What's Wrong with This Picture

Photographs by Jelle Wagenaar

Psychologists often use the famous Rorschach Inkblot test and related tools to assess personality and mental illness. But research says the instruments are frequently ineffective for those purposes.
What if you were asked to describe images you saw in an inkblot or to invent a story for an ambiguous illustration—say, of a middle-aged man looking away from a woman who was grabbing his arm? To comply, you would draw on your own emotions, experiences, memories and imagination. You would, in short, project yourself into the images. Once you did that, many practicing psychologists would assert, trained evaluators could mine your musings to reach conclusions about your personality traits, unconscious needs and overall mental health.

But how correct would they be? The answer is important because psychologists frequently apply such “projective” instruments (presenting people with ambiguous images, words or objects) as components of mental assessments, and because the outcomes can profoundly affect the lives of the respondents. The tools often serve, for instance, as aids in diagnosing mental illness, in predicting whether convicts are likely to become violent after being paroled, in evaluating the mental stability of parents engaged in custody battles, and in discerning whether children have been sexually molested.

We recently reviewed a large body of research into how well projective methods work, concentrating on three of the most extensively used and best-studied instruments. Overall our findings are unsettling.

Butterflies or Bison?
The famous Rorschach inkblot test—which asks people to describe what they see in a series of 10 inkblots—is by far the most popular of the projective methods, given to hundreds of thousands, or perhaps millions, of people every year. The research discussed below refers to the modern, rehabilitated version, not to the original construction, introduced in the 1920s by Swiss psychiatrist Hermann Rorschach.

The initial tool came under severe attack in the 1950s and 1960s, in part because it lacked standardized procedures and a set of norms (averaged results from the general population). Standardization is important because seemingly trivial differences in the way an instrument is administered can affect a person’s responses to it. Norms provide a reference point for determining when someone’s responses fall outside an acceptable range.

In the 1970s John E. Exner, Jr., then at Long Island University, ostensibly corrected those problems in the early Rorschach test by introducing what he called the Comprehensive System. This set of instructions established detailed rules for delivering the inkblot exam and for interpreting the responses, and it provided norms for children and adults.

In spite of the Comprehensive System’s current popularity, it generally falls short on two crucial criteria that were also problematic for the original Rorschach: scoring reliability and validity. A tool possessing scoring reliability yields similar results regardless of who grades and tabulates the responses. A valid technique measures what it aims to measure: its results are consistent with those produced by other trustworthy instruments or are able to predict behavior, or both.

To understand the Rorschach’s scoring reliability defects, it helps to know something about how reactions to the inkblots are interpreted. First, a psychologist rates the collected reactions on more than 100 characteristics, or variables. The evaluator, for instance, records whether the person looked at whole blot or just parts, notes whether the detected images were unusual or typical of most test takers, and indicates which aspects of the
be obtained in other, more practical ways, such as by conducting interviews or administering objective personality tests. (Objective tests seek answers to relatively clear-cut questions, such as, "I frequently have thoughts of hurting myself—true or false?".) This shortcoming of projective tools makes the costs in money and time hard to justify.

What to Do?

Some mental health professionals disagree with our conclusions. They argue that projective tools have a long history of constructive use and, when administered and interpreted properly, can cut through the veneer of respondents' self-reports to provide a picture of the deepest recesses of the mind. Critics have also asserted that we have emphasized negative findings to the exclusion of positive ones.

Yet we remain confident in our conclusions. In fact, as negative as our overall findings are, they may paint an overly rosy picture of projective techniques because of the so-called file drawer effect. As is well known, scientific journals are more likely to publish reports demonstrating that some procedure works than reports finding failure. Consequently, researchers often quietly file away their negative data, which may never again see the light of day.

We find it troubling that psychologists commonly administer projective instruments in situations for which their value has not been well established by multiple studies; too many people can suffer if erroneous diagnostic judgments influence therapy plans, custody rulings or criminal court decisions. Based on our findings, we strongly urge psychologists to curtail their use of most projective techniques and, when they do select such instruments, to limit themselves to scoring and interpreting the small number of variables that have been proved trustworthy.

Our results also offer a broader lesson for practicing clinicians, psychology students and the public at large: even seasoned professionals can be fooled by their intuitions and their faith in tools that lack strong evidence of effectiveness. When a substantial body of research demonstrates that old intuitions are wrong, it is time to adopt new ways of thinking.

### HOW OFTEN THE TOOLS ARE USED

#### Popularity Poll

In 1995 a survey asked 412 randomly selected clinical psychologists in the American Psychological Association how often they used various projective and nonprojective assessment tools, including those listed below. Projective instruments present people with ambiguous pictures, words or things; the other measures are less open-ended. The number of clinicians who use projective methods might have declined slightly since 1995, but these techniques remain widely used.

<table>
<thead>
<tr>
<th>Projective Techniques</th>
<th>Use Always or Frequently</th>
<th>Use At Least Occasionally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rorschach</td>
<td>43%</td>
<td>82%</td>
</tr>
<tr>
<td>Human Figure Drawings</td>
<td>39%</td>
<td>80%</td>
</tr>
<tr>
<td>Thematic Apperception Test (TAT)</td>
<td>34%</td>
<td>82%</td>
</tr>
<tr>
<td>Sentence Completion Tests</td>
<td>34%</td>
<td>84%</td>
</tr>
<tr>
<td>CAT (Children’s version of the TAT)</td>
<td>6%</td>
<td>42%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonprojective Techniques*</th>
<th>Use Always or Frequently</th>
<th>Use At Least Occasionally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weschler Adult Intelligence Scale (WAIS)</td>
<td>59%</td>
<td>93%</td>
</tr>
<tr>
<td>Minnesota Multiphasic Personality Inventory-2 (MMPI-2)</td>
<td>58%</td>
<td>85%</td>
</tr>
<tr>
<td>Weschler Intelligence Scale for Children (WISC)</td>
<td>42%</td>
<td>69%</td>
</tr>
<tr>
<td>Beck Depression Inventory</td>
<td>21%</td>
<td>71%</td>
</tr>
</tbody>
</table>

* Those listed are the most commonly used nonprojective tests for assessing adult IQ (WAIS), personality (MMPI-2), childhood IQ (WISC) and depression (Beck Depression Inventory).


### MORE TO EXPLORE


inky swirls (such as form or color) most determined what the respondent reported seeing.

Then he or she compiles the findings into a psychological profile of the individual. As part of that interpretive process, psychologists might conclude that focusing on minor details (such as stray splatters) in the blots, instead of on the whole images, signals obsessiveness in a patient and that seeing things in the white spaces within the larger blots, instead of in the inked areas, reveals a negative, contrary streak.

For the scoring of any variable to be considered highly reliable, two different assessors should be very likely to produce similar ratings when examining any given person’s responses. Recent investigations demonstrate, however, that strong agreement is achieved for only about half the characteristics examined by those who score Rorschach responses; evaluators might well come up with quite different ratings for the remaining variables.

Equally troubling, analyses of the Rorschach’s validity indicate that it is poorly equipped to identify most psychiatric conditions—with the notable exceptions of schizophrenia and other disturbances marked by disordered thoughts, such as bipolar disorder (manic-depressive). Despite claims by some Rorschach proponents, the method does not consistently detect depression, anxiety disorders or psychopathic personality (a condition characterized by dishonesty, callousness and lack of guilt).

Moreover, although psychologists frequently administer the Rorschach to assess propensities toward violence, impulsiveness and criminal behavior, most research suggests it is not valid for these purposes either. Similarly, no compelling evidence supports the use for detecting sexual abuse in children.

Other problems have surfaced as well. Some evidence suggests that the Rorschach norms meant to distinguish mental health from mental illness are unrepresentative of the U.S. population and mistakenly make many adults and children seem maladjusted. For instance, in a 1999 study of 123 adult volunteers at a California blood bank, one in six had scores supposedly indicative of schizophrenia.

The inkblot results may be even more misleading for minorities. Several investigations have shown that scores for African-Americans, Native Americans, Native Alaskans, Hispanics, and Central and South Americans differ markedly from the norms. Together the collected research raises serious doubts about the use of the Rorschach inkblots in the psychotherapy office and in the courtroom.

Doubts about TAT

ANOTHER PROJECTIVE TOOL—the Thematic Apperception Test (TAT)—may be as problematic as the Rorschach. This method asks respondents to formulate a story based on ambiguous scenes in drawings on cards. Among the 31 cards available to psychologists are ones depicting a boy contemplating a violin, a distraught woman clutching an open door, and the man and woman who were mentioned at the start of this article. One card, the epitome of ambiguity, is totally blank.

The TAT has been called “a clinician’s delight and a statistician’s nightmare,” in part because its administration is usually not standardized: different clinicians present different numbers and selections of cards to respondents. Also, most clinicians interpret people’s stories intuitively instead of following a well-tested scoring procedure. Indeed, a recent survey of nearly 100 North

Once deemed an “x-ray of the mind,” the Rorschach inkblot test remains the most famous—and infamous—projective psychological technique. An examiner hands 10 symmetrical inkblots one at a time in a set order to a viewer, who says what each blot resembles. Five blots contain color; five are black and gray. Respondents can recall images. The one above is an inverted Andy Warhol rendering; the actual Rorschach blots cannot be published.

Responses to the inkblots purportedly reveal aspects of a person’s personality and mental health. Advocates believe, for instance, that references to moving animals—such as the dinosaurs mentioned above—often indicate impulsiveness, whereas allusions to a blot’s “blackness”—in the spaceship—often indicate depressiveness.

Swiss psychiatrist Hermann Rorschach probably got the idea of showing inkblots from a European parlor game. The test debuted in 1921 and reached high status by 1945. But critical backlash began taking shape in the 1950s, as researchers found that psychologists often interpreted the same responses differently and that particular responses did not correlate well with specific mental illnesses or personality traits.

Today the Comprehensive System, more a measure of weaknesses, is widely used to score and interpret Rorschach responses. But it has been criticized on similar grounds. Moreover, several recent findings indicate the Comprehensive System incorrectly labels many normal respondents as pathological.
American psychologists presume in twin and family courts discovered that 3 percent relied on a standardized TAT scoring system. Unfortunately, some evidence suggests that clinicians who interpret the TAT in an intuitive way are likely to overdiagnose psychological disturbance.

Many standardized scoring systems are available for the TAT, but some of the more popular ones display weak "test-retest" reliability: they tend to yield inconsistent scores from one picture-viewing session to the next. Their validity is frequently questionable as well; studies that find positive results are often contradicted by other investigations. For example, several scoring systems have proved unable to differentiate normal individuals from those who are psychotic or depressed.

A few standardized scoring systems do appear to do a good job of discerning certain aspects of personality—namely the need to achieve and a person's perceptions of others, a property called "object relations." But many times individuals who display a high need to achieve do not score well on measures of actual achievement, so the ability of that variable to predict a person's behavior may be limited. These scoring systems currently lack norms and so are

THEMATIC APPERCEPTION TEST

The Thematic Apperception Test (TAT), created by Harvard University psychiatrist Henry A. Murray and his student Christiana Morgan in the 1930s, is among the most commonly used projective measures. Examiners present individuals with a subset (typically five to 12) of 31 cards displaying pictures of ambiguous situations, mostly featuring people. Respondents then construct a story about each picture, describing the events that are occurring, what led up to them, what the characters are thinking and feeling, and what will happen later. Many variations of the TAT are in use, such as the Children's Apperception Test, featuring animals interacting in ambiguous situations, and the Blacky Test, featuring the adventures of a black dog and its family.

Psychologists have several ways of interpreting responses to the TAT. One promising approach—developed by Boston University psychologist Drew Westen—relies on a specific scoring system to assess people's perceptions of others ("object relations"). According to that approach, if someone wove a story about an older woman plotting against a younger person in response to the image visible in the photograph at the right, the story would imply that the respondent tends to see malevolence in others—but only if similar themes turned up in stories told about other cards.

Surveys show, however, that most practitioners do not use systematic scoring systems to interpret TAT stories, relying instead on their intuitions. Unfortunately, research indicates that such "impressionistic" interpretations of the TAT are of doubtful validity and may make the TAT a projective exercise for both examiner and examinee.
OTHER PROJECTIVE TOOLS

What's the Score?

Psychologists have dozens of projective methods to choose from beyond the
Rorschach Test, the TAT and figure drawings. As the sampling below indicates,
some stand up well to the scrutiny of research, but many do not.

Hand Test
Subjects say what hands pictured in various positions might be doing. This method is
used to assess aggression, anxiety and other personality traits, but it has not been
well studied.

Handwriting Analysis (Graphology)
Interpreters rely on specific “signs” in a person’s handwriting to assess personality
characteristics. Though useless, the method is still used to screen prospective
employees.

Lüscher Color Test
People rank colored cards in order of preference to reveal personality traits.
Most studies find the technique to lack merit.

Play with Anatomically Correct Dolls
Research finds that sexually abused children often play with the dolls’ genitalia;
however, that behavior is not diagnostic, because many nonabused children
do the same thing.

Rosenzweig Picture Frustration Study
After one cartoon character makes a provocative remark to another, a viewer
decides how the second character should respond. This instrument, featured in the
movie A Clockwork Orange, successfully predicts aggression in children.

Sentence Completion Test
Test takers finish a sentence, such as, “If only I could...” Most versions are poorly
studied, but one developed by Jane Loewinger of Washington University is valid for
measuring aspects of ego development, such as morality and empathy.

Szondi Test
From photographs of patients with various psychiatric disorders, viewers select
the ones they like most and least. This technique assumes that the selections reveal
something about the choosers’ needs, but research has discredited it.

Even when projective methods assess what they claim to
measure, they
to informa'
that can be obtained in other, more practical ways.
MISLEADING SIGNS

Psychologists have many projective drawing instruments at their disposal, but the Draw-a-Person Test is among the most popular—especially for assessing children and adolescents. A clinician asks the child to draw someone of the same sex and then someone of the opposite sex in any way that he or she wishes. (A variation involves asking the child to draw a person, house, and tree.) Those who employ the test believe that the drawings reveal meaningful information about the child’s personality or mental health.

In a sketch of a man, for example, small feet would supposedly indicate insecurity or instability—a small head, inadequacy. Large hands or teeth would be considered signs of aggression; short arms, a sign of shyness. And feminine features—such as long eyelashes or darkly colored lips—would allegedly suggest sex-role confusion.

Yet research consistently shows that such “signs” bear virtually no relation to personality or mental illness. Scientists have denounced these sign interpretations as “phrenology for the 20th century,” recalling the 19th-century pseudoscience of inferring people’s personalities from the pattern of bumps on their skulls.

Still, the sign approach remains widely used. Some psychologists even claim they can identify sexual abuse from certain key signs. For instance, in the child’s drawing at the right, alleged signs of abuse include a person older than the child, a partially unclothed body, a hand near the genitals, a hand hidden in a pocket, a large nose, and a mustache. In reality, the connection between these signs and sexual abuse remains dubious, at best.

Our literature review, then, indicates that, as usually administered, the Rorschach, TAT and human figure drawings are useful only in very limited circumstances. The same is true for many other projective techniques, some of which are described in the box on the preceding page.

We have also found that even when the methods assess what they claim to measure, they tend to lack what psychologists call “incremental validity”: they rarely add much to information that can