

Using the Rorschach and Thematic Apperception Tests in Personality Assessment

John Q. Student

Doane College-Lincoln

Clinical Assessment, COU 612

Spring, 2002

*NOTE: This paper has been updated to reflect the 2010 APA Publication Manual changes included in the 2010 Quick Guide to APA Style.*

### Abstract

Despite criticism over many years because of limited reliability and validity, projective techniques continue to be widely used in personality assessment. The Rorschach Inkblot Test and the Thematic Apperception Test have both been shown to have adequate reliability when scoring criteria are clearly designated and when interpreters are well trained. Both tests have shown validity in specific areas of personality and motivation, but neither is generally effective in identification of psychiatric disorders. It is suggested that they may play a role as additional tools to be used with standardized assessments, rather than instead of them..

### Using the Rorschach and Thematic Apperception Tests in Personality Assessment

There are many personality assessment procedures available, which vary with respect to structure and content. Certain tests are highly structured with carefully selected questions and a limited range of possible responses that are objectively administered and scored. Other tests, referred to as projective personality tests, are less structured and allow for a variety of responses. It is the hope of these tests to uncover the major complexities of personality in order to better understand the entire individual (Rose, Kaser-Boyd, & Maloney, 2001).

According to Frank (1948), responses to projective techniques reflect needs, motives, feelings, experiences, and thought processes automatically, which is known as the projective hypothesis. All projective techniques allow individual interpretation of task demands and in organizing responses (Teglasi, 2001). Responses are usually in the form of completed sentences, associations, or descriptive or storytelling responses (Anastasi, 1988).

Despite much criticism over the years, projective techniques continue to be administered by professionals today. A survey conducted by Watkins, Campbell, Nieberding, and Hallmark (1995) of clinical psychologists revealed that the most popular personality tests were the Minnesota Multiphasic Personality Inventory (MMPI-2), Rorschach Inkblot Test, and the Thematic Apperception Test (TAT). This survey also revealed that both the Rorschach and the TAT ranked in the top 10 of the most frequently used assessment tools. This paper will examine these two most widely-used projective personality tests.

### **Rorschach Inkblot Test**

The Rorschach Inkblot Test was created by H. Rorschach, based on his variation of the child's game Blotto (Hess, Zachar, & Kramer, 2001). Since its early development, the Rorschach

has received critical and controversial reviews. Much of the development of the test has occurred in the years after Rorschach's death in 1922 (Rose et al., 2001).

The Rorschach consists of 10 plates, each containing a symmetrical inkblot. Five blots appear in black and white, two appear in black, white, and red, and three are composed of pastel colors (Hess et al., 2001). Each inkblot is then presented one at a time and the responder is asked to respond to each inkblot separately. Responders are instructed to describe what the inkblot might be. If only one response is given, the examiner might ask if the inkblot looks like anything else, often resulting in more than one response (Hess et al.).

In the 1950s and 1960s, the test was criticized because it lacked standardized procedures and a set of norms (Lilienfeld, Wood, & Garb, 2001). During this time many psychologists adopted Rorschach's system for scoring the assessment. However, others developed their own interpretive systems of scoring the test (Hess et al., 2001). This led Exner (1974) to develop the Comprehensive System, which established rules for delivering the test and for interpreting the responses, as well as providing norms for both children and adults.

### **Reliability**

Tests with open-ended responses, such as the Rorschach, must be scored subjectively. This raises a question for consistency between various raters, known as inter-rater reliability (Hess et al., 2001). Finding evidence for inter-rater reliability for the Rorschach is not easy because there are so many scoring categories. In regard to the Comprehensive System, trained scorers agree with expert scorers between 88% and 99% of the time (Hess et al.). Other research reveals that strong agreement is achieved for only about half of the characteristics examined by those who score Rorschach responses (Lilienfeld et al., 2001). Inter-rater reliability may also be low in examiners who are not well trained or experienced (Rose et al., 2001).

Test-retest reliability, or the ability of a test to measure a trait that is supposed to be stable over time (Hess et al., 2001), is also important. Many variables in the Comprehensive System show reliability estimates at or above .70 at both 1-year and 3-year intervals (Exner, 1974). Exner and Weiner (1995) have noted however that test-retest reliability estimates of the Comprehensive System with children are low. These reliability estimates increase as participants reach later adolescence. Exner and Weiner attribute this to the rapid developmental changes that occur during childhood. The test-retest reliability of the Comprehensive System is significantly higher.

### **Validity**

Surveys of clinical settings such as Piotrowski and Keller's (1989) study of 413 outpatient mental health facilities, found that more than 80% of the responding agencies used the Rorschach. Watkins et al. (1995) survey of 412 clinical psychologists revealed that 82% of the respondents used the Rorschach. Because the Rorschach is an ambiguous, unstructured task, it produces data that may not be produced by self-report measures (Rose et al., 2001). The Rorschach has also been the topic of many studies. It ranks second to the MMPI in research articles published between 1974 and 1994 (Rose et al.).

Meta-analytic studies have found that validity coefficients for the Rorschach are not significantly different from those for the MMPI (Atkinson, 1986; Parker, Hanson, & Hunsley, 1988). Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neuleib (1999) also reported similar findings. They found the Rorschach to have a mean effect size of .26 and those of the MMPI to have a mean effect size of .37, both respectable for personality tests. Much of the criticism of the Rorschach results from research conducted before the use of the Comprehensive System (Rose et al., 2001).

Despite these findings, the Rorschach is not accurate in identifying most psychiatric

conditions, with the exceptions of schizophrenia and bipolar disorder (Lilienfeld et al., 2001). The Rorschach does not consistently detect depression, anxiety disorders, or psychopathic personality. Other evidence suggests that the Rorschach norms meant to distinguish mental health from mental illness are unrepresentative of the United States population and make many adults and children seem maladjusted (Lilienfeld et al.). Results from the inkblots may be even more misleading for minorities. Results have revealed that scores from African-Americans, Native Americans, Native Alaskans, Hispanics, and Central and South Americans differ significantly from the norms (Lilienfeld et al.).

### **Administration, Scoring, and Interpretation**

Administration procedures for the Rorschach are fairly simple and standardized (Rose et al., 2001). Administration procedures are the same for both children and adults, and scoring rules are also standardized. The Rorschach provides data about cognitive and emotional variables that can be analyzed quantitatively and qualitatively and interpreted from different theoretical perspectives. This test can also be integrated with data from other psychological tests because the language is consistent with general psychological descriptions (Rose et al.).

One weakness of interpretation is that advanced training is required (Rose et al., 2001). Interpretation is also the least well-taught subject in basic Rorschach training. Examiners are also required to become familiar with Rorschach research, especially for the population he or she is testing, to become better prepared to administer the test (Rose et al.).

### **Thematic Apperception Test**

By the time the TAT was being developed, the Rorschach was beginning to gain popularity. The creators of the TAT, C. Morgan and H. Murray, utilized storytelling about

pictured scenes to elicit motives, intentions, and expectations (Teglasi, 2001). Murray developed a theoretically based system for interpreting the stories told, but the variability of the storytelling technique eventually led to many interpretive approaches for the TAT pictures (Teglasi). Other storytelling assessment instruments include Robert's Apperception Test for Children, Tell-Me-A-Story, and the Children's Apperception Test.

The TAT consists of 31 black-and-white picture cards, most containing people, and one card completely blank (Hood & Johnson, 1997). Examples include a boy thinking about a violin and a confused woman holding an open door. Typically, 20 cards are presented in a test administration. The selection of cards presented to respondents varies with clinicians (Lilienfeld et al., 2001). Examinees are asked to tell a story about each picture including as much detail as possible (Hood & Johnson).

### **Reliability**

Inter-rater reliability of the TAT is similar to that of the Rorschach. Because storytelling responses are open-ended, the task of finding evidence for inter-rater reliability is not easy. Inter-rater reliability for thematic techniques tends to be high when interpretive criteria are clearly designated and interpreters are well trained in the rating procedure. Under these conditions, agreement between raters often exceeds the .80 to .85 range, which is generally considered adequate (Lundy, 1985).

There are two important considerations regarding test-retest reliability: whether the focus is on similarity of story content or similarity of the clinician's judgment (Lundy, 1985), and whether the personality construct under consideration is relatively stable or fluctuating (Cramer, 1996). The reliability of the specific content is less relevant than consistency of the interpretive meaning of the response (Teglasi, 2001). Stories of the TAT tend to remain stable, at least for

short periods of time (Locraft & Teglassi, 1997). However, other research has indicated that some of the standardized scoring systems display weak test-retest reliability over longer periods of time (Lilienfeld et al., 2001).

### **Validity**

Tomkins (1947) did not believe it was appropriate to ask if the TAT is a valid test. Rather, he thought it was important to question whether inferences based on the test are likely to be true. In other words, to what extent do scores from a particular TAT measure relate to other criteria in a predicted manner?

To demonstrate forms of construct validity, a relationship must exist between the TAT measure and some independent measure of personality (Cramer, 1996). Measures of the TAT are often criticized because correlations obtained are not as high as those obtained from self-report questionnaires. Researchers of the TAT note that although questionnaire scores tend to correlate well with scores from other similar questionnaires, they are less successful in predicting real behavior. Research has indicated that TAT scores do correlate with independent measures of behavior in meaningful ways (Cramer). The TAT and self-report measures of achievement motivation both correlate with external criteria, but the patterns of correlations are different. These differences support the conclusion that self-report and storytelling measure different achievement-related constructs, and they cannot be used interchangeably (Teglassi, 2001).

Other studies have questioned TAT validity. Several scoring systems have been unable to differentiate normal individuals from those who are psychotic or depressed. Other standardized scoring systems for the TAT do appear to be accurate in determining aspects of personality, notably the need to achieve and a person's perceptions of others (Lilienfeld et al., 2001).

### **Administration, Scoring, and Interpretation**

The TAT is confusing in part because its administration is usually not standardized. Most clinicians also interpret people's stories intuitively instead of following a well-tested scoring procedure. Some evidence suggests that clinicians who interpret the TAT in an intuitive way are likely to overdiagnose psychological disturbance (Lilienfeld et al., 2001).

The value of interpretation is determined by its effect. According to Cramer (1996), the goal is to understand the interpretation of each story. Rather than searching for evidence from the past that might validate an interpretation of the TAT, looking to the future might explain aspects of the storyteller's life or show the interpretation to be false. The criterion for a good interpretation is whether it will produce new meanings, clarify current struggles, or demonstrate new relationships.

### **Discussion**

Projective techniques are widely used in personality assessment. Piotrowski and Zalewski (1993) indicate that many academic psychologists have expressed the belief that knowledge of projective testing is not as important as it used to be and that use of projective tests will likely decline in the future. However, most PhD and PsyD clinical psychology training programs include formal instruction in the use of the Rorschach, and the majority of the internship sites approved by the American Psychological Association place a high value on knowledge of the method (Durand, Blanchard, & Mindell, 1988).

For the Rorschach, how reliable and valid the results are appears to depend, at least in part, on whether or not the examiners are well trained and experienced, and whether or not they are using Exner's (1974) Comprehensive System (Hess et al., 2001; Rose et al., 2001). When the

Comprehensive System is used and examiners are well trained, reliabilities are acceptable, and meta-analytic studies have shown validity of the Rorschach to be comparable to that of the MMPI. However, the Rorschach appears unable to identify accurately most psychiatric disorders, except for schizophrenia and bipolar disorder (Lilienfeld et al., 2001).

The TAT has been shown to have adequate reliability when interpretive criteria are clearly designated and interpreters are well trained (Lundy, 1985). In the area of validity, Cramer (1996) has shown that TAT scores correlate with independent measures of behavior in meaningful ways, particularly in relation to achievement motivation. However, Lilienfeld et al. (2001) have shown that several scoring systems have been unable to differentiate normal individuals from those who are psychotic or depressed.

Perhaps the issue is not using projective techniques versus using self-report measures. Recent trends have focused on the use of the Rorschach in addition to the MMPI-2, rather than as an alternative, which allows for the integration of the strengths of these two approaches (Weiner, 1993). Some mental health professionals argue that projective techniques have a long history of constructive use and, when administered and interpreted properly, can provide a clear picture of the mind, just as good, if not better than self-reports (Lilienfeld et al., 2001).

Although much criticism and confusion surrounds the Rorschach and the TAT regarding reliability and validity, the tests continue to be used. The TAT is used as a clinical assessment tool and research instrument to study motivation and fantasy (Butcher & Rouse, 1996). Although much of the research on the TAT centers around nonclinical studies with normal populations, clinical samples are still emphasized. Research continues to be conducted on norms and on the development of scoring methods (Butcher & Rouse).

The future of the Rorschach and the TAT are uncertain. They continue to be popular

psychological assessment instruments, although they have been heavily criticized. Because of the interpretation needed for these projective instruments, opinions and research vary greatly. Will these instruments give way to objective self-reports, or will they be able to survive in the world of psychological testing?

### References

- Anastasi, A. (1988). *Psychological testing* (6<sup>th</sup> ed.). New York: Macmillan.
- Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychology, 27*, 238-347.
- Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47*, 87-111.
- Cramer, P. (1996). *Storytelling, narrative, and the Thematic Apperception Test*. New York: Guilford.
- Durand, V. M., Blanchard, E. B., & Mindell, J. A. (1988). Training in projective testing: Survey of clinical training directors and internship directors. *Professional Psychology: Research and Practice, 19*, 236-238.
- Exner, J. E. (1974). *The Rorschach: A comprehensive system* (Vol. 1). New York: Wiley.
- Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system, assessment of children and adolescents* (Vol. 3, 2<sup>nd</sup> ed.). New York: Wiley.
- Frank, L. D. (1948). *Projective methods*. Springfield, IL: Thomas.
- Hess, A. K., Zachar, P., & Kramer, J. (2001). Review of Rorschach Inkblot Test. From B. S. Plake & J. C. Impara (Eds.), *The fourteenth mental measurement yearbook*. (14122595).
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*, 278-296.
- Hood, A. B., & Johnson, R. W. (1997). *Assessment in counseling: A guide to the use of psychological assessment procedures* (2<sup>nd</sup> ed.). Alexandria, VA: American Counseling Association.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2001). What's wrong with this picture? *Scientific American, 284*, 80-87.
- Locraft, C., & Teglassi, H. (1997). Teacher rated empathic behavior and children's TAT stories. *Journal of School Psychology, 35*, 217-237.
- Lundy, A. (1985). The reliability of the Thematic Apperception Test. *Journal of Personality Assessment, 49*, 141-149.

- Parker, K. C., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach and WAIS: A meta-analytic comparison of reliability, stability and validity. *Psychological Bulletin, 103*, 367-373.
- Piotrowski, C., & Keller, J. W. (1989). Psychodiagnostic testing in APA-approved clinical psychology programs. *Professional Psychology, 15*, 450-456.
- Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical psychology programs. *Journal of Personality Assessment, 61*, 394-405.
- Rose, T., Kaser-Boyd, N., & Maloney, M.P. (2001). *Essentials of Rorschach assessment*. New York: Wiley.
- Teglasi, H. (2001). *Essentials of TAT and other storytelling techniques assessment*. New York: Wiley.
- Tomkins, S. (1947). *The Thematic Apperception Test*. New York: Grune & Stratton.
- Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*, 54-60.
- Weiner, I. B. (1993). Clinical considerations in the conjoint use of the Rorschach and the MMPI. *Journal of Personality Assessment, 60*, 148-152.