A Thesis

Entitled

Development and Preliminary Validation of a Brief Behavioral Measure of Problems in Thought Organization and Perception

by

Joshua J. Eblin, B.A.

Submitted to the Graduate Faculty as partial fulfillment of the requirements for

The Master of Arts Degree in Psychology

____________________________
Gregory Meyer, Ph.D., Committee Chair

____________________________
Joni Mihura, Ph.D., Committee Member

____________________________
Donald Viglione, Ph.D., Committee Member

____________________________
Jeanne Brockmyer, Ph.D., Committee Member

____________________________
Patricia Komuniecki, Ph.D., Dean
College of Graduate Studies

The University of Toledo

May 2012
An Abstract of

Development and Preliminary Validation of a Brief Behavioral Measure of Problems in Thought Organization and Perception

by

Joshua J. Eblin, B.A.

Submitted to the Graduate Faculty as partial fulfillment of the requirements for
The Master of Arts Degree in Psychology

The University of Toledo
May 2012

The present research represents the first step in a planned series of studies that will ultimately culminate in a Rorschach-based system that assesses solely for problems in thinking and perception. The system is provisionally dubbed the Thinking and Perception Behavioral Assessment System (TP-BAS) and is designed for use in clinical practice and research. The TP-BAS will yield time savings, the ability to conduct test-retest evaluations, and dimensional behavioral assessment data for the identification of problems in thinking and perception. The most recently developed system for conducting broadband Rorschach assessments is the Rorschach Performance Assessment System (R-PAS; Meyer, Viglione, Mihura, Erard, & Erdberg, 2011). The aim of the present research was to develop series of “short-forms” of R-PAS Thinking and Perception variables. Each short-form makes use of only the Thinking and Perception variables. One short-form consisted of one 10-card option, another short-form consisted of two 5-card options, another short-form consisted of two 4-card options, and the final short-form consisted of three 3-card options. Thus, short-forms of the R-PAS were created in two ways. First, the number of variables was reduced to just the Thinking and Perception variables. Second,
alternatives to using the full Rorschach card set were developed by reducing the number of cards to be administered in all but one of the TP-BAS card sets. It was first determined which R-PAS Perception and Thinking variables meaningfully differ in the frequency with which they occur across cards. Second, on a card-by-card basis, the mean and standard deviation were computed for the variables that were found in to meaningfully differ in frequency across cards in step 1. Third, TP-BAS card sets were assembled by balancing their relative difficulty, as determined in steps 1 and 2. Fourth, card sets and series were compared to one another regarding part-whole correlations with the full-form, interrater reliability, and construct validity. Diagnostic efficiency statistics for the TP-BAS are also reported using data from a composite measure of perceptual and thinking ability (i.e., the Thought and Perception Composite) as an example. The present analyses suggest that each series has unique merits and limitations, with the 4-card series perhaps being slightly less optimal than the 5- and 3-card series. In general, part-whole correlations, reliability coefficients, validity coefficients, and overall correct classification rates (OCCs) are acceptable for each card set and slightly decrease as a function of decreasing the number of cards in each respective short-form. Implications and future uses of the TP-BAS for research and clinical assessment are discussed.
Acknowledgements

This project would not have been possible without the thoughtful contributions and feedback from each of my committee members. Their feedback has greatly shaped my understanding of the Rorschach assessment of thought disorder and psychosis. Additionally, the chair of my committee, Dr. Gregory Meyer, always made himself available for assistance throughout the duration of this project and provided prompt and thorough responses to my inquiries. I am also grateful to the respondents who allowed for their Rorschach protocols to be used in this research and to the examiners who collected the data. Finally, I would like to thank my family and friends for their continued support throughout the completion of this project, which was successfully defended on December 14, 2011.
Table of Contents

Abstract iii

Acknowledgements v

Table of Contents vi

List of Tables viii

List of Figures ix

List of Abbreviations x

Preface xii

I. Introduction and Literature Review 1
   A. Definitions 1
   B. Rorschach Assessment 3
   C. Broadband Rorschach Systems 4
   D. Rorschach Assessment of Problems in Thinking 8
   E. Rorschach Assessment of Problems in Perception 15
   F. Combined Rorschach Assessment of Problems in Thinking and Perception 16
   G. Why the Rorschach? 19
   H. The DSM-5 Dimensional Assessment of Psychosis 23
   I. A Brief Review of Select Short-Form Development Efforts 30
      a. Short-Form Development Sin #1 37
      b. Short-Form Development Sin #2 38
      c. Short-Form Development Sin #3 39
II. Development and Validation of the Thinking and Perception Behavior Assessment System

A. Statement of the Problem 43

B. Purpose of the Present Research 44

C. Samples 45

D. Selecting Preliminary Short-Form Card Sets 46

E. Preliminary Evaluation of the TP-BAS 48

F. Interrater Reliability 49

G. Examination of Short-Form Validity 50

H. Criterion Measures 50

I. Post-Hoc Power Analyses 52

J. Selecting the Optimal Series of Short-Form Card Sets 53

K. Results 55

III. Discussion and Implications 72

A. Discussion 72

B. Limitations & Future Directions 77

C. Conclusion 85

References 88
List of Tables

Table 1. Delta Index – Summary of Scoring Values……………………………………… 11
Table 2. Thought Disorder Index Criteria & Formulae…………………………………… 13
Table 3. Formulae for Rorschach Indices of Problems in Thinking & Perception……… 18
Table 4. A Continuum of Ego Deficiency for DSM Diagnostic Codes………………… 52
Table 5. Post-Hoc Power Analyses……………………………………………………… 53
Table 6. Diagnostic Efficiency Statistics Illustration…………………………………… 55
Table 7. Dean, Chicago, and R-PAS Samples – Variable Mean by Card……………… 57
Table 8. Dean, Chicago, and R-PAS Samples - Final Card Selections………………… 58
Table 9. Combined Dean, Chicago, and R-PAS Samples – Card Set by Protocol-Level Mean of Variable…………………………………………………………… 60
Table 10. Dean, Chicago, and R-PAS Samples – Protocol-Level Full-Form to Short-Form Part-Whole Correlations………………………………………………… 62
Table 11. Dean Sample – Protocol-Level TP-BAS Validity; Criterion = Thought Disorder Summary Scale…………………………………………………………… 64
Table 12. Chicago & R-PAS Samples – Protocol-Level TP-BAS Validity; Criterion = Ego Dysfunction Severity Scale………………………………………………… 65
Table 13. Protocol-Level TP-BAS Interrater Reliability – Meyer et al. (2011) Sample...66
Table 14. TP-Comp TP-BAS Diagnostic Efficiency – Chicago and R-PAS Samples….71
List of Figures

Figure 1. Dean Sample - Mean Part-Whole Correlations and Mean Validity.............67

Figure 2. Chicago and R-PAS Samples - Mean Part-Whole Correlations and Mean Validity.................................................................68

Figure 3. Meyer et al., (2011) Sample - Mean TP-BAS Interrater Reliability............69
List of Abbreviations

CON........... Contamination
CS............. Comprehensive System

DR1........... Deviant Response – Level 1
DR2........... Deviant Response – Level 2
DV1........... Deviant Verbalization – Level 1
DV2........... Deviant Verbalization – Level 2

EII........... Ego Impairment Index

FAB1........... Fabulized Combination – Level 1
FAB2........... Fabulized Combination – Level 2
FQ............. Form Quality
FQo........... Ordinary form quality
FQu........... Unusual form quality
FQ-........... Minus form quality

GHR........... Good Human Representation

INC1........... Incongruous Combination – Level 1
INC2........... Incongruous Combination – Level 2

Lv2........... A CS variable: the unweighted sum of the level 2 Cognitive Codes

M.............. The protocol-level sum of human movement responses also coded FQ-
MMPI-2....... Minnesota Multiphasic Personality Inventory – Second Edition

P.............. Popular
PEC.......... Peculiar Logic
PHR........... Poor Human Representation
PTI.......... Perceptual Thinking Index

R.............. Total number of responses in protocol
R-PAS........ Rorschach Performance Assessment System

Sum6........... A CS variable: the unweighted sum of the Cognitive Codes

TDI........... Thought Disorder Index
TP-Comp...... Thought and Perception Composite
WDA%........ A CS variable: the percentage of responses given to common card locations that also have good FQ

WD-%.......... Responses to W and D locations with an FQ-code divided by the total number of responses to W and D locations; provides an index of poor form quality to the common location areas

WSumCog…. Weighted sum of 6 Cognitive Codes

X+%............ A CS variable: conventional form percent, FQo divided by R
XA%............ A CS variable: the percentage of responses in which there is an appropriate use of form
Preface

Problems in thinking and perception are important components of a diverse array of severe psychological disorders and are core characteristics of schizophrenia and the other psychoses (e.g., schizoaffective disorder, bipolar disorder, major depressive disorder with psychotic features, and substance induced psychosis). The psychoses are among the most debilitating psychiatric disorders (Morrison, 1995). For example, individuals with schizophrenia often lack steady employment, housing, and meaningful interpersonal relationships, and have difficulties with self care (Combs & Meuser, 2007). The level of distress and dysfunction associated with the psychoses makes their reliable and valid assessment of clinical importance. Notably, the psychoses are best characterized as continuous rather than categorical syndromes (Allardyce, Gaebel, Zielasek, & van Os, 2007; Dutta et al., 2007; Regier, 2007; van Os & Tamminga, 2007). As such, the forthcoming DSM-5 diagnostic criteria for the psychoses will be dimensional in nature (American Psychiatric Association, 2011b).

The Rorschach has its roots in the assessment of problems in thinking and perception (Rorschach, 1942) and has demonstrated over the years that the assessment of psychosis is its primary strength (Mihura, Meyer, Bombel, & Dumitrascu, 2011; Wood, Lilienfeld, Garb, & Nezworski, 2000). The test also validly assesses for an array of other personality traits (Mihura et al., 2011) and is most commonly used in contemporary practice as a broadband behavioral personality assessment measure. R-PAS’s utility is somewhat limited when the purpose of an assessment is only to assess for psychosis as
the system requires administration of all 10 Rorschach cards and the coding of many variables that are not directly relevant to such assessments. Several other Rorschach thought disorder scoring systems do exist (e.g., Carpenter et al., 1993; Holt, 1956, 1977; Holt & Havel, 1960; Johnston & Holzman, 1979; Rapaport, Gill, & Schafer, 1968; Watkins & Stauffacher, 1952) but are limited for various reasons (e.g., methodological limitations in their derivation, time to teach, learn, and score). As such, in the present research, a series of short-form Rorschach card sets that make use of only of the R-PAS Thinking and Perception variables are developed. These short-form card sets are intended to assess for problems in thinking and perception in a reliable, valid, and efficient manner.
Chapter One

Introduction and Literature Review

Definitions

One way to conceptualize psychosis is as a decompensation of one’s “reality testing” ability, a concept initially introduced by Sigmund Freud in 1895. In his initial conceptualization, Freud indicated that the essence of reality testing is the capacity to recognize the distinction between one’s internal ideas and his or her perceptions of the external environment. Others (i.e., Balint, 1942; Hartmann, 1953) have since added that reality testing also encompasses the ability to find meaning in an accurately perceived environmental stimulus.

Psychotic symptoms are classified as being either positive or negative. “The positive symptoms appear to reflect an excess or distortion in normal functions, whereas the negative symptoms appear to reflect a diminution or loss of normal functions” (American Psychiatric Association, 2000, p. 299). The positive symptoms include delusions, hallucinations, disorganized speech (i.e., thought disorder), and “grossly disorganized or catatonic behavior” (American Psychiatric Association, 2000, p. 299). The negative symptoms include affective flattening, alogia (i.e., poverty of speech), avolition (i.e., lack of motivation), anhedonia (i.e., loss of interest and/or pleasure), and dysphoric mood (American Psychiatric Association, 2000).

Though the DSM-IV-TR does caution that a universally accepted definition of psychosis does not exist, it states that “The narrowest definition of psychosis is restricted to delusions or prominent hallucinations, with the hallucinations occurring in the absence of insight into their pathological nature” (American Psychiatric Association, 2000, p.
Delusions typically constitute misinterpretations of perceptions and/or experiences that lead to false beliefs that may take many forms (e.g., paranoid, persecutory, grandiose, or ideas of reference in which the individual has unfounded suspicions that they are being talked about; American Psychiatric Association, 2000). Hallucinations are perceptions occurring in the absence of external stimuli and may manifest via any of the five senses. Some of the most common are auditory hallucinations wherein the afflicted individual hears voices (Combs & Meuser, 2007).

Thought disorder is another important component of a wide variety of disorders. Until the mid-70s, thought disorder was believed to be such a central component of schizophrenia that the two were practically treated as if they were synonymous. Since that time, some researchers have asserted that thought disorder can exist independent of psychosis (Kleiger, 1990). In fact, considerable research efforts have been devoted to investigating the phenomenon as an independent construct (e.g., Kleiger, 1990; Johnston & Holzman, 1979). Supposing that they are truly independent, it is fair to say that thought disorder is not a sufficient criterion for diagnosing psychosis, though many individuals with diagnoses of psychosis do have thought disorder.

Thinking can be disordered in content and/or form (Bachman & Cannon, 2005). With regard to disordered thought content, an individual may experience odd or nonsensical thoughts and beliefs (i.e., delusions). With regard to thought form, the structure of one’s speech may demonstrate looseness of associations in thinking (i.e., “word salad”). Kleiger offers a more sophisticated definition of disturbed thought structure by stating that:
A more comprehensive definition of thought disorder would be one that encompasses a broader perspective that includes not only traditional concepts such as impaired pace and flow of associations, but also such factors as errors in syntax, word usage, syllogistic reasoning, inappropriate levels of abstracting, failure to maintain conceptual boundaries, and a breakdown in the discrimination of internal perceptions from external ones (Kleiger, 1999, p. 6).

There has been considerable debate regarding the nature and structure of thought disorder and psychosis. In particular, some researchers still consider thought disorder to be a manifestation of psychosis that does not warrant consideration as an independent entity (American Psychiatric Association, 2000). To complicate matters further, many psychiatric diagnoses do not require, but may include, the presence of psychotic processes (e.g., major depressive disorder). In the present research, novel definitions are not offered for either construct in an attempt to settle the aforementioned disputes; rather, the ambiguity is dealt with by treating thought disorder and psychosis as separate, albeit not mutually exclusive, entities. Doing so allows for the most specificity and clarity of presentation. It is notable that the behavioral manifestations of psychosis and thought disorder very often co-occur and are conceptually overlapping. As such, the two conditions are often difficult to tease apart in clinical cases.

**Rorschach Assessment**

Standard Rorschach assessment is comprised of 10 inkblot cards to which respondents are prompted to answer the open-ended question “What might this be?” The test is best thought of as a performance-based task that provides examiners with samples
of respondent behavior that can be recorded, coded, and interpreted for assessment purposes (Meyer & Kurtz, 2006). Responses to the 10 inkblot stimuli can vary substantially, both within and between different respondents, because it is a relatively unstructured test. Ultimately, all of the responses given by the respondent are aggregated according to scoring formulas to form indices of personality processes and mental health. As such, inferences can be made on the basis of Rorschach data regarding the personality traits and degree of mental health of the respondent being evaluated.

The study of thought disorder and psychosis and Rorschach assessment share a number of features in common, most notably their conceptualization and their history (Kleiger, 1999). Emil Kraepelin (1896) and Eugen Bleuler (1911) were the first to identify and study what is currently referred to as “schizophrenia.” The Rorschach’s originator, Hermann Rorschach, was a student of Eugen Bleuler. Under Bleuler, Rorschach studied the clinical phenomena of hallucinations and developed his test as a means of differentially diagnosing schizophrenia from other psychopathologies (Leichtman, 1996). In his initial system, Rorschach included several codes to capture disturbed thinking and perception that reflect contemporary codes used to measure the same phenomenon. For example, his system included indices of poor form, bizarre content, “confabulatory-combined whole responses” (currently referred to as fabulized combinations), and “contaminated whole responses” (currently referred to as contaminations; Rorschach, 1942).

**Broadband Rorschach Systems**

From 1942 until the Comprehensive System (CS) was initially published in 1974 (Exner, 1974) there were five commonly used Rorschach systems available (i.e., Beck,
1944, 1945; Hertz, 1951; Klopfer & Kelly, 1942; Piotrowki, 1957; Rapaport, Gill, & Shafer, 1946, 1968). Each system provided unique guidelines for the administration, scoring, and interpretation of the test. As its name implies, the CS was intended to synthesize the extant systems into one unitary system that contained what Exner believed were the most valid Rorschach practices across the available systems. Although not without its critics, the CS did make notable strides in the restoration of scientific credibility to the Rorschach (Anastasi, 1988). The CS dominated Rorschach practice and research from its initial publication until the 2011 publication of the Rorschach Performance Assessment System (R-PAS; Meyer et al., 2011).

R-PAS (Meyer et al., 2011) is largely an outgrowth of research endeavors initiated by Exner’s Rorschach Research Council (RRC; Exner, 1997). That research was initially launched with the intention of refining CS administration and coding procedures and four of the five R-PAS systematizers served on the RRC. However, the RRC does not retain legal ownership of the CS and, consequently, was unable to publish their updates to the CS under the name “Comprehensive System” after Exner’s death.

The four primary refinements to emerge from the RRC’s work are the use of internationally collected and consolidated normative data, the introduction of R-optimized administration procedures (Dean et al., 2007), the fine-tuning of the variables included on the basis of validity evidence (Mihura et al., 2011), and updates made to the form quality (FQ) tables. R-PAS normative data are used in the present research and, as such, are described in the Methods section. R-optimization procedures were first tested by Dean et al. (2007) and later modified and adopted into R-PAS (Meyer et al., 2011) to deal with the long-standing problem of large variability in the number of responses given
when standard CS administration procedures are followed. Specifically, when CS administration guidelines are used, the respondent is able to give either very few responses (14 is a minimum) or an unwieldy number of responses (e.g., 60 or more). Too few responses causes concerns about the sensitivity and specificity of the protocol (Dean et al., 2007) and too many responses can lead to an excess of administration and scoring time, introduce the potential problem of overpathologizing respondents (Meyer et al., 2001) and lead to lower stability in protocols (Sultan & Meyer, 2009). Using R-optimized procedures, examiners initially ask the respondent for two or maybe three responses for each card. If the respondent then goes on to provide only one response on a given card the examiner prompts the respondent for an additional response. Alternatively, the card is removed if the respondent spontaneously provides four responses to a given card (Meyer et al., 2011).

With regard to variable selection, most of the decision making was based on the Mihura et al. (2012) meta-analysis of the CS variables. By and large, only those variables with acceptable to good validity were retained in R-PAS.¹ Mihura et al. (2012) report effect sizes for both introspective and external validity criteria. Specifically, the introspective criteria were all self-report questionnaires and the external criteria consisted of “DSM diagnoses, observer or chart ratings, other performance-based personality or cognitive measures, offender status, and more objectively-assessed measures like amount of education” (Mihura et al., 2011, p. 17). One of the overarching findings of the meta-analysis was that Rorschach data overwhelmingly correlate more strongly with externally

¹ Non-CS variables that were included in R-PAS were those that have demonstrated good validity in the Rorschach literature, including the Rorschach Oral Dependency Scale (ROD; Bornstein & Masling, 2005; Bornstein, 1999), Mutuality of Autonomy Scale (MOA; Urist, 1977), and the Ego Impairment Index (EII; Perry & Viglione, 1991; Viglione, Perry, & Meyer, 2003; 2007).
assessed criteria than they do with self-reported traits; this was always true of the Perception and Thinking variables (as classified in the present research).

The derivation of R-PAS FQ tables was conceptually based largely on CS FQ derivation procedures. As such, this section begins with a broad description of FQ coding practices. Second, deviations in the R-PAS FQ derivation procedures from those of the CS are outlined. FQ is an index of the extent to which a response object conforms to the contours of the inkblot location to which it is given as well as the frequency with which that particular response is given at a particular card location. For example, if a respondent indicates that a particular inkblot looks like a bear, the FQ of that response would be the extent to which the blot actually does look like a bear. Responses are classified as “ordinary” (FQo), “unusual” (FQu), or “minus” (FQ-) (Meyer et al., 2011). To be classified as FQo the response must be given frequently and have contours that exist in the blot. A response is considered to be FQu or FQ- if it is infrequently given. The key distinction between FQu and FQ- responses is that FQu responses are quickly and readily perceptible when intentionally looked for and FQ- responses are not (Exner, 2003).

CS and R-PAS FQ derivation processes differ in several ways. Only three judges were used to make CS FQ response fit ratings, ratings were only made for infrequently reported objects (i.e., those occurring less than 2% of the time), and response objects were rated dichotomously regarding whether they were quickly and readily perceptible. In R-PAS, an average of 9.9 judges rated all response objects regardless of frequency and ratings were made on a dimensional scale. Specifically, judges evaluated response objects using Form Accuracy (FA) ratings. FA ratings are made using a five-point scale ranging from “1 = No. I can’t see it at all. Clearly, it’s a distortion” to “5 = Definitely. I think it
looks exactly or almost exactly like that” (Meyer et al., 2011, p. 179). Additionally, all CS judges were from the United States whereas R-PAS judges were from Brazil, China, Finland, Israel, Italy, Japan, Portugal, Romania, Taiwan, Turkey, and the United States (Meyer et al., 2011).

R-PAS FQ codes were determined in light of the average fit rating of judges, the frequency with which the response object was reported by respondents in five contemporary samples from Argentina, Brazil, Italy, Japan, and Spain, and the FQ code assigned in the CS, Beck, and Hertz FQ tables (Meyer et al., 2011). After these modified steps were taken, 39.9% of the FQ ratings differ across the CS and R-PAS. Nonetheless, validity data suggest that the performance of the CS and R-PAS FQ codes are comparable in the detection of psychotic spectrum disorders (Meyer et al., 2011).

The most important consequence of this historical lineage for purposes of the present research is that much of the extensive empirical foundation of the CS Perception and Thinking variables is applicable to the variables included in R-PAS. Each R-PAS Perception and Thinking variable is either identical to or a variant of a CS variable. Furthermore, many of R-PAS administrative, coding, and interpretive procedures are either very similar or identical to those of the CS.

**Rorschach Assessment of Problems in Thinking**

Over the last 50 or so years, several influential independent thought disorder scoring systems and indices using the Rorschach as the primary stimulus have been developed. The primary scoring systems and indices have been the Rapaport Method (Rapaport, Gill, & Schafer, 1968); Delta Index (Watkins & Stauffacher, 1952); Holt’s Primary Process Scoring System (Holt, 1956, 1977; Holt & Havel, 1960); Thought
Disorder Index (TDI; Johnston & Holzman, 1979); CS Special Scores, Schizophrenia Index (SCZI), and Perceptual Thinking Index (PTI; Exner, 2003); and R-PAS Cognitive Codes and Thought and Perception Composite (TP-Comp).2

Rapaport, Gill, & Schafer (1968) developed a Rorschach system by borrowing and modifying many coding categories of pathological verbalizations from Hermann Rorschach’s original system (Rorschach, 1942). They then added several categories of disordered thinking to be coded. Rapaport made invaluable contributions to the Rorschach study of thought disorder and much of the work in this area proceeding Rapaport’s initial work was based largely on his conceptualizations and scoring system. However, the Rapaport system proved to be unwieldy to teach, learn, and score reliably (Kleiger, 1999). This is because the scoring system included an extensive list of variables that are sometimes difficult to discern from one another.

The Delta Index (Watkins & Stauffacher, 1952) was the first Rorschach index to be solely concerned with thought disorder (Kleiger, 1999). This was in contrast to other indices that were designed to measure the broader construct of schizophrenia. The Delta Index was based largely on the Rapaport system and was intended to be an index of the extent to which primary process thinking intruded into respondents’ thinking as measured by the Rorschach. The concepts of primary and secondary process thinking were initially introduced by Freud (1895, 1900). “Freud maintained that primary process thinking involves energies that are uninhibited, free (vs. bound), and not neutralized, while

2 Other Rorschach thought disorder scoring systems that, in the interest of space, are not discussed in the present manuscript include Wynne, Singer, and Toohey’s (1978) communication deviance scoring system; Harrow and Quinlan’s system intended for use in research (Quinlan, Harrow, Tucker, & Carlson, 1972; Harrow and Quinlan, 1977, 1985); the practice-focused systems of Aronow, Reznikoff, and Moreland (1994) and Schuldberg and Boster (1985); Burstein and Loucks (1989) psychoanalytically-based system; Wagner’s TRAUT system (1998); and the Menninger Thought Disturbance Scales (Athey, Colson, & Kleiger, 1992, 1993). Interested readers are referred to Kleiger (1999) for a review of these systems.
secondary process energy is inhibited, bound, and neutralized. Freud equated primary process with both drive-dominated and wishful thought content” (Kleiger, 1999, p. 142). The Delta Index consists of 15 of the thought disorder scores included in the Rapaport method, each of which was assigned a severity weighting of .25, .50, or 1.0. Ratings were made on the basis of the authors’ clinical experience (Kleiger, 1999).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Delta Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Fabulized Responses</td>
<td>.25</td>
</tr>
<tr>
<td>2.</td>
<td>Fabulized Combinations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Spontaneously corrected or recognized</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>b. Not corrected or recognized</td>
<td>.50</td>
</tr>
<tr>
<td>3.</td>
<td>Confabulations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Extreme affect-loading or specificity</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>b. Far-fetched elaboration</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>c. DW</td>
<td>1.00</td>
</tr>
<tr>
<td>4.</td>
<td>Contamination</td>
<td>1.00</td>
</tr>
<tr>
<td>5.</td>
<td>Autistic Logic</td>
<td>1.00</td>
</tr>
<tr>
<td>6.</td>
<td>Peculiar Verbalization</td>
<td>.25</td>
</tr>
<tr>
<td>7.</td>
<td>Queer Verbalization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Usual</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>b. Extreme</td>
<td>1.00</td>
</tr>
<tr>
<td>8.</td>
<td>Vagueness</td>
<td>.25</td>
</tr>
<tr>
<td>9.</td>
<td>Confusion</td>
<td>.50</td>
</tr>
<tr>
<td>10.</td>
<td>Incoherence</td>
<td>1.00</td>
</tr>
<tr>
<td>11.</td>
<td>Over-elaborate Symbolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Moderate</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>b. Extreme</td>
<td>.50</td>
</tr>
<tr>
<td>12.</td>
<td>Relationship Verbalization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Between two percepts (same or different cards)</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>b. Within a series of cards</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Absurd Responses</td>
<td>1.00</td>
</tr>
<tr>
<td>14.</td>
<td>Deterioration Color</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Pure color</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>b. With form (CF)</td>
<td>.50</td>
</tr>
<tr>
<td>15.</td>
<td>Mangled or Distorted Concepts</td>
<td>.25</td>
</tr>
</tbody>
</table>
Holt (Holt, 1956, 1977; Holt & Havel, 1960) developed an intricate Rorschach thought disorder scoring system for use in research. It was designed to investigate adaptive and maladaptive regressive thoughts and the Freudian concept of primary process thinking (Kleiger, 1999, p. 80). Holt’s scoring system was never intended to be used in clinical practice and it does not provide an efficient means of doing so (Kleiger, 1999). Despite broadening our understanding of thought disorder, the system’s difficulty makes it less practical to teach, learn, and score (Kleiger, 1999).

The TDI was developed in 1979 by Johnston and Holzman and was developed as a method to code and score verbal samples yielded by the Rorschach and the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). Johnston and Holzman (1979) consider the TDI to give a comprehensive assessment of thought disorder wherein various types and levels of thinking difficulty are assessed and weighted to form an overall index of disordered thinking. The various types of thinking disturbance are listed in Table 2, accompanied by their relative weights. Finally, the formulae for computing the TDI are also given. $TD_w = \text{the sum of the TDI scores on the WAIS}$, $TD_R = \text{the sum of the TDI scores on the Rorschach}$, divided by the number of Rorschach responses, multiplied by 100, $A = \text{the number of responses scored at level .25}$, $B = \text{the number of responses scored at level .50}$, $C = \text{the number of responses scored at level .75}$, $D = \text{number of responses score at level 1.00}$. 
### Table 2

**Thought Disorder Index Criteria & Formulae**

<table>
<thead>
<tr>
<th>2.5 Level</th>
<th>.5 Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Inappropriate distance</td>
<td>10. Queer responses</td>
</tr>
<tr>
<td>a. Loss or increase of distance</td>
<td>11. Confusion</td>
</tr>
<tr>
<td>b. Tendency to looseness</td>
<td>12. Looseness</td>
</tr>
<tr>
<td>c. Concreteness</td>
<td>a. Distant association</td>
</tr>
<tr>
<td>d. Overspecificity</td>
<td>b. Loose association</td>
</tr>
<tr>
<td>e. Syncretistic response</td>
<td>13. Fabulized combinations, impossible to bizarre</td>
</tr>
<tr>
<td>2. Vagueness</td>
<td></td>
</tr>
<tr>
<td>3. Peculiar verbalizations and responses</td>
<td>14. Fluidity</td>
</tr>
<tr>
<td>b. Stilted, inappropriate expression</td>
<td>16. Confabulations</td>
</tr>
<tr>
<td>c. Idiosyncratic word usage</td>
<td>a. Details in one area generalized to larger area</td>
</tr>
<tr>
<td>d. Peculiar expression</td>
<td>b. Extreme elaboration</td>
</tr>
<tr>
<td>e. Peculiar response</td>
<td>c. Tendency to confabulation</td>
</tr>
<tr>
<td>4. Word-finding difficulty</td>
<td>17. Autistic logic</td>
</tr>
<tr>
<td>5. Clangs</td>
<td>a. Tendency to autistic logic</td>
</tr>
<tr>
<td>6. Perseveration</td>
<td></td>
</tr>
<tr>
<td>7. Relationship verbalizations</td>
<td>18. Contamination</td>
</tr>
<tr>
<td>8. Incongruous combinations</td>
<td>19. Incoherence</td>
</tr>
<tr>
<td>a. Composite response</td>
<td>20. Neologisms</td>
</tr>
<tr>
<td>b. Arbitrary form-color response</td>
<td></td>
</tr>
<tr>
<td>c. Inappropriate activity response</td>
<td></td>
</tr>
<tr>
<td>d. External-internal response</td>
<td></td>
</tr>
</tbody>
</table>

**Intermediate .25, .5**

9. Idiosyncratic symbolism

\[
TD_W = \sum .25(A) + .50(B) + .75(C) + 1.00(D)
\]

\[
TD_R = \sum .25(A) + .50(B) + .75(C) + 1.00(D)
\]

\[
R = \frac{TD_W}{TD_R} \times 100
\]

As previously mentioned, each of the R-PAS Cognitive Codes are derivatives of CS variables. In the CS, the individual Thinking variables were called “Cognitive Special Scores” (Exner, 2003). However, the Thinking variables are introduced here using their
R-PAS Cognitive Codes assess for speech coherence, relevance, logic, and peculiarity (Meyer et al., 2011). They include Deviant Verbalizations (DVs; levels 1 and 2), Deviant Responses (DRs; levels 1 and 2), Peculiar Logic (PEC), Incongruous Combinations (INC; levels 1 and 2), Fabulized Combinations (FAB; levels 1 and 2), and Contaminations (CONs). Level 1 distinctions designate low severity and level 2 distinctions designate high severity. PEC and CON are also more severe codes.

Descriptions from Meyer et al. (2011) of R-PAS Cognitive Codes follow. A “...DV is coded when the respondent uses a mistaken or inappropriate word or phrase to communicate or to describe a response” (p. 81). DRs “...involve confused language, task distortions, or rambling, circumstantial responses that drift from the task” (p. 81). PEC “...is coded for peculiar, strained, confused, or overly concrete reasoning” (p. 86). INCs “...are perceptually-based Cognitive Codes that involve implausible combinations of two or more blot details... [i]n a single response object” (p. 89). FABs “...are perceptually-based Cognitive Codes that involve implausible or impossible relationships between two or more distinct response objects” (p. 90). CON “...is a very rare score and is restricted to the perception of two mutually exclusive response objects or subcomponents involving the same blot area” (p. 92). The Cognitive Codes are ultimately synthesized to yield the Weighted Sum of the 6 Cognitive Codes (WSumCog), as is shown in the formula below, to form an overall index of problems in thinking. In addition, SevCog is computed by summing PEC, CON, and Level 2 scores contained within a protocol.

---

3 Interested readers are referred to the R-PAS manual for a table of CS counterparts to R-PAS variables (Meyer et al., 2011, p. 494).
\[ WSumCog = DV1 \times (1) + DV2 \times (2) + \\
INC1 \times (2) + INC2 \times (4) + \\
DR1 \times (3) + DR2 \times (6) + \\
FAB1 \times (4) + FAB2 \times (7) + \\
PEC \times (5) + CON \times (7) \]

Mihura et al. (2012) found that r values for the CS Thinking variables using external assessment validity criteria ranged from .36 to .38 and for introspective assessment from .11 to .13.\(^4\) For the purposes of interpreting these effect sizes it is important to consider Cohen’s (1988) benchmarks; small = .10, medium = .30, large = .50.

**Rorschach Assessment of Problems in Perception**

Perceptual distortion has traditionally been assessed primarily via form quality (FQ) on the Rorschach. At the protocol-level, most of the Perception variables are interpreted as percentages. FQu% is the percentage of FQu responses contained within a protocol. WD-% is the percentage of responses in a protocol that are FQ- and given to the common whole (W) and detail (D) location areas of the card. FQo% is the percentage of FQo responses contained within a protocol. Popular (P) responses are also a component of the assessment of problems in perception and are interpreted in their raw form. Responses are coded P if they are reported in 1/3 or more of protocols (Exner, 2003). Raw popular scores are the sum of Popular responses given within a protocol.

Mihura et al. (2012) found r values for the key CS Perception variables (i.e, FQ-% or XA%, X+%, and WDA%) using external assessment validity criteria ranging from .37 to .44 and from .00 to .08 for introspective assessment. More specifically, poor FQ (i.e., FQ-% or XA%) correlated with externally assessed criteria at .41 on average; good FQ

\(^4\) Mihura et al.’s (2011) r values are current as of December 2011, though the final numbers may change slightly as a result of revisions requested as part of the review process.
(i.e., X+\% and WDA\%) correlated with externally assessed criteria at .42 on average. Protocol-level Populars correlated with externally assessed criteria at .27 on average. Protocol-level M- correlated with externally assessed criteria at .19 on average. FQu\% correlated with externally assessed criteria at .32 on average; however, the contributing data for FQu\% came from only one study.

**Combined Rorschach Assessment of Problems in Thinking and Perception**

Collectively, the Perception and Thinking variables assess one’s reality testing ability which, as previously noted, is one of the defining features of psychosis. The most current index designed to assess for both problems in combination is the TP-Comp (Meyer, et al., 2011), whose origins can be traced back to the original Schizophrenia Index (SCZI; Exner, 1984, 1986). The SCZI was revised several times and then ultimately gave way to the Perceptual Thinking Index (PTI; Exner, 2003) because the SCZI yielded an unacceptably high number of false positive identifications of schizophrenia and the other psychoses. Specifically, patients with disorders not residing on the psychotic spectrum were being falsely identified as having schizophrenia or another psychotic disturbance on the basis of SCZI data (Exner, 2003). The PTI is a measure of perceptual and thinking problems that uses set cutoff values for its constituent criterion measures. The TP-Comp criteria are similar to the PTI criteria; however, one notable difference is that the various criteria are fully dimensional and cutoff values are not used. The TP-Comp was developed using a formula producing a continuous distribution of values that allows for more differentiation in the data. As such, it provides a more fine-grained analysis of disruptions in thought and perception.
The Ego Impairment Index (EII; Perry & Viglione, 1991) was developed to assess for “deficits in reality testing, reasoning, and the quality of object relations” (Exner, 2003, p. 511). The EII was initially used to study thinking problems and internal resources in a sample of individuals with depression; however, its use was later expanded to assessing psychosis (Perry, Viglione, & Braff, 1992) because many of its constituent variables assess problems in perception and thinking (Exner, 2003). Despite the fact that the EII has demonstrated validity in the assessment of psychosis (Perry et al., 1992), it is not an optimal candidate for short-form development because it is composed of numerous relatively diverse variables, some of which assess the quality of object relations. Impaired object relations are associated with psychosis; however, impaired object relations are not the essence of problems in thinking and perception.
Table 3
*Formulae for Rorschach Indices of Problems in Thinking & Perception*

**SCZI**
- X+% < .61 and S-% < .41 OR X+% < .50
- FQ-% > .29
- FQ- ≥ FQu or FQ- > FQo + FQ+
- Sum Level 2 Special Scores > 1 or FAB2 > 0
- Sum6 > 6 or WSum6 > 17
- M- > 1 or FQ-% > .40

**PTI**
- XA% < .70 and WDA% < .75
- FQ-% > .29
- LVL2 > 2 and FAB2 > 0
- R < 17 and WSum6 > 12 or R > 16 and WSum6 > 17
- M- > 1 or FQ-% > .40

**TP-Comp**
- $+ 2.159 \times (1.168) \times (\text{WD-\%}/100)$
- $+ 5.618 \times (1.194) \times (\text{FQ-\%}/100)$
- $+ 0.598 \times (1.000) \times (-1/(\text{FAB2}+1)^2)$
- $+ 0.191 \times (1.000) \times (\sqrt{\text{WSumCog}})$
- $+ 0.329 \times (1.068) \times (\sqrt{\text{M}-})$
- $- 0.017 \times (1.000) \times (\text{R})$
- $- 0.458 + (0.560)$

The final value is rounded to the tenths decimal place.

**EII-3**
- $+ 0.138 \times (1.318) \times (\text{FQ}-)$
- $+ 0.302 \times (1.000) \times (\sqrt{\text{WSumCog}})$
- $+ 0.265 \times (1.000) \times (\sqrt{\text{CritCont}})$
- $+ 0.321 \times (1.068) \times (\sqrt{\text{M}-})$
- $+ 0.287 \times (1.002) \times (\sqrt{\text{PHR}})$
- $- 0.101 \times (0.998) \times (\text{GHR})$
- $- 0.052 \times (1.000) \times (\text{R})$
- $- 0.955 + (0.268)$

The final value is rounded to the tenths decimal place.
With regard to the Thinking and Perception variables, Mihura et al. (2012) concluded that “In general, the nature of the CS variables with the most validity support targeted cognitive, perceptual, and representational processes” (p. 18). It should be noted that the Rorschach does not directly measure psychosis or, some would argue (e.g., American Psychiatric Association, 2000; Andreasen, 1982; Chaika, 1990), thought disorder. It is indisputable that the Rorschach does not directly assess psychosis as it is impossible to directly measure delusions and hallucinations. Rather, the Rorschach assesses for the underlying distorted perceptual processes that are characteristic of individuals with psychoses, even during periods when they are not overtly psychotic.

With regard to thought disorder, some argue that speech is not an appropriate approximation of thought organization (Andreasen, 1982; Chaika, 1990). While there are definitely limitations inherent in using speech as a proxy for thought, it is also true that speech is one of the best approximations of thought content that researchers have at their disposal (e.g., Harrow & Quinlan, 1985; Lanin-Kettering & Harrow, 1985; Holzman, Shenton, & Solovay, 1986). Consequently, one could argue that the positive symptoms of psychosis (i.e., hallucinations, delusions, and thought disorder) can be indirectly assessed in clinical settings via speech samples and behavior and that the Rorschach provides an excellent means for doing so.

**Why the Rorschach?**

The reasons for using the Rorschach to assess for problems in thinking and perception are numerous. With regard to contemporary clinical diagnostic procedures, it is notable that assessment data yielded by the Rorschach are very well aligned with the proposed dimensional classification of the psychoses in the DSM-5. This issue is more
fully discussed in the section that follows. The present section highlights the core reasons that the Rorschach is a highly useful instrument in the assessment of the psychoses, regardless of the time period and associated edition of the DSM under consideration.

Individuals with schizophrenia are known to have deficits in visual perception (e.g., Amador et al., 1995; Doniger, Silipo, Rabinowiez, Snodgrass, & Javitt, 2001; Doniger, Foxe, Murray, Higgins, & Javitt, 2002; Gabrovska, Laws, Sinclair, & McKenna, 2003; Heckers et al., 2000; Silverstein et al., 2006 a, b) and thinking (e.g., American Psychiatric Association, 2000; Harrow & Quinlan, 1985; Lanin-Kettering & Harrow, 1985; Holzman et al., 1986). In many respects, assessing problems in thinking and perception via the Rorschach is analogous to assessing intelligence via a formal intelligence test (e.g., the WAIS). The most important parallel between the two practices is that clinicians are assessing a latent psychological construct via behavioral tasks. Thus, it is not surprising that individuals with psychotic diagnoses display the behavioral manifestations of the associated perception and thinking deficits on the Rorschach. In line with this conceptualization, Perry, Potterat, Auslander, Kaplan, and Jeste (1996) described the Rorschach as a neuropsychological test adept at measuring one’s ability to use complex integrative perceptual processes.

It is unlikely that any researcher or clinician would argue that intelligence is better assessed via self-report than via a performance based assessment (i.e., formal intelligence test). The correlation between self-reported intelligence and measured intelligence tends not to exceed .30 (Paulhus, Lysy, & Yik, 1998). Furthermore, very few individuals report being “below average” in intelligence (McCrae, 1990), which is logically impossible given that intelligence is normally distributed, such that half the population has an IQ
below average. This suggests that individuals lack insight into the limitations of their intellectual capacities and/or fear they are of limited intellect and are ashamed to say so. In instances where insight is lacking, one can easily see why it would be difficult for individuals to accurately self-report on their IQ. Individuals with schizophrenia also generally lack insight regarding the fact that they have a psychotic diagnosis (American Psychiatric Association, 2000), suggesting that those individuals’ ability to provide an accurate self-report of their symptomatology might be substantially compromised. This suggests that a behavioral measure of psychosis and thought disorder, such as provided by the Rorschach, would be essential to fully capture the phenomenology and severity of these symptoms. Indeed, recent meta-analytic evidence overwhelmingly suggests that Rorschach data, particularly the Perception and Thinking variables, are highly correlated with externally assessed behavioral data but only minimally with self-report data (Mihura et al., 2011).

The Rorschach probes weaknesses in thinking more than structured interviews (Viglione, Perry, & Meyer, 2003) because answering closed-ended questions does not provide a sample of speech or behavior, not to mention that items may not be understood correctly by the respondent (Kleiger, 1999). The speech sample produced by respondents on the Rorschach (i.e., the observable behavior) is a very good approximation of the true variable of interest (i.e., thought organization). Due to the Rorschach’s unique ability to tap into the behavioral manifestations of psychosis and thought disorder, it has been shown in various studies (Cadenhead, Perry, & Braff, 1996; Dao, Prevatt, & Horne, 2008; Dawes, 1999; Kleiger, 1999; Meyer, 1996, 1997; Perry & Braff, 1994; Perry,
Geyer, & Braff, 1999; Zillmer & Perry, 1996) to add incrementally to other measures of psychosis and thought disorder.

Some have argued that, although the Rorschach does provide a valid assessment of psychosis, the presence of psychosis is so obvious in clinical settings that the Rorschach is not likely to add incrementally to a clinical interview (Wood et al., 2000). This suggests that the relative presence or absence of psychosis and thought disorder is an easily observable dichotomy. Generally speaking, concerns about the reliability and validity of clinical diagnoses of prodromal psychosis tend to counter this view (e.g., Carpenter, 2009; Woods et al., 2009). That is, if the presence of psychotic symptomatology were extremely obvious, one would have no reason to doubt that clinicians could reliably and validly diagnose the condition. Furthermore, the clinical interview is similar to self-report in many ways. The clinician is reliant upon the client to accurately self-report information in a verbal format.

Also in contrast to Wood et al.’s (2000) view, empirical evidence suggests that the Rorschach yields incremental information about one’s perceptual and cognitive functioning that is distinct from that provided by other measures of personality (Dao et al., 2008; Dawes, 1999; Kleiger, 1999; Meyer, 1996, 1997) and standard neuropsychological measures (Cadenhead et al., 1996; Perry & Braff, 1994; Perry et al., 1999; Zillmer & Perry, 1996). For example, the SCZI and PTI have been empirically demonstrated to add incrementally to the information provided by the MMPI-2 regarding psychosis (Dao et al., 2008; Meyer, 2000; Ritsher, 2004). This phenomenon has also been demonstrated via a case study (Rubin & Arceneaux, 2001). Meyer (1997) correlated MMPI-2 scales 8, BIZ, and PSY-5-Psy with the SCZI and found that, when response
style is ignored, correlations were notably low, averaging .11. In another study, Meyer, Riethmiller, Brooks, Benoit, and Handler (2000) correlated MMPI-2 scales 8, BIZ, and PSY-5-Psy with the SCZI and found that, when response style is ignored, correlations were again very low, averaging .06. They conclude that “the Rorschach and MMPI-2 are distinct assessment methods. Globally, the results of one test are independent of the results of the other” (p. 212). Given that both tests have good reliability and convergent validity (Parker, Hanson, & Hunsley, 1988), it appears that the tests add valid incremental information to one another.

**The DSM-5 Dimensional Assessment of Psychosis**

The DSM-5 is scheduled to be published in 2013 (American Psychiatric Association, 2011c). Significant structural changes to clinical diagnostic procedures are currently being planned for the forthcoming manual as it is becoming increasingly clear that individuals rarely fit neatly into any one diagnostic category (e.g., Allardyce et al., 2007; Dutta, Greene, Addington, McKenzie, Phillips, & Murray, 2007; van Os & Tamminga, 2007). Specifically, plans are being made to transition the DSM from a categorical scheme to a dimensional scheme wherein individuals can possess varying degrees of any given type of psychopathology. Clinicians will arrive at dimensional diagnoses by identifying the type of psychopathology present and then assigning severity weights to the diagnoses (American Psychiatric Association, 2011b). The current categorical model for the diagnosis of the psychoses is largely derived from Kraepelin’s (1899) work. However, there has been a long-standing debate about whether schizophrenia is indeed a categorical entity (Crow, 1986) and many argue that the ecological validity of the DSM-IV is restricted as a function of its arbitrary cutoff criteria.
(e.g., van Os & Tamminga, 2007). In 1933, Kasanin introduced the label “schizoaffective psychosis” to account for clinical presentations consisting of mood disturbance and psychosis. This intermediary diagnostic classification then gained empirical support from studies showing that the outcome of schizoaffective psychosis is indeed intermediate between schizophrenia and the affective disorders (Brockington, Kendell, & Wainwright, 1980; Clark & Mallett, 1963; Croughan, Welner, & Robins, 1974; Hunt & Appel, 1936) and other evidence demonstrating that manic-depressive illness and schizophrenia are not distributed in a bimodal fashion with regard to symptoms and outcome (Kendell & Brockington, 1980; Kendell & Gourlay, 1970). These observations suggest that categorical classifications are not feasible and speaks to the potential utility of dimensional diagnostic criteria and assessment instruments to assess for those criteria.

Empirical evidence for the validity of a dimensional model of psychosis comes from research in a diverse range of disciplines, such as genetics, biology, psychopharmacology, and psychology. For example, some researchers have tried to account for the heterogeneity of individuals who carry diagnoses of schizophrenia by outlining subtypes of the disorder (e.g., positive, negative, and mixed schizophrenia; deficit and nondeficit schizophrenia). However, these researchers have still found a problem with intermediate forms of the disorder (Dutta, Greene, Addington, McKenzie, Phillips, & Murray, 2007). Additionally, antipsychotic medications have proven to be efficacious in treating both bipolar disorder and schizophrenia. This suggests that faulty dopamine regulation may be the underlying pathology in both conditions (Dutta et al., 2007). Furthermore, the concurrence of schizophrenia and bipolar disorder within families is statistically significant (Dutta et al., 2007). It is also of interest that researchers
have assessed for positive schizophrenic symptomatology in the general population and found that 4% to 17.5% of the general population report delusions and hallucinations (Allardyce et al., 2007).

Other evidence that psychosis resides on a continuum of severity comes from the emerging literature demonstrating that a prodrome stage of psychosis exists and can be reliably assessed (Woods et al., 2009). Researchers and clinicians have long conceptualized schizophrenia as an episodic syndrome that fluctuates from periods of frank psychosis to periods of remission wherein residual symptoms still cause mild to moderate impairment. As part of this conceptualization, researchers and clinicians have also observed that individuals who ultimately develop full-blown psychosis often begin to gradually decompensate up to 5 years before their first psychotic break (Combs & Mueser, 2007). This period of decompensation is termed a “prodrome.” A prodrome of any given clinical syndrome is “an early or premonitory manifestation of impending disease, before specific symptoms begin” (Gennaro & Gould, 1979; as cited in Woods et al., 2009, p. 894). Woods et al. (2009) characterized the prodrome of psychosis in particular as having “evolving attenuated positive symptoms, negative symptoms, and functional impairment” (p. 894). Though prodromal schizophrenia has been recognized as far back as Bleuler (1911), the condition has only recently begun to accumulate solid empirical evidence in support of its existence.

Woods et al. (2009) empirically demonstrated that prodromal psychosis does exist and can be reliably assessed in research contexts. Each respondent included in their study was administered the Structured Interview for Prodromal Syndromes (SIPS; McGlashan, Miller, Woods, Hoffman, & Davidson, 2001; Miller et al., 2002; Rosen, Woods, Miller,
McGlashan, 2002). Their sample consisted of individuals with schizotypal personality disorder, help-seeking comparisons, familial high risk individuals, and a healthy comparison group consisting of individuals with no familial history of psychosis. The health seeking comparison group consisted of individuals who were referred for an evaluation secondary to reporting symptoms during a phone screening that appeared to possibly meet prodromal criteria. Upon a formal diagnostic interview, their symptoms did not meet prodromal criteria. The familial high risk group consisted of individuals determined to be at high risk for developing full-blown psychosis on the basis of having a first-degree relative diagnosed with a psychotic disorder.

Almost all subjects underwent a baseline interview wherein functional status, positive, negative, disorganized, and general symptoms were assessed. Follow-up data were then collected every 6 months for a duration of 30 months. Respondents classified as prodromal in their study met SIPS criteria for prodromal syndromes, with most prodromal individuals being classified as such due to the presence of attenuated positive symptoms (96%). The central dependent variable was time elapsed between baseline assessment and conversion to psychosis. “Conversion to psychosis was defined…as the presence of positive symptoms of sufficient intensity that are either seriously disorganizing or dangerous or that have been present for a month, at least half the days, at least an hour per day” (Woods et al., 2009, p. 896).

With regard to symptomatology, Woods et al. (2009) found that prodromal individuals differed meaningfully from the healthy comparison group on all measures of prodromal symptoms. Prodromal individuals’ scores occupied an intermediary position between the healthy comparison group and the schizotypal group. With regard to
functional status, the prodromal group showed higher impairment than the healthy comparison group on most functional measures. 40% of subjects identified as prodromal at baseline converted to full-blown psychosis within 2.5 years, while 0% of healthy controls, 4% of help-seeking comparisons, 0% of familial high risk, and 36% of schizotypal personality disorder subjects converted to psychosis (Woods et al., 2009).

The DSM-5 Psychosis Workgroup has proposed the inclusion of an “attenuated psychosis syndrome” to account for and diagnose prodromal psychosis among individuals at high risk for developing full-blown psychosis. Given that patients can spend up to five years in the prodromal phase (Combs & Mueser, 2007) and preventative and early intervention efforts lead to better long-term outcomes (Penn, Waldheter, Perkins, Mueser, & Lieberman, 2005), the identification of patients with prodromal psychosis and thought disorder is needed. A description of the proposed diagnostic criteria for attenuated psychosis syndrome follows.

Individuals must have “at least one of the following in attenuated form with intact reality testing, but of sufficient severity and/or frequency that it is not discounted or ignored; (i) delusions, (ii) hallucinations, (iii) disorganized speech.” The aforementioned symptoms “must be present in the past month and occur at an average frequency of at least once per week in the past month” and “have begun in or significantly worsened in the past year” and are “sufficiently distressing and disabling to the patient and/or parent/guardian to lead them to seek help” and “clinical criteria for any DSM-5 psychotic disorder have never been met” (American Psychiatric Association, 2011a).

The reality is that clinicians wishing to use the DSM-5 for diagnostic purposes will need to begin assessing symptomatology dimensionally. Furthermore, clinicians are
likely to need assessment tools to help them arrive at dimensional diagnoses. As Carpenter (2009) notes, clinician reliability and validity in diagnosing prodromal psychosis is a concern, as validity evidence for prodromal psychosis is based on formal diagnostic assessment interviews and supplemental information often not available in clinical practice. This makes it likely that clinicians will need assessment instruments that yield relevant data to help them diagnose prodromal psychosis in a valid, reliable, and efficient manner.

The Rorschach is uniquely positioned among the various available methods of assessing for thought disorder and psychosis as it is the only extensively validated behavioral assessment measure that yields dimensional data. Clinicians making use of the DSM-5 are likely to find dimensional assessment data as a helpful supplement to their clinical judgment when they are making diagnostic severity ratings. Although referencing prodromal psychosis in particular, Carpenter (2009) expressed the concern that the reliability of clinician diagnoses would be lower than those achieved in research; this is also likely to be a concern with the broad category “Schizophrenia and Other Psychotic Disorders.” Dimensional Rorschach data are expected to promote interclinician reliability in diagnosing the psychoses, much like standardized testing of intelligence facilitates the ability of teachers to differentiate children with differential aptitude for learning.

A notable body of research documents that the Rorschach is capable of detecting subtle differences in psychotic symptomatology. For example, research has reliably demonstrated that individuals with schizophrenia give a higher frequency of perceptually distorted and thought-disturbed responses on the Rorschach (Minassian, Granholm, Verney, & Perry, 2005; Perry et al., 1992). Individuals with schizotypal personality
disorder also display similarly distorted responses but to a less severe degree (Exner, 1993). This suggests that the Rorschach is able to adequately capture the fact that schizotypal personality disorder shares a number of features with schizophrenia but in a less severe form. With regard to the EII-II in particular, Perry, Minassian, Cadenhead, Sprock, and Braff (2003) found that problems in perception and thinking increase in a linear fashion as the level of psychotic psychopathology increases.

Finally, evidence suggests that the Rorschach may prove to be a valuable tool in the assessment and diagnosis of the proposed attenuated psychosis syndrome. For example, Ilonen, Heinimaa, Korkeila, Svirskis, and Salokangas (2010) found that the PTI is able to distinguish individuals with psychosis and individuals at a clinically high risk for developing psychosis from other psychiatric patients. However, their results indicate that the individuals at a high risk for developing psychosis and individuals with psychosis cannot be reliably distinguished from one another using the PTI.

Kimhy et al. (2007) administered the Rorschach to individuals at a high-risk for developing psychosis, individuals with recent onset schizophrenia, and individuals with chronic schizophrenia. They found that high-risk individuals show perceptual deficits, as measured by FQ, comparable to individuals already carrying psychotic diagnoses. On the other hand, they found that problems in thinking, as measured by WSum6, increased linearly as the level of psychotic psychopathology increased. This suggests that perceptual abilities decline before thinking abilities in the onset and progression of psychosis. Furthermore, the decline in thinking ability appears to occur as a function of the disorder’s progression. Therefore, FQ may serve as a good screening measure for high-risk status while the Cognitive Codes may serve as a method of determining
conversion from the prodromal phase to full-blown psychosis and monitoring the progression of the disorder.

In general, the Rorschach is able to detect attenuated symptoms of psychosis in individuals at a high-risk of later developing full-blown psychosis (e.g., Ilonen et al., 2010). Some evidence also suggests that prodromal individuals’ pathology, as measured by the Rorschach Thinking variables, occupies an intermediary position between that of nonpatients and psychotic patients (e.g., Kimhy et al., 2007). Though the aforementioned studies were taken from the CS literature, non-CS Rorschach literature has documented similar findings (e.g., Arboleda & Holzman, 1985; Hain, Maier, Hoechst-Janneck, & Franke, 1995; Metsänen, et al., 2004) and evidence suggests that the Rorschach is able to detect communication deviance in parents of high-risk children (e.g., Doane, Jones, Fisher, Ritzler, Singer, & Wynne, 1982; Roisko, Wahlberg, Hakko, Wynne, & Tienari, 2011; Wahlberg, K-E et al., 2000). Furthermore, Woods et al. (2009) found that most individuals (i.e., 96%) who meet SIPS criteria for prodromal psychosis do so because of the presence of positive symptoms; the Rorschach Perception and Thinking variables primarily measure the underlying pathology associated with positive psychotic symptoms. Although in need of further replication and extension to R-PAS, these observations suggest that the Rorschach may be a viable means of assessing for prodromal psychosis.

A Brief Review of Select Short-Form Development Efforts

have since created short-form versions of nearly every type of construct assessed in clinical practice (Smith et al., 2000). Consequently, a large body of literature exists documenting efforts to develop abbreviated version of extant instruments (Stanton, Sinar, Balzer, & Smith, 2002). This section begins with a brief review of four of those research efforts and end with a summary of short-form development best practices articulated by Smith et al. (2000).

The Big Five (John & Srivastava, 1999) are five empirically derived, so-called “normal,” personality domains (i.e., openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) that are widely researched and accepted as valid (Gosling, Rentfrow, and Swann, 2003). Perhaps the most well known measure of the Big Five personality traits is Costa and McCrae’s (1992) 240-item NEO Personality Inventory, Revised (NEO-PI-R). Although it is very popular, its high quantity of items makes its use prohibitive when clinicians’ and researchers’ time is limited. As such, many have resorted to using the 44-item Big-Five Inventory (BFI; John, Donahue, & Kentle, 1991). Gosling et al. (2003) developed two measures of the Big Five personality domains that are even shorter. Specifically, they developed and evaluated a 5- (FIPI; one item per Big Five domain) and a 10-item (TIPI; two items per Big Five domain) measure of the Big Five using the BFI to establish the convergent validity of their instruments. They also evaluated the test-retest reliability of their short-forms and made efforts to establish the associated nomological networks (Chronbach & Meehl, 1955) of each.

As Gosling et al. (2003) note, one common approach to item-selection in the development of short-forms is to select items with good psychometric properties. Most commonly when the psychometric approach is taken researchers select items on the basis
of their item-total correlation with the overarching factors in a measure. However, Gosling et al. (2003) opted to prioritize content validity when constructing their short-forms. Specifically, the authors collectively came to a consensus about appropriate descriptors for each of the five factors and then constructed each “item” on the FIPI by providing a short two word description or phrase accompanied by a brief elaboration to explain what was meant by the short description; item construction was similar for the TIPI though items were shorter in length. They took great care to make sure that each of the Big Five personality dimensions were adequately represented on the FIPI and TIPI by selecting non-extreme items, avoiding items that simply negate some trait, and minimizing overlap between items.

Gosling et al. (2003) found high convergent and discriminant validity between self-reported FIPI and BFI scores. They also found that test-retest correlations were high for both the BFI and the FIPI. Regarding external correlates, they found that the FIPI demonstrated correlational patterns very similar to those of the BFI on other measures of personality and social functioning. Gosling et al. (2003) found that the TIPI converges highly with the BFI. The TIPI outperformed the FIPI with regard to its correlations to external measures of personality and social functioning. They also found a strong six week test-retest correlation for the TIPI. Finally, the TIPI demonstrated patterns of correlations with external criteria almost identical to those of the BFI.

Gosling et al. (2003) ultimately conclude that both the FIPI and TIPI demonstrate acceptable convergent and discriminant validity, test-retest reliability, and patterns of external correlates. As such, they believe that the FIPI and TIPI can be used in lieu of a full-form measure of the Big Five personality domains when time constraints demand for
shorter administrations. However, they indicated that the TIPI is probably more desirable than the FIPI as it has slightly better psychometric properties, can be used in more diverse research designs as a function of two-item factors, and takes the same amount of time to complete as the FIPI (approximately 1 minute).

Donnellan, Oswald, Baird, and Lucas (2006) develop a 20-item (i.e., 4 items per Big Five personality dimension) short-form of the 50-item International Personality Item Pool (IPIP-FFM; Goldberg, 1999). They dubbed their short-form the Mini-IPIP scales. The Mini-IPIP scales were developed using 5 independent samples. Donnellan et al. (2006) conducted an exploratory factor analysis (EFA) on the 50-item IPIP-FFM. They forced a five factor solution and found that all items loaded onto their expected factors. They then selected two negatively keyed and two positively keyed items for all but one of the Big Five personality domains, taking care to select items with the lowest cross-loadings on factors to which they do not belong. They then conducted a second EFA on the 20-item Mini-IPIP scales using an independent sample and obtained a clear five factor solution. Donnellan et al. (2006) report that the Mini-IPIP scales have acceptable internal consistency reliability. They also found high convergent validity with the IPIP-FFM scales and acceptable discriminant validity among scales.

Donnellan et al. (2006) evaluated the nomological network of the Mini-IPIP by administering it in conjunction with the BFI, IPIP-FFM, Rosenberg Self-Esteem Scale (Rosenberg, 1965), the Behavioral Inhibition Scale (Carver & White, 1994), and the Behavioral Approach Scale (Carver & White, 1994). They also conducted a confirmatory factor analysis (CFA) of the MINI-IPIP and found that it has a factor structure comparable to most other Big Five measures. With regard to criterion-related validity,
they found that patterns of correlations were very similar across each Big Five measure they used.

Donnellan et al. (2006) also collected 3 week test-retest data and found that the test-retest correlations for the Mini-IPIP were both high. The authors then collected 6 to 9 month test-retest data, IPIP-FFM informant reports, and additional criterion-related validity data. They found that test-retest correlations were again high for the Mini-IPIP. They also found that IPIP-FFM and Mini-IPIP shared a similar pattern of positive association with informant ratings of the Big Five personality domains. With regard to criterion-related validity, they found that the correlational pattern demonstrated by the IPIP-FFM and Mini-IPIP were similar across external validity criteria. In conclusion, Donnellan et al. (2006) are enthusiastic about their findings and believe they have developed a short-form measure of the Big Five that warrants use in future research where time and resources are limited.

Müller, Postert, Beyer, Furniss, & Achtergarde (2010) compared 11 extant shortened versions of the Symptom Checklist 90-Revised (SCL-90-R; Derogatis, 1992), ranging from 5 to 53 items in length. The SCL-90-R is a measure of general psychopathology. It is used for screening and assessment purposes, evaluation of treatment outcome, and the assessment of symptom severity. Müller et al. (2010) reported that the measure best captures one overarching factor of general distress, as measured by its Global Severity Index. The Global Severity Index is computed as the mean of all items. They sought to identify an extant short-form of the SCL-90-R possessing optimal reliability and validity in Global Severity Index measurement.
Müller et al. (2010) correlated the short-forms with the full-form and then correlated each short-form with the Beck Depression Inventory (Beck et al., 1988; German Version: Hautzinger et al., 1995) to evaluate their validity. Their evaluation criteria for the short-forms included internal consistency reliability, probability of distinct test scores (a new method of assessing a test’s ability to distinguish subjects by their results on the test), validity indices, and screening accuracy and efficiency. They found positive and significant correlations between each short-form and the full-form Global Severity Index. Additionally, they found moderate to strong correlations between the Beck Depression Inventory and each of the short-forms. The internal consistency of each short-form was satisfactory. More variability in probability of distinct test score values was observed across the short-forms. Not surprisingly, the full-form had the best probability of distinct test score values.

Müller et al. (2010) also evaluated the sensitivity, specificity, and positive and negative predictive values of each short-form. They conclude that their results are indicative of good accuracy in screening for all short-forms except the 5 item short-form. As expected per its length, they found that the 27 item short-form had the most optimal sensitivity, specificity, and positive and negative predictive values.

Carpenter et al. (1993) developed a short-form of the TDI to provide a more efficient way of assessing for thought disorder. They chose groupings of 4 cards on the basis of card color to assemble their various short-forms. This resulted in four four-card sets with overlap among the constituent cards. Part-whole correlations were then examined by administering the full 10 card Rorschach to each subject and subsequently computing the short-form scores from the available full administrations. Responses were
scored according to the criteria listed in Table 2. Their obtained part-whole correlations between each short-form and the full-form ranged from .94 to .97. However, it is notable that each of their samples appeared to contain patients with extreme scores that produced inflated SDs. As such, their obtained correlation coefficients were inflated. Intraclass correlations, which control for the rate of agreement between forms on the basis of chance alone, were .81 to .84. The authors ultimately conclude that the TDI short-forms represent a valid and efficient way to assess for thought disorder (Carpenter et al., 1993).

As this brief review of the short-form development literature demonstrates, there are numerous distinct methodologies for developing short-forms. By and large, only sophisticated short-form development efforts are reviewed here; however, there have been less sophisticated attempts made. In fact, the practice of developing short-forms has been confronted with various criticisms and many of its critics (e.g., Levy, 1968; Smith & McCarthy, 1995; Wechsler, 1967) would argue that the transfer of validity from the full-form to the short-form is typically too readily assumed. Some critics even argue that the practice of short-form development should be entirely abandoned (e.g., Wechsler, 1967).

Despite this negative attention, several researchers (e.g., Burisch, 1984a, 1984b, 1997; Smith et al., 2000) are of the opinion that, if done carefully, short-form assessments can be both valid and valuable tools. The present review of the short-form development literature corroborates this view. Furthermore, there are many practical reasons to develop short-form instruments. For example, researchers and clinicians are able to save time, research participants and clients are less burdened by the assessment process, less burdened respondents are more likely to produce valid profiles (Donnellan et al., 2006), researchers and clinicians can use the short-forms more readily to assess treatment
progress, short-forms can be used in instances where time limitations dictate that either the short-form be used or no assessment be conducted (Gosling et al., 2003), and the current managed health care climate does not allow for the adequate reimbursement of more comprehensive assessments.

In response to the popularity of short-form development and the diversity of methodologies used to develop them, Smith et al. (2000) outlined best-practices that they advocate for use in short-form development. They do note, however, that some assessment methods do not permit the usage of all of their guidelines. For example, some of their guidelines are primarily applicable to self-report instruments and, as such, are difficult to apply to Rorschach data. Though Smith et al.’s (2000) guidelines are not an exhaustive list of all possible valid approaches to short-form development, they are some of the only published guidelines available (Donnellan et al., 2006). Smith et al. (2000) delineate nine specific sins that are commonly committed when developing short-forms and then provide steps that can be taken to avoid commission of the sins. Given that each of their guidelines is methodologically sound and defensible, attempts are made in the present research to adhere to them wherever possible. Their recommendations are also supplemented by implementing methodologically sound approaches that address their concerns in alternative ways. For example, Smith et al. (2000) articulate several guidelines for developing short-forms of multifactoral instruments. However, factor analytic evidence and methodology are not used in the present study as the same ends were accomplished in a conceptual sense using methodologies explained below.

**Sin #1: Develop a short form of an insufficiently validated measure.**
Smith et al. (2000) first point out that it is not a worthwhile investment to develop a short-form of a full-form measure with less than adequate validity evidence. This makes good sense as the end result of such an endeavor would likely be a measure equally lacking in validity. As previously noted, the assessment of problems in thinking and perception is the Rorschach’s primary strength. More specific validity data from Mihura et al.’s (2012) follow. Mihura et al. (2012) found that r values for the Thinking variables using external assessment validity criteria ranged from .36 to .38. With regard to key Perception variables, Mihura et al. (2012) found that poor FQ correlates with externally assessed criteria at .41 on average and good FQ correlates with externally assessed criteria at .42 on average. All of the variables to be included in the TP-BAS are identical to, or are derivatives of, variables that were included in Mihura et al.’s (2012) meta-analysis and rest upon the same empirical foundation.

**Sin #2: Fail to show that your short form preserves the content coverage of each factor in the measure.** Loosely speaking, the TP-BAS might be thought of as a two-factor measure, with the Perception variables comprising one factor and the Thinking variables comprising the other. The content coverage of the Perception and Thinking variables is preserved by conducting the Rorschach equivalent of a thorough “item analysis.” Some cards “pull” for certain response types as a function of card contours. A card is said to pull for certain responses or response types if its stimulus properties overwhelmingly induce the propensity to give a particular response type among most respondents. For example, cards that have relatively complex properties have a higher frequency of perceptually distorted responses (Exner & Erdberg, 2005). As such, the
distribution of response types across the cards are evaluated to allow for balance across TP-BAS card sets. This methodology is explained fully in Chapter Two.

**Sin #3: Fail to show that your short form measures each factor scale reliably.**

Smith et al. (2000) indicate that researchers should first explicitly state what type of reliability (e.g., interrater, test-retest, internal consistency) is the most important to consider for the type of assessment in which the short-form is intended to be used (e.g., self-report, behavioral). Smith et al. (2000) further specify that any reliability coefficient below .70 is unacceptable, despite the fact that a drop in reliability simply as a consequence of developing a short-form is expected. Interrater reliability analyses were conducted in the present research and the results are evaluated in light of this proposed standard. It will also be important to establish the test-retest reliability of the TP-BAS in future research.

**Sin #4: Fail to show that your short form has adequate overlapping variance with the full form, using independent administrations.** Smith et al. (2000) indicate that an important component of demonstrating the validity of a short-form is to show that it has adequate overlapping variance with the full-form, using independent administrations. That is, they recommend administering the full-form and the short-form to the same individuals on independent occasions and then correlating the obtained scores. As they indicate, a short-form of an instrument is much like an alternate form of the full-form measure. When viewed from this framework, it is clear why a high correlation between the short-form and the full-form is desirable. The archival nature of the data used in the present research does not allow me to address this concern in the present research. However, this issue will be addressed in follow-up research.
Sin #5: Fail to show empirically that your short form reproduces the factor structure of a multifactoral instrument; Sin #6: If your short form omits subfactors and preserves only overall factors, then fail to show that the short form preserves the content domains represented by the subfactors. As previously mentioned, the “factor structure” of the full-form is conceptually preserved in each TP-BAS card set. However, formal factor analytic methodology is not used.

Sin #7: Fail to show that each factor in the short form has validity on an independent sample. The TP-BAS was developed using a composite sample consisting of nonpatients, patients without psychosis, and patients with psychosis. Data in each sample were collected independently. The validity of each TP-BAS card set is evaluated using one validity criterion in two of the samples and another validity criterion in the third sample.

Sin #8: Fail to show that classification rates remain high with the short form. Smith et al. (2000) suggest that researchers should demonstrate that diagnostic accuracy is preserved when transitioning from the full-form of an instrument to a short-form. The diagnostic accuracy of each TP-BAS card set is evaluated in the Results section by reporting several diagnostic efficiency statistics using the TP-Comp as an example.

Prior research examining the diagnostic efficiency of the PTI is briefly reviewed to set the stage for a discussion of TP-BAS diagnostic efficiency. The following evidence of the PTI’s diagnostic accuracy is helpful in understanding that of the TP-BAS given that the variables included in the PTI are identical or highly similar to TP-BAS variables. The PTI has a strong ability to differentially identify schizophrenia spectrum disorders. Specifically, overall correct classification rates (OCCs) in one study ranged from .61 (PTI
= 5) to .86 (PTI ≥ 3) in differentiating individuals with schizophrenia-spectrum disorders from those with mood disorders (Dao & Prevatt, 2006). In another study, OCCs ranged from .70 (PTI = 5) to .88 (PTI ≥ 2 and PTI ≥ 3) in differentiating psychotic disorder patients from nonpatients and from .51 (PTI ≥ 1) to .69 (PTI ≥ 3 and PTI ≥ 4) in differentiating psychotic disorder patients from patients carrying diagnoses of Cluster A personality disorders or borderline personality disorder (Hilsenroth, Eudell-Simmons, DeFife, & Charnas, 2007).

**Sin #9: Fail to show that your short form offers meaningful time or resource savings for the loss in validity.** Given the fact that decreasing administration time is the primary reason for developing the short-form, time savings is of primary concern in choosing the quantity of cards to be retained (i.e., a 3-card set would be more desirable than a 5-card set). On the other hand, the principle of aggregation indicates that testing instruments increase in reliability and validity as the number of items contained in them increases (Rushton, Brainerd, & Pressley, 1983). It follows that the most reliable and valid form will be the full 10 card set and that among the TP-BAS card sets the 5-card set will be more reliable and valid than the 4- or 3-card sets. However, the purpose of the present research is to develop alternate forms with fewer cards than the full set, which will necessitate a compromise between validity and utility. The findings of the present research are evaluated in light of these considerations.

The R-PAS manual (Meyer et al., 2011) states that administration of the full-form typically takes slightly less than one hour. It follows that the TP-BAS-3, -4, and -5 card sets should take approximately 18, 24, and 30 minutes to administer, respectively. Although it is certain that TP-BAS administration will be briefer in duration than R-PAS...
administrations, this will be documented in future research by timing administrations during normative data collection.
Chapter Two

Development and Validation of the Thinking and Perception Behavior Assessment System

Statement of the Problem

Thus far, it has been demonstrated that there is a need for a brief, reliable, and valid behavioral measure that yields dimensional data to aid in the assessment of thought disorder and psychosis with the capacity for nonoverlapping test-retest evaluations. It has also been established that the Rorschach has a long history of accomplishing those ends. There are currently a number of Rorschach scoring systems that can be used to assess for thought disorder and psychosis. However, none of the systems provides each of the aforementioned desired components in a single package.

The Rapaport System, Delta Index, and Holt’s Primary Process Scoring System are all relatively difficult to learn and score and are not intended for brief assessments. R-PAS also provides a viable means of assessing for thought disorder and psychosis; however, it requires the administration of all ten cards as well as the coding of variables that are not of critical importance in assessing for thought disorder and psychosis. The nearest approximation of the type of instrument needed is Carpenter et al.’s (1993) short-form of the TDI. However, their short-form is limited for several reasons. First, they established the validity of their short-forms by simply correlating the short-forms with the full-form of the TDI. However, their short-form is limited for several reasons. First, they established the validity of their short-forms by simply correlating the short-forms with the full-form of the TDI. A more optimal design would have validated the short-forms using criterion measures external to the TDI (Smith et al., 2000). Second, the TDI and the TDI short-form only assess for problems in thinking; they do not assess perceptual distortions. Third, the full- and short-form TDIs require the use of a coding rubric which, as previously noted, is difficult and time consuming to teach, learn, and score. Finally, the
methodology used to derive the short-form was limited in that card color was the only basis for card selection. This means that their short-form card sets were not balanced on the basis of the means and standard deviations of the Thinking variables. This is undesirable because the Thinking variables form the basis of interpreting the TDI short-form card sets, not card color.

**Purpose of the Present Research**

The purpose of the present research is to develop a short-form Rorschach assessment measure that targets thought disorder and perceptual distortion in conjunction for the assessment of psychosis, is as straightforward to score as possible, has good interrater reliability, and has well documented validity. This is accomplished by developing a modified version of R-PAS administration and scoring procedures that includes only the Perception and Thinking variables. 10-, 5-, 4-, and 3-card series are constructed that are alternate forms of one another. This will facilitate test-retest data collection which could prove to be very helpful in monitoring the progression of thought disorder and psychosis within patients and in monitoring the effects of psychopharmacological and psychosocial treatments. Each card set is only scored for problems in thinking and perception. The 10-card set is one set of cards that consists of each of the standard Rorschach cards. The 5- and 4-card series are both comprised of two card sets. The 3-card series is comprised of three card sets. Each card set within a given series does not have overlapping cards; however, cards do overlap across card series. The new system is provisionally called the Thought and Perception Behavioral Assessment System (TP-BAS).
The TP-BAS is expected to retain each of the positive attributes previously reviewed with regard to Rorschach data. Specifically, it will be a behavioral measure that does not require respondent insight or introspection, it is expected to yield incremental information over self-report and structured interview methodologies, the data produced will be dimensional, and it is expected to be capable of detecting prodromal psychosis. In addition, it will offer notable time savings. Time savings will be yielded by virtue of card reduction, the reduced number of variables to be scored, and by the availability of the R-PAS online scoring program (www.r-pas.org; Meyer et al., 2011). The scoring program produces standard score values for every raw variable and index included in the system. Furthermore, the program automatically compares the individual to R-PAS reference database and then produces the standard score profile.

**Samples**

Three samples are used in the construction and preliminary validation of the TP-BAS. The Dean adult patient sample (Dean, Viglione, Perry, & Meyer, 2007), hereafter referred to as the Dean sample, consists of 61 patients who were in long-term residential treatment. Thirty of the Rorschach protocols in the Dean sample were administered using standard CS (Exner, 2003) procedures and 31 were administered using an early version of the R-Optimized administration procedures.

The University of Chicago patient sample (Hoelzle & Meyer, 2008), hereafter referred to as the Chicago sample, consists of a large pool of patient data collected from 1990 to 1994 at the University of Chicago. Although all protocols were initially collected using CS administration guidelines, statistical modeling was used to ensure this sample very closely approximates the distribution of responses that is obtained using R-
Optimized administration procedures. For the purposes of the present research, cases were removed if data were not available regarding the severity of the associated diagnosis or if testing was a repeat assessment. Finally, cases were removed if their Rorschach protocol did not meet criteria for R-Optimization modeling. After applying these exclusion criteria, 133 cases remain in the Chicago sample. Cases in this sample were given testing-based diagnoses as well as diagnoses generated for billing purposes that were independent of the testing and assigned before testing began. Billing diagnoses are used for the present research in all instances where diagnostic category is a consideration.

The R-PAS reference sample, hereafter referred to as the R-PAS sample, is based on 1,396 nonpatient protocols obtained from 15 adult samples of internationally collected data. Data were collected in Argentina, Belgium, Brazil, Denmark, Finland, France, Greece, Israel, Italy, Portugal, Romania, Spain, and the United States. Of the available 1,396 protocols, 641 were retained in the final R-PAS normative reference sample. Although all protocols were initially collected using CS administration guidelines, statistical modeling was used to ensure this sample very closely approximates the distribution of responses that is obtained using R-Optimized administration procedures. Of these 641 protocols, 118 protocols are available with verbatim responses in English and these protocols form the foundation for R-PAS form quality norms and are included in this study as the R-PAS sample.

Selecting Preliminary Short-Form Card Sets\footnote{All of the responses included in the present research had their FQ ratings recoded according to R-PAS FQ tables. Subsequently, all indices and variables that make use of FQ were recomputed.}

First, preliminary short-form card sets were chosen by carefully selecting Rorschach cards on the basis of their relative pull for cognitive disorganization and
perceptual distortion. This was accomplished by conducting a one-way ANOVA; the variable of interest was entered as the dependent variable and card number was entered as the independent variable. Given the considerable degree of statistical power in these analyses, Cohen’s $f$ effect size was also computed to determine the practical significance of the findings. Specifically, it is used in the present research to contrast the standard deviation of the means across Rorschach cards to the pooled within card standard deviation.

Based on previous research (Exner & Erdberg, 2005), it was hypothesized that all of the Perception variables would meaningfully differ across cards. Unless the results suggested some alternative considerations, it was determined that it would be most appropriate to use FQ- and FQo as the primary considerations in card selection because they offer the most parsimonious method for detecting the presence or relative absence of perceptual distortion and a lack thereof. FQn was not a consideration because it represents a lack of form. FQu was not considered because it represents the intermediate area between FQ- and FQo as it is not clearly indicative of perceptual distortion or a lack thereof. Furthermore, Mihura et al.’s (2012) recent meta-analysis found limited support for FQu’s validity. WD variables were not considered because they are subsets of the broad FQ variables. Likewise, M- was not considered because it is a subset of FQ-.

Popular was taken into consideration to the extent possible after matching sets on FQ- and FQo. Based on prior research (Carpenter et al., 1993), it was expected that fewer of the Thinking variables would differ across cards and that the magnitude of difference would be small. Specific predictions regarding which Thinking variables would differ
across cards were not made. However, the aim was to take the Thinking variables into consideration for card selection purposes as needed.

Next, the within card mean and standard deviation were computed for FQ-, FQo, and WSumCog as the primary variables, and P and the individual Cognitive Codes as secondary or supplemental variables, and subjected them to an ANOVA. For variables that meaningfully differed across cards, balanced, alternate-form, nonoverlapping (within series) TP-BAS card sets were developed. Given that the full-form begins with an achromatic card, it was also ensured that each TP-BAS card set begins with an achromatic card. After assembling card sets, the mean and standard deviation of each variable by TP-BAS card set was computed for descriptive purposes. As a note, each instance where the name “TP-BAS” is used a reference is being made to all TP-BAS card sets. To allow me to reference TP-BAS card series and card sets in isolation, the series have been named by their number and the subset number is appended to the name. For example, the first card set in the 5-card series is named “TP-BAS-5.1.”

**Preliminary Evaluation of the TP-BAS**

The full-form versions of R-PAS Perception and Thinking variables served as criterion measures in the initial evaluation of the TP-BAS card sets. Each of the Perception and Thinking indices and variables were computed for the full-form at the protocol-level for each of the cases in the Dean, Chicago, and R-PAS samples. Those same variables and indices were computed for each of the TP-BAS card sets and then correlated with the full-form scores. Those correlations are expressed as Pearson’s r coefficients. Given the principle of aggregation (Rushton et al., 1983), a step-wise decrement in part-whole correlations was expected, such that the TP-BAS-5 card sets
would have the strongest part-whole correlations followed by the TP-BAS-4 and -3 card sets, respectively.

A primary index of interest in the present research is the TP-Comp. The TP-Comp makes use of WD-%, FQ-%, FAB2, WSumCog, M-, and R. WD-% is the percentage of responses contained within a protocol that are given to common location areas and are also FQ-. FQ-% is the percentage of FQ- responses contained within a protocol. An individual’s raw FAB2 score is the sum of Fabulized Combination responses that are also of Level 2 severity given within a protocol, regardless of location. The six Cognitive Codes included in WSumCog are Deviant Verbalizations (DV1 and DV2), Incongruous Combinations (INC1 and INC2), Deviant Responses (DR1 and DR2), Fabulized Combinations (FAB1 and FAB2), Peculiar Logic (PEC), and Contaminations (CON). M- is the protocol-level sum of human movement responses that are also coded FQ-. Finally, R is the total number of responses in a protocol.

In addition to using existing R-PAS variables, a simplified alternative to the TP-Comp (i.e., the TPMZ) was examined and was computed as the mean of the z-scores for R-PAS Thinking (i.e., WSumCog and SevCog) and select Perception variables (i.e., FQ-%, WD-%, and FQo%). FQo% was reverse scored before being averaged into the mean thinking and perception computation. FQu% and Popular were omitted from the TPMZ because of their relatively lower validity in detecting psychosis (Mihura et al., 2011).

**Interrater Reliability**

Inter-rater reliability was evaluated for the full 10 card Rorschach set as well as each of the TP-BAS card sets; coefficients are expressed as intraclass correlations (ICCs). Available interrater reliability data for 93 protocols reported in the R-PAS manual
(Meyer et al., 2011) were used for these analyses. The 93 protocols were drawn from research reported in Meyer et al. (2002) and were statistically modeled to approximate protocols obtained using R-Optimization procedures. Respondents came from clinical, forensic, and nonpatient settings. Two independent judges coded the Rorschach protocols and were blind to one another’s coding. Interrater reliability data are provided in the Methods section.

**Examination of Short-Form Validity**

**Criterion Measures.** The Thought Disorder Summary Scale (TDSz; Dean et al., 2007) and the Ego Dysfunction Severity Scale (Dawes, 1999; Meyer & Resnick, 1996) were used as criterion measures to examine the validity of the TP-BAS. Specifically, validity was examined by correlating each of the Perception and Thinking variables to be included in the TP-BAS with the criterion measures. Correlations are expressed as Pearson’s r coefficients. Based on Kleiger’s (1999) summary of available research, it was hypothesized that trends of perception and thinking problems would emerge such that individuals with the highest Ego Dysfunction Severity Scale and TDSz scores would appear the most pathological on the TP-BAS, whereas the adult nonpatient sample would appear the least pathological.

**The Thought Disorder Summary Scale – Dean Sample.** A slightly modified version of the Thought Disorder Summary Scale was used as the validity criterion for the Dean sample. The original Thought Disorder Summary Scale (TDSz; Dean et al., 2007) was generated by transforming raw scores from the Scale for the Assessment of Thought, Language, and Communication (TLC; Andreasen, 1986), Magical Ideation Scale (MIS; Eckblad & Chapman, 1983), and the Scale for the Assessment of Positive Symptoms
(SAPS; Andreasen, 1986) into z scores and then summing the results. The version of the TDSz that is used in the present research modifies the formula by omitting a scale on the TLC that measures poverty of speech. This modification was made for logical reasons and it yields a small increment in validity (Meyer et al., 2011).

The TLC is a semistructured interview wherein examiners rate the level of patients’ thought disorder on a 4- or 5-point Likert scale. The MIS is a 30-item self-report measure of magical thinking. Magical ideation has proven to be predictive of psychotic vulnerability, which is reflected in the MIS’s ability to differentiate individuals prone to psychosis from those who are not. The SAPS consists of four subscales that assess for hallucinations, delusions, bizarre behavior, and thought disorder. When data were collected for the Dean et al., (2007) study, the only portion of the SAPS that was administered was the delusional subscale, which was treated as a supplement to the TLC. As such, the delusional subscale is the only portion of the SAPS that contributes to the present research.

**The Ego Dysfunction Severity Scale – Chicago and R-PAS Samples.** The Ego Dysfunction Severity Scale was developed by Meyer and Resnick (1996) to designate diagnostic severity as a function of the severity of ego dysfunction associated with diagnoses. Ego dysfunction ratings were made on a 5-point Likert scale, with 1 being the least severe level of ego dysfunction and 5 being the most severe level of ego dysfunction. The nonpatient respondents in R-PAS sample were given ego dysfunction severity ratings of 0 and were then added to the Chicago sample. Some nonpatients do certainly have mild ego dysfunction; however, it is expected that instances of ego dysfunction should be relatively rare and mild in this sample. For the Chicago sample,
Ego Dysfunction Severity Scale ratings were made on the basis of billing diagnoses. To understand what is meant by “ego dysfunction,” it is illustrative to consider examples of how various disorders are rated on the Ego Dysfunction Severity Scale. Examples are provided in Table 4 below.

Table 4

<table>
<thead>
<tr>
<th>Score</th>
<th>Diagnosis</th>
</tr>
</thead>
</table>
| 5     | Undifferentiated Schizophrenia  
       Schizoaffective Disorder  
       Bipolar I, Most recent manic, severe (nonpsychotic or psychotic) |
| 4     | Paranoid Schizophrenia  
       Bipolar I, Most recent manic, moderate  
       Schizotypal Personality Disorder |
| 3     | Major Depression, Recurrent, severe-nonpsychotic  
       Pain Disorder associated with psychological factors  
       Sexual Masochism  
       Histrionic Personality Disorder |
| 2     | Major Depression, Single, moderate/severe-nonpsychotic  
       Generalized Anxiety Disorder  
       Fetishism  
       Alcohol Abuse |
| 1     | Specific Phobia  
       Premature Ejaculation  
       Adjustment Disorder with Depressed Mood |


**Post-Hoc Power Analyses**

Given that the present analyses were conducted on archival data sets, post-hoc power analyses were conducted using G-Power software (Faul, Erdfelder, Lang, &
Buchner, 2009) to determine the minimum population effect size needed for detectability at 80% power ($P$), taking into consideration the size of each sample. Each of the analyses were conducted using the achieved sample sizes with two-tailed alpha set at $\alpha=.05$. The required $r$ represents the correlation in the population between the full-form and short-form variables and the criterion measure of interest used in the validity analyses. The results of the post hoc power analyses are presented in Table 5 below.

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Required r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dean</td>
<td>61</td>
<td>.35</td>
</tr>
<tr>
<td>Chicago &amp; R-PAS</td>
<td>251</td>
<td>.18</td>
</tr>
</tbody>
</table>

Selecting the Optimal Series of Short-Form Card Sets

Mean interrater reliability, validity, and part-whole correlations for the full-form R-PAS and TP-BAS Thinking and Perception variables (i.e., TP-Comp, TPMZ, WSumCog, SevCog, FQ-%, WDm%, and FQo%) were plotted in line graph format for each TP-BAS card series (Figures 1 and 2). The line graphs served as a visual aid in determining whether or not there is a substantial decrement to reliability and validity across the various TP-BAS card sets. Given that the Dean sample used a different validity criterion than the Chicago and R-PAS samples, the associated data were analyzed and graphed separately.
With regard to diagnostic efficiency, the sensitivity (SN), specificity (SP), positive predictive Power (PPP), negative predictive power (NPP), and overall correct classification (OCC) rates are reported for each TP-BAS card set using TP-Comp data from the Chicago and R-PAS samples as an example. The TP-Comp was chosen as an example because it makes use of many of the core variables included in the TP-BAS. Given the high similarity among the variables included on the PTI and the TP-Comp, it is reasonable to use previous PTI research to derive a TP-Comp cutoff value for positive psychotic identification in the present research.

Doa and Prevatt (2006) indicate that a PTI cutoff value of 3 is adequate to differentiate patients with psychosis from patients with mood disorders. Similarly, Hilsenroth et al. (2007) ultimately determine that a PTI cutoff value of 3 is adequate to differentiate individuals with psychotic diagnoses from nonpatients and patients with relatively less severe personality disorders. However, Hilsentroth et al. (2007) obtained an OCC value of .88 for cutoffs of both $\geq 2$ and $\geq 3$ with SN values of .82 and .76, SP values of .93 and .98, PPP values of .90 and .96, and NPP values of .86 and .83, respectively. For the purposes of the present research, SN is perhaps the most important diagnostic efficiency statistic to attend to given the disproportionately high number of nonpatients in the merged Chicago and R-PAS database. Given these sample characteristics as well as Hilsenroth et al.’s (2007) finding that PTI SN is more optimal when a cutoff of $\geq 2$ is used, a TP-Comp cutoff of $\geq 2$ was used in the present research. Diagnostic efficiency statistics are best understood via a 2 X 2 table such as the one depicted in Table 6 below.
Table 6  
*Diagnostic Efficiency Statistic Illustration*

<table>
<thead>
<tr>
<th>Test Results</th>
<th>True State in the World</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive / Condition Present</td>
<td>Negative / Condition Absent</td>
</tr>
<tr>
<td>Positive</td>
<td>A (True Positive)</td>
<td>B (False Positive)</td>
</tr>
<tr>
<td>Negative</td>
<td>C (False Negative)</td>
<td>D (True Negative)</td>
</tr>
</tbody>
</table>

*Note. Table 6 is adapted from Meyer (2007).*

The SN of a test is the percentage of individuals who actually do have a condition who are positively identified by the test \[A / (A + C)\]. On the other hand, SP is the percentage of individuals who actually do not have the condition who are correctly identified by the test as not having the condition \[D / (B + D)\]. PPP is the proportion of people labeled by the test as having the condition who actually have the condition \[A / (A + B)\]; NPP is the proportion of people labeled by the test as not having the condition who actually do not have the condition \[D / (C + D)\]; OCCs are the percentage of individuals who are correctly classified by the test. “Correctly classified” individuals are those who actually do have the condition of interest who are positively identified by the test and individuals who actually do not have the condition who are negatively identified by the test \[(A+D)/(A+B+C+D)\] (Meyer, 2007; Streiner, 2003).

**Results**

In line with my hypothesizing, the Perception variables generally differed to a meaningful degree across cards. Although relatively inconsequential for purposes of card selection, FQo was more variable across cards than was FQ-. The ANOVA results do not suggest that Perception variables other than FQ- and FQo should be used for card selection. Also in line with my hypothesizing, differences in Thinking variable means
were not as notable across cards as were the Perception variable means. As is shown in Table 7, INC1 and FAB1 differed meaningfully across cards to a small degree. INC1 is most common on Card I and FAB1 is most common on Card II. Thus, Cards I and II were not concurrently included in any particular TP-BAS card set. In sum, key variables that were found to meaningfully differ across cards (i.e., FQo, FQ-, INC1, and FAB1) are shown in bold in Table 7. All other variables are shown for descriptive purposes.
### Table 7
*Dean, Chicago, and R-PAS Samples – Variable Mean by Card*

<table>
<thead>
<tr>
<th></th>
<th>I n=814</th>
<th>II n=764</th>
<th>III n=735</th>
<th>IV n=661</th>
<th>V n=674</th>
<th>VI n=679</th>
<th>VII n=704</th>
<th>VIII n=711</th>
<th>IX n=855</th>
<th>X</th>
<th>Grand Mean</th>
<th>Cohen’s f</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FQn</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
<td>.03</td>
<td>.04</td>
<td>.03</td>
<td>.02</td>
<td>.10**</td>
</tr>
<tr>
<td>FQ-</td>
<td><strong>.10</strong></td>
<td><strong>.12</strong></td>
<td><strong>.11</strong></td>
<td><strong>.10</strong></td>
<td><strong>.08</strong></td>
<td><strong>.17</strong></td>
<td><strong>.14</strong></td>
<td><strong>.16</strong></td>
<td><strong>.23</strong></td>
<td><strong>.16</strong></td>
<td><strong>.14</strong></td>
<td><strong>.12</strong>**</td>
</tr>
<tr>
<td>FQu</td>
<td>.20</td>
<td>.32</td>
<td>.31</td>
<td>.37</td>
<td>.17</td>
<td>.40</td>
<td>.26</td>
<td>.31</td>
<td>.47</td>
<td>.34</td>
<td>.32</td>
<td><strong>.18</strong>**</td>
</tr>
<tr>
<td>FQo</td>
<td><strong>.69</strong></td>
<td><strong>.54</strong></td>
<td><strong>.57</strong></td>
<td><strong>.52</strong></td>
<td><strong>.74</strong></td>
<td><strong>.43</strong></td>
<td><strong>.59</strong></td>
<td><strong>.50</strong></td>
<td><strong>.26</strong></td>
<td><strong>.48</strong></td>
<td><strong>.53</strong></td>
<td><strong>.26</strong>**</td>
</tr>
<tr>
<td>WDn</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
<td>.03</td>
<td>.04</td>
<td>.03</td>
<td>.02</td>
<td><strong>.10</strong>**</td>
</tr>
<tr>
<td>WDm</td>
<td>.06</td>
<td>.09</td>
<td>.07</td>
<td>.07</td>
<td>.04</td>
<td>.11</td>
<td>.11</td>
<td>.11</td>
<td>.16</td>
<td>.13</td>
<td>.10</td>
<td><strong>.11</strong>**</td>
</tr>
<tr>
<td>WDu</td>
<td>.13</td>
<td>.28</td>
<td>.23</td>
<td>.29</td>
<td>.13</td>
<td>.31</td>
<td>.23</td>
<td>.26</td>
<td>.36</td>
<td>.26</td>
<td>.25</td>
<td><strong>.16</strong>**</td>
</tr>
<tr>
<td>WDo</td>
<td>.65</td>
<td>.52</td>
<td>.54</td>
<td>.51</td>
<td>.73</td>
<td>.41</td>
<td>.58</td>
<td>.47</td>
<td>.21</td>
<td>.42</td>
<td>.50</td>
<td><strong>.28</strong>**</td>
</tr>
<tr>
<td>M-</td>
<td>.01</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.03</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td><strong>.07</strong>**</td>
</tr>
<tr>
<td>Popular</td>
<td>.21</td>
<td>.12</td>
<td>.32</td>
<td>.24</td>
<td>.47</td>
<td>.21</td>
<td>.27</td>
<td>.37</td>
<td>.06</td>
<td>.12</td>
<td>.23</td>
<td><strong>.29</strong>**</td>
</tr>
<tr>
<td><strong>Thinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.03</td>
</tr>
<tr>
<td>PEC</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.03</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td><strong>.04</strong>**</td>
</tr>
<tr>
<td>DV1</td>
<td>.04</td>
<td>.03</td>
<td>.03</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.04</td>
<td>.03</td>
<td>.03</td>
<td><strong>.04</strong>**</td>
</tr>
<tr>
<td>DV2</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td><strong>.03</strong>**</td>
</tr>
<tr>
<td>DR1</td>
<td>.03</td>
<td>.04</td>
<td>.05</td>
<td>.07</td>
<td>.05</td>
<td>.04</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td><strong>.04</strong>**</td>
</tr>
<tr>
<td>DR2</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>.02</td>
<td>.01</td>
<td><strong>.05</strong>*</td>
</tr>
<tr>
<td>INC1</td>
<td><strong>.10</strong></td>
<td><strong>.02</strong></td>
<td><strong>.03</strong></td>
<td><strong>.03</strong></td>
<td><strong>.04</strong></td>
<td><strong>.02</strong></td>
<td><strong>.01</strong></td>
<td><strong>.06</strong></td>
<td><strong>.02</strong></td>
<td><strong>.05</strong></td>
<td><strong>.04</strong></td>
<td><strong>.12</strong>**</td>
</tr>
<tr>
<td>INC2</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td><strong>.03</strong>**</td>
</tr>
<tr>
<td>FAB1</td>
<td><strong>.01</strong></td>
<td><strong>.08</strong></td>
<td><strong>.04</strong></td>
<td><strong>.01</strong></td>
<td><strong>.02</strong></td>
<td><strong>.02</strong></td>
<td><strong>.04</strong></td>
<td><strong>.02</strong></td>
<td><strong>.04</strong></td>
<td><strong>.03</strong></td>
<td><strong>.11</strong>**</td>
<td></td>
</tr>
<tr>
<td>FAB2</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
<td><strong>.06</strong>**</td>
</tr>
<tr>
<td>WSumCog</td>
<td>.64</td>
<td>.79</td>
<td>.88</td>
<td>.70</td>
<td>.65</td>
<td>.64</td>
<td>.67</td>
<td>.80</td>
<td>.77</td>
<td>.82</td>
<td>.74</td>
<td>.04</td>
</tr>
</tbody>
</table>

*Note.* * Significant at the .05 level (two tailed). ** Significant at the .01 level (two tailed).
As is shown in Table 8, it was possible to generate fairly well-balanced card sets with regard to FQ-, FQo, and P. However, some considerations are worth noting. TP-BAS-5.2 has more fully colored cards than TP-BAS-5.1. Unfortunately, no satisfactory solution exists as one of the sets must have one fully colored card and the other must have two. Additionally, Cards II and III are the only cards composed of entirely black and red ink. As such, one TP-BAS-3 card set must go without a black and red card. These slight qualitative imbalances will be revisited as part of the presentation of the validity data.

Table 8
Dean, Chicago, and R-PAS Samples - Final Card Selections

<table>
<thead>
<tr>
<th>Series</th>
<th>Subset</th>
<th>Cards</th>
<th>Variable</th>
<th>FQ-</th>
<th>FQo</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>I, III, IV, VII, IX</td>
<td>3600</td>
<td>.14</td>
<td>.06</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>II, V, VI, VIII, X</td>
<td>3635</td>
<td>.14</td>
<td>.04</td>
<td>.54</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>III, VII, X</td>
<td>2269</td>
<td>.14</td>
<td>.03</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>II, IV, VII, VIII</td>
<td>2808</td>
<td>.13</td>
<td>.03</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>III, V, VI, VIII</td>
<td>2902</td>
<td>.13</td>
<td>.04</td>
<td>.56</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>II, V, IX</td>
<td>2113</td>
<td>.14</td>
<td>.08</td>
<td>.51</td>
</tr>
</tbody>
</table>

Note. Values are reported for all responses given to the cards in each set.

Table 9 provides protocol level means for each variable by card set. As previously noted, an additional variable was computed to control for R in the raw score variables (i.e., P, WSumCog, SevCog, and TP-Comp); as such, these adjusted figures provide a more meaningful basis for comparing the card sets to one another. The adjusted figures are shown in bold print in Table 9. Three of the adjusted variables were simple proportions that divided by R. However, a more complex procedure was used to roughly equate TP-Comp scores across sets. Here each of the raw variables used in the TP-Comp
formula (i.e., FAB2, WSumCog, M-) was multiplied by the inverse of the reduction in number of cards for each TP-BAS card set. Specifically, because the 5 card series was .50 as long as the full set, each raw variable was multiplied by 2 for the 5 card series. The 4 card series was .40 as long as the original so the raw scores were multiplied by 2.5; and the 3 card set was .30 as long as the original so the raw scores were multiplied by 3.33 for this series. As is evident in Table 9, all sets have roughly equivalent means for each variable.
### Table 9
**Combined Dean, Chicago, and R-PAS Samples – Card Set by Protocol-Level Mean of Variable**

<table>
<thead>
<tr>
<th>Card Series</th>
<th>10</th>
<th>5</th>
<th>4</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Card Subset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FQu%</td>
<td>30.66</td>
<td>30.94</td>
<td>30.32</td>
<td>30.56</td>
</tr>
<tr>
<td>FQo%</td>
<td>53.86</td>
<td>54.00</td>
<td>53.96</td>
<td>54.59</td>
</tr>
<tr>
<td>WDm%</td>
<td>9.76</td>
<td>9.56</td>
<td>9.89</td>
<td>9.53</td>
</tr>
<tr>
<td>Popular</td>
<td>5.38</td>
<td>2.52</td>
<td>2.86</td>
<td>2.21</td>
</tr>
<tr>
<td><strong>Thinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSumCog</td>
<td>17.10</td>
<td>8.44</td>
<td>8.66</td>
<td>6.66</td>
</tr>
<tr>
<td>WSumCog/R</td>
<td>.73</td>
<td>.73</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td>SevCog</td>
<td>1.63</td>
<td>.85</td>
<td>.78</td>
<td>.59</td>
</tr>
<tr>
<td>SevCog/R</td>
<td>.07</td>
<td>.07</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Composites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-Comp</td>
<td>1.20</td>
<td>1.06</td>
<td>1.06</td>
<td>.95</td>
</tr>
<tr>
<td>TP-Comp/R</td>
<td>1.20</td>
<td>1.30</td>
<td>1.29</td>
<td>1.22</td>
</tr>
</tbody>
</table>

**Note:** n=312 for all card sets.

- **Popular/R** = Protocol-level sum of Popular responses divided by the number of responses in the protocol.
- **WSumCog/R** = Sum of protocol-level WSumCog divided by the number of responses in the protocol.
- **SevCog/R** = Sum of protocol-level SevCog divided by the number of responses in the protocol.
- **TP-Comp-R** = Each of the raw variables used in the TP-Comp formula (i.e., FAB2, WSumCog, M-) was multiplied by the inverse of the reduction in number of cards for each TP-BAS card set. Specifically, each raw variable was multiplied by 2 for the 5 card series, 2.5 for the 4 card series, and 3.33 for the 3 card series.
As is shown in Table 10, all part-whole correlations were large and ranged from .57 to .94. Consistent with psychometric theory, part-whole correlations were largest for the TP-BAS-5 and gradually decrease as a function of card reduction. As previously noted, the TPMZ was computed as an alternative to the TP-Comp, after omitting FQu and P. FQu and P were omitted because of their relatively low validity in the detection of psychosis (Mihura et al., 2011). FQu and P are also omitted from Table 10 and thereafter for the same reason.
Table 10

Dean, Chicago, and R-PAS Samples –
Protocol-Level Full-Form to Short-Form Part-Whole Correlations

<table>
<thead>
<tr>
<th>Card Series</th>
<th>Perception</th>
<th>Thinking</th>
<th>Composites</th>
<th>Mean Part-Whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card Subset</td>
<td>FQ-%</td>
<td>WSumCog</td>
<td>TP-Comp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FQo%</td>
<td>SevCog</td>
<td>TPMZ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WDm%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.85 .82 .<strong>83</strong></td>
<td>.93 .93 .<strong>93</strong></td>
<td>.85 .83 .<strong>84</strong></td>
<td>.87 .86 .<strong>86</strong></td>
</tr>
<tr>
<td></td>
<td>.76 .76 .<strong>76</strong></td>
<td>.87 .90 .<strong>89</strong></td>
<td>.77 .79 .<strong>78</strong></td>
<td>.79 .81 .<strong>80</strong></td>
</tr>
<tr>
<td></td>
<td>.73 .67 .61 .<strong>67</strong></td>
<td>.85 .84 .82 .<strong>83</strong></td>
<td>.73 .69 .64 .<strong>68</strong></td>
<td>.76 .74 .69 .<strong>73</strong></td>
</tr>
<tr>
<td>4</td>
<td>.81 .84 .<strong>82</strong></td>
<td>.94 .92 .<strong>93</strong></td>
<td>.89 .88 .<strong>89</strong></td>
<td>.82 .85 .<strong>84</strong></td>
</tr>
<tr>
<td>3</td>
<td>.83 .79 .<strong>81</strong></td>
<td>.84 .87 .<strong>87</strong></td>
<td>.80 .80 .<strong>80</strong></td>
<td>.73 .78 .<strong>78</strong></td>
</tr>
<tr>
<td>3</td>
<td>.73 .66 .57 .<strong>65</strong></td>
<td>.81 .84 .83 .<strong>83</strong></td>
<td>.80 .80 .<strong>80</strong></td>
<td>.73 .78 .<strong>78</strong></td>
</tr>
</tbody>
</table>

Note: n=312 for all card sets.
All correlations were significant at the .01 level (two tailed).
TPMZ= Average of z scores for WSumCog, FQ-%, reverse scored FQo%, WDm%, and SevCog.
As is reported in Tables 10 and 11 and graphed Figures 1 and 2, the same general trends emerged with regard to validity across all three samples. Figures 1 and 2 also collectively depict part-whole correlations and interrater reliability data for all three samples. In all but one instance, all validity correlations were in the predicted direction. This suggests that TP-BAS measures of perception and thinking problems increase as the degree of psychotic psychopathology increases. Validity coefficients for the Thinking variables tended to be higher than for the Perception variables. The TPMZ consistently performed better than the TP-Comp. Although not observed in the Dean sample, FQo tended to have higher validity coefficients than FQ- in the combined Chicago and R-PAS samples. There also tended to be more variability in validity across TP-BAS card sets in the Dean sample than in the merged data set containing the Chicago and R-PAS samples, which is likely due to its relatively small sample size. Overall validity tended to be slightly higher in the Chicago and R-PAS samples merged than in the Dean sample. No impact on validity was observed as a function of card color variability across card sets. In line with psychometric theory, validity coefficients, part-whole correlations, and interrater reliability tended to decrease slightly as a function of card reduction.
Table 11
Dean Sample – Protocol-Level TP-BAS Validity; Criterion = Thought Disorder Summary Scale

| Card Subset | Card Series | 1   | 2   | M   | 1   | 2   | M   | 1   | 2   | 3   | M   |
|------------|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Perception | FQ-%        | .32 | .29 | .24 | .27 | .22 | .26 | .24 | .22 | .22 | .15 | .20 |
|            | FQo%        | .27 | .19 | .28 | .23 | -.03| .37 | .17 | -.27| .25 | .06 | .01 |
|            | WDrm%       | .28 | .20 | .27 | .23 | .14 | .26 | .20 | .18 | .21 | .14 | .18 |
| Thinking   | WSumCog     | .56 | .48 | .54 | .51 | .40 | .52 | .46 | .47 | .54 | .44 | .48 |
|            | SevCog      | .48 | .45 | .49 | .47 | .47 | .48 | .48 | .46 | .50 | .39 | .45 |
| Composites | TP-Comp     | .39 | .32 | .34 | .33 | .25 | .33 | .29 | .26 | .30 | .22 | .26 |
|            | TPMZ        | .48 | .42 | .48 | .45 | .29 | .49 | .39 | .40 | .47 | .33 | .40 |
| Mean Validity |          | .39 | .33 | .37 | .35 | .25 | .39 | .32 | .25 | .35 | .25 | .28 |

*Correlation is significant at the .05 level (two tailed). ** Correlation is significant at the .01 level (two tailed).
FQo% was reverse scored.

Note: n=61 for each card set.
Table 12  
Chicago & R-PAS Samples – Protocol-Level TP-BAS Validity; Criterion = Ego Dysfunction Severity Scale

<table>
<thead>
<tr>
<th>Card Series</th>
<th>1</th>
<th>2</th>
<th>M</th>
<th>1</th>
<th>2</th>
<th>M</th>
<th>1</th>
<th>2</th>
<th>M</th>
<th>1</th>
<th>2</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FQ-%</td>
<td>.28**</td>
<td>.22**</td>
<td>.22</td>
<td>.22**</td>
<td>.18**</td>
<td>.20</td>
<td>.13</td>
<td>.23**</td>
<td>.19**</td>
<td>.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FQo%</td>
<td>.32**</td>
<td>.21**</td>
<td>.32</td>
<td>.26</td>
<td>.23**</td>
<td>.28**</td>
<td>.25</td>
<td>.18**</td>
<td>.27**</td>
<td>.22**</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>WDm%</td>
<td>.19**</td>
<td>.11</td>
<td>.17</td>
<td>.14</td>
<td>.17**</td>
<td>.12</td>
<td>.15</td>
<td>.13</td>
<td>.08</td>
<td>.17**</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td><strong>Thinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSumCog</td>
<td>.61**</td>
<td>.56**</td>
<td>.58</td>
<td>.57</td>
<td>.56</td>
<td>.55</td>
<td>.56</td>
<td>.50**</td>
<td>.51**</td>
<td>.48**</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>SevCog</td>
<td>.60**</td>
<td>.53**</td>
<td>.58</td>
<td>.55</td>
<td>.52</td>
<td>.54</td>
<td>.53</td>
<td>.45**</td>
<td>.50**</td>
<td>.47**</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td><strong>Composites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-Comp</td>
<td>.50**</td>
<td>.39</td>
<td>.40</td>
<td>.40</td>
<td>.39**</td>
<td>.34</td>
<td>.36</td>
<td>.26**</td>
<td>.33**</td>
<td>.32**</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>TPMZ</td>
<td>.55**</td>
<td>.47**</td>
<td>.52</td>
<td>.49</td>
<td>.49**</td>
<td>.48</td>
<td>.49</td>
<td>.41**</td>
<td>.46**</td>
<td>.44**</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>Mean Validity</td>
<td>.43</td>
<td>.35</td>
<td>.40</td>
<td>.38</td>
<td>.37</td>
<td>.36</td>
<td>.36</td>
<td>.29</td>
<td>.34</td>
<td>.33</td>
<td>.32</td>
<td></td>
</tr>
</tbody>
</table>

* Correlation is significant at the .05 level (two tailed). ** Correlation is significant at the .01 level (two tailed).
FQo% was reverse scored.
As is shown in Table 13, the obtained interrater reliability coefficients surpass Smith et al.’s (2000) suggested cutoff value of .70.

<table>
<thead>
<tr>
<th>Card Series</th>
<th>10</th>
<th>5</th>
<th>4</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Card Subset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FQ-%</td>
<td>.94</td>
<td>.88</td>
<td>.86</td>
<td>.87</td>
</tr>
<tr>
<td>FQo%</td>
<td>.94</td>
<td>.90</td>
<td>.90</td>
<td>.90</td>
</tr>
<tr>
<td>WDm%</td>
<td>.95</td>
<td>.93</td>
<td>.90</td>
<td>.92</td>
</tr>
<tr>
<td><strong>Thinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSumCog</td>
<td>.90</td>
<td>.89</td>
<td>.87</td>
<td>.88</td>
</tr>
<tr>
<td>SevCog</td>
<td>.85</td>
<td>.84</td>
<td>.78</td>
<td>.81</td>
</tr>
<tr>
<td><strong>Composites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-Comp</td>
<td>.94</td>
<td>.91</td>
<td>.88</td>
<td>.89</td>
</tr>
<tr>
<td>TPMZ</td>
<td>.94</td>
<td>.93</td>
<td>.88</td>
<td>.90</td>
</tr>
<tr>
<td>Mean</td>
<td>.92</td>
<td>.90</td>
<td>.87</td>
<td>.88</td>
</tr>
</tbody>
</table>

Table 13
*Protocol-Level TP-BAS Interrater Reliability – Meyer et al. (2011) Sample*
Figure 1. Dean Sample - Mean Part-Whole Correlations and Mean Validity

Note: n=61 Protocols per TP-BAS Card Series
PW = part-whole correlation; V = validity coefficient
Table 10 includes part-whole correlation data from a composite dataset consisting of the Dean, Chicago, and R-PAS samples. However, Figure 1 displays part-whole correlation data for the Dean sample only.
Figure 2. Chicago and R-PAS Samples - Mean Part-Whole Correlations and Mean Validity

Note: n=251 Protocols per TP-BAS Card Series
PW = part-whole correlation; V = validity coefficient
Table 10 includes part-whole correlation data from a composite dataset consisting of the Dean, Chicago, and R-PAS samples. However, Figure 2 displays part-whole correlation data for composite dataset consisting of the Chicago and R-PAS samples (i.e., the Dean sample is omitted).
Figure 3. Meyer et al., (2011) Sample - Mean TP-BAS Interrater Reliability

Note: n=93 Protocols Per TP-BAS Card Series
IR = interrater reliability
As is shown in Table 14, OCCs for each of the TP-BAS card sets are comparable to those previously reported for the PTI (Dao & Prevatt, 2006; Hilsenroth et al., 2007). It is notable, however, that the merged Chicago and R-PAS dataset contains a disproportionately high number of nonpatients. This is likely to have elevated the obtained NPP, SP, and OCCs. The inflation of OCCs was likely driven by the inflated SP values. Given these considerations, attending to the NPP, SP, and OCCs may be slightly misleading, which makes SN and PPP important considerations when evaluating the diagnostic efficiency of the TP-BAS TP-Comp. Overall, SN and PPP drop as a function of card reduction but, when considered on a set-by-set basis, the drop in SN and PPP is not linear as might be expected. However, when one considers the mean SN of the TP-BAS-5, -4, and -3 (.387, .322, and .349, respectively) and the mean PPP of the TP-BAS-5, -4, and -3 (.510, .477, .442, respectively), a more linear trend emerges.
<table>
<thead>
<tr>
<th>Card Set</th>
<th>TP-Comp</th>
<th>Yes</th>
<th>No</th>
<th>SN</th>
<th>SP</th>
<th>PPP</th>
<th>NPP</th>
<th>OCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP-BAS10</td>
<td>≥ 2</td>
<td>33</td>
<td>21</td>
<td>.532</td>
<td>.889</td>
<td>.611</td>
<td>.853</td>
<td>.801</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>29</td>
<td>168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS5.1</td>
<td>≥ 2</td>
<td>22</td>
<td>23</td>
<td>.355</td>
<td>.878</td>
<td>.489</td>
<td>.806</td>
<td>.749</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>40</td>
<td>166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS5.2</td>
<td>≥ 2</td>
<td>26</td>
<td>23</td>
<td>.419</td>
<td>.878</td>
<td>.531</td>
<td>.822</td>
<td>.765</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>36</td>
<td>166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS4.1</td>
<td>≥ 2</td>
<td>22</td>
<td>20</td>
<td>.355</td>
<td>.894</td>
<td>.524</td>
<td>.809</td>
<td>.761</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>40</td>
<td>169</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS4.2</td>
<td>≥ 2</td>
<td>18</td>
<td>24</td>
<td>.290</td>
<td>.873</td>
<td>.429</td>
<td>.789</td>
<td>.729</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>44</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS3.1</td>
<td>≥ 2</td>
<td>15</td>
<td>31</td>
<td>.242</td>
<td>.836</td>
<td>.326</td>
<td>.771</td>
<td>.689</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>47</td>
<td>158</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS3.2</td>
<td>≥ 2</td>
<td>20</td>
<td>26</td>
<td>.323</td>
<td>.862</td>
<td>.435</td>
<td>.795</td>
<td>.729</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>42</td>
<td>163</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP-BAS3.3</td>
<td>≥ 2</td>
<td>30</td>
<td>23</td>
<td>.484</td>
<td>.878</td>
<td>.566</td>
<td>.838</td>
<td>.781</td>
</tr>
<tr>
<td></td>
<td>&lt; 2</td>
<td>32</td>
<td>166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter Three
Discussion and Implications

Discussion

Despite obtaining results that are generally consistent with the extant Rorschach validity literature, some of the present findings were unexpected. First, the mean level of FQo was not expected to differ more substantially across cards than the mean level of FQ- (Cohen’s f of .26 vs. .12). Specific predictions were not made with regard to the relationship of FQo and FQ- across cards. However, as previously noted, certain cards tend to be relatively “easy” as a function of relatively simple stimulus properties while others are more “difficult” as a function of more complex stimulus properties. Thus, it is reasonable to anticipate more FQo and less FQ- responses on easy cards and vice versa, such that a balance is created between the two codes across all 10 Rorschach cards. As such, one might reasonably expect the degree of variability in these codes to be comparable across cards.

In contrast to this reasoning, the present results suggest that some Rorschach cards tend to be exceptionally “easy” (i.e., Cards I and V) relative to the other cards as measured by FQo while the level of “difficulty” of the cards, as measured by FQ-, is more evenly distributed across cards. In other words, it appears that the Rorschach is able to detect a wider range of conventional perception as measured by FQo than is the case for detecting distorted perception as measured by FQ-. Notably, Card IX appears to be the most difficult card as measured by FQ-, FQo, and P. Finally, FQo and P tend to increase and decrease in the same pattern across cards but the shifts are not comparable in
magnitude. This suggests that FQo and P are tapping in to distinct facets of conventional Perception among respondents.

The present results suggest that the TPMZ may perform better than the TP-Comp with regard to validity. However, firm conclusions cannot be made about the relative merits of the TP-Comp and the TPMZ as no formal statistical test was conducted to determine whether a significant difference is present between the validity of the two indices. Nonetheless, the preliminary evidence presented here suggests that it may be appropriate to replace the TP-Comp with the TPMZ or a comparable measure in R-PAS and TP-BAS. Notably, an added benefit of the TPMZ is that it is much easier to compute than the TP-Comp.

Next, the FQ variables did not demonstrate validity quite as high as that reported in Mihura et al.’s (2012) recent meta-analysis. Specifically, those authors reported that, with the exception of FQu, all other FQ variables “showed significant moderate to large effect size relationships to their criteria” (p. 25), while the present analyses yielded validity coefficients in the small to medium range. It was initially suspected that the obtained validity coefficients may have been smaller than those reported in Mihura et al. (2012) because the meta-analyzed studies made use of CS FQ tables whereas the present analyses made use of R-PAS FQ tables. However, a follow-up diagnostic efficiency analysis of the PTI using Chicago sample data scored using CS FQ ratings revealed that the TP-Comp in the same sample scored using R-PAS FQ ratings performed at roughly the same level. Thus, this was ruled out as a potential explanation of the anomalous FQ validity findings.
Two explanations for the anomalous FQ findings seem viable. First, the samples used may have systematically differed in some unknown way from the samples included in the meta-analysis. Second, it may be possible that the validity coefficients obtained in the present research were reduced as a function of the criterion measures used. Specifically, respondents in the R-PAS sample were not screened for psychopathology and were assumed to have an Ego Dysfunction Severity Scale score of 0; it is possible that some of the individuals met criteria for higher Ego Dysfunction Severity Scale scores. Furthermore, Ego Dysfunction Severity Scale scores were assigned on the basis of diagnoses that were rendered before the initiation of testing and, as such, test data did not inform diagnoses or Ego Dysfunction Severity Scale ratings. Thus, the validity coefficients that obtained from the Chicago sample are an underestimate of actual TP-BAS validity. Additionally, the Ego Dysfunction Severity Scale is not a direct measure of psychosis as individuals scoring in the low range may not have any psychotic symptomatology. A more pure measure of psychosis may have yielded higher validity coefficients. It is also the case that Mihura et al. (2012) classified a variety of criteria as “externally assessed” (e.g., observer and chart ratings, non-Rorschach performance-based assessment data, offender status, and level of education) whereas the present validity criteria were restricted to DSM diagnoses and structured interview data. Finally, the validity criterion used in the Dean sample (i.e., the Thought Disorder Summary Scale) is heavily focused on problems in thinking and does not provide a direct measure of problems in perception. Despite these possible limitations of the criterion measures, it is notable that the lowered validity was not simply a function of reducing the number of
cards administered; the full-form also yielded less impressive validity coefficients for the FQ variables than expected.

Perhaps most surprisingly, the Thinking variable validity coefficients were larger than expected. Mihura et al. (2012) report r values for the Thinking variables using external assessment validity criteria ranging from .36 to .38 whereas Thinking variable r values yielded by the present analyses ranged from .48 to .61 across each sample. It is not entirely clear why the Thinking variable validity coefficients obtained in the present research were notably high. However, it is possible that the aforementioned limitation of the Dean validity criterion measure (i.e., a disproportionate emphasis on problems in thinking) contributed to the larger than expected Thinking variable validity.

Finally, it was expected that one of the TP-BAS card series would clearly possess the most optimal balance of validity and utility. Contrary to this expectation, each series has slightly unique merits, making it possible for one to reasonably argue for the retention of each. Regarding validity, the TP-BAS-4 in the Dean sample had the most variability across its component card sets. This variability was not present in the merged database consisting of the Chicago and R-PAS samples. This may be a function of the different validity criteria used; however, it is more likely due to the considerably larger sample size in the merged database. The TP-BAS-5 generally has the most validity. Thus, opting to use the TP-BAS-4 and -3 would be defensible only if the decrement in validity is a reasonable compromise that results in valuable time savings. In general, use of the TP-BAS-4 is least likely to be optimal. This is partially because of the observed variability in TP-BAS-4 validity coefficients but is primarily because the amount of time saved by transitioning from the TP-BAS-5 to the -4 is trivial. On the other hand, use of
the TP-BAS-3 offers noteworthy time savings (approximately 1 hour for the full-form and 18 minutes for the TP-BAS-3) and a relatively minimal cost in validity. In addition, if one uses the TP-BAS-3 they gain the advantage of an additional unique test-retest administration for future use as the TP-BAS-3 contains three alternate forms rather than two. It is noteworthy that the TP-BAS-3.3 had a fairly high standard deviation for FQo. However, the mean level of this variable in the TP-BAS-3.3 is comparable to the other cards sets. This suggests that the variability of FQo in TP-BAS-3.3 is not a cause for great concern. Perhaps clinicians using the TP-BAS-3 should consider using the TP-BAS-3.1 and -3.2 before using the TP-BAS-3.3.

As the TP-BAS TP-Comp diagnostic efficiency statistics indicate, each TP-BAS card series appears to be an adequate means of collecting diagnostic information about the psychoses and thought disorder. Although the TP-BAS diagnostic efficiency statistics obtained in the present research are less impressive than those obtained for the PTI in previous research (Hilsenroth et al., 2007), it is important to consider that the discrepancy is at least partially explained by the psychometric properties of the PTI and TP-Comp. As a function of regression to the mean and its dimensional distribution, TP-Comp values tend to correspond to higher PTI values. For example, the required level of pathology in Rorschach responses to elevate the PTI to a 2 corresponds to the approximate level of pathology required to elevate the TP-Comp to 1.8 (Viglione, Giromini, Gustafson, & Meyer, 2012). The implication of this for SN and PPP is that is “harder” to reach criteria for psychotic classification by TP-Comp test data and, consequently, fewer individuals who truly do have a psychotic diagnosis will reach the TP-Comp criterion. It is likely that the diagnostic efficiency statistics in the present research would have more closely
approximated those reported by Hilsenroth et al. (2007) if a TP-Comp cutoff of 1.8 had been used.

Limitations & Future Directions

The present research was not intended to establish a final version of the TP-BAS for use in clinical practice. Rather, the goal of the present research was to construct TP-BAS card series, provide preliminary reliability and validity evidence for each series, and then outline future steps to be taken to fully establish the TP-BAS as a viable clinical instrument. As such, the limitations of the present research are not fatal methodological flaws because most are essentially one and the same as the future directions proposed.

The future collection of normative data for the TP-BAS is the most important next step for developing the system to be used in clinical practice. Before normative data collection can take place, TP-BAS series to be included in the final system will need to be selected and TP-BAS systematizers will need to develop and finalize administration procedures by conducting pilot studies to examine the efficacy of preliminary procedures. As previously noted, the TP-BAS-5 and -3 are more optimal than the TP-BAS-4.

Furthermore, the TP-BAS-5 and -3 each have their own unique merits and limitations. As such, it seems reasonable to collect normative data for the TP-BAS-5 and -3, clearly articulate the strengths and weaknesses of each, and then offer clinicians and researchers the ability to choose which series best suits their needs (i.e., they will need to choose between time efficiency and the need for precision in validity).

Preliminary TP-BAS administration considerations and guidelines follow. Use of the TP-BAS will be less labor intensive than the full R-PAS administration and scoring procedures for two reasons. First, administration of fewer cards obviously results in less
administration and scoring time. Second, although the short-form Clarification Phase will mimic the full-form Clarification Phase, the only variables that will be coded are the Thinking and Perception variables. Specifically, administration of the short-form will require examiners to code Location, Cognitive Codes, and FQ. Developing clarification guidelines for the TP-BAS is arguably the most difficult task required to establish the instrument as a viable means of assessing thought disorder and psychosis.

There is currently no prototype for abbreviated Clarification Phase administration guidelines. As previously noted, the TDI short-form (Carpenter et al., 1993) is the nearest approximation of the instrument to ultimately be yielded by the present program of research. Reviewing TDI short-form administration guidelines was not of help in generating TP-BAS Clarification Phase guidelines as both the TDI (Johnston & Holzman, 1979; Solovay et al., 1986) and the TDI short-form (Carpenter et al., 1993) make use of the Rapaport method of clarification which requires that examiners inquire for enough information to score response “location, determinants, form level, and content” (Johnston & Holzman, 1979, p.63). This is more information than will be needed to score the TP-BAS. The difficulty to be encountered in deriving TP-BAS Clarification Phase guidelines will arise from the goal of simultaneously reducing the labor intensity of teaching and learning administration procedures while maintaining the richness of the verbal sample yielded from the R-PAS Clarification Phase.

Research has made it clear that eliminating the Clarification Phase entirely is not a viable option, particularly if one is interested in assessing thought disorder. Ritzler and Nalesnik (1990) compared Rorschach protocols with and without the Clarification Phase of nonpatients and patients with schizophrenia, depression, and personality disorders to
determine whether eliminating the Clarification Phase causes a decrement to protocol validity. They found that the results varied depending upon the variable under consideration. Of course, the Thinking and Perception variables are most relevant to the present research. Ritzler and Nalesnik (1990) found that those variables do in fact meaningfully differ with and without the Clarification Phase with the largest effect present in WSumCog. Specifically, they found that the presence of Cognitive Codes substantially decreased as a function of eliminating the Clarification Phase. Regarding the Thinking variables, they state that the “…inquiry effected the score of the schizophrenic group to a greater extent than for the other three diagnostic groups” (p. 651) and “without these scores, the presence and nature of serious disorders such as schizophrenia are nearly impossible to determine” (p. 652).

One might speculate that eliminating the Clarification Phase may be permissible if R-Optimized procedures were modified such that examiners prompt for more responses per card in order to obtain a larger speech sample. However, this is unlikely to be true given that the clarification task and associated speech sample qualitatively differ from that of the Response Phase in that providing a clarification requires more logic and reasoning abilities. Furthermore, this alternative seems to defeat the purpose of administering fewer cards as it would not be more time efficient.

One early option that was considered to contend with the Clarification Phase problem consisted of providing examiners with a list of key words that commonly signal the presence of a particular response characteristic (e.g., the use of color) rather than asking clarifying questions with regard to specific determinants as is done in R-PAS. Specifically, when respondents use a key word, examiners would follow-up with a
question such as “What makes it look [insert key word]?”. The rationale for this procedure was to decrease new examiner learning time by replacing complex decision making during the Clarification Phase with a premade list of common key words. However, TP-BAS systematizers ultimately ruled this out as an option as a list of key words may prove to unnecessarily restrict the range of verbalizations that get clarified and, although less labor intensive than R-PAS clarification, it may be difficult for examiners to manage a lengthy list of key words during administrations.

As of this writing, the clarification procedures most likely to be pilot tested consist of asking a standardized list of clarifying questions. These questions are expected to elicit verbalizations from respondents that are comparable to those obtained from R-PAS clarification while only requiring examiners to memorize or reference a list of approximately three questions. For example, examiners might clarify responses by asking “What makes it look like that?” or “How are you seeing it?” or “Help me see it as you do.” These clarification questions are considered taboo in the R-PAS manual because they are too broad to clarify the presence of a particular determinant. However, the goal of the TP-BAS Clarification Phase is somewhat different than that of the R-PAS Clarification Phase. In order to score a Rorschach protocol using R-PAS, examiners are required to have more detailed information about the determinants of each response. Use of the TP-BAS will not require such detail as examiners will only need to know where respondents see the percepts they report and, in a broad sense, what made the inkblot look like that to them. The primary purpose of asking the examinee what made the inkblot look like the percept that they report is to obtain a verbal sample sufficient to score the TP-BAS Thinking variables (i.e., to evaluate speech coherence, logic, and
reasoning). The aforementioned clarification questions seem optimal for the purposes of TP-BAS administrations as they are not difficult to learn, do not require complex decision making on the part of the examiner, and should yield sufficient speech samples. TP-BAS systematizers are considering renaming the Clarification Phase in the TP-BAS the “Elaboration Phase” to highlight its distinctions from the R-PAS Clarification Phase. As a note, examiners should begin TP-BAS administrations with an achromatic card and then progress through the remaining cards in numerical order. Examiners should also use R-optimized R-PAS procedures (Meyer et al., 2011) when administering the TP-BAS.

Efforts are being planned to collect normative data from patients with and without psychotic diagnoses, forensic patients, and healthy adult volunteers. Normative data collection will provide protocols coded by a more diverse array of examiners to further ensure that site specific coding conventions are not confounding TP-BAS results. Normative data collection will also address the fact that the Rorschach protocols included in the present research were not administered using TP-BAS administration guidelines and determine whether the previously established validity of 10-card Rorschach CS administration in identifying thought disorder and psychosis carries over to the TP-BAS. Administration of the TP-BAS will differ from CS administration in that fewer cards are administered and the range of allowable responses is smaller as a function of R-Optimized administration procedures. Aside from decreasing administration time, the use of fewer cards impacts the range of stimuli and the order of the cards that are presented to respondents. It is possible that responses will be influenced by the aforementioned modifications to be made to the administration guidelines.
As part of normative data collection, TP-BAS administrations should be timed so that the projected time savings can be evaluated for accuracy. Respondents will also be administered self-report instruments that assess for psychosis (e.g., MMPI) and structured or semi-structured interviews (e.g., TLC) in order to establish the convergent and discriminant validity of the TP-BAS on independent samples. The collection of normative data will also determine whether the lower than expected FQ validity coefficients replicate using various other validity criterion measures. If they are replicated, it may be determined that the Perception variables should be removed from the TP-BAS. However, the elimination of FQ from the TP-BAS does not seem optimal for several reasons. Interpretively and practically speaking, Kimhy et al.’s (2007) finding that perceptual impairment is a steady trait like characteristic of individuals prone to psychotic processes suggests that FQ is a potentially valuable tool in identifying prodromal psychosis. Furthermore, FQ provides data regarding perceptual difficulties that are impossible to obtain from other speech samples or interview methods. Perhaps most important is the fact that Mihura et al.’s (2012) meta-analysis suggests that FQ has strong validity. Given that their study was based on a much larger empirical evidence base than the one used in the present research, it is more useful to use their data as a point of reference for evaluating the global validity of FQ.

Several research steps can be taken to improve the soundness of the TP-BAS after normative data are collected and published. First, overlapping variance between the full 10 card set and each TP-BAS card set should be established by administering the full-form and TP-BAS card sets to the same individuals during independent assessment sessions. Additionally, test-retest reliability data should be collected by administering
each TP-BAS card set to the same individuals on at least two separate occasions. Regarding diagnostic efficiency, future research should investigate whether using a TP-
BAS TP-Comp cutoff of $\geq 1.8$ is more optimal than a cutoff of $\geq 2$. Furthermore, future research should investigate the diagnostic efficiency of each TP-BAS variable in turn. As previously noted, TP-Comp data were simply used in the present research to illustrate the approximate magnitude of decrements in diagnostic efficiency as a function of card reduction across TP-BAS card sets. The TPMZ appears to be a particularly good candidate for future investigations of TP-BAS diagnostic efficiency as it consistently outperformed the TP-Comp with regard to validity and, consequently, it is likely that the TPMZ would also outperform the TP-Comp with regard to its diagnostic efficiency.

As Ilonen et al. (2010) note, very few studies have been conducted using the Rorschach to detect prodromal psychosis. As such, this is an area that definitely warrants future research attention and the TP-BAS appears to be a good alternative to full-form administrations for such research. For example, future research might seek to establish the concurrent validity of the TP-BAS by administering it in conjunction with other measures of prodromal psychosis, such as the SIPS. Future researchers might seek to establish the diagnostic accuracy of the TP-BAS in detecting prodromal psychosis by conducting a longitudinal study wherein baseline and follow-up TP-BAS data are collected for individuals initially diagnosed with attenuated psychosis syndrome to determine what percentage go on to develop full-blown psychosis. Such a longitudinal study could also provide useful data about the normative characteristics of TP-BAS data at each stage in the development and progression of psychosis.
It will also be important to investigate those components of the TP-BAS that are modified versions of previously existing variables and indices. For example, given that prior research suggests that the PTI cannot reliably distinguish individuals at a clinically high risk for developing psychosis from those that have already converted to psychosis (Ilonen et al., 2010), future research should investigate whether the TP-Comp or TPMZ are better able to distinguish these two groups. As a note, it was possible to equate Thinking and Perception variable means within TP-BAS card sets; however the standard deviations are more variable. Future research might investigate whether TP-BAS card sets with larger standard deviations provide a more valid assessment of psychosis than those with smaller standard deviations using a regression analysis.

Finally, future research might seek to identify a single optimal 3-card Rorschach short-form via regression analysis. The advantage of having an optimal set is that it would be the most valid set and would work well for those interested in a single brief measurement of an examinee’s perceptual and thinking abilities. The research design would consist of entering individual variables and/or overarching indices as predictor variables and a validity criterion as the outcome variable, on a card-by-card basis. The regression models yielded for each variable and/or index could then be compared with regard to how well each predictor variable performs, on a card-by-card basis. The most difficult part of such analyses would be to decide whether to use individual variables or overarching indices as predictors. For example, one may need to choose between using the TP-Comp or its constituent variables (e.g., FQ-, M-, FAB2) as predictors. Entering individual variables would allow for a more fine-grained analysis and precise means of constructing the three card set; however, this methodology is also likely to create
situations where difficult decisions with no clear answers must be made. Specifically, the results of such analyses are not likely to yield a pattern of results where three individual cards are clearly the most optimal as one or more variables is likely to deviate from whatever general trend might be present in variable validity. On the other hand, indices are likely to provide more coherent and parsimonious results at the expense of precision.

Regardless of whether one uses individual variables or overarching indices as predictor variables, regression methodology would allow for the derivation of the most optimal (i.e., highest prediction of validity criterion) set of three Rorschach cards to be used in a short-form as card selection would be based solely on predictive validity. This is distinct from the methodology used in the present research as card sets were equated based on the mean and standard deviation of the Perception variables and subsequently validated the resulting card sets using external validity criteria. In short, the present methodology placed a higher premium on short-form utility in the form of test-retest capability while regression methodology would place a higher premium on utility in the form of predictive validity. Neither of the aforementioned methodologies is clearly better than the other and, as such, exploration of both is warranted.

Conclusion

Harsh criticisms have been leveled against the Rorschach throughout the history of its existence. Concerns about its validity have caused considerable skepticism regarding its credibility and usefulness in clinical practice. Garb (1999) even called for a moratorium on the use of the Rorschach in clinical practice. He said that the instrument should be subjected to empirical scrutiny before being used again in clinical practice. Clearly, psychological assessment instruments should demonstrate acceptable reliability
and validity before they are used in clinical settings (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Admittedly, the Rorschach, like all psychological assessment instruments, brings with it a unique blend of unreliability and limited validity. However, evidence suggests that singling out the Rorschach for criticism is not warranted (Meyer, 2004; Mihura et al., 2011).

To be fair, critiques of the Rorschach’s validity and utility contain a certain degree of truth. Based on a systemic review of the literature it turns out there are Rorschach variables and indices with minimal empirical support that have traditionally been taught in training programs and used in practice (Mihura et al., 2011). Furthermore, it is understandable that many training programs are of the opinion that teaching the CS (Exner, 2003) is unwieldy (Groth-Marnat, 2009). R-PAS offers substantial improvements in both areas. However, for those training programs and clinicians who want even shorter administration and scoring times, the TP-BAS promises to deliver.

Fortunately, Rorschach critics and proponents agree about one thing: the Rorschach does provide a valid assessment of psychosis (Wood, Nezworski, & Stejskal, 1996a, 1996b). Specifically, critics acknowledge that patients with psychotic diagnoses look more pathological on Rorschach Thinking and Perception variables. The results of the present research corroborate past research documenting the validity of the Rorschach Perception and Thinking variables in detecting psychosis and suggest that those findings will generalize to the TP-BAS upon its completion. Therefore, Rorschach critics and proponents should recognize and appreciate the merits of the TP-BAS, insofar as the final system disseminated for use in clinical practice is validated on independent samples.
Thus, it is hoped that the instrument ultimately yielded by the culmination of the present research and the future collection of normative data will be widely accepted and adopted in practice and research.

The development of the TP-BAS is occurring at a pivotal point in mental health assessment and treatment history, namely the development of the DSM-5. The TP-BAS is uniquely positioned to aid in the assessment, diagnosis, monitoring, and treatment of the psychoses in ways that are consistent with the proposed revisions to the DSM classification scheme. Specifically, the system will provide dimensional behavioral data that do not rely on self-report, client introspection, thorough reviews of client charts, or client understanding of interview questions. Furthermore, it should facilitate interclinician diagnostic reliability. The TP-BAS will accomplish all of these ends in approximately 18 to 30 minutes, depending upon which card series is used. Furthermore, the online scoring program for R-PAS will also be made available for the TP-BAS to simplify the scoring and interpretive processes.
References


from http://www.dsm5.org/about/Pages/Timeline.aspx.


Hartmann, H. (1953). The metapsychology of schizophrenia. *The Psychoanalytic Study of*
the Child, 8, 177-198.


doi:10.1093/schbul/12.3.360


Meyer, G. J. (2004). The reliability and validity of the Rorschach and TAT compared to other psychological and medical procedures: An analysis of systematically


Parker, K. C., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A


Impairment Index incorporating the Human Representation Variable. *Journal of Personality Assessment, 81*, 149-156. doi: 0.1207/S15327752JPA8102_06


