

Grant Aram Killian, Ph.D.

*Assistant Professor of Psychology, Nova University, Fort Lauderdale,
Florida.*

HOUSE-TREE-PERSON TECHNIQUE

*John N. Buck. Los Angeles, California: Western Psychological
Services.*

Introduction

The earliest written article on analyzing drawings was by Cook (1885) who recognized that there were successive stages in the development of children's drawings. During the turn of the century and the 40 years following Cook's publication, numerous articles on drawings appeared and were reviewed by Goodenough (1926), who wrote the first text using a scoring procedure for assessing children's intelligence by adding up the number of appropriate human details. The total number of appropriate details was then converted to an IQ score.

The House-Tree-Person (H-T-P) projective technique developed by John Buck (1948) was originally an outgrowth of the Goodenough (1926) scale utilized to assess intellectual functioning. Like Machover (1949), who was also interested in projective drawings as an appraisal of children's intelligence, Buck felt artistic creativity represented a stream of personality characteristics that flowed onto graphic art. They believed that through drawings, subjects objectified unconscious difficulties by sketching the inner image of primary process. By allowing subjects to respond by their own construction to stimuli that are familiar and ambiguous, they assumed that subjects would project a self-portrait that could be used to assess personality dynamics. Since it was assumed that the content and quality of the H-T-P was not attributable to the stimulus itself, they believed it had to be rooted in the individual's basic personality. Thus, in the interpretative process, the three objects assume symbolic aspects of the subject's world: the *House* mirrors the subject's home life and intrafamilial relationships; the *Tree* reflects the elemental relationships that the subject experiences within his or her environment; and the *Person* echoes the subject's interpersonal relationships. Since the H-T-P was an outcropping of an intelligence test, Buck (1948) developed a quantitative scoring system to appraise gross classification levels of intelligence along with a qualitative interpretive analysis to appraise global personality characteristics. Unfortunately, the original standardization research falls short of today's standards for acceptable methods of test development and construction.

The sample size used in developing norms for the quantitative scoring system

This reviewer wishes to acknowledge and thank Evelyn Villafana for her assistance in the preparation of this review.

was most meager ($N = 140$) and no attempt was made to select randomly a stratified sample of subjects from the general population (Buck, 1981). Criteria for inclusion in the study were nebulous. Twenty adults were selected for each of seven intellectual levels (imbecile, moron, borderline, dull average, average, above average, and superior) based on an assessment of the person's "complete clinical picture . . . and not a score on one or more standard intelligence tests" (Buck, 1981, p. 8). Furthermore, no attempt was made to select an equal number of males and females (e.g., the sample at the imbecile level consisted of 5 males and 15 females, and the sample at the superior level consisted of 19 males and 1 female) and no attempt was made to control for age across and within the seven intellectual levels (ages ranged from 13.6 to 29 years). Finally, two separate collection methods were used during the standardization study: the drawings of subjects who had less than above average intelligence (high school or less) were obtained by individual examination, whereas the drawings of subjects who had above average intelligence and more (college students) were obtained by group examination.

The standardization data for the qualitative analysis are equally suspect (Buck, 1981). The normative sample consisted of 150 adult subjects: 52 patients who were seen at the University of Virginia Hospital and 98 who were seen at the Lynchburg State Colony Hospital or at the Colony's Mental Hygiene Clinics in other cities. Sex, age, race, IQ scores, and other demographic information about the sample were not presented. Eight gross classification groups of unequal numbers were formed (adult maladjustment = 10, epilepsy with personality maladjustment = 29, psychopathic personality = 22, psychoneurosis = 53, prepsychotic state = 3, mental deficiency with psychosis = 6, organic psychosis = 11, and functional psychosis = 16); however, criteria for inclusion in the various groups were not reported.

To date, no new normative studies on the H-T-P have been conducted. But despite the questionable underpinnings of the instrument, research on the H-T-P has expanded into several different countries and a variety of modifications or variations have emerged over the years, as well as a Post-Drawing Interrogation Form for children under age 15.

Diamond (1954) developed a projective test that combined aspects of the Thematic Apperception Test (TAT) and the H-T-P. The subject is instructed to make up (i.e., write) a story involving a *House*, a *Tree* and a *Person*, and is told that the three objects have personality as well as the ability to speak to one another. The subject is further instructed to describe what kind of house, tree, and person the characters are and how they feel about each other.

The Draw-A-Family test, which according to Hammer (1978) has no known authorship, is a projective technique that is often used in combination with the H-T-P. In this test, the subject is simply asked to draw a picture of his or her family. The drawing is used to assess the individual's perception of him- or herself in the family system and/or the relationship to parental and sibling figures. Interpretations focus on omitted figures, exaggerated figures, insignificant small figures, spatial placement of figures, activity of figures, facial expressions, and content and

movement of all the figures. Figures in the Draw-A-Family test are analyzed in isolation and in relation to the overall gestalt. Both views are presumed to reflect subjects' perceptions of their relationships within the family.

The Man-Woman Drawing, a modification of the Draw-A-Person technique developed by Machover (1949), has now also become by tradition part of the H-T-P administration. After drawing a person the subject is then requested to draw a person of the opposite sex. Hammer (1978) was the first to combine both techniques into what he has called the H-T-P-P which includes a drawn person of each sex.

According to Hammer (1978) the Draw-A-Person-In-The-Rain also has no known authorship and is a technique that assesses adaption and withdrawal under unpleasant environmental conditions. Subjects are told to draw a person in the rain, and interpretations focus on the use of protective shields, the position of an umbrella, the amount of figural exposure with shields, facial expressions, additional environmental phenomena (e.g., clouds or rainbows), and other figures, as well as content and movement within the drawing.

Hammer (1978) developed a test called Draw-A-Member-Of-A-Minority Group to assess the projection of negative traits in oneself onto a member of a minority group. In this drawing, projected attributes and prejudices are assumed to reflect unconscious negative traits in oneself and not attributes of the minority group.

The Rosenberg Draw-A-Person Test (1948) is conducted with carbon paper for both male and female drawings. Once the drawing is completed, the examiner retains the carbon copy and asks the subject to alter, erase, or cross out the original drawing. After changes have been completed, a post-drawing inquiry focuses on the modifications.

Caligor (1951) developed an Eight-Figure Redrawing technique based on the Rosenberg test. In this test, subjects make a series of eight human figure drawings and are then instructed to change each of them. Instead of using carbon, subjects are given onionskin paper that they can see through which is then placed on each previous drawing. Interpretations are based on serial change and deeper personality levels are assumed to underlie each progressively repeated sketch. The principle assumption underlying this method is that repetition of task will lead to deeper unconscious identifications.

The Animal-Drawing-Story technique developed by Levy and Levy (1978) requires 1) the drawing of any animal, 2) a pet name for the animal, 3) naming the kind of animal drawn, and 4) an optional imaginative story about the animal. Normative data was based on 7,346 drawings obtained from selected adult males and females, institutionalized male and female psychotics, male and female adolescents, and male prisoners. After completion of the test, the animal drawings are analyzed normatively, formally, and symbolically.

Harrower (1978) developed the Most-Unpleasant-Concept-Test, another projective technique, that requires subjects to graphically sketch the most unpleasant thing that they can imagine. Subjects are allowed to pencil draw the image in actuality, schematically, or symbolically, and after completion subjects describe and give free associations to the drawing. Carbon paper provides additional information about erasures, pressure, and shading. Results are interpreted in terms of the subject's reactions, content, and whether the concept was an internal

external event. The technique is based on the qualitative analysis of the results of 500 subjects in psychotherapy who were given a full battery of psychological tests.

Although the Goodenough (1926) Draw-A-Man Test was developed 22 years before the H-T-P, it served as the groundwork. Harris (1963) revised the Goodenough test in order to develop an alternate form to the Man scale, to extend the scale to adolescents, to develop norms on a more representative sample, to establish a quality scale for quick approximation for point scores, and to extend the test to a Self drawing. The text of this revision (Harris, 1963) includes the manual, general scoring instructions, scoring and practice examples, a short scoring guide, conversion tables, plates for quality scales, percentile ranks, and a test booklet.

The H-T-P is a projective technique that utilizes pencil and crayon freehand drawings of a House, a Tree, and a Person. During the administration of the drawings the subject is given almost complete freedom in sketching the three objects.

The test materials suggested by Buck (1981) to administer the H-T-P include:

1. A four-page 7" x 8½" Drawing Form. The first page provides for identification information, the second has *House* printed at the top, the third has *Tree* printed at the top, and the fourth has *Person* printed at the top. Two drawing forms are needed for each subject: one folder is used for the freehand pencil drawings and the other is used during the crayon drawings. A Two-Copy Drawing Form is available for group testing which records erasures. The chemically treated paper provides a second copy that records all lines drawn even if they are erased on the original paper, and may thereby provide further clinical application for the practitioner;
2. A four-page H-T-P Post-Drawing Interrogation (P-D-I) Folder which is used after the achromatic and chromatic drawing phase (an abbreviated version of the P-D-I can be used after the chromatic drawings). In addition, there is a set of revised questions that are recommended for subjects under 15 years of age;
3. A four-page H-T-P Scoring Folder that is used for quantitative scoring;
4. Several pencils with erasers that are recommended along with a set of 8 or more crayons (red, green, blue, yellow, brown, black, purple, and orange);
5. The H-T-P Manual; and,
6. A stopwatch.

In most settings, all of these materials are not used and the majority of clinicians agree with Ellis (1970), who "sees no particular value in employing the standard H-T-P booklets for the actual drawing. Any ordinary size white paper serves the same purpose" (p. 592). In terms of the P-D-I, Ellis goes on to state that "The time and effort spent in making this postdrawing interrogation on an individual basis is questionably expended, as against utilizing this time for a general psychological interview" (p. 107).

The H-T-P test consists of two drawing phases with each phase followed by a structured interview. The first step consists of having the subject sketch a free-

hand pencil drawing of a House, a Tree, and a Person, which is then followed by the examiner asking 60 questions from the P-D-I. After the subject answers the questions about the three achromatic pencil drawings, the subject is again requested to produce a freehand drawing of the three objects, but this time with the eight or more crayons. This is followed by the same structured questions of the P-D-I or a shorter version if the subject is fatigued or if undue time has elapsed.

The P-D-I consists of 60 questions varying from direct and concrete to indirect and abstract. Preceding the number of each question are the letters H, T, or P, which respectively deal with the House, the Tree, and the Person. In addition, the questions are followed by one or more letters to indicate A for Association, P for Pressure, and R for Reality-Testing, which are used for scoring purposes. To prevent the possibility of an answer set and reduce the chance of a subject remembering a response to a previous drawing, the different types of questions have been intermittently spaced.

During the drawing phase the examiner's participation is minimal and is primarily focused on recording: 1) the order in which the subject sketches parts of each drawing; 2) all spontaneous comments as they relate to details; 3) any emotions expressed by the subject; 4) the initial latency period; 5) any intra-whole pauses; and 6) the total time used to sketch each object. During the questioning of the P-D-I, which is not intended to be a rigidly structured procedure, the examiner becomes more active, asking the subject to respond to the 60 questions on the H-T-P and to any additional questions that may seem clinically pertinent.

Although Buck's original normative group was arbitrarily set at 15 years or older, more recent research on the H-T-P has found that children as young as 4 can perform the task (Jolles, 1952; Beck, 1955; Bieliauskas, 1960); hence, no specific age limitations are given in the revised manual (Buck, 1981). In addition, since the difficulty level of the drawing is set by the subject, no specific mental limitations are given in the revised manual (Buck, 1981). Consequently, the test can be used for ages as young as 4 and over, and can be used with hearing-impaired (Davis & Hoopes, 1975) and handicapped children (Johnson & Wawrzaszek, 1963). In the original normative group, subjects ranged from "imbecile" with an IQ score of 25 to "superior" with an IQ score of 140.

Once the P-D-I has been administered and the interview has been completed, the examiner records items of detail, proportion, and perspective in the Scoring Folder. After completing the elaborate scoring tables by examining the plates containing sketches that illustrate quantitative scoring points, the examiner derives an IQ figure for the percentage of raw G, a net weighted score, a weighted "good" score, and a weighted "flaw" score, which then comprise the items for the profile configuration.

The H-T-P requires few materials, minimal space, and a small flat desktop surface; consequently, any quiet setting with illumination can be appropriate as long as the subject is comfortably seated with sufficient room to draw. The H-T-P can be administered in an individual or group format. For group testing the same materials are used with the possible exception of the "Two-Copy Drawing Form," which permits permanent recording of erasures. After completing the achromatic drawings, the P-D-I is distributed and the examinees are instructed to answer all 60 questions, and the procedure then continues as in the individual examination

method. Empirical studies are scant comparing the two methods. According to Buck (1981) group administrations may have merit, but are less informative than individual testing. In contrast, Cassel et al. (1958), Cowden et al. (1955), Ellis (1970), and Hammer (1978) suggest that group testing is less time-consuming and seems to provide richer diagnostic and prognostic data.

The testing procedure is simple and straightforward. A trained psychologist is not necessary to administer the separate phases of the test and either a secretary, an aide, a teacher, or a mental health counselor can be quickly trained to administer the test. The examiner first presents the drawing form sheet to the subject with only the second page showing with the word *House* at the top and states:

Take one of these pencils, please. I want you to draw me as good a picture of a house as you can. You may draw any kind of house; you wish, it's entirely up to you. You may erase as much as you like, it will not be counted against you. And you may take as long as you wish. Just draw me as good a house as you can (Buck, 1981, p. 18).

When a subject seems concerned about his or her drawing ability, the examiner should assure the person that it is not a test of artistic ability. Once completed, the examiner then repeats the drawing instructions for the Tree and the Person. After the drawing phase is terminated, the examiner administers the P-D-I in order to determine what the House, Tree, and Person meant to the subject. Since this part of the test is not intended as a rigid structured interview, the examiner may always conduct further questions that seem clinically appropriate.

In general, normal adults will use between 30 and 50 minutes for the drawing of the H-T-P series (either the achromatic or chromatic) and a similar amount of time in discussing the P-D-I; however, these time ranges depend on the subject's degree of adjustment and level of intelligence. In extreme cases, all six drawings could be done in less than 5 minutes or in the case of obsessive-compulsive styles, more than an hour could be taken for each series.

In most settings, clinicians do not administer the chromatic drawing since there is no scientific evidence that it provides any additional clinical information. Moreover, the standard procedure does not provide for the drawing of persons of the opposite sex or family. Greater insight could be achieved by eliminating the chromatic phase and including sketches of the family and the opposite sex. Ellis (1970) strongly advises "clinical psychologists who use the H-T-P to take Buck's administration and scoring procedures with decided skepticism and to adopt testing methodology to their own realistic work schedules" (p. 593).

The H-T-P technique was designed to foster projection, and through the interpretations of drawings to assess an individual's efficiency, sensitivity, maturity, flexibility, personality integration, and level of interaction with the environment by allowing subjects to paint a picture of their world where each drawing is assumed to represent aspects of a self-portrait. As a projective prognostic tool, the drawings are intended to lay bare symbolically conscious and unconscious mental processes that aid in the identification of suppressed or repressed dispositions and conflicts and outline the various resistances and defense mechanisms to these dispositions. As a therapeutic tool, the drawings are intended to function as a

springboard for the elaboration of fruitful associations, furnishing a broader picture of the personality to further psychological insight. The House, the Tree, and the Person were selected because they are familiar, can be easily drawn, and promote open dialogue with all ages and personality types.

Practical Applications/Uses

The H-T-P has become one of the standard projective tools in the psychological test battery and frequently serves as the introductory test in the battery by providing a minimally threatening and maximally absorbing beginning to psychological assessment. Serving as an easy bridge to the psychological examination with minimal contact with the examiner, the subject is not intimidated by specific questions concerning intelligence or by threatening inkblots, which may disturb unconscious conflicts.

As a diagnostic tool, the H-T-P through tradition has been given a secure position in a prognostic battery, although there is question about its contribution to the entire clinical picture. Wyatt (1949) maintains that the H-T-P harnesses deeper and more primary conflicts that are less differentiated than those obtained through other projective tests, while Hammer (1955) argues that the H-T-P provides a "grosser personality picture" (p. 17). For him, the Rorschach provides the "richer personality picture" (Hammer, 1978, p. 600) except when subjects are guarded, then the H-T-P proves to be the more revealing device (Hammer, 1954; Landesberg, 1953). On the other hand, as a screening tool for detecting the onset of incipient psychopathology, Hammer (1955) and Zucker (1948) see the H-T-P as a more sensitive prognostic tool than the Rorschach. However, this is refuted by recent research, which suggests that the H-T-P is one of the least useful tools for discriminating between normal subjects and psychiatric patients, and that if employed in a test battery the H-T-P is more likely to increase error and reduce the probability of making an accurate diagnosis (Wildman & Wildman, 1975).

As a screening device, Buck (1981) and Boring and Topper (1949) recommend the H-T-P be employed for measuring therapeutic change; as treatment progresses the size of the drawings change appropriately and intra-whole proportions improve. As an initial screening device (Buck, 1981) the H-T-P can also be used 1) in a group testing format to identify adjustment levels; 2) to evaluate personality integration and adjustment prior to training programs, therapy (Hammer, 1978), employment, and school enrollment (Beck, 1960); and 3) as an evaluation tool to assist research.

As a therapeutic tool, the H-T-P can facilitate free associations that will enhance increasingly deeper levels of insight (Buck, 1981). In art therapy it can be employed with adolescents or children and can be a supplementary tool in analytic group therapy (Naumberg, 1978).

In short, as a diagnostic, therapeutic and screening device, the H-T-P seems to have many practical uses and is easy to administer; however, because of the extremely complex quantitative scoring system and the subjective nature of the qualitative analyses, psychological interpretation of the technique requires a trained and experienced clinician. Although a useful therapeutic tool which can

be used in numerous settings, the scoring method and the sophistication required to interpret the H-T-P seem to preclude its diagnostic utility in non-clinical settings unless a trained professional is available for interpretation. Thus, despite its value and ease of administration in numerous settings, scoring and interpretation constraints preclude its use as a prognostic or diagnostic device by school counselors, educators, or social workers who are not trained in quantitative and qualitative scoring. Its utility as a diagnostic or screening tool is thus limited to the private practitioner or clinician in a mental health service. Moreover, if the research findings of Wildman and Wildman (1975) and others persist, it seems the H-T-P will have even less value since in a test battery it decreases the probability of making an accurate diagnosis. In the future, unless there is research support, the H-T-P may be relegated to the position of only a therapeutic tool for art therapy, or be limited as a diagnostic tool for nonverbal patients. For more frequently than not, the H-T-P tells the clinician what he or she already knows, or other instruments are used to develop or support hypotheses to interpret the H-T-P drawings.

The elaborate quantitative scoring system presented by Buck (1981) involves several phases before the final IQ scores are calculated. Initially, the examiner uses the scoring tables, item by item, assessing spontaneous and omitted items for all the achromatic and chromatic drawings in terms of detail, proportion, and perspective. Two major classes ("flawed" score or "good" score) are subdivided into intelligence levels totaling eight possible scores (each with a symbolic designation) that can be given to each characteristic of a drawing. "Flawed" scores from most to least flawed are rated as "very inferior (D3)," "imbecile (D2)," and "moron (D1)," while "good" scores are "borderline (A1)," "dull average (A2)," "average (A3)," "above average (S1)," and "superior (S2)." After scoring all sketches for all possible drawn or omitted characteristics, the eight grand raw scores (D3, D2, D1, A1, A2, A3, S1, S2) are converted to grand total weighted scores. The examiner then calculates the percentage of raw G and enters all possible scores onto the tabulation sheet, where IQ figures are derived from the percentage of raw G score, the net weighted score, the weighted "good" score, and the weighted "flawed" score. If there is significant scatter or more than one intelligence classification level between "good" and "flawed" scores, then this is suggestive of repression or deterioration.

This extensive quantitative scoring of the H-T-P, however, is rarely if ever used because it is cumbersome and scorer reliability is less than adequate (Bieliauskas, 1956). Despite a clear need for revision in the quantitative scoring, Bieliauskas's (1956) suggestion for refinements in aspects of the H-T-P scoring have been disregarded. Scoring instructions in the manual continue to be complex, ambiguous, and lacking in clarity, and "in any set of drawings the examiner may find items for which no scoring or only some scoring is provided" (Buck, 1981, p. 34). In confounding cases, Buck (1981) suggests, "If it is learned that a drawn whole is a stereotype or a reproduction of a learned figure ('Teacher makes us draw them that way'), the examiner may treat the figure qualitatively only" (p. 35). But this is untenable; if a drawing has been invalidated for quantitative scoring because it is a learned whole that has no psychological precursors, how can the same sketch have valid psychological meaning for projective interpretation? Further, Buck's (1981) own caveats bring to question the validity of the quantitative scoring: "In

constructing this relatively objective quantitative scoring system, it was very difficult to divorce the measurement of 'architectural artistry' which is presumed to be a highly specialized and specific ability from the appraisal of good proportional relationships" (p. 35). On this point, Bieliauskas and Bristow (1959) found that the drawings of art-trained subjects significantly received more favorable scores. In another study evaluating the feasibility of the quantitative system for children in grades 2-5 Bieliauskas and Moens (1961) found negative results, indicating that the scoring was not applicable to children.

Because of these preceding points, the value and meaning of the H-T-P IQ scores seem questionable. Consequently, a longer than average training period is required to learn the quantitative scoring method, and the "technique cannot be properly mastered from manuals and journal articles" (Harriman, 1970, p. 858); however, once mastered, according to Buck (1981) an "experienced examiner usually can score and interpret a full achromatic-chromatic H-T-P in one hour and a half or less" (p. 251). Krugman (1970) and others feel "the method of scoring the H-T-P for intelligence is so complicated . . . that it seems doubtful whether anyone but the author of the test can achieve as high a correlation with a standardized intelligence scale. Furthermore, the time required for scoring the objective part seems prohibitive . . ." (p. 345). Hayworth (1970) also feels the "scoring criteria are so detailed, qualified, and ambiguous that the reliability of scoring is questionable, and no data are offered on this aspect. The time spent in such scoring would be better spent in administering a standard intelligence test" (p. 1240).

Unfortunately, due to these complexities and the continual need for plate comparisons in the manual, computer scoring is impossible. In addition, three inherent difficulties in the scoring involve the use of the manual. First, the comparison plates needed to illustrate scoring points are extremely small. Second, assessing the relative quality value based on the eight possible intelligence levels for each characteristic of the House, Tree, and Person for detail, proportion, and perspective is cumbersome and time-consuming. And third, the manual does not clearly explain the tabulation sheet used to plot mean raw scores, percentage scores, and derived IQ scores. Faced with all these shortcomings led Ellis (1970) to state, "Considering the unknown reliability and, especially, validity of the H-T-P intelligence estimations and personality interpretations at the present time, and considering the time available to the psychologist in a normal clinical situation, it is unlikely that the H-T-P is normally worth this many hours of a busy psychologist's time" (p. 592). According to Harris (1963):

Buck's manual is not clear as to procedure of evaluation, or wholly satisfactory as a guide to interpretation . . . Buck's own statistical criteria for denoting certain characteristics as unusual while consistently applied, appear to have no basis in statistical logic. (p. 49).

The interpretation of the H-T-P is based on both the quantitative scoring and qualitative analysis. In terms of the quantitative scoring, interpretations involve five steps (Buck, 1981): 1) differences between the IQ scores; 2) appraisal of mean score patterns; 3) evaluation of detail, proportion, and perspective; 4) comparison of the "good" and "flawed" scores; and 5) comparison of the achromatic and

chromatic drawings. In terms of the projective analysis of the drawings, interpretations involve the evaluation of: 1) detail, proportion and perspective; 2) elapsed time, line quality, attitude toward the task, color choice, and drive; 3) clinical analysis of the P-D-I; and 4) the subject's concepts of each sketch encompassing both graphic and verbal productions.

Several limitations affect the reliability and level of training needed to interpret the H-T-P (Buck, 1981):

- 1) No single sign itself is an infallible indication of any strength or weakness in the S.
- 2) No H-T-P sign has but one meaning.
- 3) The significance of a sign may differ markedly from one constellation to another.
- 4) The amount of diagnostic and prognostic data derivable from each of the points of analysis may vary greatly from S to S.
- 5) Colors do not have any absolute and universal meaning.
- 6) Nothing in the quantitative scoring system can be taken automatically at face value (p. 80).

Moreover, each sketch can symbolize multiple concepts. For example, the House could represent (Buck, 1981): 1) home as it is now, or as the subject 2) would like it to be; 3) an unsatisfying or 4) a satisfying past home. As a self-portrait, aspects of the house hypothetically can represent: 1) the subject's psychosexual maturity; 2) the subject's accessibility; 3) the subject's contact at the level of reality; 4) the subject's intra-personal balance; 5) the degree of rigidity of the subject's personality; 6) the relative roles of the psychological past and future in the subject's psychological field; and 7) the subject's attitude toward his or her family and/or the subject's interpretation of the family's feeling toward him or her.

The Person may potentially represent an individual in the subject's environment whom the subject most likes, dislikes, or feels ambivalent towards. As a self-portrait, aspects of the Person hypothetically may represent: 1) the subject as he or she is now, 2) feels now, 3) would like to be; 4) the subject's concept of his or her sexual role; 5) the subject's attitude toward interpersonal relationships in general or 6) toward a specific relation; and 7) certain specific fears and/or obsessions.

The Tree could represent either the subject or some other person, and as a self-portrait aspects of the Tree could represent: 1) the subject's subconscious picture of self in relation to his or her psychological field; 2) the subject's subconscious picture of his or her development; 3) the subject's psychosexual level; 4) the subject's contact with reality; or 5) the subject's feeling of intra-personal balance.

Because of these complexities and the multiple meanings that can be given to various aspects of a drawing, a high degree of training and experience is required to properly interpret the sketches. To simplify the task, Wenck (1984) and Jolles (1983) have considerably reduced the task by providing a catalog of examples for various interpretative points. Although most of the interpretations are hypothetical, anyone interested in using the H-T-P from a projective standpoint will find the task of projective interpretation less overwhelming. However, beginners should use caution with these texts since there is little experimental support for

the interpretations of these signs, and these interpretations should only be considered as tentative hypotheses that need further support from other sources. It is worth quoting Ellis's (1970) observation that "Buck frequently contradicts his own warnings and makes rash general and specific interpretative statements about the H-T-P which, as yet, are not backed up by any factual evidence whatever. He presents, in fairly dogmatic form, hypothesis after hypothesis which may logically seem to be true but which have not yet been psychologically and scientifically established" (p. 593).

Technical Aspects

In both the original manual (Buck, 1948) and the revised manual (Buck, 1981) evidence regarding reliability and validity are conspicuously missing. According to Ellis (1970) the original manual "displays incredible naivete, fanaticism, and arrant disregard for any attempt at scientific validation of the material presented" (p. 592). In terms of the revised manual, perhaps the most accurate denunciation comes from the developer himself (Buck, 1981):

There is almost no statistical proof of the validity of the qualitative scoring points and their interpretations which would satisfy . . . "the tenets of research design and scientific method . . ." And it appears unlikely, in fact, that such evidence will be available for several reasons, two of which are: 1. The fact that almost no *H-T-P* scoring point has a single implication . . . 2. The fact that a given characteristic or trait may be expressed in the *H-T-P* in many ways . . . It is the author's belief that the validity of the principle of the *H-T-P* method as a whole has been satisfactorily established (although the evidence is almost wholly clinical). The evidence of the validity of the individual differential items and their interpretations is less well-established but is certainly sufficient to justify the conclusion that the *H-T-P* is a mature clinical instrument . . . As for reliability, no significant data are offered at this time . . . The *H-T-P* does not possess a high order of statistically defined reliability . . . (p. 164).

One wonders how a test can be valid and not reliable: if a test is not accurate or consistent (reliability) how can a test actually measure what it purports to measure (validity)! A test is simply not valid in general; tests are valid for a specific purpose. In test construction the types of reliability (test-retest, alternate form, and internal consistency) and validity (content, criterion-related, and construct) should be presented. For the *H-T-P*, reliability studies are meager and are limited to test-retest reliability or interrater reliability. Although alternate form reliability is not possible with the *H-T-P*, internal consistency reliability could be performed, but as yet no studies are available to assess the extent to which items on a drawing correlate among themselves.

Reliability studies using the *H-T-P* IQs would be relatively easy to perform; unfortunately, most clinicians do not use these scores and question their validity. Since IQ scores are frequently not utilized, investigations on test-retest reliability lack objective criteria in determining similarity of reproduction between different

administrations. In addition, even if IQ scores were used to assess reliability, scores could show a positive correlation yet form could be radically different, producing alternate clinical interpretations. The most reasonable solution would seem to be internal consistency reliability; however, this requires interrater reliability which currently seems difficult to achieve on the H-T-P.

In evaluating the drawings of 32 paranoid schizophrenics, Fisher and Fisher (1950) found little interrater reliability. Their results indicated that trained psychologists had no greater interrater reliability than untrained raters. Lehner and Gunderson (1952) used the Draw-A-Person technique with normal subjects on 21 graphic traits and found a "relatively higher" agreement between raters on these 21 traits. However, with 43 college students, Bieliauskas (1956) compared judges on "flaw score" and "percent of raw G" and found lower than required correlations for interrater reliability.

In terms of test-retest reliability, Gasorek (1951) found conflicting evidence in children's drawings for consistency and reliability of formal and structural properties. Lehner's and Gunderson's (1952) study on test-retest reliability found results similar to Gasorek's. When limiting interrater reliability to just 21 items, they found interrater reliability to be greater than test-retest reliability.

If Buck (1981) is correct that "the H-T-P does not possess a high order of statistically defined reliability" (p. 164), then validation studies seem to have little if any meaning. If this projective test is not consistent in its measurement, and if raters cannot consistently agree on scoring, then it cannot possibly measure what it purports to measure. Buck incorrectly assumes that validity can exist without reliability. Thus, the following review of validation studies seems to have little value, except from a historical perspective.

The validity studies conducted by Buck (1981) show evidence that the projective technique indeed reflects intellectual functioning and not just non-intellective factors. The correlation coefficients between the H-T-P percent of raw G IQ and IQs of other tests are as follows: Otis, Higher Examination (.41); Stanford-Binet, Forms L-M (.45); Wechsler-Bellevue, Verbal (.70), Performance (.72), Full scale (.75). However, more recently Hellkamp and Johnson (1970) found nonsignificant correlations between the H-T-P IQs and the Wechsler Adult Intelligence Scale and Raven's Coloured Progressive Matrices. Since this last study used psychometrically sound instruments, it raises serious questions concerning the actual meaning of these IQs.

The most frequent validation studies attempt to correlate graphic traits of a particular group to those of another defined group. Typically, groups are defined in terms of psychotic disorders, personality disorders, organic disorders, or physical characteristics.

Singer (1950) found inconclusive results when comparing 40 college students and 34 schizophrenic patients given the H-T-P. On the other hand, Holzberg and Wexler (1950) using a 174-item checklist found statistically significant differences between 38 schizophrenic women and 78 nurses. More recently, Wildman and Wildman (1975) showed that the H-T-P, out of three projective tests, discriminated the least between 10 nurses and 10 female patients. Twenty sets of protocols were given to 6 blind clinical psychologists; the MMPI had the highest hit rate, discriminating with 88% accuracy, while the H-T-P discriminated with only 53% accuracy.

Demming (1949) found no statistically significant differences between the H-T-Ps of 20 psychopathic patients and 20 normal controls that were matched for intelligence. Similarly, Royal (1949) and Blum (1954) did not find a significant difference between normals and neurotics on the Draw-A-Person test. Giedt and Lehner (1951) also found no significant difference between normals and neurotics in terms of the age assigned to the Person drawing. Gravitz (1969) grouped 200 normal adult males and females into those high and low on depression. Using the Depression (D) score of the Minnesota Multiphasic Personality Inventory it was hypothesized, based on Buck's theory, that subjects with high D scores would tend to draw smaller figures than those with low D scores. Statistical analyses failed to show any significant differences in the size of drawings based on D scores. Marzolf and Kirchner (1972) gave 1,054 college students the H-T-P and the 16 Personality Factor Questionnaire. All drawings were scored based on Buck's 108 drawing characteristics, but there were no significant relationships between drawing characteristics and personality traits.

Beck (1955) investigated the H-T-Ps of 25 organic and 13 non-organic mentally retarded children and found no significant differences in the drawings of the House. Michael-Smith (1953) studied 25 matched pairs of children with normal and abnormal EEG patterns. H-T-P signs indicative of organicity were negative except for "line quality." A follow-up study was recommended and was conducted by Bieliauskas and Kirkham (1958), who used 18 criteria to examine the H-T-P drawings of 20 organic and 20 non-organic subjects matched for sex, age, and IQs, but found that the H-T-P signs of organicity were not valid since the H-T-P signs failed to hold for either group. Williams (1964) matched 20 schizophrenic, 20 organic, and 20 normal controls on the same 18 criteria for organicity and 32 signs for schizophrenia, and again no statistically significant differences were found.

Waxenberg (1955) compared three groups of females, comprised of 20 asthmatics, 20 non-psychosomatics, and 20 with histories of ulcerative colitis, on the H-T-P drawings, the Thematic Apperception Test, the Bender-Gestalt, and the Rorschach, and found no significant differences on all tests between the three groups. Silverstein and Robinson (1956) compared 22 children with orthopedic disabilities with 44 healthy children on the H-T-P; findings showed no significant differences between the two groups. Wawrzaszek et al. (1958) compared 41 matched pairs of handicapped and non-handicapped children, and no significant differences were found between the two groups. Davis and Hoopes (1975) compared the Human Figure Drawings of 30 deaf and 80 hearing children between ages 7.5 - 10.5 on 19 items concerning the characteristics of the ear and mouth. No significant differences between deaf and hearing children were found related to the ear and mouth drawings, except that there was more frequent shading around the mouth for the hearing children. There were however significant differences in the branch structure of the Tree: deaf children tended to imply a branch system, while hearing children drew out the branch system.

Other validity studies have been conducted in an attempt to verify certain specific hypotheses that Buck (1981) formulated. The overwhelming majority of the research has shown that many of these clinical interpretations are not valid. According to Hayworth (1970), "The most recent research studies designed to test

various hypotheses connected with the H-T-P's rationale have generally reported non-significant findings" (p. 1241).

Critique

From the present review, it seems that the psychometric properties of the H-T-P fall short of today's standards and the test seems to lack the required focused relevance for the practicing clinician or experimental psychopathologist. Psychological studies should be directed at valid assessment techniques that can reliably differentiate processes and functions that may be clearly implicated in various disorders (Killian et al., 1984). The use of unreliable techniques that are recommended only by their availability and familiarity should be abandoned. Instead, reliable and valid tests such as the Stroop Color and Word Test could be easily administered and scored, yet provide significant information about processes and functions in various disorders (Killian, in press). On one final note:

The "figure drawing" approach, as loosely described by Machover, Buck, Jolles, and others, appears more simple and direct, but permits the interpreter to "project" as much as his subject! (Harris, 1963, p. 51)

References

- Beck, H. S. (1955). A study of the applicability of the H-T-P with respect to the drawn house. *Journal of Clinical Psychology, 11*, 63-66.
- Beck, H. S. (1960). The house-drawing test as a predictor of first-grade achievement. *Journal of Experimental Psychology, 2*, 197-200.
- Bieliauskas, V. J. (1956). Scorer's reliability in the quantitative scoring of the H-T-P technique. *Journal of Clinical Psychology, 12*, 366-369.
- Bieliauskas, V. J. (1960). Sexual identification in children's drawings of the human figure. *Journal of Psychology, 16*, 42-44.
- Bieliauskas, V. J., & Bristow, R. B. (1959). The effect of formal art training upon the quantitative scoring on the H-T-P. *Journal of Clinical Personality, 15*, 57-59.
- Bieliauskas, V. J., & Kirkham, S. L. (1958). An evaluation of the "organic" signs in the H-T-P drawings. *Journal of Clinical Psychology, 14*, 50-54.
- Bieliauskas, V. J., & Moens, J. F. (1961). An investigation of the validity of the H-T-P as an intelligence test for children. *Journal of Clinical Psychology, 17*, 176-180.
- Blum, R. (1954). The validity of the Machover DAP technique. *Journal of Clinical Psychology, 10*, 121-125.
- Boring, R. O., & Topper, R. C. (1949). *A psychodiagnostic screening technique*. Tuscaloosa, AL: Veterans Administration Hospital.
- Buck, J. (1948). The H-T-P technique, a qualitative and quantitative scoring method. *Journal of Clinical Psychology Monograph Supplement No. 5*, 1-120.
- Buck, J. (1981). *The House-Tree-Person technique: A revised manual*. Los Angeles: Western Psychological Services.
- Galigor, L. (1951). The determination of the individual's unconscious concept of his masculinity-femininity identification. *Journal of Projective Technique, 15*, 494-509.