

The Draw-A-Person Test: An Indicator of Children's Cognitive and Socioemotional Adaptation?

J. TER LAAK

Department of Developmental Psychology

M. DE GOEDE

Department of Methods and Statistics

A. ALEVA

P. VAN RIJSWIJK

Department of Developmental Psychology

Utrecht University, The Netherlands

ABSTRACT. The authors examined aspects of reliability and validity of the Goodenough–Harris Draw-A-Person Test (DAP; D. B. Harris, 1963). The participants were 115 seven- to nine-year-old students attending regular or special education schools. Three judges, with a modest degree of training similar to that found among practicing clinicians, rated the students' human figure drawings on developmental and personality variables. The authors found that counting details and determining developmental level in the DAP test could be carried out reliably by judges with limited experience. However, the reliability of judgments of children's social and emotional development and personality was insufficient. Older students and students attending regular schools received significantly higher scores than did younger students or students attending special education schools. The authors found that the success of the DAP test as an indicator of cognitive level, socioemotional development, and personality is limited when global judgments are used. The authors concluded that more specific, reliable, valid, and useful scoring systems are needed for the DAP test.

Key words: adaptation, children, human figure drawings

THE DRAW-A-MAN TEST was introduced by Goodenough (1926) to assess children's mental development. She standardized the scoring by using the drawings of 2,300 four- to 10-year-old children. The test was revised and renamed by Machover (1949) as Draw-A-Person (DAP). In 1961, its goal was redefined to assess psychological maturity. The scoring was sufficiently reliable and correlated well with intelligence test scores.

Machover (1949) found that the DAP test could be used as an indicator of a child's conceptual level, and he introduced the instrument as a test that could be

used to infer personality characteristics. The drawings were considered to be either a direct or a symbolic expression of the child's personality, or both. In early perusals of the DAP test, Swensen (1957, 1968) criticized the lack of empirical support of Machover's use of the DAP test to make inferences about personality and concluded that the DAP test contained a rough indicator of the level of adjustment.

The DAP test thus became a controversial tool. Nevertheless, the instrument is among the top 10 tools used by practitioners (Cummings, 1986; Yama, 1990). One reason for this popularity is that researchers have found that scoring the developmental level of human figure drawings and counting the number of details portrayed provide successful indexes of children's levels of cognitive development and psychometric intelligence. This scoring emphasizes norms and quantification methods. Scribner and Handler (1987) called this the *thinking approach*.

The results of some empirical research have supported the claim that the DAP test measures children's levels of cognitive development and intelligence. Chappell and Steitz (1993) found a significant relationship between Piagetian stages of development and levels of human figure drawing. Cox and Howarth (1989) reported significant differences in drawing by normal children and by children with developmental delays lagging 4 years behind normal. Fabry and Bertinetti (1990) found substantial and significant correlations between the number of details in DAP test results and WISC-R (Wechsler, 1989) Performance IQ (.69), Verbal IQ (.45), and General IQ (.62), respectively, in 6- to 10-year-old children with behavioral and emotional problems. Naglieri, McNeish, and Bardos (1991) applied this thinking approach to assess disturbances in children and developed a 55-item scoring procedure. Naglieri and Pfeiffer (1992) also used DAP test scoring to assess emotional disturbance in 7- to 17-year-old normal students and in those with conduct and oppositional disorders who attended a psychiatric day-care treatment center. The clinical group produced more signs associated with emotional disturbance than did the normal group. McNeish and Naglieri (1993) used the DAP procedure to screen for emotional disturbance in normal and seriously emotionally disturbed 11-year-old children. They reported a significant difference in emotional adaptation between the two groups.

In addition to using the thinking approach to measure the levels of development and intelligence, researchers have used the DAP test to assess personality (Koppitz, 1968, 1984). This *global approach* emphasizes the evaluation of specific characteristics of the content and quality of drawings. Judges inferred the children's social and emotional adaptation from their drawings. This global approach was used to screen for social and emotional problems in seriously disturbed children because these children were difficult to question and interview.

The authors thank the judges, A. Dierckxsens, P. van Rijswijk, and G. van Rijswijk, for their assistance.

Address correspondence to M. de Goede, Heidelberglaan 2, 3508 TC Utrecht, The Netherlands; m.degoede@fss.uu.nl (e-mail).

Some attempts failed, whereas others were successful. For example, Kahill (1984) showed that the specific contents of drawings did not correlate with specific personality disorders. Groth-Marnat and Roberts (1998) used quantitative scoring of human figures and other drawings, but were not able to predict self-esteem in 18- to 47-year-old students enrolled in a college-level psychology course. However, Yama (1990) found the evaluation of the level of bizarreness in drawings of a group of seriously disturbed young Vietnamese refugees to be related to their level of adaptation (i.e., the number of changes they had had from one adoptive family to another). Tharinger and Stark (1990) reported substantial correlations between the evaluation of drawings and self-esteem and of the level of family functioning.

Despite these miscellaneous results since the publication of Machover's book (1949), the DAP test is still dominated by the global approach. Researchers have assumed that children's drawings reflect their basic personality and adaptation. By focusing on outstanding features and signs in children's human figure drawings, psychologists have tried to assess disturbance in children. In studies by Swensen (1957); Roback (1968); Chapman and Chapman (1969); Motta, Little, and Tobin (1993); and Smith and Dumont (1995), this method of inferring children's characteristics did not seem to be valid because individual parts of drawings did not prove to have any specific meaning.

Researchers apply the two approaches to the DAP test (the thinking approach and the global approach) to two domains—developmental level (intelligence level) and personality level (emotional adaptation), respectively (see Table 1). The use of the thinking approach for the DAP test has been successful in distinguishing cognitive and developmental levels. Use of Naglieri et al.'s (1991) thinking approach was successful in distinguishing seriously disturbed children from normal children but not, for example, in distinguishing relevant differences between normal and delinquent groups (Sullivan, 1998) or within a group of normal students who varied in levels of self-esteem (Groth-Marnat & Roberts, 1998).

The results of our perusal of the literature suggested that the issue of reliability and validity of human figure drawings as an indicator of cognition and

TABLE 1. Validity of Thinking and Global Approaches in Assessing Developmental Level (IQ) and Personality and Emotional Disturbances

Variable	Thinking approach	Global approach
Developmental level	high level of validity ^a	miscellaneous results
Personality and emotional adaptation	moderate validity	negligible level of validity

^aLevel of reliability and validity (Cohen, 1977; Rosenthal, 2000).

socioemotional adaptation is still a relevant one (Groth-Marnat & Roberts, 1998; Smith & Dumont, 1995).

The reliability, validity, and usefulness of the DAP test can be determined with reference to various factors assessed in the drawings: (a) the type of approaches (thinking, global), (b) the groups of participants, (c) the judges, and (d) the domains (i.e., cognitive developmental level, personality, socioemotional adaptation). In this study, we concentrated on the issues of reliability, validity, and usefulness of human figure drawings by children. We asked the judges to rate drawings for several variables (e.g., developmental level, cognitive development, socioemotional adaptation, drawing skills) for 7- to 9-year-old students attending regular education schools as well as students attending special education schools.

Our aim in this study was to examine aspects of reliability and validity of some developmental variables in the children's human figure drawings. We also designed this study to collect empirical data as a basis for making suggestions to professionals who use this controversial diagnostic tool. Our research question, with respect to reliability, was: How reliable are the thinking and global approaches in the scoring of children's human figure drawings (i.e., what are the correlations of estimates made by independent judges?)

With respect to the validity of the DAP test, we considered aspects of construct and criterion validity. We formulated the following four research questions:

1. Is there one dimension underlying children's human figure drawings, or should more dimensions be distinguished?
2. Are differences in psychological development between various groups of children reflected in the judgments of their human figure drawings as made by independent judges? This is the known groups criterion.
3. What is the relationship between the scores as assessed by independent judges and the scores of the same variables made by another comparable measurement tool (i.e., teacher's judgments of their students)?
4. Is it possible to predict children's school achievement based on their human figure drawings?

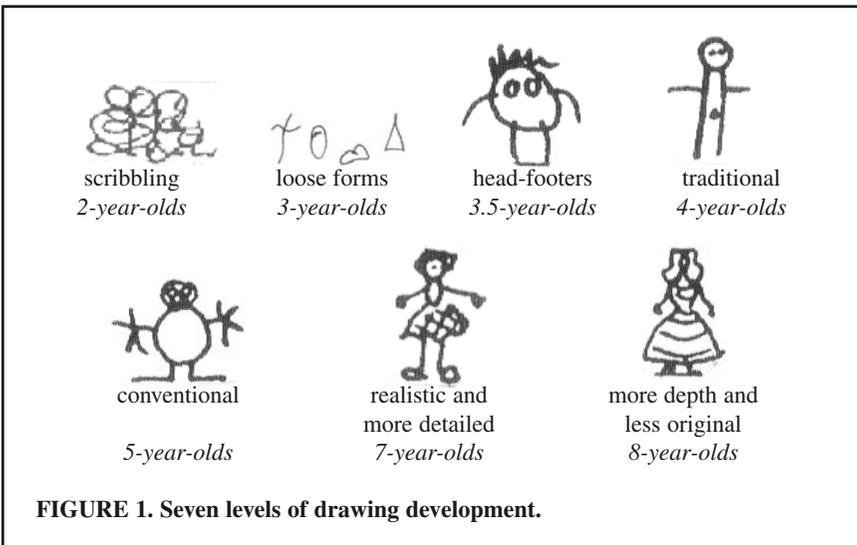
We formulated hypotheses as answers to these questions and used terms such as high, moderate, and low reliability. We considered scores to have high reliability if the correlations were .80 or higher. We considered correlations between .60 and .80 to be moderate, and those with correlations of .60 or less to be of low reliability (Cohen, 1977; Nunnally & Bernstein, 1994).

Reliability. Judging the developmental level and counting the number of details in a drawing require accuracy. These tasks can be defined objectively and do not require much inference. These results lead to two hypotheses: Judging children's human figure drawings with respect to their developmental level is highly reliable (Hypothesis 1), and scoring the number of details in children's human figure drawings is reliable (Hypothesis 2). However, judging children's level of cog-

nitive, social, and emotional development on the basis of general characteristics of their human figure drawings requires inference. Judging cognitive, social, and emotional development as well as drawing skill is moderately reliable (Hypothesis 3). Measuring children’s personality characteristics (e.g., impulsiveness, self-image) on the basis of their human figure drawings goes a step further in drawing inferences, which will result in a low reliability. This result leads to a fourth hypothesis: The reliability of assessing children’s impulsiveness and self-image from their human figure drawings is low (Hypothesis 4).

Construct validity. Clinicians often distinguish many different factors and behaviors in describing their clients. These factors and behaviors usually show high correlations (see Garb, 1998). For example, there is a high correlation between the scoring of developmental level and number of details; a scribble (low) contains fewer details than does an identifiable human figure with depth (high; see Figure 1). Elaborating on the assumption of one important undifferentiated construct, we suggested a fifth hypothesis: All the variables that can be derived from children’s human figure drawings refer to one construct (i.e., level of adjustment; Hypothesis 5; Swensen, 1957).

An important indication of the validity, considering the known groups criterion, would be that categories of children who differ in intelligence and school achievement should receive different scores for related relevant dimensions or variables (see Cronbach & Meehl, 1955). There are evident differences between students attending regular schools and those attending special education schools, and within these two categories of students, there are also differences between first and third graders. This implied a sixth hypothesis: Differences in children’s



age and type of schooling are reflected in the judgments of human figure drawings (Hypothesis 6; see Shangnessy, Zechmeister, & Zechmeister, 2000).

Predictive validity. Teachers can estimate developmental variables and level of school achievement for their students, and these results can be considered as a criterion for the judges. So, judges' assessment of human figure drawings should correlate with the teachers' assessment of the same variables. Our assumption was that teachers' observations could be considered as valid measuring criteria. Predictions of people's personality are moderately accurate (see Grove, Zald, Lebow, Snitz, & Nelson, 2000). To a certain extent, this result also holds for teachers (ter Laak, de Goede, & Brugman, 2001). This conclusion led to our first exploratory question: How well do the assessments made by independent judges using the DAP test correlate with teachers' assessments?

Researchers have reported inconsistent results with regard to the predictive validity of human figure drawings. Another exploratory question was: Is there a correlation between children's academic achievement and independent judgments of their DAP test results?

Method

Participants

Participants included children from two regular education and two special education schools in Gouda, a city of approximately 100,000 inhabitants in the central area of The Netherlands. The area and city are considered to be typically Dutch. The four schools all served the same socioeconomic status groups—lower middle class to middle class. In the two regular schools, all the children spoke fluent Dutch and 12% had Turkish or Moroccan parents, whereas in the special education schools 6% of the students had Turkish or Moroccan parents, which is characteristic for this type of Dutch school. We asked two special education schools to participate to get a sufficient number of participants, because the number of children in a special education school class is only half that of a regular school class (28 to 32 students). The special education schools are for children with an intelligence level in the normal range (no less than 90) who have learning difficulties, mild behavioral problems, or both. It is difficult to involve such children in activities with classes of approximately 30 students. However, current Dutch education policy is to try and keep these children in regular education schools but to provide special help for them. In total, 115 first- and third-grade students (aged 7 to 9 years) participated in our study (see Table 2).

The students attending the special education schools were, on average, 4 months older than the students in the regular education schools. There were more boys than girls in these samples, from both the regular and the special education schools. The age difference was accidental, but the second difference, in

TABLE 2. Distribution of the Participants by Type of Education, Gender, and Age, in Absolute Numbers

Gender	Regular education				Special education			
	Age		<i>n</i>	%	Age		<i>n</i>	%
	7 years	9 years			7 years	9 years		
Boys	21	19	40	59	15	20	35	75
Girls	11	17	28	41	4	8	12	25
Total	32 (48%)	36 (52%)	68	100	19 (40%)	28 (60%)	47	100

the distribution of students by gender, reflects the actual situation—more boys than girls attend special education classes.

Measurements

Developmental level. We distinguished seven developmental levels, which were described verbally and illustrated with examples. Level 1 was a scribble, and Level 7 showed a detailed conventional human figure with depth (see Figure 1). The levels comprised an ordinal scale, and we borrowed them from the theoretical and empirical work summarized by Barrett and Eames (1996), Cox and Parkin (1986), Thomas and Silk (1990), and Meykens and Cluckers (1996). Empirical analysis of children's drawings between 1 and 10 years of age showed a regular pattern (Figure 1). There are also theories about perceptual development that support this regular pattern from simple designs to more detailed and complex figures and from the use of two to three dimensions (depth; Fischer, 1980).

Number of details. The judges had to decide whether 30 details of body, head, clothes, hair, and space proportions were present. In addition, they had to evaluate the quality of the drawing. This procedure was borrowed (a) from the DAP test (Harris, 1963), (b) from a Dutch school readiness test (Mönks, Rost, & Coffie, 1972), and (c) from a toddlers' intelligence test containing human figure drawings as a subtest (Baarda, 1978).

Level of cognitive development (range 1 to 9). The judges did not know the ages or the school achievement levels of the students. We asked the judges to grade the drawings on a scale of 1 to 9 (marks regularly used in these schools) for the students' level of cognitive development after inspecting the drawings. We asked them to infer the level of cognitive development from the drawings in the same way they would from children's language utterances. A higher level would

accompany more elements, more integrated elements, and the expression of a real human figure.

Level of social development (range 1 to 9). The judges inferred this level from the posture and the openness of the human figure. Was the figure open and accessible, and to what extent? Did it search for contact or was it introverted?

Level of social–emotional development (range 1 to 9). The judges inferred this result from the form of the human figures. Was the figure firmly positioned? Did the figure stand firmly with its feet on the ground? How confident was the facial expression?

Level of drawing skill (range 1 to 9). We instructed the judges to recognize that there would be huge individual differences in drawing skills because of differences in talent and training. We asked the judges to give a mark for the drawing skill of the student. How skilled or talented was the drawer?

Level of impulsiveness (range 1 to 9). The judges assessed the level of impulsiveness from the lines in the drawing, the size of the drawing, and the proportion of the A4-sized paper (29.5 cm × 21 cm) used for the drawing. Did the drawing go over the edge of the paper? Was the drawing an appropriate size for the piece of paper? Were the lines drawn straight, firmly, and consciously chosen, or were they the result of (random) heavy scratching?

Level of self-image (range 1 to 9). The judges inferred a positive self-image from a human figure that stood firmly on its feet with an impression of self-confidence. Did the human figure give the impression of a self-confident person?

Procedure

Teachers asked their students to draw a human figure during school time. This activity is a regular task for students attending these schools. Three female university graduate students from the departments of developmental psychology (Judge I), pedagogy (Judge II), and history of arts (Judge III) judged the drawings independently. The drawings had no student name and gave no information on age or type of school. Judge I had 6 months of experience judging children's drawings but had no formal training. Judges II and III had no experience but had received a short training (4 hr each day for 4 days—16 hr in total) in how to judge human figure drawings, which was taught by an experienced school psychologist, who was their supervisor. We chose this procedure of judging and estimating because it resembled how several clinicians using the DAP test in a practical setting are trained. There is no explicit or formal training in judging drawings. The task is learned by practicing the skills and by discussing inferences with a supervisor.

Moreover, people with little or no experience seemed to be able to judge figure drawings just as accurately as experienced clinical psychologists (see Scribner & Handler, 1987). The judges first scored the drawing's developmental level, then they scored the number of details, and last they scored the other variables.

Teachers gave the drawings to the judges. The judges asked the teachers to rate their students' cognitive development, social and emotional development, drawing skill, and school achievement on a Likert-type scale of 1 to 9: *very poor* (1), *average* (5), and *excellent* (9). The judges estimated one developmental variable per day for 8 days to prevent interference from previous judgments.

Analysis

We used one-way analysis of variance (ANOVA) and Pearson product-moment correlations to compute the interjudge differences and reliability. Computing average differences between judges was relevant because we used different levels as diagnostic labels for developmental stages (e.g., in the statement, "This child is in the scribbling phase"). We conducted principal component analysis (PCA) with varimax rotation using the sums of the three judges' scores for each of the eight variables. We conducted the same analysis on the teachers' scores for cognitive, emotional, and social development and for drawing skill. We conducted 2 (age group) \times 2 (type of education) ANOVAs to determine the known groups criterion validity of the DAP test. Because more boys than girls attend special education classes, we addressed the possible relationship of gender to the dependent variable. However, this effect proved limited, and therefore we did not include the results in the analysis. Boys scored lower than did girls only on cognitive development and drawing skill. We computed Pearson product-moment correlations between academic achievement and the variables assessed by the three judges and the four teachers. We conducted a hierarchical regression analysis with academic achievement as the dependent variable and the variables assessed by the judges as predictors.

Results

Reliability. Hypotheses 1 and 2 stated that judging children's human figure drawings with respect to their developmental level is highly reliable and that scoring the number of details in children's human figure drawings is reliable. The mean interjudge correlations were .79 and .87. We therefore accepted Hypotheses 1 and 2. The interjudge reliability scores for cognitive, social, and emotional development and for drawing skill varied from .35 to .65. The reliability of the cognitive development was moderate (.64), and the evaluation of drawing skill approached a moderate to low level (.58). The other correlations were low. We accepted Hypotheses 3 and 4; that is, we predicted correctly that the correlations would be moderate to low (see Table 3).

Hypothesis 5 stated that the judgments reflected only one dimension, interpreted as level of adjustment. PCA with varimax rotation on the eight variables resulted in a two-factor solution for the judges. The results seemed to differentiate between two factors: (a) cognitive development and (b) socioemotional adaptation. Therefore, we rejected Hypothesis 5.

According to Hypothesis 6, first graders should have scored lower on the eight variables than did third graders, and students in regular education classes should have scored higher than did students in special education classes (see Table 4). With the exception of emotional development and impulsiveness, the differences in the variables between the two age groups, as well as between the regular and special education school groups, were significant at the $p < .05$ level, and in the predicted direction. So, we accepted Hypothesis 6.

Correlations of the judgments by the judges and teachers (see the first exploratory question) indicated agreement (Funder, 2001). In this study, we considered the teachers' judgments to be the criterion for the estimations by the judges. The correlations of cognitive, social, and emotional development scores by judges and teachers were significant ($p < .05$; $N = 115$) but modest ($r_s = .35, .23, \text{ and } .29$, respectively). The estimation of drawing skill by judges and teachers correlated significantly ($r = .59, p < .05$; $N = 115$) but moderately.

TABLE 3. Mean Scores of the Three Judges on Eight Developmental Variables Based on Human Figure Drawings by All Children, Plus the Mean Correlations (Product-Moment Correlations) for the Pairs of Judges (r_{I-II} , r_{I-III} , r_{II-III}) on These Variables

Developmental variable	Mean scores of judges				Intercorrelations of the mean scores			
	I	II	III	F^a	r_{I-II}	r_{I-III}	r_{II-III}	Mean r_{I-III}
1 Developmental level	5.2	5.0	5.2	6.6	.78	.83	.76	.79
2 Number of details	16.6	19.0	16.0	6.4	.84	.90	.89	.87
3 Cognitive development	4.5	5.2	4.9	9.2	.64	.63	.67	.65
4 Social development	4.9	6.3	4.9	5.0	.48	.51	.53	.51
5 Emotional development	5.0	6.4	4.7	8.1	.39	.50	.46	.45
6 Drawing skill	4.0	5.2	4.9	4.3	.58	.64	.72	.65
7 Impulsiveness	4.8	5.8	5.3	2.3	.56	.13	.35	.35
8 Self-image	4.8	6.3	5.1	5.7	.39	.47	.39	.41

^a $df = 2.113$ for the variables 1 to 8.

TABLE 4. Differences in Eight Development Variables Derived from Human Figure Drawings Between Students' Age and School Type

Variable	Age <i>F</i> (1, 111)	Type of school <i>F</i> (1, 111)
1 Scoring of developmental level	19.13	9.93
2 Scoring of number of details	22.52	9.56
3 Evaluation of level of cognitive development	23.22	17.58
4 Interpretation and evaluation of level of emotional development	12.97	9.06
5 Interpretation and evaluation of level of social development	5.39	6.73
6 Evaluation of drawing skill	22.78	8.74
7 Interpretation of impulsiveness	14.60	1.02 <i>ns</i>
8 Interpretation of self-image	17.02	9.21

Note. All *F*s are significant at $p < .05$, except the type of school effect for impulsiveness (Variable 7).

The teachers judged five developmental variables (see Table 5). These variables clearly reflected one dimension, which could be labeled as the level of adjustment to school. The teachers' estimation of the cognitive development predicted the children's school achievement almost perfectly ($\beta = .92$). The other variables included in this analysis (social and emotional development, drawing skill) were of no importance (all β s $< .08$).

The judges differed significantly in their average assessment of the levels of human figure drawings. Judge II gave systematically higher scores than did the other two judges on (a) the number of details; (b) cognitive, social, and emotional development; (c) impulsiveness; and (d) self-image. The only exception was the level of drawing skill, on which there was better agreement. The judges seemed to differ more among one another than to have common ground. The assumed convergent validity turned out to be a discrepant validity (Shangnessy et al., 2000), so this result answers our first exploratory question.

Predictive validity. In our second exploratory question, we wondered whether it was possible to predict children's academic achievements from independent judgments of their human figure drawings. The results of multiple regression analysis suggested that social and emotional development and drawing skill did not play any part in predicting academic achievement. Cognitive development, however, could not be neglected ($\beta = .32$), but it was not a particularly strong factor. We excluded the other variables (emotional development, social development, impulsiveness, self-image, drawing skill) from the hierarchical regression analysis because of the low beta values (all β s $< .10$). The response to this exploratory ques-

TABLE 5. Principal Components Analysis of Eight Variables Estimated by the Judges and of the Variables Assessed by the Teachers

Variable	Judges		Teachers factor 69%
	Factor I 45%	Factor II 34%	
1 Drawing skill	.90		.70
2 Number of details	.90		^a
3 Cognitive development	.83		.90
4 Developmental level	.76		^a
5 Impulsiveness	.70		^a
6 Social development		.91	.85
7 Emotional development		.87	.85
8 Self-image		.84	^a
9 School achievement	^a	^a	.89

^aNot included in analysis; only loadings > .50 are presented.

tion was clear. Independent judges were not adequately able to predict children's school achievement solely on the basis of scoring their human figure drawings.

Discussion

Researchers have suggested that the thinking approach to human figure drawings offers useful indicators of children's cognitive development and intelligence, whereas the global approach is insufficient in predicting children's cognitive, social, and emotional adaptation. The literature results have suggested that the two approaches were not distributed equally over the two domains of developmental level and of personality and emotional adaptation. Researchers have used counting and judging results to indicate cognition. They have used global inference to assess personality and emotional disturbance, with the exception of Naglieri et al.'s (1991) procedure.

The judges' estimation was based on the content of the drawing as an indicator of cognitive development, and drawing skill yielded moderate reliability. The inference of personality characteristics, such as self-image and impulsiveness, yielded unreliable results.

The average difference between judges—in particular with Judge II—was probably an indication of Judge II's response set (i.e., she was lenient in judging children's human figure drawings). This judge was a pedagogy student with a concentration in helping disturbed children, which is known for its child-orientation, even when using intelligence tests to evaluate children. Given our results, it made sense to compute and to check for differences in average estima-

tions between the judges, especially when the means were interpreted as the presence of a developmental level of cognition or social and emotional adaptation.

Although we expected one dimension, the estimation of the judges yielded two dimensions referring to maximum and typical performance (Cronbach, 1990). Because the first dimension comprised the reliable variables, the two-dimensional structure could possibly be a consequence of the differences in reliability of the variables. However, one also could argue that the judges used just these two dimensions, one referring to cognitive developmental level (maximum performance) and one related to personality and social adaptation (typical performance).

The judges were able to infer the differences in age and school type from the drawings. Our results resembled the findings of Cox and Howarth (1989), although in our study, the age range was smaller (2.5 years vs. 4 years). Our results also resembled those of McNeish and Naglieri (1993), who compared normal and mildly disturbed children.

Predictive validity can be characterized as low because the three judges were not successful in predicting academic achievement. The results cannot be considered as good enough or sufficient. Estimation of drawing skill by teachers and judges showed a significant but moderate correlation, from the perspective of reliability requirements proposed by Nunnally and Bernstein (1994). The high reliability for cognitive development judgments of teachers was probably the consequence of repeating the same judgment in different words. The estimation of the children's drawings by the judges and the teachers' marks for the different developmental variables showed significant, but not substantial, correlations.

Teachers' marks reflected one dimension representing the child's adaptation to the student role (i.e., the extent to which the student is achieving at school in an acceptable manner, developing normally cognitively, adapting socially and emotionally). The results of the judges and the teachers did not allow us to consider the variables as expressions of a set of differentiated dimensions in this sample. Only the obvious distinction between maximum and typical performance was present, and then only for the judges. Because the teachers knew their students' ages and type of school, we did not use the known groups criterion for teachers. In predicting school achievement, teachers probably used their evaluation of the cognitive level. Other variables played no role in this prediction. Thus, we concluded:

1. In using the DAP test to assess development variables, only the estimation of the level of cognitive development is sufficiently reliable.
2. The developmental variables based on the DAP test do not seem to have been derived from one dimension. There was empirical evidence for one general dimension, such as level of adjustment, that covered all the developmental variables.
3. The DAP test can be considered as good (or good enough) because of its sensitivity to children's age and school type, but the predictive validity is only

moderate. This result implies that the success of using the global approach for assessing human figure drawings as indicative of cognition and social and emotional development is limited. In our opinion, the results of our study can be generalized to other countries if the DAP test is used in comparable situations (i.e., with normal or mildly disturbed students as participants and with teachers).

It is common for practitioners to estimate cognitive, social, emotional, and personality variables based on the global approach (i.e., inspection and apprehension of the drawings using only some general descriptions and questions). Researchers have often argued that the item-by-item procedure of the thinking approach is lengthy and laborious with insufficient benefit, but our results suggested that the extra effort yields worthwhile results.

What are the implications of our research for professionals who use the DAP test in assessing and diagnosing children? We found that judges (even those with little experience in evaluating drawings of 7- to 9-year-old students attending either regular education schools or special education schools) can reliably count details and determine the developmental level of students. It is clear that the global approach to human figure drawings cannot be used to assess children's social and emotional developmental or personality characteristics, such as impulsiveness and self-image. However, this global approach is still used widely. If one decides to use the DAP test for assessing a child's adaptation, one could plead for extensive training, but Scribner and Handler (1987) did not find the assessments of more experienced professionals to be better than those without experience. So, the global approach does not yield better results after training. School achievement can only partly be predicted by the judges' estimation of cognitive level from human figure drawings.

The DAP test can be used for assessing a child's developmental level, but not for determining personality and emotional adaptation. A specific scoring system is necessary to achieve better results (see Naglieri et al., 1991). Such a system is worthwhile, but time must be invested in the training to use such a system. Thinking, training, and a specific empirically based scoring procedure can help to enhance the reliability and validity of scores based on the DAP test (Garb, 1998; Naglieri, 1988).

The systematic differences in developmental level caused, for example, by leniency or judges' higher or lower anchor values, also can be prevented by giving clear information on the full range of the scales used and by describing when to use which mark.

In our study, we provided some knowledge on the reliability, validity, and usefulness of the DAP test in predicting academic achievement. However, teachers assessed academic achievements. For practical limitations, we were unable to measure the students' academic achievements by school achievement tests. We noted that most teachers assess achievement using marks (ranging from 1 to 9), codes with a short description of their meaning, or extended written reports. Although the teachers' judgments are not completely accurate, teachers are often

the best source of information on their students' school achievements. This result encouraged us to use their estimations.

Finally, the DAP test is good enough for assessing the developmental level, but has serious drawbacks in assessing social adaptation. Our results could be considered as an invitation and stimulation to pay closer attention to assessing adequately the socioemotional development based on children's drawing of a human figure.

REFERENCES

- Baarda, D. B. (1978). *Utrechtse Korte Kleuter Intelligentie Test* [Utrecht Shortened Toddler Intelligence Test]. Lisse, The Netherlands: Swets & Zeitlinger.
- Barrett, M., & Eames, K. (1996). Sequential developments in children's human figure drawing. *British Journal of Developmental Psychology*, *14*, 219–236.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to use valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*, 271–280.
- Chappell, P. A., & Steitz, J. A. (1993). Young children's human figure drawings and cognitive development. *Perceptual & Motor Skills*, *76*, 611–617.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Cox, M. V., & Howarth, C. (1989). Human figure drawing of normal children and those with severe learning difficulties. *British Journal of Developmental Psychology*, *7*, 333–339.
- Cox, M. V., & Parkin, C. E. (1986). Young children's human figure drawing: Cross-sectional and longitudinal studies. *Educational Psychology*, *6*, 353–368.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.
- Cummings, J. A. (1986). Projective drawings. In H. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 199–244). New York: Guilford Press.
- Fabry, J. J., & Bertinetti, J. F. (1990). A construct validation study of the human figure drawing test. *Perceptual & Motor Skills*, *70*, 465–466.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, *87*, 477–531.
- Funder, D. C. (2001). *The personality puzzle* (2nd ed.). London: Norton.
- Garb, H. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Goodenough, F. (1926). *Measurement of intelligence by drawings*. New York: Yonkers World Book.
- Groth-Marnat, G., & Roberts, L. (1998). Human figure drawings and house–tree–person drawings as indicators of self-esteem: A quantitative approach. *Journal of Clinical Psychology*, *54*, 219–222.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical predictions: A meta-analysis. *Psychological Assessment*, *12*, 19–30.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt Brace.
- Kahill, S. (1984). Human figure drawing in adults: An update of the empirical evidence 1967–1982. *Canadian Journal of Psychology*, *25*, 269–292.
- Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. New York: Grune & Stratton.

- Koppitz, E. M. (1984). *Psychological evaluation of human figure drawing by middle school pupils*. New York: Grune & Stratton.
- Laak, J. ter, de Goede, M., & Brugman, G. M. (2001). Teachers' judgments of pupils: Agreement and accuracy. *Social Behavior & Personality*, *29*, 257–270.
- Machover, K. (1949). *Personality projections in the drawing of the human figure*. Springfield, IL: Thomas.
- McNeish, T. J., & Naglieri, J. A. (1993). Identification of individuals with serious emotional disturbance using the Draw-A-Person. *Journal of Special Education*, *27*, 115–121.
- Meykens, S., & Cluckers, G. (1996). *Kindertekeningen in ontwikkelings: Psychologisch en diagnostisch perspectief* [Children's drawings from a developmental and psychodiagnostic perspective]. Leuven, The Netherlands: Acco.
- Mönks, F. J., Rost, H., & Coffie, N. (1972). *Nijmeegse School Bekwaamheids Test (NST): Handleiding en Monografie. Vorm A en B* [Nijmegen School Readiness Test: Manual and Monography, Vol. 1 & II]. Nijmegen, The Netherlands: Berkhout.
- Motta, R., Little, S., & Tobin, M. (1993). The use and abuse of human figure drawings. *School Psychology Quarterly*, *8*, 162–169.
- Naglieri, J. A. (1988). *Draw-a-person: A quantitative scoring system*. San Antonio, TX: Psychological Corporation.
- Naglieri, J. A., McNeish, T. J., & Bardos, A. N. (1991). *Draw-a-person: Screening procedure for emotional disturbance*. Austin, TX: PRO-ED.
- Naglieri, J. A., & Pfeiffer, S. I. (1992). Validity of the draw-a-person: Screening procedure for emotional disturbance with a socially-emotionally disturbed sample. *Psychological Assessment*, *4*, 156–159.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.
- Roback, H. B. (1968). Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. *Psychological Bulletin*, *70*, 20–44.
- Rosenthal, R. (2000). Effect sizes in behavioral and biomedical research, estimation, and interpretation. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy*. Thousand Oaks, CA: Sage.
- Scribner, C. M., & Handler, L. (1987). The interpreters' personality in draw-a-person interpretation: A study of interpersonal style. *Journal of Personality Assessment*, *51*, 112–122.
- Shangnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2000). *Research Methods in Psychology* (5th ed.). Boston: Wiley.
- Smith, D., & Dumont, F. (1995). A cautionary study: Unwarranted interpretations of the draw-a-person test. *Professional Psychology: Research and Practice*, *26*, 298–303.
- Sullivan, M. J. (1998). Discriminant validity of the draw-a-person screening procedure for emotional disturbance for incarcerated juvenile delinquents in special education. *Dissertation Abstracts International Section—A: Humanities and Social Sciences*. 1998, Vol. 58 (10-A): 3736 IS: 0419-4209.
- Swensen, C. H. (1957). Empirical evaluations of human figure drawings. *Psychological Bulletin*, *54*, 431–466.
- Swensen, C. H. (1968). Empirical evaluations of human figure drawings 1957–1966. *Psychological Bulletin*, *70*, 20–44.
- Tharinger, D. J., & Stark, K. (1990). A qualitative versus a quantitative approach to evaluating draw-a-family and kinetic family drawing: A study of mood and anxiety disordered children. *Psychological Assessment*, *2*, 265–375.
- Thomas, G. V., & Silk, A. M. J. (1990). *An introduction to the psychology of children's drawings*. New York: Harvester Wheatsheaf.

- Wechsler, D. (1989). *Manual for the Wechsler Intelligence Test for Children-Revised II*. New York: Psychological Corp.
- Yama, M. F. (1990). The usefulness of human figure drawing as an index of overall adjustment. *Journal of Personality Assessment*, *54*, 78–86.

Received October 13, 2004

Copyright of Journal of Genetic Psychology is the property of Heldref Publications and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.