

ORIGINAL ARTICLE

## *Reading Personality: Assessing “Big Three” Traits with the Sentence Completion Method*

*Stephen P. Joy*

Department of Psychology, Albertus Magnus College

### BACKGROUND

Performance-based personality assessment has advantages, and the sentence completion method is one of the most reliable of these techniques. We sought to derive rating scales for three major personality traits (Extraversion, Neuroticism, and Psychoticism) based on Rotter Incomplete Sentences Blank (RISB) protocols.

### PARTICIPANTS AND PROCEDURE

Four successive samples of students ( $N = 231$ ) completed the RISB and other measures. Rating scales were derived empirically and cross-validated against the Eysenck Personality Questionnaire (EPQ-R). Additional validity evidence was obtained regarding symptomatic distress (SCL-90-R scales) and creativity (based on originality scores and expert judgments of student-generated drawings and poetry).

### RESULTS

The scales could be applied reliably by undergraduate research assistants (mean individual intraclass correlation

coefficients [ICCs] of .75, .79, and .71, respectively) and correlated well with self-reported traits (mean  $r$  values of .51, .57, and .58). Reliability of the finalized system was higher, with ICCs of .84, .83, and .79. RISB-rated traits correlated predictably with symptomatic distress. RISB-rated Psychoticism correlated strongly with the originality and judged creativity of drawings and poems. Psychoticism ratings were more strongly related to criterion measures than was the original self-report.

### CONCLUSIONS

The sentence completion method can be used to rate personality traits reliably. These ratings correlate fairly strongly with the results of self-report inventories but sometimes correlate more strongly with theoretically relevant variables than do self-report scores.

### KEY WORDS

Rotter Incomplete Sentences Blank extraversion; neuroticism; psychoticism

---

CORRESPONDING AUTHOR – Stephen P. Joy, Department of Psychology, Albertus Magnus College, 700 Prospect Street, New Haven CT 06511, e-mail: [sjoy@albertus.edu](mailto:sjoy@albertus.edu)

AUTHORS' CONTRIBUTION – A: Study design · B: Data collection · C: Statistical analysis · D: Data interpretation · E: Manuscript preparation · F: Literature search · G: Funds collection

TO CITE THIS ARTICLE – Joy, S. P. (2017). Reading Personality: Assessing “Big Three” Traits with the Sentence Completion Method. *Current Issues in Personality Psychology*, 5(4), 215–231.

RECEIVED 23.03.2017 · REVIEWED 2.06.2017 · ACCEPTED 4.06.2017 · PUBLISHED 4.09.2017

## BACKGROUND

Performance-based personality measures enable test-takers to construct responses expressing their thoughts and feelings, but extracting the information contained in test protocols is challenging. Calls to restrict their use are common (e.g., Lilienfeld, Wood, & Garb, 2000), and assuredly it is easier to score self-report measures. Yet when both self-report and performance-based tests are administered, they may yield overlapping but distinct information, providing a more valid prediction than either alone. This statement is based on several lines of research.

First, motives, such as the need for achievement, have been studied using the picture-story exercise (Smith, Atkinson, McClelland, & Veroff, 1992) as well as self-report measures. The two types of test tend not to correlate well with each other; meta-analyses have reported means of  $r = .09$  (Spangler, 1992) and  $r = .13$  (Kollner & Schultheiss, 2014). McClelland (1985; McClelland, Koestner, & Weinberger, 1989) argued that both approaches are valid but predict different types of behavior. He (1985) compared self-report measures with respondent behaviors and performance-based measures with operant behaviors. Later, self-reports were attributed to explicit, performance-based tests to implicit, mental operations (McClelland et al., 1989). Self-reports should do a better job of predicting choices under well-defined conditions, while the picture-story technique should be superior at predicting longer-term engagement in an activity. Some evidence supports this position. As one example, among athletes, self-reported achievement motivation predicts the distance from which a player will take a shot, but a picture-story test predicts how much a player contributes to the team during a series of games (Schultheiss, Yankova, Dirlikov, & Schad, 2009).

Second, dependency has been measured using the performance-based Rorschach Oral Dependency Scale (Masling, Rabie, & Blondheim, 1967) as well as self-reports. They correlate moderately, with mean  $r = .35$  (Bornstein, 1999; Bornstein, Rossner, & Hill, 1994), and are similar in their ability to predict behavior; Bornstein's (1999) meta-analysis reported mean  $r$  values of .37 and .31, respectively. Yet they have different properties. Only self-reported dependency is affected by gender or instructional set; only performance-based dependency is affected by mood (Bornstein, 2002).

Third, studies of implicit processing by cognitive science began with memory (Schachter & Graf, 1986) but expanded to affective variables. The advent of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) revolutionized the study of attitudes. Self-reported and implicit attitudes correlate, on average, at  $r = .24$  (Hoffman, Gawronski, Gschwendner, Le, & Schmitt, 2005). IAT measures have been developed for self-esteem (Greenwald

& Farnham, 2000), anxiety (Egloff & Schmukle, 2002), and aggression (Richetin, South Richardson, & Mason, 2010) with similar results; the IAT correlates modestly with self-reports and with observable behavior not accounted for by self-reports.

In sum, both classic "projective" tests and novel experimental techniques yield valid predictions of behavior largely independent of those made by self-report inventories. However, it is not always easy to identify a task as explicit vs. implicit. Many may involve both processes, just as recognition memory involves both explicit ("remembering") and implicit ("familiarity") memory systems (Mandler, 1980). Consider interviews. A structured interview administered by an epidemiological researcher is explicit; a clinical interview conducted by a freewheeling gestalt therapist is mostly implicit. But many interviews probably tap into both processes, gathering declarative evidence while also eliciting affective reactions that are observed by the clinician.

Even detractors of performance-based personality assessment (Lilienfeld, Wood, & Garb, 2000) often make an exception for properly scored sentence completion tests. There is ample evidence that these instruments can be scored reliably and made to yield valid information with meaningful behavioral correlates (Hy & Loevinger, 1996; Rotter, Lah, & Rafferty, 1992); they are, in addition, relatively easy to master.

The status of sentence completion measures along the explicit-implicit continuum is not known. They *are* performance-based tasks, and the samples of verbiage they elicit resemble the TAT-type storytelling technique, albeit writ small. In keeping with this, they have been used to measure work-related motives (Miner, 1964) with validity similar to that of other approaches (a mean effect size of  $r = .20$ ; Collins, Hanges, & Locke, 2004). On the other hand, the units of verbal behavior can be so small, the prompts (stems) so straightforward, that explicit responses are likely in many instances. Completing a sentence beginning with "a mother" is a far simpler task than writing a story in response to a picture of a man at a drafting board. As with word associations, there may sometimes be a few responses so dominant that the task is virtually a multiple-choice one: almost as much a self-report as a constructed response. It seems probable that the sentence completion method draws upon both implicit and explicit processes.

This paper introduces new scoring systems for an existing sentence completion measure, the Rotter Incomplete Sentences Blank (RISB; Rotter et al., 1992). Introduced 70 years ago (Rotter & Willerman, 1947), the RISB is more often used clinically than all other sentence completion measures (Holaday, Smith, & Sherry, 2000), most likely due to its existing well-validated scoring system.

Standard RISB scoring assesses Adjustment: a product of the interaction between the individual's resour-

es and environmental demands. Each RISB response is rated separately; the item scores are then summed. Inter-rater reliability averages .93 (Rotter et al., 1992). It correctly classifies 85% of clinical vs. control cases and correlates well with other adjustment-related measures. Recent studies support its validity with clinic-referred adolescents (Weis, Toolis, & Cerankosky, 2008) and adult psychiatric patients (McCloskey, 2014), including evidence of incremental validity when added to a standard assessment (McCloskey, 2014; Torstrick, McDermut, Gokberk, Bivona, & Walton, 2015).

However, despite these qualities, most users do not bother to score the RISB (Holaday et al., 2000). Qualitative interpretation is the rule. It is intended to be used this way, but objective scores and intuitive interpretations ought to be complementary; this disconnect between the two approaches is troubling. One suspects that users perceive the scoring system as too limited. Test scores should function as a scaffold upon which clinical intuition may build, yet only limited elaboration upon a single score is possible. Scoring systems for additional variables would be helpful.

An obvious direction in which to extend the RISB is the rating of major personality traits. These are fairly stable over time and influence behavior across many settings. Their assessment and study comprises a large portion of contemporary personality research. This is nearly always done using self-report inventories designed solely for the purpose. If the RISB could measure them, it would offer several benefits. For clinicians who have limited room for formal personality assessment in their everyday practice, the RISB could be made to serve “double duty”; in addition to its use as a qualitative measure, it could (up to a point) substitute for an additional self-report inventory. Furthermore, a client’s idiosyncratic ways of expressing each trait (and its relationship with other personality features) could be explored qualitatively. The less transparent nature of the sentence completion method might sometimes be valuable as well. Perhaps most important, convergent (and discriminant) evidence could be obtained by using both types of test in more comprehensive assessments. That is, adding the RISB to a self-report inventory would sometimes merely strengthen one’s confidence in drawing clinical inferences about client traits, but on other occasions (when the two methods yielded discrepant results) it would mandate more careful consideration of how, and to what extent, the individual expresses the trait in question. This process will be facilitated as additional data are collected on the correlates of each type of trait measure.

### THREE MAJOR PERSONALITY TRAITS

The present investigation utilizes the simplest trait model, Eysenck’s, which comprises three major

traits: Extraversion, Neuroticism, and Psychoticism (Eysenck & Eysenck, 1994). The first two of these have won widespread acceptance and will be described but briefly. The third is less familiar to many psychologists and will be discussed more fully.

Extraversion (E) represents one end of a continuum bounded at the other extreme by Introversion. High E people are sociable, outgoing, and active. They tend to have many friends and acquaintances and to make new ones easily, enjoying social situations such as parties. They prefer team work to solitary pursuits, tending to be assertive in interpersonal settings. Craving stimulation, they shift activities frequently. Positive emotions dominate, and they learn best through reward. Low E people tend to be quiet and reserved, approach tasks more cautiously, and to have a few close friends rather than a wide range of acquaintances. Low E is characterized not by intense social anxiety or negative affect, but by less interest in socializing and less intense positive emotions. When stable, low E people are viewed as serene and even-tempered. In Gray’s revised Reinforcement Sensitivity Theory (RST; Gray & McNaughton, 2000), E is associated with a strong Behavioral Activation System (BAS; Pickering & Corr, 2008).

Neuroticism (N) represents one end of a continuum whose opposite pole is emotional stability. High N people often struggle with anxiety, depression, and (presumably stress-related) physical complaints. They worry about their own adequacy and are pessimistic about the future, anticipating (and learning from) punishing rather than rewarding outcomes. They react strongly to stress and are slow to return to a baseline level of arousal that is already uncomfortable. Low N people are “even keeled” sorts who do not react strongly to stressors or fret over anticipated pain. In Gray’s RST, N is associated with a strong Behavioral Inhibition System (BIS; conflict) and a relatively strong Fight-Flight-Freeze System (FFFS; Fear; Pickering & Corr, 2008).

Psychoticism (P) was originally conceptualized as a personality dimension underlying psychotic, schizoid, and psychopathic presentations, much as N underlay the so-called neuroses. Because it was deemed a “normal” dimension of personality, manifesting psychotic or quasi-psychotic symptoms only when decompensated, Eysenck devised P scales that did not include symptoms or other pathological content (Eysenck & Eysenck, 1994). This sets the P scale apart from most efforts at measuring schizotypal or psychosis-prone traits, such as the Chapman scales (Chapman & Chapman, 1996) or the Schizotypal Personality Questionnaire (Raine, 1991).

High P individuals tend to be cold, impersonal, unempathic and egocentric (Eysenck & Eysenck, 1976). They view others chiefly as a means to an end. They have little regard for social norms or authority. Their basic interpersonal stance is hostile; aggres-

sion, whether “naked” or veiled as competitiveness, is typical. Hostility and/or marked ambivalence may apply even to family members. There may be an inhumane, even cruel, quality to their relationships; they may enjoy deceiving others, and suspect others of harboring equally malign intentions. Grandiose aspirations may be present, though perhaps not the self-discipline needed to achieve these goals. They are attracted by odd or unusual ideas, art forms, and so on. In Gray’s RST, P is associated with a weak BIS and a strong BAS “Fun Seeking” sub-system, though not with BAS as a whole (Heym & Lawrence, 2010; Pickering & Corr, 2008).

Extensive research has compared high P and low P samples on measures known to elicit different performances from people with schizophrenia as opposed to controls. The reasoning was that if a similar pattern of differences emerged (i.e., schizophrenia: control :: high P: low P), similar mechanisms might be at work. This held true for many tasks (e.g., eye tracking, latent inhibition, negative priming) as well as several physiological variables (Eysenck, 1992).

Chapman and Chapman (1994) reported mixed but generally positive findings from a sample of students re-evaluated after a 10-year interval of whom 26 had obtained very high scores on Eysenck’s P scale, while 310 obtained low scores. None developed a psychotic disorder, but the high P group obtained higher scores on schizotypal and paranoid personality disorders (assessed via structured interview) and reported more psychotic-like experiences, including visual illusions, aberrant beliefs, and thought transmission.

On a more positive note, P is associated with creativity. A link between creative “genius” and psychosis (or “madness”) has been suspected since ancient times, and continues to attract research interest (Becker, 2001). People with psychotic disorders produce unusual responses on word association tests (Rapaport, 1946), and similar tasks have been used to measure creative thinking (Benedek, Konen, & Neubauer, 2012). High scorers on P scales also tend to produce unusual word associations (Merten, 1993), so a link between P and creativity seems logical.

This line of investigation was lent impetus by the impressive results of an early study (Woody & Claridge, 1977); many more studies followed, using a variety of criterion measures. One meta-analysis (Feist, 1998) found that scientists and artists tend to obtain elevated P scores. Another (Acar & Runco, 2012) reported the mean correlation of P with creativity measures as  $r = .16$ : small but significant, and larger for outcomes measured in terms of uniqueness.

Criminals, drug addicts, and personality disordered people also tend to obtain elevated P scores (Eysenck & Eysenck, 1976). Indeed, it is sometimes argued that the P scale measures subclinical psychopathy more than it does psychosis-proneness (Clark

& Watson, 2008). Most authorities recognize two facets of psychopathy: one made up of affective and social qualities (e.g., lack of empathy, callousness), the other of impulsive and antisocial behavior (Hare, 2003; Skeem, Poythress, Edens, Lilienfeld, & Cale, 2003). Hare (1982) reported that, among male prison inmates, P correlated only with the second factor, but in other populations P correlates with both factors (Heym, Ferguson, & Lawrence, 2013).

Nothing quite like P appears in trait models featuring five or more factors, but two “Big Five” traits correlate negatively with P: Agreeableness (A) and Conscientiousness (C). Costa and McCrae (1995) reported that when EPQ-R scores were subjected to a five-factor solution, P loaded on both A and C in the .34-.49 range (which loading was stronger depended on the rotational strategy/rotational strategy). Similarly, when NEO-PI-R scores were placed on the three-factor P-E-N model, A and C each loaded on P with an average loading of about .50. More recently, Heaven et al. (2013) reported correlations with P of  $r = -.42$  for A and  $r = -.34$  for C in a large sample of adolescents. One may argue that P is an amalgam of these traits. Conversely, one could argue that A and C are facets of a single superordinate trait: P. This paper takes no position with regard to that issue.

In any case, characteristics of high A and high C people are likely to reflect the opposite pole of the P dimension. According to Costa and McCrae (1992), the six “facets” comprising A are trust, straightforwardness, altruism, compliance, modesty, and tender-mindedness. The six facets of the high C person are competence, order, dutifulness, achievement striving, self-discipline, and deliberation. In other words, high A/low P people care for and about other people and their feelings; high C/low P people are diligent in the pursuit of conventional goals. The common denominator is *socialization*; low P individuals are committed to the values and norms of the community, caring about their fellow humans (at least, “in group” members) and their welfare while following the rules prevailing in their society.

The present study, in short, is an attempt to measure these three traits (E, N, and P) using the RISB.

## SCALE DEVELOPMENT AND INITIAL VALIDATION METHOD

### PARTICIPANTS

All participants were undergraduates who received extra credit for their work; the large majority were traditional students enrolled in introductory psychology, but the third sample also included older students drawn from several classes. Detailed demographics are not available, but the student body at the college is ethnically, socioeconomically, and academically diverse.

Sample #1 ( $N = 45$ , 78% female) was used for initial scale development (see below). Sample #2 ( $N = 44$ , 70% female) was used for cross-validation and scale revision. Sample #3 ( $N = 84$ , 73% female) was used for further cross-validation and fine-tuning of the system. Sample #4 ( $N = 58$ , 74% female) was used to evaluate the reliability of the final system. Altogether, then, 231 individuals (170 females, 61 males) were included in this series of studies.

## INSTRUMENTS

All participants completed the RISB (Rotter et al., 1992), and all those in the first three samples also completed the Eysenck Personality Questionnaire – Revised (EPQ-R; Eysenck & Eysenck, 1994).

The RISB, discussed above, is a 40-item sentence completion measure. Most stems are brief (e.g., “I like...”). A single line after each stem encourages relatively concise responses.

The EPQ-R (Eysenck & Eysenck, 1994) is a 100-item self-report measure scored on four scales: Psychoticism (27 items,  $\alpha = .67$ ), Extraversion (22 items,  $\alpha = .85$ ), Neuroticism (24 items,  $\alpha = .86$ ), and “Lie” (21 items,  $\alpha = .77$ ). The Lie scale measures defensiveness or social desirability. The last 6 items are not scored. The four scales are largely orthogonal, but modest negative correlations obtain between E and N ( $r = -.28$ ) and between P and L ( $r = -.21$ ).

## RATERS

The RISB protocols were rated by undergraduate research assistants who earned college credits for their work. The first six began by learning the traditional RISB Adjustment scoring system. Each one’s ratings correlated with the author’s at  $r = .90$  or above: results comparable to those obtained with graduate students or professional raters.

## DEVELOPMENT OF RATING SCALES AND RELIABILITY ANALYSIS

All three sets of criteria were developed using an empirical approach informed by theoretical understanding of the constructs involved. The P scale will be used to illustrate this process. First, the EPQ-R protocols from sample #1 were scored. Then the RISB protocols for the 12 individuals with the highest P scores and those for the 12 with the lowest P scores were examined in search of themes or responses that occurred more frequently in one group. (These 24 protocols contained 960 personal statements.) Slight differences were noted only if highly consistent with the core features of the P trait. For instance,

if even one high P person wrote “I like frightening people,” or one low P person wrote “I am very kind, caring, and considerate,” it was considered noteworthy. Themes less closely related to the core features of P needed to occur several times more in one group than the other. For example, several high P participants wrote completions for item #23 (“My mind”) indicating a lack of control over their mental processes (e.g., “is about to explode;” “is all over the place”). Although not a defining feature of P, thought disorder and cognitive dysfunction are relevant to a trait associated with psychosis. Eventually a set of general criteria, supplemented by limited item-level guidelines, was developed.

Having studied the system, a research assistant scored all the protocols from sample #1, followed by those from sample #2. Sample #2 was critical because of the empirical approach taken to scale development; to some degree, the first set of ratings capitalized on chance, and the extent to which this occurred would appear as validity shrinkage. Convergent validity was evaluated by correlating the RISB P ratings with the EPQ-R P scale, discriminant validity by correlating the RISB P ratings with the other EPQ-R scales. The same procedure was followed in developing the E and N rating scales. One research assistant worked on each of the three scales.

After these initial studies, three new research assistants were trained to use all three scales and rated the protocols from the first two samples. The scales then underwent minor revisions. Next, the protocols from sample #3 were scored. The use of multiple raters enabled estimation of inter-rater reliability using the intraclass correlation coefficient (ICC; Shrout & Fleiss, 1979). The ICC is a more conservative procedure than simply using the Spearman-Brown Prophecy Formula in that it takes not only the correlation between raters, but also the closeness of the actual ratings, into account. Two judges whose ratings correlated well, but whose mean ratings were quite different (owing to systematic bias) would obtain lower values using the ICC. However, in practice the two methods yield reasonably comparable results. The “individual” ICC represents the reliability of the average judge acting alone, while the “averaged” ICC represents that of the judges taken as a group – inevitably a higher value.

At least four students rated the cases in samples #1 and #2 on each variable; three students rated those in sample #3. Table 1 shows the results. In general, all three scales exhibited adequate reliability, with median ICC = .76 for individual raters (much higher for pooled or averaged ratings, which were used in subsequent data analysis).

When all the ratings had been collected, the author fine-tuned the scoring criteria, then rescored every RISB protocol. As a result, the correlations in the first sample remained the same or shrank

Table 1

*Reliability Analysis of RISB Trait Scores (Intraclass Correlation Coefficients)*

	Psychoticism		Extraversion		Neuroticism	
	Single	Average	Single	Average	Single	Average
Sample #1	.76	.93	.86	.94	.90	.96
Sample #2	.81	.95	.79	.95	.71	.91
Sample #3	.56	.79	.59	.81	.74	.90
Sample #4	.79	.92	.84	.94	.83	.94

slightly; those in later samples grew somewhat stronger due to elimination of spurious criteria. The internal consistencies of these RISB ratings were acceptable, ranging from  $\alpha = .71$  to  $\alpha = .79$  for Psychoticism, from  $\alpha = .57$  to  $\alpha = .77$  for Extraversion, and from  $\alpha = .77$  to  $\alpha = .82$  for Neuroticism. These final scores, which correlated at about  $r = .90$  with the mean student ratings, were not truly "blind," so they are not cited here. However, it makes little sense to present already superseded original criteria. Therefore, we present them as they now stand. Table 2 displays the criteria for Psychoticism; Table 3, those for Extraversion; Table 4, those for Neuroticism. Note that a sentence may be scored on more than one scale.

Finally, two new research assistants were trained to use the revised system with its more extensive scoring examples. They began with practice cases, then scored a new set of protocols (sample #4). Reliability generally improved with the added item-level support (Table 1), even though (unlike their predecessors) they were not first trained in the traditional RISB Adjustment system.

## RESULTS

### DESCRIPTIVE STATISTICS

Table 5 displays descriptive statistics for the three traits as measured by the EPQ-R. They are fairly similar across samples. Values for E and N resemble those in the standardization sample. Those for P are a bit higher: not surprising, since P scores decline with age (Eysenck & Eysenck, 1976) and the present sample comprised mainly 18-21 year olds. Table 5 also displays descriptive statistics for the RISB scales, which tend to track the EPQ-R scores fairly closely.

### CORRELATION OF ROTTER INCOMPLETE SENTENCES BLANK RATINGS WITH SELF-REPORT

Table 6 presents the principal findings of the study: the correlations between personality trait scores de-

rived from sentence completions and those obtained via self-report.

The RISB Psychoticism (P) scale correlated very strongly with its EPQ-R counterpart in the initial sample and displayed a moderate degree of validity shrinkage upon cross-validation. This scale was left essentially unchanged at that stage, and results for the third sample are quite similar to those in sample #2. RISB P scores are clearly independent of self-reported E and N scores, with no correlations exceeding .20; one sample showed a significant negative correlation with the L scale, but this is also true of the EPQ-R (Eysenck & Eysenck, 1994).

The RISB Extraversion (E) scale also correlated very strongly with its EPQ-R equivalent in sample #1 and, despite validity shrinkage, remained a strong correlate of self-reported E in the cross-validation sample. Its correlation with self-reported N was a bit higher than ideal in sample #1, but dwindled comfortably in sample #2.

The RISB Neuroticism (N) scale correlated strongly with its parallel EPQ-R scale in sample #1. Validity shrinkage was modest. However, the RISB N scale showed a worrisome tendency toward strong negative correlations with self-reported E, mainly in sample #2. The RISB E and N scales were then revised with an eye to reducing this overlap. In sample #3, the correlation between RISB-rated and self-reported N grew stronger, and RISB N scores were no longer so strongly correlated with self-reported E. RISB E scores also became much more independent of self-reported N. Unfortunately, the correlation between RISB and EPQ-R E scores weakened somewhat. This issue was revisited in the final revision of the rating system.

### CORRELATIONS AMONG THE THREE TRAITS WITHIN EACH INSTRUMENT

The correlations among P, E, and N on the EPQ-R were similar to expected values (Eysenck & Eysenck, 1994). P was largely independent of E (mean  $r = -.07$  across samples) and N (mean  $r = .08$ ). Scores on E and N displayed the expected negative correlation (mean  $r = -.25$ ).

Table 2.

*Psychoticism Rating Scale for Sentence Completions: General Guidelines***Indicators of High P (Score +1) – Look for references to:**

- Being “different,” “weird,” or alienated from others
- Desire to explore or learn, especially about unusual or even arcane things (not, for example, learning to ski or vacationing in Florida)
- Escape into other worlds, imaginary realms; getting lost in fantasy
- Difficulty controlling one’s mind (not just “wandering” or “daydreams” unless amplified); inability to concentrate, etc.
- Primitive drive states:
  - Sexual desires or interests (without love/affection specified)
  - Aggression (e.g., yelling, anger, hitting, smashing things, especially if enjoyed)
  - Hunger (including a desire for food, but not simple statements of liking foods)
- Past or present alcohol/drug use, promiscuity, or other major norm-violating behaviors (not just single episodes, especially not if associated with guilt)
- Use of explicitly vulgar language (sexual, scatological, or blasphemous)
- Dehumanizing, unempathic views of people: a fundamental hostility or contempt
- Strong concern with personal independence and/or with power dynamics in relationships
- Negativity or marked ambivalence about social norms (e.g., marriage, family, school, or laws). Do not score critical remarks directed solely at a particular school or workplace.
- Viewing oneself (or people generally) as untrustworthy or deceitful (not in a context of describing types of people one dislikes, and not solely regarding the opposite sex)
- Overtly denying such affects as fear, anxiety, worry, pain, or suffering (those that imply vulnerability or weakness and involve withdrawing from threats)
- Desire to achieve some kind of greatness (a sense of personal destiny, sometimes with a grandiose quality) and/or an intense fear of not doing so, of facing life as a mediocrity. Distinguish carefully from aspirations for conventional success: jobs, money, etc.
- UNUSUAL OR STRIKING RESPONSES: typically “unique,” often involving vivid images or peculiar word choice. Often these responses are difficult to score using the standard Adjustment system, but convey a sense of “strangeness” that makes the reader pause. They are particularly likely in the protocols of those very high in Psychoticism.

**Indicators of Low P (Score –1)**

- Look for references to:
  - Being “nice,” “sweet,” “kind,” etc.
  - Being helpful, selfless and giving: even too much so (e.g., neglecting one’s own wishes)
  - Tender, loving feelings toward family members and friends; feelings of love and caring generally. Do not score responses alluding to *being* loved by others (which could equally well reflect narcissism), or to simply enjoying times spent with friends.
  - References to love in marriage and family, an expectation of emotional closeness in these relationships that is a major source of happiness
  - Concern over seeing other people suffer or failing to take care of them when needed
  - Concern over losing close family members (in response to stems probing for fears or worries: i.e., harm coming to others is the first thing that comes to mind)
  - People in general as being good, trustworthy, or likeable; bland assurances about the state of the world and the future
  - Denying affects such as hate, anger, etc. in oneself
  - Being responsible, diligent, rule-abiding, etc.
  - Certain very conventional goals (e.g., jobs in respected fields, financial security)
  - Some very common, stereotypical responses offered with little or no elaboration

---

*Note.* Some brief, unelaborated responses that otherwise fit these criteria occur frequently in the protocols of people at all levels of P and should not be scored. Examples include harsh criticism of men (or women) and the completion “I regret nothing.” These guidelines are elaborated more fully in the *Manual*, which also provides item-specific guidelines and examples of completions scored as +1, –1, or 0. Accurate scoring is more likely when the *Manual* is consulted. This also applies to Extraversion and Neuroticism ratings.

Table 3.

*Extraversion Rating Scale for Sentence Completions: General Guidelines***Indicators of High E (Score +1) – Look for references to:**

- Positive emotions (e.g., happiness), especially if energetic in nature (e.g., excitement); do not score emotions like “calm” or “relaxed” [see low N]
- Having fun, enjoying oneself, etc., especially when expressed vigorously (e.g., “a blast”)
- Spending time/doing things with friends (especially in the plural), with one’s boyfriend or girlfriend, or with family groups; do not score if the response makes it clear that the person named is the only friend, or one of very few with whom a close relationship exists
- Enjoying sports, especially team sports (e.g., soccer), whether as player or fan; rarely will a sporting activity be described as isolating and unadventurous – if so, do not score
- Being popular, well-liked, outgoing, or socially prominent
- Positive feelings about people in general
- Expressions of confidence in one’s interpersonal influence or social dominance
- Concerns about liars or unreliable people
- Concern over losing important/loved people (e.g., due to death, bad things happening to them, or just losing touch)

**Indicators of Low E (Score -1)**

Look for references to:

- Being shy, withdrawn, quiet, or reserved, or lacking social confidence; sometimes also scored as high N, but try to differentiate from pure anxiety reactions (only scored for N)
- Preference for engaging in solitary activities such as writing, drawing, playing a musical instrument (unless performing or playing with a group) or other quiet activities
- Not being popular or socially prominent (but not “outcast” responses)
- Desire to be alone, out of the attentional spotlight
- Concern over being alone too much, over not having friends now or in the future
- Specific individuals or small groups of people as caring or as sources of good feelings; the response must make it clear that they are unusual or exceptional. Remember: usually references to social outlets are scored for high E. If in doubt, do not score

The RISB P, E, and N ratings correlated more strongly with one another. That between P and E averaged  $r = -.44$ ; that between P and N averaged  $r = .47$ ; that between E and N averaged  $r = -.47$ . However, these correlations generally did not prevent the RISB from displaying adequate convergent and discriminant validity vis-à-vis self-report.

**GENDER DIFFERENCES**

Although the males in sample #1 obtained higher scores on RISB-rated P and E, these differences fell short of statistical significance. The females obtained higher RISB N scores:  $t(43) = -2.33$ ,  $p = .025$ . Similarly, the only near-significant difference on the EPQ-R in sample #1 was for females to score higher on N:  $t(43) = -1.73$ ,  $p = .087$ .

RISB ratings of E and N were similar across gender in sample #2. Males, however, obtained significantly higher scores on P:  $t(42) = 2.64$ ,  $p = .012$ . They also self-reported higher levels of P:  $t(42) = 2.05$ ,  $p = .047$ .

In sample #3, males again obtained higher P scores on the RISB:  $t(81) = 3.22$ ,  $p = .002$ , but there were no differences on the other scales. Females self-reported higher levels of E:  $t(79) = -2.36$ ,  $p = .020$ . They also

had higher N scores, but not to a statistically significant degree.

In sample #4, males again obtained higher P scores, but not to a statistically significant degree. This time it was the males who obtained higher E scores:  $t(56) = 2.79$ ,  $p < .050$ . Females obtained significantly higher N scores:  $t(56) = -4.04$ ,  $p < .010$ .

Self-reports typically yield higher N scores among females and higher P scores among males (Eysenck & Eysenck, 1994). The present results are consistent with this. The RISB scales also show a tendency for males to obtain higher P scores and for females to obtain higher N scores. In short, the pattern of gender differences is roughly similar across methods.

**ADDITIONAL VALIDITY STUDY #1:  
SYMPTOMATIC DISTRESS**

Personality traits can predispose a person to, or help to protect one against, psychological disturbances. Neuroticism is strongly associated with many forms of distress, especially depression and anxiety. Psychoticism has a more selective relationship with psychological problems; people high in P are likely to experience (and act on) feelings of rage, feel alienated

Table 4.

*Neuroticism Rating Scale for Sentence Completions: General Guidelines***Indicators of High N (score +1) – Look for references to:**

- Experiencing negative emotions or symptoms (e.g., fear, anxiety, “stress,” depression, or guilt; “tired” is *not* scored); internalizing psychopathology (e.g., Panic Disorder)
- Being unable to cope, overwhelmed by life; to being inadequate or unable to do things one would like to do (e.g., sports, dancing); wishing to avoid real-life responsibilities
- Unhappiness with one’s home life, family members, or romantic partner (not limited to “lost love” in a single relationship)
- Desire to avoid or escape from painful/negative situations: to find safety/security
- Love as unattainable, as desired but unlikely to be found; to a strong desire to be loved or cared for that implies such caring is not part of the respondent’s actual life
- Being frightened or pushed around by people. Do not score being “annoyed” by them; being “bothered” *may* be scored *only if* the context implies fear or submission
- People at large, or entire groups of people, as mean, cruel, “catty,” or rejecting of the respondent; to oneself as a social isolate/outcast (not by preference)
- A desire to relieve stress, find relaxation, implying that this is not the respondent’s usual state – especially if unaccompanied by references to positive affects/enjoyment;
- Expressed desire to be better, psychologically healthier, get over one’s problems; or to have life somehow be changed for the better

**Indicators of Low N (Score –1)**

Look for references to:

- Describing oneself or one’s life as happy, optimal, etc. (but not references to highly active, upbeat emotions)
- Describing oneself as calm, composed, serene: able to handle stress; to one’s mind as sharp, focused, clear
- Clear-cut optimism (not qualified)
- References to specific personal victories, triumphs, etc.; specific happy memories (not generic, and not just conversations)

Table 5

*Descriptive Statistics for EPQ-R and RISB Trait Scales*

	Sample #1		Sample #2		Sample #3		Sample #4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
EPQ-R								
Psychoticism	6.64	3.37	8.09	3.48	7.32	3.45	–	–
Extraversion	15.63	4.22	15.48	5.06	15.69	4.25	–	–
Neuroticism	13.75	5.21	14.89	5.12	14.32	5.41	–	–
RISB								
Psychoticism	1.05	6.38	4.09	5.59	3.69	4.11	1.73	4.51
Extraversion	2.27	7.45	2.60	5.18	3.55	3.88	3.79	3.80
Neuroticism	-0.16	7.40	0.98	5.00	2.43	5.42	4.10	5.13

Note. The RISB figures shown here are based on the mean scores assigned by the raters working on each sample.

from others, and maybe to manifest bizarre ideas. Extraversion, by contrast, tends to exert a positive effect on personal adjustment, especially in the interpersonal domain. It is hypothesized that scores on the new RISB personality scales will relate to measures of psychological problems in accordance with these well-established facts.

**METHOD**

A subset ( $n = 67$ ) of participants from samples #2, #3, and #4 completed the SCL-90-R (Derogatis, 1983), which entails rating one’s current (past week) level of distress due to each of 90 symptoms using a 5-point Likert-type scale. Widely used in clinical

Table 6  
Correlations between RISB Trait Ratings and Self-Reported Personality Traits

EPQ-R Scale	RISB Personality Trait Ratings								
	Psychoticism			Extraversion			Neuroticism		
	P1	P2	P3	E1	E2	E3	N1	N2	N3
Psychoticism	.72***	.53***	.50***	.02	-.34*	.36**	.06	.28 <sup>+</sup>	.36**
Extraversion	-.13	-.25	-.19 <sup>+</sup>	.80***	.48***	.24*	-.42**	-.46**	-.19 <sup>+</sup>
Neuroticism	.16	-.04	.20 <sup>+</sup>	-.44**	-.27 <sup>+</sup>	-.19 <sup>+</sup>	.61***	.47***	.62***
Lie	-.31*	.08	-.07	.00	-.23	.09	-.27 <sup>+</sup>	.06	-.22*

Note. P1, P2, P3 = Psychoticism score from samples #1, 2, and 3 respectively; E1, E2, E3 = Extraversion scores; N1, N2, N3 = Neuroticism scores.

<sup>+</sup> =  $p < .100$ ; \* =  $p < .050$ ; \*\* =  $p < .010$ ; \*\*\* =  $p < .001$ .

Table 7  
Correlations between Personality Traits and Current Symptomatic Distress

SCL-90R Scale	RISB Trait Scales			EPQ-R Trait Scales		
	P <sup>1</sup>	E <sup>2</sup>	N <sup>3</sup>	P <sup>1</sup>	E <sup>2</sup>	N <sup>3</sup>
Somatization	.16	-.14	.25 <sup>+</sup>	-.17	.18	.33*
Obsessive-Compulsive	.10	-.06	.44**	-.15	.11	.46**
Interpersonal Sensitivity	.15	-.41**	.56**	-.13	-.17	.56**
Depression	.22 <sup>+</sup>	-.34**	.64**	-.08	-.20	.64**
Anxiety	.19	-.26*	.41**	-.14	-.01	.40**
Hostility	.28*	-.16	.35**	.03	-.01	.32*
Phobic Anxiety	.09	-.16	.28*	-.11	-.06	.30*
Paranoid Ideation	.31*	-.27*	.44**	.15	.08	.50**
Psychoticism	.22 <sup>+</sup>	-.33**	.40**	-.05	-.11	.36*

Note. 1. Psychoticism; 2. Extraversion; 3. Neuroticism.

<sup>+</sup> =  $p < .100$ ; \* =  $p < .050$ ; \*\* =  $p < .010$ .

research, the SCL-90-R yields scores on 9 subscales, though they tend to correlate strongly with one another. EPQ-R results were available for most ( $n = 49$ ) of these participants.

Correlations between the RISB and EPQ-R P, E, and N scores and the SCL-90-R were calculated. It was hypothesized that (1) N would correlate positively with most or all of the subscales, (2) P would correlate positively with only a few subscales, mainly Hostility, Paranoid Ideation, and Psychoticism, and (3) E would correlate negatively with Interpersonal Sensitivity and perhaps also with Paranoid Ideation or Psychoticism (which contain items relating to social alienation). Item-level analyses were intended to explore the personality constructs more fully.

Note that although Derogatis (1983) cites Eysenck as an influence upon the SCL-90-R Psychoticism subscale, the content of the two inventories is quite different. As noted earlier, Eysenck's P scale excludes pathological content. By contrast, the SCL-90-R is a measure of psychopathology. Some of its 10 items

are overt psychotic symptoms; the others relate to social alienation or to peculiar, but not psychotic, concerns.

## RESULTS

Table 7 shows the results of the subscale analyses. As expected, RISB-rated Neuroticism correlated significantly with nearly every SCL-90-R subscale and trended in the same direction with the remaining one. Unsurprisingly, Depression and Interpersonal Sensitivity were the strongest correlates. Those high on N are unhappy people plagued by concerns that others do not like or esteem them. These results closely tracked those for the self-report EPQ-R.

RISB-measured Psychoticism also behaved as expected, correlating with Paranoid Ideation, Hostility, and (at the .10 level) SCL-90-R Psychoticism. An additional weak correlation with Depression also emerged. People high in RISB-measured P are hostile,

Table 8

*Correlations between RISB Personality Trait Scores and Originality/Creativity Measures*

	Psychoticism	Extraversion	Neuroticism
House-Tree-Person Drawings			
Unusual House	.34*	-.23	.01
Unusual Tree	.36**	-.26 <sup>+</sup>	.39**
Unusual Person	.36**	-.25 <sup>+</sup>	.22
Total Originality	.48**	-.33*	.28*
Common Details	.07	-.21	.33*
Judged Proficiency	.47**	-.43**	.35*
Judged Creativity	.40**	-.42**	.33*
Associative Poems			
Word Count	.44**	-.04	.29 <sup>+</sup>
Infrequency Ratio	.43**	.03	-.07
Unique: Repeated	.35*	-.30 <sup>+</sup>	.26
Affective Blends	.56**	-.28 <sup>+</sup>	.21
Verbal Contrasts	.40*	-.21	.12
Arousal Composite	.71**	-.31 <sup>+</sup>	.35*
Judged Poetic Value	.31*	-.17	.13
Chromatic Drawings			
Unusual Artifact	.56**	-.42**	.24
Unusual Life Form	.16	-.26 <sup>+</sup>	.15
Unusual Person	.03	-.17	.19
Originality Composite	.35*	-.38*	.26 <sup>+</sup>
Judged Proficiency	.27 <sup>+</sup>	-.35*	.21
Judged Creativity	.29*	-.11	.08

*Note.* *Unusual House* = uncommon features or “type” of house; *Unusual Tree* = uncommon features or “type” of tree; *Unusual Person* = atypical rendering of the human figure; *Total Originality* = sum of z scores for the preceding variables; *Common Details* = number of common details included in House drawing; *Judged Proficiency & Judged Creativity* = mean ratings of technical quality and creativity of drawings assigned by 2 art therapy graduate students.

*Word Count* = # of words in poem; *Infrequency Ratio* = infrequency of words used in English (controlling for word count);

*Unique: Repeated* = ratio of words used only once to words used more than once; *Affective Blends* = presence of both positive

and negative emotions in the poem; *Verbal Contrasts* = use of contrasting pairs of words in the poem; *Arousal Composite* = score

derived by converting all individual indicators to z scores and combining them; *Judged Poetic Value* = mean rating assigned by two English professors working independently.

*Unusual Artifact* = less frequently selected subjects for “artifact” drawing; *Unusual Life Form* = less frequently selected subjects

for “life form” drawing; *Unusual Person* = human figure rendered in uncommon ways (see text); *Originality Composite* = sum

of z scores for the three preceding scores; *Judged Proficiency & Judged Creativity* = technical quality and creativity of drawings

(mean of ratings provided by 3 art therapy graduate students).

\*  $p < .100$ ; \*  $p < .050$ ; \*\*  $p < .010$ .

vigilant, and alienated. In this case, the RISB measure appears to be *more* sensitive to P-related pathology than the self-report EPQ-R, which failed to correlate with any SCL-90-R scales, even those to which it is theoretically related. Using Steiger’s Z test, the RISB correlated significantly more strongly with Hostility ( $Z = 2.30, p = .021$ ) and Psychoticism ( $Z = 2.47, p = .013$ ), though the difference for Paranoid Ideation fell short of statistical significance ( $Z = 1.50, p = .130$ ).

RISB-measured Extraversion provided moderate protection against symptomatic distress. As predicted, Interpersonal Sensitivity was the strongest (negative) correlate. Correlations with 4 additional subscales (Depression, Anxiety, Psychoticism, and Paranoid Ideation) also reached statistical significance. The RISB E scale was more strongly correlated with theoretically related constructs than was the self-report EPQ-R. This was clearly the case for Inter-

personal Sensitivity ( $Z = -2.07, p = .038$ ) and Paranoid Ideation ( $Z = -2.89, p = .004$ ), though the difference for Psychoticism was marginal ( $Z = -1.85, p = .064$ ).

At the item level, N correlated significantly with 48 symptoms and showed trends for 10 more. Correlations with 7 of these symptoms reached or exceeded  $r = .50$ : feeling lonely, feeling self-conscious, worrying too much, feeling blue, being afraid that people will take advantage of you, recurrent unpleasant thoughts that won't go away, and trouble concentrating.

No correlations with the other RISB scales were so strong; the SCL-90R is primarily a measure of Neuroticism-related experiences. However, there were some noteworthy findings concerning the RISB Extraversion and Psychoticism scales.

Correlations with E reached or exceeded  $r = .35$  for 6 items, all in the negative direction. In order of magnitude, people high in RISB-measured Extraversion are *less* likely to be uneasy in crowds, never feel close to another person, be afraid others will take advantage of them, feel lonely, feel blue, or suffer from recurrent unpleasant thoughts.

Correlations with P equaled or exceeded  $r = .35$  for just 3 items. In order of magnitude, people high in RISB-measured Psychoticism are more likely to have ideas or beliefs others do not share, never feel close to another person, and have urges to break or smash things. Other significant correlates included feeling alone even when with people, feeling critical of others, uncontrollable temper outbursts, feeling tense or keyed up, feeling watched or talked about by others, and recurrent unpleasant thoughts.

## ADDITIONAL VALIDITY STUDY #2: PSYCHOTICISM AND CREATIVITY

As noted earlier, evidence linking trait Psychoticism with creativity has emerged in many studies (Acar & Runco, 2012; Eysenck, 1995). It is therefore hypothesized that RISB P scores will correlate positively with performance on tasks requiring originality and creativity.

### METHOD

Many members of samples #2 and #4 participated in studies of creativity. One group ( $n = 54$ ) completed a set of House-Tree-Person drawings; most of them ( $n = 40$ ) also wrote a short poem. Another group ( $n = 45$ ) completed a series of drawings using colored pencils. Details of the creativity tasks and analyses from these studies have been reported previously (Joy, 2005, 2008); briefer descriptions will suffice for present purposes.

The House-Tree-Person (HTP; Buck, 1992) was used here to elicit artistic productions. Drawings were rendered in #2 pencil on unlined sheets of

8.5" x 11" paper. They were analyzed in two ways. First, unusual features were tabulated: elements that occurred infrequently in the sample, such as a house presented as a floor plan or a tree with a swing. These were summed to yield a composite originality score. (To control for expenditure of effort, a separate score was derived for *common* details in the house drawings, such as shrubbery.) Second, ratings of the technical proficiency and creativity displayed by each artist were made by two art therapy graduate students with reliabilities of .76 and .77, respectively. (All reliability figures reported here for creativity ratings are values for pooled judgments. Further details concerning reliability may be found in the cited studies.)

The poems were written following a structured procedure. Participants began with a stimulus word ("door"). They were instructed to write down the first word they associated with the stimulus word, then a word suggested by their first association, and so forth until they had a list of 6 words. They were then to write a poem using each word in its own line.

The poems, too, were analyzed in two ways. First, objective indicators of originality and arousal potential (Martindale, 1990) were tabulated. These included (1) the number of words in the poem, (2) the infrequency with which those words are used in written English, (3) the ratio of "unique" words (those used only once) to repeated words, (4) the rarity of the word associations, (5) blends of positive and negative emotions, and (6) contrasting pairs of words (e.g., "open" vs. "shut"). An originality composite was derived by combining the indicators. Second, two English professors rated the poems on overall poetic value. This rating had effective reliability of .68.

The drawings in the other creativity study were executed on 9" x 12" drawing paper using sets of 12 colored pencils. Each participant completed 3 drawings: one was something made by humans, one was something living (other than a person), and one was a person. These directives were modeled on those for the HTP, but modified to allow more scope for originality.

As with the HTP, these drawings were evaluated in two ways. First, originality was rated. For the first two, this meant awarding more points to infrequently chosen themes. For example, many participants drew a car or a flower, but very few drew a factory or a paramecium. For the human figure, such variables as depicting a person engaged in an activity or from a perspective other than full-face were scored. Second, the drawings were evaluated for technical proficiency and creativity by 3 art therapy graduate students with reliabilities of .84 and .77, respectively.

### RESULTS

Table 8 displays the correlations between RISB trait scores and the main HTP variables. As expected, Psy-

choticism correlated significantly with all originality scores and ratings of both proficiency and creativity. This was not due simply to investing effort, as shown by the lack of a correlation with the “common details” score. The pattern of results is actually much stronger than for self-reported P, which correlated at only  $r = .31$  ( $p < .050$ ) with the originality composite and nonsignificantly with the judges’ ratings (Joy, 2008). Even after partialing out self-reported P, the RISB P score correlated significantly with most HTP ratings, including the total originality score (partial  $r = .49$ ), judged technique ( $r = .46$ ), and judged creativity ( $r = .41$ , all  $p < .010$ ).

N showed a weak relationship with originality, perhaps due to investment of effort (see the “common details” score). E tended to correlate negatively with originality and creativity; introverted people may be more able or willing to generate original artwork. Perusal of the RISB protocols suggests that many high E participants would rather have been playing basketball.

Table 8 also displays the correlations between RISB trait scores and the originality and creativity ratings made of the associative poems. P correlated significantly with every indicator of originality; the correlation with the composite was a remarkable  $r = .71$ . P also correlated significantly with the judged quality of the poems. As with the HTP, the RISB P measure was more strongly related to originality and creativity than was self-report. The EPQ-R P scale correlated with the originality composite at  $r = .40$  ( $p < .050$ ), but failed to correlate significantly with the judged value of the poems. After partialing out self-reported P, the RISB P score still correlated significantly with the originality composite (partial  $r = .58$ ,  $p < .010$ ) and several of its components; it also displayed a strong trend ( $r = .31$ ,  $p < .060$ ) on judged poetic merit.

N exhibited a weak tendency to correlate with originality, though when all the indicators were summed together the relationship reached statistical significance. E displayed a weak tendency to correlate negatively with originality. Neither correlated with judged poetic merit.

Table 8 also displays the correlations between RISB trait scores and chromatic drawings. P correlated significantly with unusual choices of man-made artifacts and with the judged quality of the drawings. N showed only a weak tendency to correlate with originality and was unrelated to judged creativity. E tended to correlate negatively with originality and technical proficiency, but was unrelated to judged creativity.

## DISCUSSION

The present findings demonstrate that accurate measurements of personality traits can be derived from

the Rotter Incomplete Sentences Blank (RISB). Even students can achieve solid inter-rater reliability. Strong correlations with self-reports for each trait support the system’s convergent validity; weaker correlations with the other traits (and with defensiveness) provide evidence of discriminant validity.

There are, of course, issues with the data. To begin with, the samples are relatively small. To some extent this is inevitable. Unlike self-reports (which can be computer-scored *en masse*), each RISB protocol requires the attentive work of a human judge. Rating 231 protocols for 3 traits requires 27,720 individual judgments. The undergraduate (and mostly female) nature of the samples is also less than ideal. This is a common failing, but one looks to see results from other populations in future studies. It is *possible* that some criteria for making ratings will differ. It is *likely* that members of other populations will have some different ways of expressing the criteria. It is *nearly certain* that the distributions of scores will differ. Studies of adolescents and of older adults may reveal ways in which these expressive traits develop across the lifespan. Studies of participants drawn from other cultures are also likely to be informative about the extent to which the personal expression of traits holds true or varies across national or linguistic contexts. And of course, studies of persons diagnosed with various forms of psychopathology will be essential for the effective use of the system in clinical settings. In this regard, the present work represents a potential beginning rather than a final summation.

The relatively high correlations among the three RISB scales also are of concern, though each correlated more strongly with its self-report equivalent than with the other self-report scales. It is, incidentally, possible that personality traits *do* correlate more strongly with one another where expressive behavior is concerned. On a self-report, every person must produce a scorable response to every item. But on performance-based tasks, some individuals project much more personality than others. This may be true of sentence completions just as it is in parties or meetings. We do not customarily think of a general factor underlying all personality traits, but in everyday life we experience some people as having “more personality” than others.

All that having been said, the present results may be even more impressive when one goes beyond the correlations with the EPQ-R and examines other criterion measures.

Evidence from the SCL-90R indicates that the three traits, as measured by the RISB, are related to psychological dysfunction in ways consistent with their definitions. First, Neuroticism is associated with a wide range of distressing symptoms: high N people are plagued with dysphoric moods and feel rejected, in addition to experiencing many other woes. Second, the impact of Psychoticism on mental health is

subtler and more selective, but certain vulnerabilities were apparent. When high P people experience adjustment problems, they tend to feel alienated or estranged from others, to feel powerful aggressive urges they cannot readily control, and to entertain grandiose and persecutory ideas. Finally, Extraversion appears to be somewhat protective against psychological suffering. This is most noteworthy in the social realm; high E people feel comfortable with and connected to their fellow humans. Whether because of the support provided by these relationships or due to their natural energy and robustness, they are also less likely to complain of anxiety-related complaints. None of the above findings are especially surprising, but they are important in that they support the construct validity of the RISB personality scales.

Evidence from several art and writing tasks indicates that Psychoticism, as measured by the RISB, is related to originality of conception and skill in the execution of creative works. First, high P people produce more original drawings. Given a choice, they tend to select unusual themes. Constrained to a single theme, they often treat it in unusual ways. Judges consider their works to exhibit higher levels of creativity. Second, high P people craft more original, arousing poems. They make unusual associations, employ uncommon words, surprise readers with novel words (rather than repeating themselves), and invoke contrasts. This tendency is pronounced; the correlation with the poetic originality composite approaches the theoretical ceiling imposed by the reliabilities of the measures. Expert judges also affirm that poems written by high P people tend to be of superior quality. Equally important, neither E nor N displayed a comparable association with artistic or poetic creativity.

Again, most of this confirms what is already believed about trait Psychoticism. Other than Openness to Experience, it is the higher-order personality factor most often linked to creativity (Acar & Runco, 2012; Eysenck, 1995). But here the RISB is not only behaving in accordance with theoretical predictions; the RISB P scale *outperformed* its self-report parent, which yielded weaker correlations with originality and none with judged creativity.

Given that the reliability of this RISB measure exceeds that of the self-report and that it correlates more strongly with criterion measures, it may shed further light on the P construct. Some RISB Psychoticism criteria are not simple reflections of the content of the self-report used to define them or of the basic definition of P. The sexual and hunger themes are a good example. They make sense (as with aggression, they express primitive drives), but nothing in the EPQ-R determined this. The same may be said of the frequent references to alcohol and drug use; they fit with what we know about high P individuals, but were not “built in” to the self-report scale.

More noteworthy still are the many completions that reflect preoccupation with odd ideas, “escape” into a world of the imagination, and/or concern over a perceived loss of control over one’s own mental processes. These suggest a tendency to withdraw from consensual reality, preferring a privately constructed version of the world to the confusing one in which those strange creatures, other people, dwell. Sometimes the tendency is subtle (as in, “*my mind is a maze*”), sometimes blatant (as in, “*my mind travels in circles and goes down dead ends,*” or even “*I want to know if anyone else truly exists, or if I even do, if everything is just something I made up*”). In some cases, we may see an unlocking of creativity – these are students! In other cases, the result may be more pathological: a schizotypal personality or an incipient psychotic break. While this vulnerability is integral to the original conceptualization of P, thought disorder plays no role in the self-report scale. This also weakens the argument that P is “really” just psychopathy; it appears that subclinical psychopathy and mild thought disorder do overlap.

This fits not only with Eysenck’s work, but also with Meloy’s (1988) psychodynamic model of psychopathy. Meloy suggests that the egocentric arrogance typical of psychopaths is often accompanied by signs of thought disorder and quasi-delusional ideas. Primitive defenses such as splitting serve a grandiose self-structure; if the psychopath is exposed and defeated, the resulting collapse of self-esteem (the “worthless” side of the split) also may feature transient psychotic symptoms.

A validated method for deriving personality trait scores from RISB protocols adds materially to the data supplied by the test. It also may facilitate interpretation. Completions may be grouped according to the scales on which they were scored and patterns identified – as well as interactions, such as items scored both for high P and high N. Thus, there could be a stronger link between objective findings and clinical inferences than is presently the case. And of course, the information in an RISB protocol is not limited to traits. Use of this scoring system does not replace qualitative analysis of RISB protocols – it simply adds a new layer of meaning.

On a theoretical level, sentence completions (as illustrated by the present findings with regard to P) can enrich our understanding. In this respect, a performance-based measure may be superior to a self-report. Self-report scales are closed systems. We can examine their correlations with one another, constructing nomological networks replete with statistical pyrotechnics, but we will learn nothing *new* about the constructs until we go beyond self-report (McClelland, 1972). Qualitative study of an RISB protocol, however, may suggest new insights. It is akin to the difference between open-ended and closed-ended questions: each is valuable for certain purposes.

Further work obviously can be done. With the *Manual* in place, there is ample scope for research in new populations (clinical, forensic, cross-cultural, etc.) and with new criterion measures (structured interviews, other self-report and performance-based tests, etc.) that will enhance our understanding of those populations and the clinical utility of the RISB. The sentence completion method may be due for renewed attention.

## ACKNOWLEDGMENTS

The author would like to acknowledge the invaluable assistance of eight talented undergraduate students, who mastered an initially rudimentary scoring system and provided invaluable feedback: Natasha Parlato, Michelle Battista, Michael Lennon, Albi Beshi, Heather Mills, Stephanie Strosahl, Erika Donoso, and Emili Dubar.

## REFERENCES

- Acar, S., & Runco, M. A. (2012). Psychoticism and creativity: A meta-analytic review. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 341–350.
- Becker, G. (2001). The association of creativity and psychopathology: Its cultural-historical origins. *Creativity Research Journal*, 13, 45–53.
- Benedek, M., Konen, T., & Neubauer, A. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 273–281.
- Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment*, 11, 48–57.
- Bornstein, R. F. (2002). A process dissociation approach to objective-projective test score interrelationships. *Journal of Personality Assessment*, 78, 47–68.
- Bornstein, R. F., Rossner, S. C., & Hill, E. L. (1994). Retest reliability of scores on objective and projective measures of dependency: Relationship to life events and intertest interval. *Journal of Personality Assessment*, 62, 398–415.
- Buck, J. N. (1992). *House-tree-person manual and interpretive guide*. Los Angeles, CA: Western Psychological Services.
- Chapman, J. P., & Chapman, L. J. (1994). Does the Eysenck Psychoticism scale predict psychosis: A ten-year longitudinal study. *Personality and Individual Differences*, 17, 369–375.
- Chapman, J. P., & Chapman, L. J. (1996). The psychometric assessment of psychosis proneness. In S. Matthyse, D. Levy, J. Kagan, & F. Benes (eds.), *Psychopathology: The emerging science of mental disorder* (pp. 313–333). New York: Cambridge University Press.
- Clark, L. A., & Watson, D. (2008). Temperament: An organizing paradigm for trait psychology. In O. P. John, R. W. Robins, & L. Pervin (eds.), *Handbook of personality: Theory and research* (pp. 265–286). New York: Guilford.
- Collins, C. J., Hanges, P. J., & Locke, E. A. (2004). The relation of achievement motivation to entrepreneurial behavior: A meta-analysis. *Human Performance*, 17, 95–117.
- Costa, P. T., & McCrae, R. R. (1992). *NEO Personality Inventory - Revised (NEO-PI-R)*. Odessa, FL: PAR.
- Derogatis, L. (1983). *SCL-90-R Manual*. Baltimore, MD: Clinical Psychometrics Press.
- Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology*, 83, 1441–1455.
- Eysenck, H. J. (1992). The definition and measurement of psychoticism. *Personality and Individual Differences*, 13, 757–785.
- Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality*. London: Hodder & Stoughton.
- Eysenck, H. J., & Eysenck, S. B. G. (1994). *Manual for the Eysenck Personality Questionnaire*. San Diego, CA: EdITS.
- Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review*, 2, 290–309.
- Gray, J. A., & McNaughton, M. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system* (2<sup>nd</sup> ed.). Oxford: Oxford University Press.
- Greenwald, A. G., & Farnham, S. D. (2000). Use of the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022–1038.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Hare, R. D. (1982). Psychopathy and the personality dimensions of psychoticism, extraversion, and neuroticism. *Personality and Individual Differences*, 3, 35–42.
- Hare, R. D. (2003). *Manual: The Psychopathy Checklist - Revised*. Toronto, Ontario: Multi-Health Systems.
- Heaven, P. C. L., Ciarrochi, J., Leeson, P., & Barkus, E. (2013). Agreeableness, conscientiousness, and psychoticism: Distinctive influences of three personality dimensions in adolescence. *British Journal of Psychology*, 104, 481–494.
- Heym, N., Ferguson, E., & Lawrence, C. (2013). The P-psychopathy continuum: Facets of Psychoticism and their associations with psychopathic

- tendencies. *Personality and Individual Differences*, 54, 773–778.
- Heym, N., & Lawrence, C. (2010). The role of Gray's revised RST in the P-psychopathy continuum: The relationships of Psychoticism with a lack of fear and anxiety, and increased impulsivity. *Personality and Individual Differences*, 49, 874–879.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- Holaday, M., Smith, D. A., & Sherry, A. (2000). Sentence completion tests: A review of the literature and results of a survey of members of the Society for Personality Assessment. *Journal of Personality Assessment*, 74, 371–383.
- Hy, L. X., & Loevinger, J. (1996). *Measuring ego development* (2<sup>nd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Joy, S. P. (2005). Innovation motivation and artistic creativity. *Journal of Creative Behavior*, 39, 35–56.
- Joy, S. P. (2008). Personality and creativity in art and writing: Innovation motivation, psychoticism, and (mal)adjustment. *Creativity Research Journal*, 20, 262–273.
- Kollner, M. G., & Schultheiss, O. C. (2014). Meta-analytic evidence of low convergence between implicit and explicit measures of the needs for achievement, affiliation, and power. *Frontiers in Psychology*, 5, article #826.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27–66.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252–271.
- Martindale, C. (1990). *The clockwork muse: The predictability of artistic styles*. New York: Basic Books.
- Masling, J. M., Rabie, L., & Blondheim, S. H. (1967). Obesity, level of aspiration, and Rorschach and TAT measures of oral dependence. *Journal of Consulting Psychology*, 31, 233–239.
- McClelland, D. C. (1972). Opinions predict opinions: So what else is new? *Journal of Consulting and Clinical Psychology*, 38, 325–326.
- McClelland, D. C. (1985). How motives, skills and values determine what people do. *American Psychologist*, 40, 812–825.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, 96, 690–702.
- McCloskey, L. C. (2014). Construct and incremental validity of the RISB in adult psychiatric outpatients. *Psychological Reports: Measures & Statistics*, 114, 1–13.
- Meloy, R. (1988). *The psychopathic mind: Origins, dynamics and treatment*. Northvale, NJ: Aronson.
- Merten, T. (1993). Word association responses and psychoticism. *Personality and Individual Differences*, 14, 837–839.
- Miner, J. B. (1964). *Scoring guide for the Miner Sentence Completion Scale*. Atlanta, GA: Organizational Measurement Systems Press.
- Pickering, A., & Corr, P. (2008). J. A. Gray's Reinforcement Sensitivity Theory (RST) of personality. In Boyle, G. J., Matthews, G., & Saklofske, D. H. (eds.), *Sage handbook of personality theory and assessment, volume one*. Thousand Oaks, CA: Sage.
- Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, 17, 555–564.
- Rapaport, D. (1946). The word association test. In D. Rapaport, M. Gill, & R. Schafer, *Diagnostic psychological testing: The theory, statistical evaluation, and diagnostic application of a battery of tests volume II*. Chicago, IL: Year Book Publishers.
- Richetin, J., South Richardson, D., & Mason, G. D. (2010). Predictive validity of IAT aggressiveness in the context of provocation. *Social Psychology*, 41, 27–34.
- Rotter, J. B., Lah, M. I., & Rafferty, J. E. (1992). *Rotter Incomplete Sentences Blank Manual, Second Edition*. San Antonio, TX: Psychological Corporation.
- Rotter, J. B., & Willerman, B. (1947). The Incomplete Sentences Test as a method of studying personality. *Journal of Consulting Psychology*, 11, 43–48.
- Schachter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 432–444.
- Schultheiss, O. C., Yankova, D., Dirlikov, B., & Schad, D. J. (2009). Are implicit and explicit motive measures statistically independent? A fair and balanced test using the picture story exercise and a cue- and response-matched questionnaire measure. *Journal of Personality Assessment*, 91, 72–81.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Skeem, J. L., Poythress, N., Edens, J. F., Lilienfeld, S. O., & Cale, E. M. (2003). Psychopathic personality or personalities? Exploring potential variants of psychopathy and their implications for risk assessment. *Aggression and Violent Behavior*, 8, 513–546.
- Smith, C. P., Atkinson, J. A., McClelland, D. C., & Veroff, J. (eds.). (1992). *Handbook of thematic content analysis*. New York: Cambridge University Press.
- Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, 112, 140–152.
- Torstrick, A., McDermut, W., Gokberk, A., Bivona, T., & Walton, K. E. (2015). Associations between the Rotter Incomplete Sentences Blank and measures

- of personality and psychopathology. *Journal of Personality Assessment*, 97, 494–505.
- Weis, R., Toolis, E. E., & Cerankosky, B. C. (2008). Construct validity of the Rotter Incomplete Sentences Blank with clinic-referred and nonreferred adolescents. *Journal of Personality Assessment*, 90, 564–573.
- Woody, E., & Claridge, G. (1977). Psychoticism and thinking. *British Journal of Social and Department of Psychology Clinical Psychology*, 16, 241–248.