (1994). *Reviving Ophelia.* New York: Ballantine.

Scribner, C., & Handler, L. (1987). The interpreter's personality in Draw-a-Person interpretation: A study of interpersonal style. *Journal of Personality Assessment, 51,* 112–122.

Silver, R. (2002). *Three art assessments.* New York: Brunner-Routledge.