agnostic draw-a-person techniques, seriously. Challenging, novel, fun to try out in a class, helpful in keeping a pedagogue’s intellectual joints supple—yes, such techniques cannot be disregarded as literary psychology. As for tangible evidence to substantiate the admirable enthusiasms of those who make them up, these psychodiagnostic techniques are still in the ovum stage of development. Caligor’s technique is not an exception. Yet he toiled valiantly for more than a decade, and a reviewer feels chastened not to be able to recommend this test as the royal road to a valid diagnosis of personality.

Brit J Psychol 48:319–20 N ’57. H. C. Gunsburg. * The new technique is intriguing—the subject goes from one drawing to the next whilst the preceding drawing is visible to him through the transparent paper; he may trace it and virtually repeat it, modify it or ignore it completely. His handling of this situation, as well as the various details of the drawing, the placement and line quality are considered valid diagnostic clues. Dr Caligor provides a scoring system based on objectively scorable dimensions and attains a high interscores agreement. The scores are stated “in terms of a deviation from a statistical norm approximately the mode scores of normal subjects.” Unfortunately the book contains no statistical material and the only relevant reference given by the author refers to a study of items differentiating between 21 college males and 21 hospitalized paranoid schizophrenics—this could scarcely be called a normative study serving as a basis for a complete test manual to be used in clinical practice. Even more disappointing is the systematic list of suggested interpretations for the various scoring categories which are, no doubt, based on wide clinical experience, but which would be more convincing and valid if supported by some evidence. Though Dr Caligor’s technique shares this fault with the systems devised by his predecessors, one feels that nowadays, having demonstrated the wealth of clinical material obtainable from drawings, investigators should approach this field more systematically and not by-pass such important factors as sex, age, intelligence, cultural background, etc. In the meantime, this new publication’s main merit is the presentation of a novel and promising projective technique and of a scoring method which may well be used for testing experimentally the validity of drawing interpretations.

[132]

*Family Relations Test: An Objective Technique for Exploring Emotional Attitudes in Children. Ages 3-7, 7-15; 1957; individual; 1 form [’57]; 2 levels; 132s. per set of test materials; 12s. 6d. per manual; postpaid within the United Kingdom; (20–25) minutes; Eva Bene and James Anthony; distributed by National Foundation for Educational Research in England and Wales.

a) YOUNGER CHILDREN. Ages 3-7; 40 item cards; 3s. 6d. per 10 record booklets [’57].

b) OLDER CHILDREN. Ages 7-15; 86 item cards; 3s. 6d. per 10 record booklets; 2s. od. per 10 scoring blanks.

REFERENCE


JOHN E. BELL, Acting Chief, Mental Health Services, United States Public Health Service, San Francisco, California.

The test materials consists of 20 cardboard figures representing people of various ages from babyhood to old age. These are relatively ambiguous and permit a child to select figures to represent each member of his family including himself. In addition, a figure standing for “Nobody” is included in the materials. The figures are attached to cardboard boxes with slots in the top. In the form for children 8 years old and above there are 86 cards containing statements reflecting feelings of like and dislike, stronger feelings of love and hate, and attitudes relating to parental overprotection and overindulgence. In the form for younger children there are 40 similar cards.

After selecting figures to represent his own family, the child places each card in a box behind the figure for which the statement is most appropriate. If the statement applies to none, it is deposited in the box attached to “Nobody.” The child’s test performance is tallied on a scoring sheet; evaluation of the results and behavior notes are entered on a separate record sheet.

The distinct advantage of the test is its relative objectivity. There are a limited number of standardized responses that can be made in the test situation. This permits a formal analysis to be undertaken and facilitates the establishing of norms and the conducting of statistical studies of the test performance with various subjects under different conditions. Little research with the technique is presently available.

The test is subtitled “An Objective Technique for Exploring Emotional Attitudes in Children.” Its objectivity is, however, only par-
tial. The test involves a rather complex set of choices. Each card may apply to one family member other than the self, several family members other than the self, the self alone, the self along with one or more other family members, or nobody. Functionally, the child has to hold in mind all these possibilities and make appropriate choices for each item in terms of them. It may appear that the full range of possible responses is considered each time, but this demands that attention to the instructions has been of a high quality and that memory for them and a set for carrying them out are kept alive. Thus, while the actual manipulation of the cards engenders objectivity, the task itself is sufficiently complex that one may not be sure that the test conditions are comparable from one subject to another, or from one testing to another with the same subject. This has real importance, since the deductions about the performance are based on tallies of the responses without regard to the quality of the subject's attention and memory, his conceptual ability, and the processes by which he makes decisions.

An additional factor confounding the results grows out of the implicit assumption that the task of perceiving the self in the family situation is comparable to the task of perceiving the other. In practice, different processes are involved in evaluating the pertinence of the test items to the self than in assessing their relevance to others. The value of the test figures of other members of the family for symbolizing those individuals differs from the value of the self figure for personifying the self. In the former instances there is a closer parallel between the object nature of the test figure and the family member; in the latter instance there is a subject-object confusion in the approach to the figure and variability in the amount of distance of the self as embodied there. It might be constructive, then, to test the comparability of performances when the self-figure is included and excluded. The test items would not seem to lend themselves especially well to a test of reactions to the self where the simple choice of "applies" or "does not apply" would be required of the child, although this might be examined.

It is apparent that actualizing the family members by the pictures, concretizing the test task by placing items in the slots, and limiting the responses by standardized items represent a new combination of features in attitude testing. The clinical illustrations in the manual demonstrate that the test has real merit for rapid assessment of latent and overt attitudes to the family. It does not permit discrimination between the felt and the expressed attitudes, but it reduces the range of observations required additionally to produce a realistic picture of family relations.

Dale B. Harris, Professor of Psychology, and Director, Institute of Child Development and Welfare, University of Minnesota, Minneapolis, Minnesota.

This is an ingenious projective "test" which records the subject's reactions through the sorting of cards on which stimulus items appear. Thus it has the virtue of presenting identical stimulus material to all subjects while preserving some of the flexibility considered essential to projective devices. The emphasis, however, is on the reaction to the items printed on cards. The schematic human figures are primarily vehicles to facilitate the child's reaction to the content of the printed items; they are not designed to elicit elaborate fantasy. The authors believe this technique appeals directly to the child's interest in manipulating materials and his tendencies to respond covertly, and to express emotion through play. The items are presented to the child after he has identified the members of his own family circle from among the role figures.

The items devised for older children are of this type: "This person in the family is sometimes a bit too fussy." Items are grouped into several categories as follows: mild positive (affectionate) feelings coming from the child, strong positive (sexualized) feelings coming from the child, mild negative feelings coming from the child (the example above is taken from this category), and strong negative (hostile) feelings coming from the child. Four additional groups of items in the same patterns of affect represent feelings going towards the child. An example of the fourth type of feeling toward the child is "This person in the family hits me a lot." Three additional groups of items represent maternal over-protection, paternal over-indulgence, and maternal overindulgence.

The items for young children are expressed more simply and represent five classes only: positive and negative feelings coming from the child, positive and negative feelings going toward the child, and dependence. "N...[name of child subject] wants you to tuck him (her) into
Child bed at night. Who should tuck N....in at night?" represents this last category.

In theory the test helps the child express conscious attitudes, including those very private feelings which he would find difficult to state directly. The test admittedly does not investigate unconscious attitudes. The authors believe, however, that it is important to investigate the child’s phenomenal world, or as they speak of it, his “psychic reality.”

Scoring is accomplished by tallying the items assigned to particular role figures, excluding the items assigned to “Nobody.” The balance among proportions of items in the several degrees and directions of affect assigned to the several family roles provides the basis for interpretation. The manual offers profiles for a number of briefly described examples in each of the following personality patterns: idealizing tendency, paranoid tendency, and egocentric states, both auto-aggressive and auto-erotic. Other dynamic mechanisms revealed by the use of items include reaction formation, projection, regression, displacement, idealization and denial. The authors attach considerable significance to the child’s selection and treatment of significant figures, to his deviation from a theoretical frequency of items expected to be assigned to the usual family roles, to the balance he achieves between self love and self hate items (which indicates his egocentric state), to the relationship between positive and negative outgoing and positive and negative incoming affect items used by the child (which indicates his ambivalence, or lack of it, toward family figures).

The authors rest the case for the test’s validity on the concept of construct validity, on comparison of test results with extensive case history material for several small groups of children (which showed considerable correspondence), on comparisons of results of mutual feelings reported in sets of siblings, where agreement of 64 per cent satisfied the 5 per cent level of confidence, and on the congruence of test findings with predictions made independently from psychiatric diagnoses in several small samples of cases. Some data are quoted to show that results are independent of the sex of the examiner. Split-half reliabilities for combinations of affect categories vary from .68 to .90, number of cases not reported.

As frequently the case with tests of this type, no norms are given beyond a few illustrative cases and interpretations: The test is ingenious and simple, and the questions are phrased in children's language and represent common personal and family experiences; none are too threatening, on the surface at least. The device of sending “messages” should appeal to many children; the test certainly deserves further study.


The Family Relations Test (FRT) is a semistandardized play situation which permits the child to express his emotional attitudes toward members of his family and the attitudes he believes that members of his family have toward him.

The test materials consist of 20 cardboard figures “representing people of various ages, shapes, and sizes, sufficiently stereotyped to stand for members of any child’s family, yet ambiguous enough to become, under suggestion, a specific family.” Each figure is attached to a red cardboard box into which can be inserted small cards which bear various expressions of attitudes: positive feelings, negative feelings, dependence, maternal overprotection, and paternal overindulgence, some expressed as emanating from the child toward family figures and some expressed as emanating from family figures toward the child. There are two sets of cards, 40 for use with younger children and 86 for use with older children.

The subject is asked to select from the 20 figures a figure to represent each member of his family, including himself. Another figure, Nobody, is introduced by the examiner to receive those attitudes which the child will not assign to any member of the family. The statement on each card is then read aloud by the examiner and the card given to the child, who is instructed to deposit it in the box attached to the family figure to whom it best applies. If the statement does not fit anybody, the card is put in Nobody. If the statement fits several people, the examiner makes a note of it. The cards are collected from the boxes and are tabulated on a special scoring form, the scoring consisting of counting the number of items of each kind of feeling assigned by the child to each member of his family. The test takes between 20 and 25 minutes to administer.

The test would seem to have possibilities, con-
considering that there are few, if any, other objective techniques which serve the functions for which it is designed and that projective techniques are of doubtful validity. A good deal of clinical wisdom as well as an accumulation of experience with the FRT would, however, seem to be necessary for making judicious interpretations from the test material. Unfortunately neither the manual nor the one article (7) on the test presents any normative data. Apparently the test has never been given to normal children; at least, only clinical patients are described in the reports of the test’s use. The evidence for the test’s validity is too meager and unsystematic to provide an adequate basis for evaluation. From a statistical point of view the reliability evidence is not impressive. Also some of the statistical procedures and computations in the manual and the article are both inappropriate and incorrect. For example, a 2 x 2 contingency table is presented in the manual (p. 48) as evidence of a significant relationship between an independent rating and the FRT regarding sibling conflicts. The “measure of agreement” is given as 64 per cent. When the appropriate test, chi square, is performed, however, it shows the results to be quite nonsignificant ($\chi^2 = .292$). In another instance (p. 46) the authors have slighted the actual significance of their data. Simply dividing the sum of the diagonal frequencies of the contingency table by the total frequencies, the authors report 64 per cent agreement and state that this result is significant at the 5 per cent level. The 5 per cent significance level was probably based on a chi square test (not given by the authors), but actually the chi square is significant at the 5 per cent level only if it is interpreted as a one-tailed test, a rather unusual procedure in the case of chi square. A more appropriate test of the significance of these data is by means of a test of trend,1 a more refined and powerful test than chi square. When a test of trend was performed, the results show a relationship significant beyond the 0.1 per cent level.

The FRT may be a potentially useful test in the clinic, though this still remains to be demonstrated; at present it must be regarded as being in the trial stage. It can be recommended for use by those who are primarily interested in investigating the test itself. It is not a finished product about which there is sufficient information to warrant its being recommended for routine clinical assessment of child-family relationships.

---

1 Armotage, P. “Tests for Linear Trends in Proportions and Frequencies.” Biometrics 11:375-86 S ’55. *

---

*DOROTHY H. EICHERN, Assistant Research Psychologist, Institute of Child Welfare, University of California, Berkeley, California.

Adequacy of standardization varies with the subtest, category of scoring, and age group. The first three tasks—cutting out a circle, heart, and star—are scored for “quality” (“edge-cutting,” “form-cutting,” and “symmetry”) and “quantity” (number of scraps). “Quality” scores measure “manual dexterity,” “artistic ability,” and “level of aspiration.” “Quantity” scores indicate “emotionality.” These tasks have been administered to 327 Viennese girls aged 8-15 years, 233 parochial school children from one American city, 141 public school children from two cities, and 30 adults. However, the only statistical data reported for the “quality” scores are the means for 134 boys and 145 girls, aged 8-13 years, drawn from the American samples. Validation of the “quantity” scores as an indication of “emotionality” consists of one table listing the per cent of each of three “adjustment” groups (good, average, and poor) producing 15 or more scraps. Adjustment was rated by teachers. The sample is some portion of the American groups, but the frequencies from which the percentages were derived are not given.

The fourth task, a projective cutout, has not been standardized.

The fifth task, Terman’s ball and field problem, is used to assess “emotionality.” Solutions are assigned to one of 10 categories (5 positive and 5 negative or “problematic”). Validation is based primarily on 165 solutions by 157 children, aged 7-15 years—65 by neurotic children and 100 by “emotionally stable” children (25 of high intelligence; 39, average; 24, low; and 12, mentally defective). The proportion of positive solutions was significantly lower for the neurotics than for any of the first three “normal” groups. Mystified that the percentage passing a subtest of the Stanford-Binet should be almost identical for these three groups, the re-
and communication vectors that actually exist in his family may be reflected, but the test offers no opportunity for verifying the authenticity of these forms of behavior. Thus, interpretations of the family life must be cautiously advanced and tested against reality through observation of other members of the family, especially as they interact. Because the test carries a title that might imply that it is a potential source of information about families, it might be well to recognize that it is essentially a test of a child's attitudes toward his parents and sibs. It does not tap attitudes of all members of a family, and thus does not yield data on habits of mind in all the various interrelationships, e.g., between parents, between parents and children. More precisely, it might be called a test of attitudes toward family rather than a test of family attitudes.

[164] Thematic Apperception Test. Ages 4 and over; 1936-43; commonly known as TAT; individual; 1 form (43); no data on reliability; $6 per set of test materials; 50¢ per manual (43); cash orders postpaid; 100(120) minutes in 2 sessions 1 day apart; Henry A. Murray; Harvard University Press. * (TAT Summary Record Blank). 1952; $1.55 per set of 35 record blanks and manual, postage extra; 20¢ per specimen set, postpaid; Pauline G. Vorhaus; World Book Co. *)

REFERENCES
102-309. See 3:1476.
Carlson, Norah. The Relationship Between Empathy and the Content as Reflected in the TAT. Master’s thesis, University of Toronto (Toronto, Ont., Canada), 1952.
Hasson, Leonard; Cameron, Norman; and Magaret, A. How the Production of a Response Under Mental Distraction: An Investigation of Continuity in Normal and Schizophrenic Language. Abstract. Am Psychol 7:351-2 Ji 52. *
"Facial Similarity Between Subject and Central Figure in the TAT as an Influence on Projection." J Abn & Social Psychol 48:341-4 Jl 53. *(PA 28:2686)


446. Husman, Harris Frederick. "Personality Indicators of Indices, Extent of Personality Factors in Fantasy Responses, and the Ambiguity of TAT Cards." J Consul Psychol 18:345-8 O 54. *(PA 29:5793)

447. La Forge, Rolfe; Leary, Timothy F.; Naboisek, Herbert; Coffey, Hubert S.; and Freedman, Kerwin B. "The Interpersonal Dimension of Personality: II, An Objective Study of the TAT." Personal. 12:33-40 D 54. *(PA 29:5513)


449. Levine, Phyllis R. "Projective Tests in a Vocational

LEONARD D. ERON, Director of Research, Rip Van Winkle Foundation, Hudson, New York.

It was the impression of the reviewer in The Fourth Mental Measurements Yearbook that in the busy clinic the use of the TAT was largely a luxury, since the material obtained by use of this rather "tedious and time-consuming tech- nique" could be more efficiently obtained in a personal interview. He saw the possibility, however, that "with the development of useful scoring systems, such as that of Aron (154), and the establishment of empirically verified principles of interpretation, the test will achieve distinctive value as a psychodiagnostic instrument."

Over the last five years much research with the TAT has been published, but it is still doubt- ful that its utility as an efficient clinical tool has been established and it is uncertain whether the amount of time necessary to evaluate all the nuances of the TAT protocol could not better be spent in other pursuits in behalf of the patient. Aron's scoring system, which is an elaboration and development of the Murray scheme described in the test manual, has proved no boon to the clinician. She reports that a minimum of 10 minutes is required for scoring each story, making it hardly likely that her system can be used economically in the clinic. One complication is that needs and press are scored for everybody in the story, not just the hero. Although this may eliminate the arbitrary judgment sometimes needed in selecting a hero, it makes the
whole procedure that much more unwieldy. The lack of normative and validational data, except for a few hints in studies with very small numbers which did not permit statistical differentiation of groups, as well as the questionable method used in establishing reliability of scoring categories, also makes one hesitate to recommend the Aron scoring procedure for clinical use. It is unlikely that a busy clinician could make more than an impressionistic analysis of the patient’s TAT protocol, especially when it is just one of a battery of tests used in psychodiagnosis. As a timesaving device, group administration of the TAT has been attempted and it has been found that the stories yielded in group administration do not differ significantly in very many ways from stories obtained in routine individual administration (260, 497). Multiple choice answers and objective scoring have been tried (40, 477, 487), but not with too much success, at least for clinical use. In the clinic where patients usually are seen individually and where the clinician is interested in analyzing more than just one or two needs or traits which are being manipulated experimentally or otherwise, it is doubtful that the group method can be adapted in such a way that it would serve as an efficient, timesaving method and, at the same time, give sufficient information about the subject to be of value.

Benton’s second condition has not been fulfilled either. Unfortunately, research has not yielded verification of principles of interpretation which have been reported as successful in the clinic. For example, the traditional “signs” of anxiety in TAT stories have been shown to have little or no relationship to independent “clinical” observation of anxiety in the subjects (451, 580). The same can be said of many other “signs” which have been reported to be clinically useful but which, on independent empirical testing, fall short of validation (204, 503).

Although the usefulness of the TAT as a routine clinical tool has not yet been demonstrated, as a research technique it has had wide and successful application. A number of scoring schemes of good reliability have been introduced (204, 473, 481) and rating scales have been developed which make TAT productions amenable to sound statistical handling without sacrificing too much in the way of clinical judgment (206, 360, 372, 468, 573). These scales have been used in a variety of investigations, both into personality processes, and into the nature of the psychological act of telling stories in response to pictures (299, 362, 438, 491, 516, 521, 529).

It cannot be assumed that, because the subject is presented with an ambiguous picture about which he is instructed to make up a story, the content of his productions will be determined solely, or even chiefly, by his own needs and attitudes. In fact, a number of studies have shown either an inverse or, at best, a curvilinear relationship between degree of ambiguity of stimulus picture and extent of personality factors involved in the response (204, 308, 446). It has been amply demonstrated that each of the pictures has its own “pull” in terms of the thematic content and emotional tone of stories told in response to it. Most of the pictures routinely elicit sad stories, and there are reliable differences among the pictures as to the degree of dysphoric affect, productivity of material, themes, level of response, need systems, etc., which they evoke (204, 206, 251, 369, 585).

The outcome of the stories, however, is one variable which seems to be based more on interpersonal dynamics (204, 206). Stereotyped responses for each of the cards have been described and a fair amount of normative data is now available (147, 175, 204, 388, 538).

Aside from the stimuli properties of the cards themselves, a number of other variables extraneous to the personality content of the individual subject contribute to a determination of both the formal and the content aspects of the productions. The interaction between the examiner and the subject is important. Although it is assumed that the results are a representation of an individual’s private fantasy, it has been shown that the mere presence of an examiner, whether the stories are orally administered or written down by the subject himself, is an inhibiting factor in the production of strongly emotional material on the TAT (521). However, the more the subject is made to feel that he is in a permissive, accepting, noncritical, nonevaluative situation, the more likely is he to contribute fantasies which approximate his unshared ideation and imagery. The examiner can no doubt control some of this atmosphere by the instructions he gives and the manner in which he gives them, by the extra-test comments he makes, and by his general demeanor. There are other factors, however, which are immutable and cannot be changed by the examiner, e.g., sex, age, race, social status, and intel-
All of these variables have been shown to affect TAT productions, especially when there are differences in them between the subject and the experimenter (331, 403, 411, 462, 501, 549).

Quite apart from the kinds of stories elicited by different examiners is the effect of the examiner on the interpretation of the stories which are made. No systematic study of this kind of confounding has been reported, although many authors have warned of the danger of the experimenter's injecting his own theoretical bias, personality shortcomings, and predilections into the interpretations. Davenport (326) found little agreement among six clinical psychologists in their application of 207 statements previously rated for ambiguity, etc., to each of six records from heterogeneous subjects. The judges tended to apply statements rated as universal and loaded with psychoanalytic terminology to any subject, while avoiding use of more specific statements; and they rarely selected statements about positive assets or traits of personality, even though some of the TAT records were from normal individuals.

As in any psychological test, the cooperativeness of the subject is important, and it cannot be assumed that the "cover story" given by the examiner, e.g., "This is a test of intelligence," takes care of the attitudes, set, and preconceived notions of the subject. TAT productions have been shown to be susceptible to distortion when the subject makes a conscious effort to give a specific kind of picture of himself. Individuals can influence the diagnosis of their personalities made by experienced TAT examiners and, to some extent, can manipulate their answers in accordance with their purpose in taking the test (208). However, the subject need not be consciously aware of any effort to distort his stories; he may be set in such a way that it is inevitable that stories will fit in with his predominant attitudes. Differences in TAT stories have been related to physiologically controlled needs such as hunger (302, 534), sex (326, 471, 505), and sleep deprivation,1 hypothetically produced attitudes such as sadness and criticalness,2 and psychologically induced motivations such as need for achievement (170) and need for affiliation (427). Similarly, a number of experiments have shown that conditions directly antecedent to the test administration will affect the productions (496). Although most of these studies have dealt with the effect of specific frustration (21, 29, 258, 479, 482), there is evidence that the immediate surroundings in general have their influence, too (204, 281).

Despite all these seemingly extraneous influences, there still remains a large portion of the individual's idiosyncratic, deep seated motivation that seems to be reflected in his TAT productions. However, the exact nature of this relationship between overt behavior and fantasy as represented by the TAT has yet to be delineated (476). Much of the research concerned with this correlation has centered around the variable of aggression and, indeed, the relationship is not uncomplicated. There is no one-to-one relationship between amount of aggressive need depicted on the TAT and the overt, or even covert, behavior of the subject. The "sign" approach advocated by a number of writers, by which one can supposedly translate what the subject says or fails to say or the way he says it to how he will act (e.g., avoidance of the gun in pictures 3 and 8 means that the subject has to inhibit strong aggressive tendencies, or the use of forceful language or the fantasizing of death or failure in nonheroes signifies the tendency to act out aggression), has been demonstrated to be a failure (547, 582). However, when a theory of behavior is used to posit the relationship between TAT fantasy and overt behavior, results are more successful. For example, Pittluck (305) reasoned that both the aggressive drive and the anxiety opposing expression of this drive must be taken into account when predicting the likelihood of overt aggressive behavior in any individual. The indications of anxiety which she obtained from TAT stories included rejection or denial of aggression, excusing of the aggression by placing it in a socially acceptable context, noncompletion of aggressions planned by a fantasy character, and displacement of the aggression to nonhuman objects. These mechanisms are considered to be defensive in purpose; by their use the aggressive response becomes a compromise between aggressive impulses and the anxiety opposing their expression. It was found that the tendency to use these mechanisms in TAT stories was negatively related to the tendency to act out. The subjects who used more defense mechanisms in proportion to their out-

going, aggressive fantasies tended to act out less than the subjects who used proportionally fewer such mechanisms. In addition, the subjects who used proportionally more unmodified, primitive responses in fantasy tended to act out more than patients with proportionally fewer fantasies of this nature. Therefore, Pittluck concluded that measures of aggressive fantasy can provide direct clues to overt aggressive behavior if these measures stress not the absolute frequency of aggressive responses but the extent to which such responses are free from modifications which are the result of anxiety.

According to behavior theory, anxiety about a given behavior usually results from the association of punishment with that behavior sometime in the past. Mussen and Naylor (455) found that subjects who anticipated punishment for aggression in their TAT stories demonstrated less overt aggression than subjects who did not anticipate such punishment. A further refinement of this relationship, which makes for more efficient prediction from TAT to behavior, is found in a study by Purcell (553) who showed that anticipated internal punishment must be distinguished from retaliatory punishment since the latter variable did not differentiate antisocial from non-antisocial subjects while the former did.

This relationship between aggressive fantasy and overt behavior was more efficiently related to actual learning conditions by Lesser 3 who compared the relationship between these two variables among boys whose mothers encouraged expression of aggression as contrasted with boys whose mothers discouraged such behavior. Where aggressive behavior had maternal encouragement, there was significant positive relationship between aggression scores obtained from stories and behavioral ratings obtained from peers; but where mothers were relatively discouraging of aggression, there was a negative relationship of about the same magnitude. If both groups had been pooled, the correlation would have been no better than zero. If both groups had been pooled, the correlation would have been no better than zero.

The foregoing studies, which have placed analysis of TAT behavior solidly in the main line of current psychological theory, seem to support a positive or representational type of relationship between fantasy and behavior. However, they have dealt only with outwardly directed aggression. An investigation by Davids, Henry, McArthur, and McNamara (475) on inwardly directed aggression invokes cultural pressures to explain the negative relationship found between TAT stories and such behavior. The investigators reason that overt expression of this type of need (intra-aggression) is made difficult in western culture while its expression in fantasy is facilitated. Therefore, there would not necessarily be a relationship between the two methods of expression of this need; indeed, if the need were strong enough and it were difficult to find expression for it in overt behavior, it might very likely then be expressed in fantasy, here represented by TAT stories. This type of validation study, in which one variable at a time is rigorously defined and systematically manipulated or observed in carefully selected subjects who form clearly defined criterion groups to whom the TAT is then administered, seems to yield positive results. Other types, which depend on retrospective accounts (318, 524) or concurrent clinical evaluation (201, 204, 503), are less successful.

In summary, it seems the TAT cannot be used in the clinic as a standardized procedure in the same sense as an intelligence test, although, as one more impressionistic tool in the armamentarium of the clinician, it may have some practical utility. However, the research possibilities of the TAT are manifold. Much of what occurs in the psychological act of telling stories in response to pictures has been clearly delineated. The effect of order of presentation, picture content, presence or absence of color; the influence of the age, sex, race, intelligence, social status, etc., of both the subject and the experimenter; the immediately preceding experience, the set of the subject, the setting in which the experiment is conducted, the method of administration—all have been investigated and their effect assessed. The crucial question of just how TAT fantasy is related to overt behavior has not been so clearly demonstrated. Most of the work has been done in the area of aggression; and the consensus here is that there is a representative relationship between TAT fantasy and behavior, at least for outwardly directed aggression, if a number of modifying mechanisms such as anxiety, and other variables such as learning conditions, are taken into account. For aggression directed inward, the evidence from one study is that the relationship is compensational, and this has been tentatively related to cultural pressures pro-

hibiting overt expression. It should be clearly understood, however, that in none of the studies relating TAT behavior to overt behavior is the obtained relationship ever high enough to permit prediction in individual cases with any degree of confidence. In clinical situations such predictions should not be made without corroboration from additional sources including other test data, interview material, and behavioral cues.


The TAT has now been with us for 23 years and has become one of the three or four best known and most widely used clinical psychological tests. Anyone entering the field had better begin with general reviews of the TAT literature (181, 320, 563), for there are now close to a thousand references on the TAT. Henry (538) gives a very extensive and up-to-date bibliography.

The TAT is not a test that anyone can use after merely studying the manual or a few books on interpretation. In untrained and inexperienced hands it can do more harm than good. It is a test for trained clinical psychologists. Its technique is best learned through practice in a clinical setting under the supervision of a seasoned clinician who is skilled in projective techniques. While it is possible to be a good clinician without knowing the TAT, it is not possible to use the TAT judiciously without being a good clinician. Experience with the TAT is usually gained as a part of the psychologist's clinical training, and expertise with the test seems to be associated with training along "dynamic" or psychoanalytic lines as well as experience in psychotherapy.

ADMINISTRATION. The TAT is perhaps the least standardized of all psychological tests as regards administration, scoring, and interpretation. The instructions to the subject given in Murray's original manual are roughly followed, but few clinicians ever use all 20 cards on one subject. From their own experience clinicians come to have favorite pictures and they sometimes add a few others they think relevant for the subject they are examining. Seldom are more than 10 pictures used. Clinicians have various methods for eliciting fantasy material. Some even ask the subject, "What is the one thing that could not be happening in this picture?" This is claimed to get at repressed psychic content better than the usual method of administration (541). It apparently makes little difference if the stories are obtained orally or are written by the subject, either alone or in a group (260, 497). The thematically "richest" TAT stories the reviewer has seen were written by subjects in a group situation (575).

SCORING. In addition to Murray's original schema and its later variations for scoring "needs" and "presses," there are a number of other scoring schemes (389, 430, 473, 481). In actual practice, however, formal scoring is little used. It is usually thought to be too time-consuming and often seems to miss the individual essence of the subject's production as well as the holistic impression the clinician wishes to obtain. In addition to the themas, attitudes, motivations, and defenses revealed in the stories, the clinician's analysis is based also on the so-called "formal" aspects, such as style, structure, the subject's complaisance with instructions, language characteristics, logical coherence, realism, bizarreness, emotional tone, productivity, and fluency. Perhaps the chief value of the schemes of scoring or tabulating various aspects of TAT productions is for students learning the TAT. Since the several scoring methods analyze the material from somewhat different angles, practice with them is a means of developing sensitivity to the many facets of TAT material that enter into interpretation.

INTERPRETATION. Rather meagre normative data on content and formal characteristics have been published (204, 388), but TAT interpretation is not based on the comparison of "scores" with standard norms. In practice the only "norms" are those held subjectively by the clinician from his own experience with the test. Analysis of as many as 50 to 100 TAT records may be required before one begins to have subjective "norms" for the TAT. It is largely for this reason that clinicians are reluctant to change to new sets of pictures, such as the Symonds Picture-Story Test or Bellak's Children's Apperception Test, for which subjective "norms" have not been accumulated. Murray, the inventor of the TAT, has restated and elaborated some of his ideas on interpretation (278), and there are now a number of good manuals on the art of TAT interpretation (430, 512, 538). There is no best or one authentic method of TAT interpretation. This fact is
demonstrated in Shneidman's book (290) in which each of 15 TAT experts analyzes the same protocol and explains his own method of interpretation. The clinician brings to the task of interpretation all his psychological knowledge, clinical experience, sensitivity, and intuition. The more actual experience the examiner has had with patients, especially if gained through psychotherapy, the more knowledge he has of dynamic psychology, psychoanalysis, and other projective techniques, the more meaning will he derive from the TAT. It is generally agreed that the TAT should not be interpreted “blind,” for then it is too apt to miss the mark by far and have no value in “elucidating” the case history material. TAT interpretations tend more to ring true when they are made in conjunction with the case history and with impressions gained from interviews and other tests.

RELIABILITY. The question of reliability has been quite neglected in the TAT literature.\(^1\) Reliability of scoring, of internal consistency, of test-retest, and of interpretation must be evaluated separately.

In searching the TAT literature, the writer has found only 15 estimates of scoring reliability based on sound statistical methods and presented in the form of the product-moment correlation coefficient so as to be strictly comparable to the usual measures of test reliability. These reliability coefficients range from .54 to .91, with an average of .77. For reliability of scoring (i.e., interscorer agreement), these figures must be considered quite low. Scoring reliability below .80 is generally considered unacceptable in scoring essay examinations, for example.

There is a widely held misconception that split-half or internal consistency reliability is meaningless in the TAT. Actually it is no less meaningful in the case of the TAT than for any other test comprised of a number of elements which are combined into some kind of “score.” A proper coefficient of internal consistency for any TAT variable may be obtained by the Kuder-Richardson formula or by a rank correlation method. When the proper technique was applied (524) to 10 of the major Murray TAT variables (Achievement, Aggression, Autonomy, etc.) the internal consistency reliability of the various themes ranged from −.07 to +.34, with a mean of .13. These reliabilities are typical of most internal consistency measures on the TAT (497). What they mean in practice is that any scoring system based on the addition of themes elicited by various pictures is fallacious. A theme on one card is not sufficiently correlated with the same theme on another card to justify an additive treatment of TAT variables. It would be like adding together pounds, gallons, and inches. Each card seems to be a unique test in itself and is correlated little, if at all, with other cards (248). This fact casts serious doubt on the validity of many methods of TAT interpretation.

Test-retest reliability estimates are rare and are usually more a measure of the subject’s memory for his first productions. When subjects were required to make up different stories on retest, the reliability coefficients of only 3 out of 17 scored variables were significantly greater than zero (497). McClelland (406) reports a retest reliability (1 week interval) of .22 for his quantitatively scored n Achievement.

Reliability of interpretation is a more important consideration. Friedeman (572) found the correlations (from a Q-sort) between different interpreters’ ideas about the characteristics of the TAT “hero” to average .74, with a range from .37 to .88 for various protocols. This study unfortunately tells us nothing about the discriminating power of the TAT with respect to subjects, but indicates only the fact that there is some agreement between interpreters about the manifest characteristics of the central figure in the stories. Davenport (329) had six clinicians rate six TAT records on 207 typical interpretive statements as they applied to each record. The major finding was the lack of reliable discrimination. There was little agreement among the judges in the differential use of the statements for the six TAT records. The judges tended to apply statements rated as universal to almost any patient while avoiding the use of more specific statements. They rarely made statements about positive aspects of personality even though normal subjects were used.

VALIDITY. With such low reliability it is not surprising to find that the validity of the TAT is practically nil. But in discussing validity, one must distinguish two main classes of variables derived from the TAT protocol: thematic material and formal characteristics (style, lan-
As to be of no practical value in clinical prediction. The median correlations between two raters using case history material and the TAT interpreter were .19 and .28. Most of the significant correlations were based on formal characteristics of the TAT material. In terms of predictive power, Winch and More (502) found that the TAT adds nothing significant to information gained in an interview. Murray (278) and others have argued that the real proof of the TAT would be the correspondence between TAT material and the deeper layers of personality which are revealed only in the process of psychotherapy. Murray has mentioned one case in which the TAT "adumbrated all the chief trends which five months of analysis were able to reveal." Studies based on larger samples have not found much correspondence between TAT and therapy material. Meyer and Tolman (502) sought a correspondence between attitudes concerning parents expressed in psychotherapeutic interviews and in TAT protocols. There was "no predictability from TAT to therapy as to whether or not parents were discussed, and when they were discussed, no similarity was found between those attitudes and images given in TAT stories and in psychotherapy." Saxe (233) had a TAT clinician rate a personality questionnaire tapping typical TAT variables on 20 patients. After the patients had undergone four months of psychotherapy, the therapist rated the patients on the same questionnaire. There was greater than chance (5 per cent level) agreement in only half of the cases.

If the TAT is short on actual validity, it certainly is not lacking in what might be called "subjective validity" (akin to "faith validity"). This is a feeling gained by the clinician using the TAT that it contributes something to his...
understanding of the case. Some psychologists have a greater capacity than others for experiencing subjective validity. This capacity seems to be associated with training and experience in psychoanalysis, psychotherapy, and projective techniques in general. The TAT also provides the clinical psychologist with the kind of dynamically interpretable material that can be appreciated by the psychoanalytically oriented psychiatrist to whom the clinician addresses his report. Thus clinicians are heard to speak of the TAT as being “useful” rather than as having validity. It is probably for these reasons that the TAT survives in clinical practice.

SUMMARY. The TAT is a nonstandardized assessment technique which is best left to clinical psychologists who have had special training in its use. While research has shown the TAT to have low reliability and negligible validity, many clinical psychologists continue to use it, apparently with some satisfaction.

For a review by Arthur L. Benton, see 4:136; for reviews by Arthur L. Benton, Julian B. Rotter, and J. R. Wittenborn, see 3:103 (1 excerpt); for related reviews, see B63, B204, B395, 4:137-41, 3:104, and 3:104a.

REFERENCE


Mary D. Ainsworth, Associate Professor of Psychology, The Johns Hopkins University, Baltimore, Maryland.

Impressed with the fruitfulness of the thematic apperception approach, Lee attempted to use the standard TAT materials for the investigation of the personality of Zulu subjects, but found that they were not adequately stimulated to imaginative production. He therefore devised his own set of pictures for use with African subjects. In order to "cross the cultural gap" and to arrive at pictorial materials that would be stimulating, he based his pictures on fantasy productions collected from Bantu inmates of a mental hospital.

His version of the TAT consists of 22 cards, 8 for males, 8 for females, and 6 (including a blank card) for both males and females. In some respects the cards impress this reviewer as being more ambiguous than the cards of the standard TAT. The faces of the figures are either highly ambiguous in expression or hidden from view. The backgrounds include little detail. The line of the drawings is more sketchy and hence less structured than that of the standard TAT drawings. On the other hand, many of the figures are portrayed in vigorous action or exaggerated posture that seems less ambiguous than that of the figures of the standard TAT and might be expected to be highly provocative of kinaesthetic empathy.

There seems to have been no attempt systematically to vary the number, sex, and apparent age of the figures in order to sample various types of relationships. Two cards, both in the female series, seem designed to elicit stories of parent-child relations, presumably from the mother's viewpoint. One card portrays a heterosexual situation. However, most of the pictures present single figures, and only six show two or more figures together.

Although the pictures were originally designed for the Zulu and contain some characteristically Zulu features, Lee reports that they have been used effectively among other African peoples (he specifies the tribes) and among both educated and uneducated subjects. Nevertheless, it may not be assumed that his Thematic Apperception Test for African Subjects is therefore applicable to all African peoples. The fact that 12 of the cards depict near-naked figures would undoubtedly be a disadvantage with tribes such as the Ganda of East Africa who traditionally clothe themselves from top to toe and consider it immodest to display the feet when sitting.

The 42-page manual provides in concise form an excellent guide for the administration and interpretation of a TAT-type test. The initial instructions contain all the essential points included in Murray's original instructions, but are worded more simply and might well be adapted to good effect in administering the standard TAT. Lee recommends that a recall phase be included at the conclusion of the storytelling phase, in which the subject is asked to recall as many of the pictures as he can in as much detail as possible. He further recommends a follow-up interview when the subject is asked to explain the sources of his plots.

Lee's suggestions for analysis and interpretation emphasize the form as well as the con-