JOURNAL OF PERSONALITY ASSESSMENT, *89*(2), 149–161

A Method to Optimize the Response Range While Maintaining Rorschach Comprehensive System Validity

Kristin L. Dean and Donald J. Viglione
Department of Clinical Psychology
California School of Professional Psychology at Alliant International University

William Perry
Department of Psychiatry
University of California at San Diego

Gregory J. Meyer
Department of Psychology
University of Toledo

Brief and lengthy Rorschach records have been identified as common problems in protocol administration. Clinicians have debated how to prevent overly short and long records, but they have been reluctant to alter standardized administration for fear of introducing bias. The present study examines a nonintrusive method for constraining responses by prompting for an extra response when only one is offered per card and by removing the card after four responses are given. Among patients who typically produce brief records, consisting of a residential sample of civil and forensic patients with a range of disordered thinking, the alternative administration method demonstrated improved Comprehensive System validity in assessing thought disorder and eliminated the need to readminister the test due to fewer than 14 responses. The findings have clinical implications for protocol administration with thought-disordered populations that typically produce brief records.

For years, Rorschach researchers and practitioners have debated whether test interpretation should be modified according to the length of the record, or the number of responses (*R*; Exner, 1995). Concerns about the validity of brief records and lengthy records have been expressed (see Exner, 1995, for a review), but no recent study has addressed the specific relationship between protocol length and the validity of the Rorschach in assessing thought disorder. Clinical and forensic settings frequently yield impoverished protocols that make test interpretation difficult (e.g., Auslander, 2000; Viglione, Perry, & Meyer, 2003), and there are data suggesting that brief records are less informative and reliable than longer records (Exner, 1993). Excessively long protocols also are problematic, resulting in skewed data and potential misinterpretation. Accordingly, it is worthwhile to examine whether restricting the response range by decreasing the number of excessively short and long protocols will maintain or even improve the interpretive yield of the test.

PROTOCOL LENGTH

Exner's (1993) former Comprehensive System (CS) normative reference group produced a mean *R* of 22.67 ($SD = 4.23$), while Exner and Erdberg's (2005) contemporary sample of 450 nonpatients produces a mean *R* of 23.36 ($SD = 5.68$). Another normative approximation sample from Shaffer, Erdberg, and Haroian (1999) had a greater proportion of brief records and also greater variability in *R*, with a mean of 20.81 and standard deviation of 7.47. Many patient populations provide an even lower frequency of responses with a greater variance. For instance, among late-life schizophrenic patients Auslander (2000) reported the mean *R* for males was 20.18 ($SD = 8.25$) and for females was 15.59 ($SD = 5.81$). In the data set utilized by Viglione, Perry, and Meyer (2003) to revise the Ego Impairment Index, 26 of the 389 records contained fewer than 14 responses. The authors reported that the majority of those records came from

individuals with schizophrenia. Response productivity is particularly important in clinical populations that often produce fewer responses than nonpatient populations.

The lack of response production in clinical populations could reduce the interpretive yield of the test and compromise test validity. Viglione (1999) and Meyer (1993) have reported that protocol length can affect the false negative and false positive rates of scores assessing thought disturbance, particularly the Schizophrenia Index (SCZI). Short protocols are deemed to have low sensitivity and are susceptible to high false negative rates, particularly among records that are impoverished or form dominant. Studying the association between response productivity and variables associated with thought disturbance, including the Perceptual Thinking Index (PTI) and Ego Impairment Index (EII), could provide information about decreasing the false negative rates in clinical samples.

RESPONSE COMPLEXITY

Response complexity is associated with cognitive flexibility, sophistication, and problem-solving skills (Viglione, 1999). Research indicates that the complexity of the protocol may influence interpretative inferences and predictive yield (Mcguire, Kinder, Curtiss, & Viglione, 1995), such that predictive power increases when complexity is considered. For example, Kates (1994) found that the EII (Perry & Viglione, 1991) was related to behavioral problems among children with less complex records. The finding did not hold true for the more complex, nonconstricted group. Another study found that guardedness related to decreased complexity was negatively correlated with sexual responding (Morgan & Viglione, 1992). Thus, the differential findings of complex respondents compared with less complex respondents suggest that the predictive ability of the Rorschach may be moderated by response complexity.

PURPOSE OF STUDY

There has been a call by many authors to scientifically examine the effects of protocol length and complexity on Rorschach validity (Kinder, 1992; Lipgar, 1992; Meyer, 1992; Viglione, 1996). The aim of the present study was to examine the effectiveness of an alternative administration method for constraining response production. The alternative administration guidelines were designed to preemptively increase *R* in records that were likely to be overly short and decrease *R* in records that were likely to be overly long. When applied to our sample, which was selected because it was a group likely to produce a substantial number of brief records, we hypothesized that providing respondents with extra response prompts would increase protocol length, resulting in greater interpretive yield, while maintaining the validity of the CS. Specifically, in this sample we expected that the alternative administration method would produce more protocols with at least 18 responses during the initial response phase and also produce less variability in *R* (i.e., a smaller standard deviation and variance). At the same time, we expected that the Rorschach scores in both the standard CS administration and alternative administration conditions would be significantly associated with criterion measures of thought disorder. We also expected that validity for predicting thought disorder criteria would be maximized with protocols containing 18 to 28 responses.[1] Our final hypothesis predicted that when complexity was considered as a moderator variable in exploratory analyses, the association between Rorschach variables and thought disorder scores would improve.

METHOD

Participants

Participants included 61 adults, ages 18 to 66, with a mean age of 37 ($SD = 9.9$) and an average of 11 years of education. Sixty-four percent were male. All were in long-term residential treatment at either a state psychiatric facility or a state prison, and carried an Axis I diagnosis in accordance with the *Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition* (*DSM-IV*; American Psychiatric Association [APA], 1994), as diagnosed by the treating psychiatrist. The participants were randomly assigned to one of two groups. The standard administration group ($n = 30$) received the traditional CS administration, and the experimental group ($n = 31$) received an alternative administration. There were no differences between the two groups in mean age (standard group $M = 36.58$, $SD = 9.79$; alternative group $M = 37.17$, $SD = 9.99$; Cohen's $d = .06$) or education (standard group $M = 11.58$, $SD = 2.32$; alternative group $M = 10.70$, $SD = 2.14$; Cohen's $d = .40$). There also were no differences in the proportion of participants who were ethnic minorities, had psychotic disorder diagnoses, or were married (see Table 1). Written informed consent was obtained for all participants, and the institutional review board (IRB) committees at the Utah State Hospital and North Carolina Department of Correction approved the study.

[1] This range was used for two reasons, both based on Exner and Erdberg's (2005) reference sample of 450 adults. First, in this sample the median value for R is 23.0, the *SD* is 5.68, and the interquartile range (IQR), a measure of dispersion or variability that is not influenced by outliers, is 5.0. Thus, the range from 18 to 28 roughly corresponds to the median $\pm$ 1 *SD* and it exactly corresponds to the median $\pm$ 1 IQR. Second, in this reference sample the cut-off values roughly separate the lowest 10% and highest 10% of the distribution from the central 80%. Specifically, the range from 14 to 17 includes 10.2% of the sample ($n = 46$) and the range from 29 and up includes 12.2% of the sample ($n = 55$), leaving the central range from 18 to 28 to include 77.6% of the sample.

TABLE 1
Frequencies and Group Comparisons on Demographic Variables

Category (Frequency)	*Alternative (n = 31)*	*Standard (n = 30)*	*Chi Square*	*p Value*	*r*
Gender ($N = 61$)			.40	.53	.08
Male	21	18			
Female	10	12			
Marital Status			.17	.68	.05
Married	4	5			
Other (Single/Divorced/Separated)	27 (22/5/0)	25 (14/4/7)			
Diagnosis			1.05	.31	.13
Psychotic Disorder (i.e., Schizophrenia/ Schizoaffective/Delusional/Psychosis NOS)	18 (16/1/1/0)	13 (10/1/1/1)			
Affective or Other Disorder (i.e., Adjustment/Anxiety/ Bipolar/Depression/Mood NOS/Substance Abuse)	13 (2/0/5/5/0/1)	16 (4/1/4/6/1/1)			
Ethnicity			1.31	.25	.15
Caucasian	19	14			
Other (African American/Latino/Native American)	12 (11/0/1)	16 (14/1/1)			

Rorschach Thought Disorder Variables

Three Rorschach variables were used to measure thought disorder, including the SCZI, the PTI, and the EII-2. Although these three variables overlap because their formulas contain some of the same CS variables, they were used because each might provide unique information.

Schizophrenia Index. The original SCZI correctly identified schizophrenic respondents with 72% to 89% accuracy, depending on the sample (Exner, 1993), with inter-rater agreement between 92% and 99% (Kleiger, 1999). The initially reported false positive rates were around 12%, although rates for the revised SCZI have been reported as between 0% and 11% (Exner, 1993).

Perceptual Thinking Index. The Perceptual Thinking Index (PTI) is the latest revision of the SCZI, and it was developed in an effort to improve its validity in detecting cognitive impairments and to reduce overly high scores among children (Exner, 2000). Two new variables are included in its formula: XA% (Form Appropriate), which indicates the percent of responses with reasonable form and is computed as the sum of FQ+, FQo, and FQu divided by *R*; and WDA% (Form Appropriate—Common Areas), which indicates the percent of appropriate form responses given to common detail areas and is calculated as the sum of FQ+, FQo, and FQu responses to W and D locations divided by the number of W and D responses (Exner, 2001). The PTI also includes age adjustments for *R* in those age 13 years and younger.

Previous findings indicated the SCZI and PTI have a similar distribution of scores. In an initial comparison of 110 individuals diagnosed with schizophrenia, 62 had SCZI values of 5 or 6, while 61 had PTI scores of 4 or 5 (Exner, 2000). One study found the PTI was slightly superior to the SCZI in differentiating child and adolescent groups, however, suggesting the PTI may be more valid for these populations (Smith, Baity, Knowles, & Hilsenroth, 2001). More recently, Dao and Prevatt (2006) examined the effectiveness of the PTI to distinguish adult inpatients diagnosed with a psychotic disorder versus those with a mood disorder and no psychotic features. The authors reported good internal consistency among the PTI criteria (*KR-20* = .75) and significant differences between the two groups on total PTI scores (Cohen's $d = 1.62$).

Ego Impairment Index-2. The Ego Impairment Index (EII: Perry & Viglione, 1991) initially was designed to assess general level of psychological disturbance and later applied to measure thought disturbance more specifically. The EII has predicted treatment outcome, level of adjustment, and thought disorder in various patient populations with disorders ranging from depression to schizophrenia (Perry, McDougall, & Viglione, 1995; Perry, Minassian, Cadenhead, Sprock, & Braff, 2003; Perry & Viglione, 1991; Perry, Viglione, & Braff, 1992). Specifically, in a sample of depressed patients, baseline EII scores demonstrated incremental validity in predicting treatment outcome ($R^2 = .20$) beyond self-report measures (Perry & Viglione, 1991). Although originally derived and validated on a sample of depressed patients, the EII has a stable factor structure (e.g., factor scores computed with schizophrenia patients correlated highly [$r = .98$] with the original factor; Perry et al., 1992) and can be reliably scored, producing kappa coefficients between .88 and .97 (Perry & Braff, 1994).

The EII-2, which is a slightly modified version of the original EII (Viglione, Perry, & Meyer, 2003), consists of form quality minus responses (FQ-), WSum6, critical contents (including anatomy, blood, explosions, fire, food, sex, x-ray, aggressive, and morbid responses), poor human movement responses (M-), poor human content (Poor HRV), good human content (Good HRV), and the number of responses (*R*). The good and poor human contents were originally from the Human Experience Variable, which was modified and now

is referred to as the Human Representational Variable (HRV; Viglione, Perry, Jansak, Meyer, & Exner, 2003). Developed by Perry and Viglione (1991), the HRV looks at the quality of human content responses (Exner, 2001).

Thought Disorder Criterion Measures

Scale for the Assessment of Thought, Language, and Communication (TLC). The TLC (Andreasen, 1984) rating scale uses a semistructured interview and observational data to assess disordered thinking and communication. Several studies have demonstrated the TLC is a valid measure of thought disorder, differentiating patient populations and yielding strong inter-rater score reliability (Andreasen, 1979, 1986; Cuesta & Peralta, 1993; Harvey, 1983). Examiners rate patients on subtypes of thought disorder on a 4- or 5- point Likert scale. The ratings are given subsequent to a semistructured interview designed to elicit verbalizations from the respondent that do not focus on their pathology. The patient interview was designed to last approximately 45 minutes (Andreasen, 1979); however, shorter interview times have been utilized while successfully ascertaining a thought disorder (Docherty, Schnur, & Harvey, 1988; Harvey & Serper, 1990; McPherson & Harvey, 1996). The present study shortened the interview to 20 minutes and followed Harvey and colleagues' (1984) suggestion to modify the frequency criterion of the TLC. The number of test items was also shortened to 10 to reflect two main factors, Verbal Productivity and Disconnection, which were developed by Harvey and colleagues (1992) through exploratory factor analysis and subsequently used by Auslander and colleagues (2002) in detecting thought disorder in late-life schizophrenia patients. The semistructured interview created by the Neuropsychology Core of the Geropsychiatry Clinical Research Center (GPCRC) at the University of California, San Diego, was utilized. All interviews were audio recorded and coded blind to ensure accuracy, as suggested by Andreasen (1979).

Magical Ideation Scale (MIS). Eckblad and Chapman (1983) developed the 30-item self-report Magical Ideation Scale based on Meehl's assertion that magical ideation is a symptom of schizotypy or schizophrenia proneness. The MIS has differentiated college students with and without a proneness to psychosis (Eckblad and Chapman, 1983) and has demonstrated stability in predicting symptoms and level of functioning 10 years later (Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994). The MIS is correlated with other measures of thought disorder (Chapman, Chapman, & Miller, 1982), and high scores are associated with psychotic symptoms (Chapman et al., 1994). Convergent and discriminant validity have been demonstrated among personality disordered patient populations, with the MIS correlating highly with schizotypal traits ($r = .68$) but not with other nonpsychotic traits (Bailey, West, Widiger, & Freiman, 1993). The scale successfully has differentiated schizophrenic and affective disorder patients from normal control subjects, accounting for 26% of the predicted variance among psychotic patients (George & Neufeld, 1987).

Scale for the Assessment of Positive Symptoms (SAPS). Andreason (1986) developed this 34-item scale that includes four domains: hallucinations, delusions, bizarre behavior, and positive formal thought disorder. The present study used the delusional subscale to assess content that may not be captured by the TLC. The following types of delusions were scored: persecutory, jealousy, sin or guilt, grandiose, religious, somatic, ideas of reference, being controlled, mind reading, thought broadcasting, thought insertion, and thought withdrawal. The scores from each item were summed for a total composite score. The items were scored from participant verbalizations during the 20-minute, audio-recorded TLC structured interview.

The SAPS has demonstrated good reliability and validity in assessing thought disturbance in a variety of diagnostic groups. One study of inter-rater reliability on a number of measures of positive and negative symptoms in a schizophrenic sample indicated that the SAPS achieved the highest intraclass correlation coefficient (ICC) at .80 (Peralta, Cuesta, & de Leon, 1995). In another study, kappa coefficients were calculated to measure inter-rater agreement. Among a sample of manic and schizophrenic patients, inter-rater agreement reached $\kappa = .84$ (Walker, Harvey, & Perlman, 1988). The SAPS also has demonstrated good criterion validity with various diagnostic groups, including schizophrenic and schizotypal groups, and nonpatients (Auslander, 2000).

Exploratory Variable

Complexity is defined as "the amount of productivity, precision, differentiation, and integration involved in the aggregate of all the responses" (Viglione, 1999). The Complexity Index provides information about the cognitive flexibility, investment, and problem-solving ability of respondents, and is believed to provide a general assessment of one's cognitive flexibility and motivation (Mcguire et al., 1995; Viglione, 1999). The index is based on location/developmental quality, determinants, and content complexity, and is calculated for each Rorschach response and then summed for a total composite score (Viglione & Meyer, 1998; see Appendix A for the complete formula). Researchers have suggested that variables such as the SCZI will be more accurate if complexity is considered (Mcguire et al., 1995). The Complexity Index was calculated in exploratory analyses to determine its effect on moderating the relationship between criterion measures and Rorschach variables in assessing thought disorder. Because complexity is associated with the number of responses, increased accuracy was expected in protocols containing between 18 and 28 responses.

Procedure

Participants were assigned randomly to one of two conditions. In the standard administration condition, examiners administered the Rorschach using the regular CS procedures that were in place at the start of the study (Exner, 1993), which was prior to the slightly modified guidelines provided in Exner (2001). Specifically, if necessary participants were given a single prompt for more responses on Card I only.

In the experimental condition examiners prompted participants for another response whenever only one response was provided to a card. Examiners gave this prompt up to three times, if necessary, except to cards V and IX. Card V contains well-structured, simplistic percepts that often do not elicit multiple responses, while Card IX is the most rejected card and produces the fewest number of popular responses (Weiner, 2003). A total of three prompts were allowed during administration to produce extra responses. Even if participants provided only one response to a card after the three prompts were offered, no additional prompts were given in order to simplify the administration instructions and to minimize multiple prompts from the examiner.

The prompts themselves were identical to the initial prompt outlined by Exner (2001), except the conditions were changed; if the participant provided only one response to any card, not just the first card, the examiner said, "Take your time and look some more. I'm sure you'll find something else too" (p. 6). The prompt was used regardless of how many responses were provided to the first card, any time the respondent provided only one response to a card.

The alternative administration condition also attempted to reduce lengthy protocols by allowing only four responses per card. After the respondent provided the fourth response to any card, the examiner removed the card. This technique to reduce unusually long protocols has been discussed by the Rorschach Research Council (October 1999) as a method of narrowing the range of responses to increase test utility.

Regardless of the condition, a minimum of 14 responses was required. Exner (1993) has indicated that protocols of 13 responses or fewer are not interpretively valid. If participants gave fewer than 14 responses, they were retested according to standard CS procedures. All participants ultimately produced valid records of 14 or more responses.

Following the Rorschach administration, and during a separate session, participants were interviewed for the TLC and the SAPS, and then completed the MIS.

Examiner Variables

The first author administered 56 of the 61 Rorschach protocols and all of the criterion measures. The author is well trained in Rorschach administration and scoring utilizing the CS. The other five Rorschach protocols were administered by a supervising clinical psychologist and a graduate intern, both of whom were trained in the CS. To decrease potential bias, all measures were scored blind, without identifying information and after all measures were administered. Interscorer reliability for the CS was analyzed on 20 randomly selected protocols by comparing scores between the initial administrator and an independent scorer. The inter-rater scorers were graduate research assistants, who were also well trained in the CS and received supervision from the second author. One-way random effects intraclass correlation correlations (ICC) for absolute agreement of a single rater were calculated for the primary Rorschach variables under investigation. The protocol-level summary scores for the PTI, SCZI, EII–2, Complexity, and *R* had ICC scores of .70, .67, .78, .92, and .99, respectively. The difference in mean scores on the SCZI, PTI, and EII-2 between the first author and the inter-rater scorers was not statistically significant (all t scores $<$ 1.0; Cohen's $d = .24$, .05, and .29, respectively).

The semistructured interview for the TLC and SAPS was transcribed from audiotape and then later scored without identifying information by the administrator. An independent scorer rated 22 randomly selected interviews. Inter-rater reliability was calculated by comparing the scores of the two independent judges. The ICC's for the TLC and SAPS total scores were .95 and .89, respectively, and the two scorers did not differ in their mean values (ts < 1.0; Cohen's $d = .22$ and .24, respectively).

RESULTS

Group Comparisons

The two groups were compared for differences in scores on the thought disorder criterion measures using one-way ANOVAs. The results revealed no significant differences (all ps $> .17$), with similar mean scores on the TLC, the MIS, and the delusional subscale of the SAPS (see Table 2). The groups also were compared on the predictor measures of thought disorder, which included the three Rorschach variable clusters. One-way ANOVAs revealed no significant differences between the groups on the SCZI, PTI, or EII–2 scores (all ps $> .49$).

The distributions of the predictor and criterion variables in each group were examined, as were statistical assumptions. The variable scatterplots of the standardized predicted and observed residuals revealed one EII–2 outlier, with an otherwise linear relationship between variables. To maintain a normal distribution, the score was changed to .01 points above the next highest EII–2 score, from 7.10 to 3.50. The resulting distribution was sufficiently normal (skewness = .860, kurtosis = 1.137). The TLC and SAPS distributions were leptokurtotic (kurtosis = 5.833 and 3.397, respectively), but for the purpose of regression analyses a standardized thought disorder score was calculated. The Thought Disorder Summary Scale (TDSz) was determined by transforming the raw total scores from each of the criteria measures (the TLC, MIS,

TABLE 2
Group Comparisons of Mean, Standard Deviation, Range, and Distribution on Criterion Measures and Rorschach Variables

		Mean	*SD*	*Cohen's d*	*Median*	*Minimum*	*Maximum*	*Skew*	*Kurtosis*
MIS	Alt	11.23	6.14	0.36	12	1	23	0.033	−0.293
	Std	13.47	6.64		14	2	26	0.127	−0.735
TLC	Alt	6.48	6.96	0.02	5.0	0	29	2.291	5.833
	Std	6.60	6.88		4.5	0	29	1.686	2.894
SAPS	Alt	2.23	3.61	0.12	0	0	13	1.633	1.926
	Std	1.80	3.93		0	0	13	2.148	3.397
TDSz	Alt	−.205	3.53	0.12	−0.90	−4.87	10.55	1.64	3.072
	Std	.212	3.62		−1.13	−5.03	10.39	1.43	1.619
EII–2	Alt	.242	1.18	0.10	.020	−1.65	3.50	1.088	1.274
	Std	.358	1.17		.315	−2.24	3.49	0.676	1.542
*EII–2	Alt	.358	1.61	0.00	.020	−1.65	7.10	2.665	9.879
	Std	.357	1.17		.315	−2.24	3.49	0.676	1.542
EII	Alt	.547	1.31	0.02	.1900	−1.15	4.34	1.625	3.006
	Std	.566	1.24		.5450	−1.48	4.18	1.000	1.551
SCZI	Alt	2.10	1.45	0.16	2	0	5	.457	−0.338
	Std	2.33	1.45		2	0	5	.393	−1.203
PTI	Alt	1.06	1.21	0.18	1	0	4	1.078	0.428
	Std	1.30	1.49		1	0	4	0.653	−1.180
R	Alt	21.77	6.95	0.34	20.00	14	40	1.249	0.856
	Std	19.37	7.43		16.50	14	45	2.104	4.687
Complexity	Alt	2.6509	.6630	0.18	2.6667	1.57	4.0	.352	−.293
	Std	2.7787	.7871		2.5479	1.11	4.36	.389	−.057

Note: SD = standard deviation; Alt = Alternative administration group; Std = Standard administration group; MIS = Magical Ideation Scale; TLC = Scale for the Assessment of Thought, Language, and Communication; SAPS = Scale for the Assessment of Positive Symptoms, Delusional Subscale; TDSz = Thought Disorder Sum of z scores; EII–2 = Revised Ego Impairment Index with outlier changed; *EII–2 = Revised Ego Impairment Index with outlier retained; EII = Original Ego Impairment Index with outlier changed; SCZI = Schizophrenia Index; PTI = Perceptual Thinking Index; *R* = Number of Responses; Complexity = Complexity Composite Score.

and SAPS) into z scores. The z scores were then summed to produce a total thought disorder score. Variable distributions and effect size differences are presented in Table 2.

Number of Responses and Prompts by Group

The average number of responses produced in the overall sample was 20.59, with a mode of 18. Almost half of the protocols, 46%, contained fewer than 18 responses. During the initial response phase before any necessary readministration for $R < 14$, the standard administration group averaged 18.73 responses, with a mode of 12. The alternative administration group averaged 21.97 responses, with a mode of 17.

Card removal due to more than four and five responses was rarely necessary, occurring only three times in the alternative group and once in the standard group. The response range in the standard group was slightly larger (14 to 45) than the alternative group (14 to 40). Standard CS rules for protocol readministration were followed in both groups when respondents produced fewer than 14 responses during the initial response phase. Readministration was necessary for 7 participants in the standard group but no participants in the alternative group.

Just over half of all participants, 54%, provided only one response to the first card and were prompted for another response. As expected, the two groups were very similar in the number of Card I prompts; 15 participants in the standard group and 18 in the alternative group provided only one initial response. In the alternative group, only 4 participants received no prompts, and more than one prompt was offered in 22 of the protocols. In examining total prompts per record, three prompts were offered most often ($n = 13$ protocols), followed by two ($n = 8$), then one prompt ($n = 6$). Prompts were offered most frequently to Card I, which was prompted in 18 of 31 protocols, followed by Card IV ($n = 12$), Card II ($n = 10$), and Card III ($n = 7$).

Primary Hypotheses

The first hypothesis predicted that using an alternative administration method would increase the number of protocols with greater than 17 responses after the initial response phase. A nonparametric 2 × 2 Pearson chi-square analysis with a one-tailed Fisher's Exact Test indicated this did not reach statistical significance, $\chi^2(1, N = 61) = 2.755$, $p = .080$; $r = .18$. An examination of frequencies, however, indicated a trend in this direction. The alternative group had 20 protocols of 18 or more responses, versus 11 with fewer than 18, whereas the standard group had 13 protocols with 18 or more responses and 17 records with fewer than 18. Power analyses indicated that to detect a medium-sized difference, 60

TABLE 3
Correlations Between Rorschach Predictors and Thought Disorder Measures Across All Participants ($N = 61$)

Scale	*TLC*	*SAPS*	*TDSz*	*EII-2*	*SCZI*	*PTI*
MIS	.14	.25*	.42**	.32*	.31*	.35**
TLC		.75***	.91***	.37**	.27*	.27*
SAPS			.84***	.46***	.35**	.32*
TDSz				.47***	.40**	.38**
EII–2					.82***	.79***
SCZI						.90***

Note. MIS = Magical Ideation Scale; TLC = Scale for the Assessment of Thought, Language, and Communication; SAPS = Scale for the Assessment of Positive Symptoms; TDSz = Thought Disorder Sum of z scores; EII–2 = Ego Impairment Index with outlier changed; SCZI = Schizophrenia Index; PTI = Perceptual Thinking Index.
2-tailed significance levels: $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

TABLE 4
Sample Specific Correlations Between Rorschach Predictors and the Thought Disorder Summary Scale (TDSz) and the Magnitude of Difference

Scale	*All ($N = 61$)*	*Alternative ($n = 31$)*	*Standard ($n = 30$)*	*Difference z*
EII–2	.47***	.56**	.38*	0.86
SCZI	.40**	.53**	.26	1.20
PTI	.38**	.44**	.33*	0.48

Note. Alternative = Alternative administration group; Standard = Standard administration group; EII–2 = Ego Impairment Index with outlier changed; SCZI = Schizophrenia Index; PTI = Perceptual Thinking Index.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

participants per condition would be required for a 2-tailed alpha = .05 and approximately 35 participants per condition for a 1-tailed alpha = .05 (Cohen, 1988). It also was predicted that variability in the alternative group would be significantly lower than in the standard group. The variance in the standard group was 63.52 compared with 48.30 in the alternative group. Although in the expected direction, the difference was not statistically significant, $F(29, 30) = 1.32$, $p = .230$.

The second set of hypotheses predicted that each of the Rorschach variables would be associated with the criterion thought disorder scores in both the standard and alternative groups. To highlight the relationship between each criterion measure with the Rorschach predictors, Pearson correlations are presented in Table 3. These analyses indicated significant correlations between the criterion measures, the MIS, TLC, and SAPS, and all three Rorschach predictors, the SCZI, PTI, and EII–2. Linear regressions demonstrated that the three Rorschach variables were significant predictors of the composite thought disorder summary scale (TDSz) for the entire sample, $F(3, 60) = 5.468$, $p = .002$, Adjusted $R^2 = .183$, Adjusted $R = .428$. When the groups were analyzed separately, the Rorschach predictors did not reach statistical significance in the standard group, $F(3, 29) = 1.772$, $p = .177$, Adjusted $R^2 = .074$, Adjusted $R = .272$ but were significant in the alternative group, $F(3, 30) = 4.367$, $p = .010$, Adjusted $R^2 = .267$, Adjusted $R = .517$.

Each Rorschach thought disorder variable was analyzed separately, as each was predicted to be associated with the composite thought disorder scale in both groups (see Table 4). The SCZI was not significantly correlated with the total thought disorder score in the standard group, but it was significantly correlated in the alternative group. The PTI and EII-2 were significantly associated in both the standard and alternative groups. The magnitude of the difference between the correlations in each group was tested but did not reach significance (Table 4). Power analyses indicated that to have 80% power to detect a modest difference (e.g., $q = .25$) in the magnitude of validity coefficients for the EII–2, PTI, or SCZI across the two conditions, we would need approximately 300 participants in each condition for a 2-tailed alpha = .05 and about 200 participants per condition for a 1-tailed alpha (Cohen, 1988).

The final primary hypothesis predicted that in a regression in which all three Rorschach predictors were entered, the residuals, or degree of error remaining in the prediction, would correlate with the number of responses, with less error and therefore greater accuracy in the 18- to 28-response range. The obtained correlation between the residuals and R itself did not reach statistical significance ($r = .235$, $p = .068$). The mean residual scores from linear regression analyses of the three Rorschach predictors and total thought disorder score were compared between protocols containing18 to 28 responses and protocols outside that range. ANOVA revealed no significant difference in mean residual scores between the groups, $F(1, 60) = 1.874$, $p = .176$, $d = .35$.

Supplementary Analyses

Complexity. The exploratory hypothesis predicted that the Complexity Index would contribute significantly to the predictions of the total thought disorder score. Regression analyses were examined in the entire sample, with each of the three Rorschach predictors individually entered in the initial block and the complexity score added in a second block to assess for added predictive power. Results are presented in Table 5.

With the SCZI predicting thought disorder, the Complexity beta value of .225 did not reach significance ($p = .060$) and the Adjusted R value increased just slightly, from .396 to .455. When the PTI was the sole predictor, the complexity beta weight was .253 ($p = .034$), and it contributed significantly to the prediction, increasing the Adjusted R from .381 to .457. Finally, when the EII–2 was the predictor, complexity did not contribute to the prediction, with a beta weight of .163 ($p = .168$) that only slightly increased the Adjusted

TABLE 5
Regression Analyses of Complexity Contribution to Rorschach Variables in Predicting Thought Disorder Summary Scale (TDSz) Across All Participants ($N = 61$)

	Full Model			*Individual Predictors*			
Model	*R*	*Adjusted R^2*	*Std. Error of the Estimate*	*R^2 Change*	*F Change*	*df*	*p Value*
SCZI	.396	.143	3.2900	.157	10.982	1, 59	.002
Complexity	.455	.180	3.2175	.050	3.689	1, 58	.060
PTI	.381	.131	3.3127	.145	10.029	1, 59	.002
Complexity	.457	.182	3.2136	.064	4.693	1, 58	.034
EII–2	.521	.259	3.0580	.272	22.006	1, 59	.001
Complexity	.539	.266	3.0447	.019	1.517	1, 58	.223

Note. EII–2 = Ego Impairment Index with outlier changed; SCZI = Schizophrenia Index; PTI = Perceptual Thinking Index.

R from .472 to .498. These analyses were repeated with the original outlier, which had been changed from 7.10 to 3.50 in an effort to maintain normality. With the outlying score unchanged, the EII–2 complexity beta weight was .140 ($p = .223$) and the Adjusted *R* changed from .521 to .539.

Regression analyses were conducted for protocols in the range of 18 to 28 responses, and protocols outside that range. All three Rorschach predictors were in the first block with complexity added in the second block. In the "optimal" 18 to 28 range, the Complexity Index did not significantly contribute to the prediction, $F(1, 20) = .876$, $p = .361$. For protocols outside that range, however, complexity did significantly improve the prediction, $F(1, 31) = 4.357$, $p = .045$, with an *R* change value of .285. Thus, as anticipated, adjusting for protocol complexity improves prediction when protocols are in the nonoptimal range of *R*.

Protocol readministration. Following the administration of the Rorschach protocols, it became apparent that in the alternative administration group there was never a need to readminister the test due to obtaining fewer than 14 responses during the response phase. There were seven readministrations, however, in the standard administration group. Post-hoc chi square analysis with a Fisher's Exact Test indicated the difference was significant, $\chi^2(1, N = 61) = 8.171$, $p = .005$, and associated with a medium effect size ($r = .33$, computed from *p*).

DISCUSSION

The purpose of the present study was to examine an alternative method for constraining Rorschach response productivity while maintaining the integrity and validity of the CS. To increase *R* in protocols likely to be short the examiner prompted for an extra response when only one was offered on a card, and to decrease *R* in protocols likely to be long the examiner removed the card after four responses. This alternative administration was studied experimentally in a residential sample of impaired patients who were expected to produce a high frequency of brief records, with validity evaluated against external criteria of thought disorder.

The alternative administration method resulted in a nonsignificant trend ($r = .18$) for an increase in protocols with 18 or more responses, an effect that may have reached significance with a larger sample. Supplemental analyses showed, however, that the alternative administration procedures significantly reduced administration time by obviating the need for protocol readministration due to an initial $R < 14$ ($r = .33$). In addition, although the overall sample demonstrated significant associations between Rorschach variables and thought disorder criterion scores, as expected, validity was enhanced in the alternative administration group and weaker in the standard administration group. The hypothesis that protocols with 18–28 responses would have less error of prediction than protocols outside that range was not supported. However, exploratory analyses showed a moderating role for protocol complexity. The Complexity Index enhanced overall predictions for the PTI, with trends in the expected direction for the SCZI and EII–2. Perhaps more importantly, however, adjusting for protocol complexity improved prediction for protocols that were in a nonoptimal range on *R* (i.e., < 18 or > 28), and the moderating adjustment was not useful (or necessary) for protocols with 18–28 responses.

Despite the guidelines in the CS (Exner, 2001, 2003) to address impoverished protocols, brief records in inpatient and forensic settings are not uncommon and often yield little useful information (e.g., Viglione, 1999; Viglione, Perry, & Meyer, 2003). Sensitivity is weakened, increasing the chance of false negative findings. For example, individuals with a known history of thought disorder who provide short and simplistic records are less likely to produce the expected elevations on the PTI, the SCZI, or the EII–2. Despite the problem of brief records, because of a lack of research support examiners have been reluctant to encourage extra responses for fear of introducing examiner bias and producing records that are not representative of the respondent's true internal process. Based on the results of the present study, however, it is recommended that when administering the Rorschach to an individual at risk for producing a brief record, prompting for another response whenever only one is spontaneously offered will improve test validity and reduce the need to readminister the test due to fewer than 14 responses. Issues relevant to administration and test validity are addressed more completely below.

Administration

In this sample the standard administration group produced brief records overall, with more than half of participants

producing $R < 18$ records. This finding is consistent with those observed in many inpatient settings (e.g., Auslander, 2000) and spoken of among clinicians but not thoroughly researched. Although the CS workbook (Exner, 2001) provides guidelines to reduce brief records among resistant responders, these guidelines are not entirely effective, perhaps because not all brief records are produced out of resistance. Individuals with cognitive deficits, mental illness, or developmental delays produce brief records (Klopfer & Davidson, 1962; Wagner, Young, & Wagner, 1992), and administration guidelines may be an important factor in their response productivity. Individuals who are less savvy to testing situations or less cognitively flexible may not pick up on subtle cues or even the one or two direct prompts offered in the CS (Exner, 2001). Among the standard group participants in this study, many did not appear to extrapolate the prompt on Card I to the other cards. For instance, when no prompt was offered on Card II after receiving a prompt to Card I, participants may have failed to realize that another response was desired. Our alternate administration method allows examiners to offer minor encouragement when needed to respondents who are likely to produce brief records. In the alternative administration group it appeared that receiving these additional prompts provided the extra cues they needed to produce longer protocols.

Response prompting. Results suggest that prompting during the response phase reduced the frequency of brief protocols. Of the 31 protocols in the alternative administration group, 21 were given more than one prompt. On average these respondents produced 3.24 more responses during the initial response phase, with a mode of 17 responses, compared with a mode of 12 in the standard group. The number of responses produced during the initial response phase in the standard group again reflects what clinicians in inpatient settings experience; many protocols contain an extra response to the now standard prompts offered to Cards I and IV, but they are otherwise "one response per card" records. This pattern of responding frequently necessitates readministration, which was no longer necessary using our modified guidelines. We also observed a trend toward obtaining more productive final records with the alternative administration method. Given the limited sample size, this trend may become significant in larger samples.

Readministration. As already noted, the alternative administration method eliminated the need for protocol readministration (effect size $r = .33$). In accordance with CS procedures, all respondents, regardless of group, were readministered the test if they produced fewer than 14 responses during the response phase. Those that received the alternative administration never experienced readministration, however, while 23% of participants in the standard group did. Providing extra prompts in the alternative procedure resulted in more efficient administration. Anecdotally, clinicians complain about readministration procedures, citing the length of time it requires, the frustration and confusion that both the examiner and the respondent experience, and the questionable validity of results when the response phase is repeated and the original responses are not offered again and thus are not used in scoring. One common reaction, even among cooperative respondents, is either not repeating their original percepts or reporting them in an abbreviated fashion. The respondent may think that because the examiner already heard the first set of responses, they either do not need to be repeated, or they can be referred to without the detail provided during the initial response phase. For example, during the initial administration a respondent may state, "Two people eating dinner." During readministration she may say, "Oh yeah, and those people." The examiner is left uncertain whether to include the descriptive content from the initial response phase or to rely solely on the verbatim statement made during the readministrated response phase.

Maintaining Comprehensive System Validity

The present sample included participants in residential psychiatric treatment in order to capture a range of scores on the thought disorder measures. As outlined in Table 2, scores on the PTI ranged from 0 to 4, scores on the SCZI ranged from 0 to 5, and scores on the EII-2 ranged from –2.24 to 7.10 (before the transformation), indicating a broad range of thought disturbance. Among those diagnosed with a psychotic disorder, the mean scores on the PTI, SCZI, and EII-2 were 1.48, 2.55, and .46, respectively, while mean scores for participants diagnosed with a mood disorder were .87, 1.87, and .13, respectively. Obtaining a broad spectrum of thought disorder was necessary to demonstrate that the test is consistent with other measures of thought disturbance both when impairment is significant and when it is mild or moderate, and results once again supported the validity of the Rorschach as a measure of thought disorder.

We hypothesized that Rorschach variables would effectively predict thought disorder scores, as measured by communication style, delusional content in speech, and participant self-report of magical ideation, regardless of the level of impairment, in both the standard and alternative administration groups. Indeed, in the overall sample the combination of Rorschach variables was significantly associated with thought disorder scores (Adjusted $R = .43$), supporting the well-known validity of the test as a measure of thought disorder. When analyzed separately, however, the two groups differed. In the standard group, the three Rorschach variables did not reach statistical significance in predicting thought disorder scores (Adjusted $R = .27$), but in the alternative administration group, despite the same sample size, the three variables were significant, with a large effect size (Adjusted $R = .52$).

Because the Rorschach predictor variables underperformed in the standard group, it strengthens the argument for using the alternative method of administration. The differing results suggest that the Rorschach's utility improved when more responses were obtained. The strong correlations and significant predictions in the alternative group also are inconsistent with the argument that prompting for additional responses may produce responses that are less meaningful or representative of the respondent. To the contrary, these results suggest that the prompting examiner is not imposing on the test process, but rather that this guidance helps respondents to effectively produce more responses that are representative of their internal state, which in this case related to thought disorder indicators. The additional data obtained by eliciting extra responses enhanced the validity of the test. Providing more responses is tantamount to administering a longer test with more items to increase reliability. Indeed, Exner (2003) has demonstrated that brief protocols have less reliability.

The goal of the alternate procedure is to maximize the extent to which the normative CS values will apply to all collected protocols. Exner's reference samples generally have smaller standard deviations for *R* than found in other reference samples (e.g., Shaffer, Erdberg, & Haroian, 1999). The modified administration procedures will help ensure a more cohesive range on *R*. Although the participants in this study were patients, the fact that the *SD* for *R* in the alternative group is closer to the *SD* in Exner's reference sample than the standard group means that the normative reference values are more applicable to that group.

Response process. An examination of the response process is necessary to address the theoretical issues involved in response validity when prompting for extra responses. It has been argued that such prompting imposes structure on the test and alters the situation, which then distorts the results because the respondent will generate an invalid response. Such arguments against providing additional prompts may be based on studies demonstrating the ability of the examiner to influence the type of responses given by reinforcing the use of certain contents or determinants (Boulay, 1969; Hersen & Greaves, 1971). The administration procedure used in the present study did not reinforce types of answers, but rather encouraged another response when the respondent failed to extrapolate from the initial prompt. Encouraging additional responses in a nondirective, standardized way is not the same as influencing or reinforcing specific aspects of the response itself.

The response process (see Exner, 1993) is a complex phenomenon that occurs according to the respondent's own perceptions, classifications, decision-making processes, and psychological traits and states. Respondents generate many more perceptions to the stimulus than they report during the response phase (Exner, Armbruster, and Mittman, 1978). An examination of the response process calls into question the idea that asking for another response creates an invalid answer. If the prompt is nondirect and standardized, the respondent essentially is being asked to report, and possibly process further, a percept that has already been perceived, not to create a new one based on the examiner's distorting influence. The additional response that is offered is a valid representation of the respondent's internal processes because it is something that was generated by the respondent. As such, these additional responses do not invalidate the CS.

Rorschach utility. The question of validity is directly related to test utility. Even though a record may contain the minimum number of allowable responses (14), it may not be interpretively useful. Brief records lack sensitivity and negative predictive power, with potentially reduced yield. If the test cannot provide incremental validity over other assessment methods due to too few responses being generated, then it lacks clinical utility. Other implications are monetary, such as the cost of administering the test and the examiner's time and resources, all of which increase substantially when the test must be readministered due to an insufficient number of responses. The issues of cost effectiveness and cost benefit are becoming increasingly important in psychological assessment as clinicians are justifying their role in clinical decision making (Yates & Taub, 2003). Although this study was conducted with a sample prone to frequent brief records, the issues are similar with overly long records, which tend to lack specificity and positive predictive power, are time consuming to administer and score, and often are frustrating or exhausting for both the examiner and client. The goal of constraining Rorschach response productivity is to increase test utility by obtaining a more optimal range of data from which to make informed clinical decisions.

Response range. One goal of the alternative administration method was to decrease the response range in an effort to decrease standardized error. We hypothesized that with a smaller range and lower standard deviation, the regression predictions would be improved among protocols containing between 18 and 28 responses. One method employed was to prompt for additional responses when needed, but the other was to remove cards after four responses were given to decrease excessively long protocols. In this sample, removing the cards occurred on only four protocols. As mentioned previously, this sample generally produced brief records, so it is not surprising that card removal was not often necessary. Despite the low rate of card removal, there was a nonsignificant trend toward a more restricted response range (i.e., less variability) in the alternative administration group, particularly during the first response phase (range of 14 to 40 responses in the alternative group, versus 10 to 45 in the standard group). Even following readministration, the

response range in the standard group remained slightly larger (14–45).

We hypothesized that protocols falling within the range of 18 to 28 responses would yield less error and therefore more validly assess thought disorder, based on the premise that the Rorschach thought disorder predictors would be more accurate in that range. In this sample there was no statistical difference in error terms or in concurrent validity among protocols in this range compared with protocols outside this range, however, suggesting that protocols in the 18 to 28 range were not necessarily more accurate. A larger sample size may be necessary to detect the small effect sizes expected from such an analysis, however, given that residuals are examined without the influence of the Rorschach predictors. In our sample the results were in the expected direction, but the effect was small to medium in size ($d = .35$). An additional consideration is that the number of responses in this sample was positively skewed, so that more records were on the low end of responses (<18) than on the high end (>28). Most protocols in the sample, 59%, fell outside the 18-to-28 range, though the vast majority, 46% of all records, contained fewer than 18 responses. Samples that produce a more balanced number of short and long records may be more adequately equipped to demonstrate that less error is associated with protocols that have *R* between 18 and 28.

Response complexity. Utilizing the Complexity Index significantly improved predictions of thought disorder for the PTI, with nonsignificant trends present for the SCZI and EII-2. Regression analyses for the entire sample indicated that the EII-2 had a substantial association with the criterion by itself (Adjusted $R = .521$), and this may explain why complexity did not improve the prediction. The SCZI (Adjusted $R = .396$) and PTI (Adjusted $R = .381$) had associations with the criterion of similar magnitude, yet complexity improved prediction for the PTI but not the SCZI.

Complexity also improved predictions in protocols falling outside the range of 18 to 28 responses but did not influence predictions for protocols in the central range of 18 to 28 responses. These findings suggest that protocols of moderate length, which are also those for which the CS reference values are most appropriate, do not require corrective adjustments for protocol complexity when predicting thought disorder criterion measures. It was necessary to adjust for protocol complexity, however, to optimally predict these criteria using short (<18) or long (>28) protocols.

Conclusions and Suggestions for Future Research

Continued work should address some of the limitations of this study, including the limited number of examiners and participants, and the generalizability of the findings to other populations. Although precautions were taken to standardize the scoring procedures and maintain blindedness from one source of data to another, and even though the ICC interrater reliabilities between the first author and the other raters ranged from good to excellent (Cicchetti, 1994; Shrout & Fliess, 1979), it would be helpful for future research to use multiple examiners who are unaware of the hypotheses. The results of this study have ecological validity for the kinds of conditions commonly encountered in residential treatment situations. They may not translate as well, however, to outpatient populations or to settings that do not typically produce brief records. Furthermore, since the present study focused on test validity relative to thought disorder criteria, the results may not generalize to other constructs or other diagnostic groups. Thus, future research fruitfully could examine outpatients across a broad spectrum of psychopathology and assess validity for other Rorschach variables and criterion constructs. A final limitation is that we selected a sample that should produce brief protocols even though our modified administration guidelines are designed to constrain *R* at both the high end and the low end. Further research in samples thought to produce overly long and complex records would be useful.

The present study is an important step in understanding how administration procedures can influence response productivity and Rorschach validity for assessing thought disorder. Differences in yield between the administration groups were noted, as were differences in validity for assessing thought disorder. The overall results suggest that test utility is improved with modified administration guidelines and that raw scores unmoderated or uncorrected by complexity are most valid when *R* is in the range from 18 to 28. Although we prompted for additional responses on all cards except V and IX, for simplicity examiners could prompt on every card, as there is no evident disadvantage associated with this strategy. Thus, we encourage examiners to prompt regularly for more responses when just one is offered to a card and to regularly remove a card after four responses have been obtained.

ACKNOWLEDGMENTS

Kristin L. Dean is now with the University of Tennessee-Cherokee Health Systems Center of Excellence. This study was completed for Kristin L. Dean's doctoral dissertation at the California School of Professional Psychology at Alliant International University. Preliminary results were presented at the 2003 annual meeting of the Society of Personality Assessment, San Francisco, California. The authors thank the staff and residents at Central Prison, the North Carolina Correctional Institute for Women, and the Utah State Hospital for their support.

REFERENCES

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Andreasen, N. C. (1979). Thought, language, and communication. *Archives of General Psychiatry, 39*, 789–793.

Andreasen, N. C. (1986). Scale for the assessment of thought, language, and communication. *Schizophrenia Bulletin, 12*, 473–482.

Andreasen, N. C., & Grove, W. M. (1986). Evaluation of positive and negative symptoms in schizophrenia. *Psychiatrie & Psychobiologie, 1*(12), 108–121.

Auslander, L. A. (2000). Thought disturbance differences between male and female late-life schizophrenia patients and its relationship to quality of life. Unpublished doctoral dissertation, California School of Professional Psychology.

Auslander, L. A., Perry, W., & Jeste, D. V. (2002). Assessing disturbed thinking and cognition using the Ego Impairment Index in older schizophrenia patients: Paranoid vs. nonparanoid distinction. *Schizophrenia Research, 53*, 199–207.

Bailey, B., West, K. Y., Widiger, T. A., & Freiman, K. (1993). The convergent and discriminant validity of the Chapman scales. *Journal of Personality Assessment, 61*, 121–135.

Boulay, M. A. (1969). Verbal reinforcement and Rorschach productivity. *Journal of Clinical Psychology, 25*, 310.

Chapman, L. J., Chapman, J. P., Kwapil, T. R., Eckblad, M., & Zinser, M. C. (1994). Putatively psychosis-prone subjects 10 years later. *Journal of Abnormal Psychology, 103*, 171–183.

Chapman, L. J., Chapman, J. P., & Miller, E. N. (1982). Reliabilities and intercorrelations of eight measures of proneness to psychosis. *Journal of Consulting & Clinical Psychology, 50*, 187–195.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement, 12*, 425–434.

Cuesta, M. J., & Peralta, V. (1993). Does formal thought disorder differ among patients with schizophrenic, schizophreniform and manic schizoaffective disorders? *Schizophrenia Research, 10*, 151–158.

Dao, T. K., & Prevatt, F. (2006). A psychometric evaluation of the Rorschach Comprehensive System's Perceptual Thinking Index. *Journal of Personality Assessment, 86*, 180–189.

Docherty, N., Schnur, M., & Harvey, P. D. (1988). Reference performance and positive and negative thought disorder: A follow-up study of manics and schizophrenics. *Journal of Abnormal Psychology, 94*, 437–442.

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting & Clinical Psychology, 51*, 215–225.

Exner, J. E. (1993). *The Rorschach: A Comprehensive System* (3rd ed., vol. 1). New York: John Wiley & Sons.

Exner, J. E. (Ed.). (1995). *Issues and methods in Rorschach research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Exner, J. E. (2000). *A primer for Rorschach interpretation*. Asheville, NC: Rorschach Workshops.

Exner, J. E. (2001). *A Rorschach workbook for the Comprehensive System* (5th ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E. (2003). *The Rorschach: A Comprehensive System: Vol. 1. Basic Foundations* (4th ed.). New York: Wiley.

Exner, J. E., Armbruster, G. L., & Mittman, B. (1978). The Rorschach response process. *Journal of Personality Assessment, 42*, 474–482.

Exner, J. E., & Erdberg, P. (2005). *The Rorschach: A Comprehensive System Vol.: 2. Advanced Interpretation* (3rd ed.). Hoboken, NJ: Wiley.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.

George, L., & Neufeld, R. W. (1987). Magical ideation and schizophrenia. *Journal of Consulting & Clinical Psychology, 55*, 778–779.

Harvey, P. D. (1983). Speech competence in manic and schizophrenic psychoses: The association between clinically rated thought disorder and cohesion and reference performance. *Journal of Abnormal Psychology, 92*, 368–377.

Harvey, P. D., Earle-Boyer, E. A., & Wielgus, M. S. (1984). The consistency of thought disorder in mania and schizophrenia: An assessment of acute psychotics. *Journal of Nervous and Mental Disease, 172*, 458–463.

Harvey, P. D., Lenzenweger, M. F., Keefe, R. S., Pogge, D. L., Serper, M. R., & Mohs, R. C. (1992). Empirical assessment of the factorial structure of clinical symptoms in schizophrenia patients: Formal thought disorder. *Psychiatry Research, 44*, 141–151.

Harvey, P. D., & Serper, M. R. (1990). Linguistic and cognitive failures in schizophrenia. *Journal of Nervous and Mental Disease, 178*, 487–493.

Hersen, M., & Greaves, S. T. (1971). Rorschach productivity as related to verbal reinforcement. *Journal of Personality Assessment, 35*, 436–441.

Kates, J. A. (1994). *A validation study of the Rorschach Ego Impairment Index*. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego.

Kinder, B. N. (1992). The problems of R in clinical settings and in research: Suggestions for the future. *Journal of Personality Assessment, 58*, 252–259.

Kleiger, J. H. (1999). *Disordered thinking and the Rorschach: Theory, research, and differential diagnosis*. Hillsdale, NJ: The Analytic Press, Inc.

Klopfer, B., & Davidson, H. H. (1962). *The Rorschach technique: An introductory manual*. New York: Harcourt, Brace & World.

Lipgar, R. M. (1992). The problem of R in the Rorschach: The value of varying responses. *Journal of Personality Assessment, 58*, 223–230.

Mcguire, H., Kinder, B. N., Curtiss, G., & Viglione, D. J. (1995). Some special issues in data analysis. In J. E. Exner (Ed.), *Issues and methods in Rorschach research* (pp. 227–250). Mahwah, NJ: Lawrence Erlbaum Associates.

McPherson, L., & Harvey, P. D. (1996). Discourse connectedness in manic and schizophrenic patients: Associations with derailment and other clinical thought disorders. *Cognitive Neuropsychiatry, 1*, 41–53.

Meyer, G. J. (1992). Response frequency problems in the Rorschach: Clinical and research implications with suggestions for the future. *Journal of Personality Assessment, 58*, 231–244.

Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI–2 criteria. *Journal of Personality Assessment, 60*, 153–180.

Morgan, L., & Viglione, D. J. (1992). Sexual disturbances, Rorschach sexual responses, and mediating factors. *Psychological Assessment, 4*, 530–536.

Peralta, V., Cuesta, M. J., & de Leon, J. (1995). Positive and negative symptoms/syndromes in schizophrenia: Reliability and validity of different diagnostic systems. *Psychological Medicine, 25*, 43–50.

Perry, W., & Braff, D. L. (1994). Information-processing deficits and thought disorder in schizophrenia. *American Journal of Psychiatry, 151*, 363–367.

Perry, W., McDougall, A., & Viglione, D. (1995). A five-year follow-up on the temporal stability of the Ego Impairment Index. *Journal of Personality Assessment, 64*, 112–118.

Perry, W., Minassian, A., Cadenhead, K., Sprock, J., & Braff, D. (2003). The use of the Ego Impairment Index across the schizophrenia spectrum. *Journal of Personality Assessment, 80*, 50–57.

Perry, W., & Viglione, D. J. (1991). The Ego Impairment Index as a predictor of outcome in melancholic depressed patients treated with tricyclic antidepressants. *Journal of Personality Assessment, 56*, 487–501.

Perry, W., Viglione, D. J., & Braff, D. (1992). The Ego Impairment Index and schizophrenia: A validation study. *Journal of Personality Assessment, 59*, 165–175.

Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current non-patient data for the Rorschach, WAIS-R, and MMPI–2. *Journal of Personality Assessment, 73*, 305–316.

Shrout, P. E., & Fliess, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Smith, S. R., Baity, M. R., Knowles, E. S., & Hilsenroth, M. J. (2001). Assessment of disordered thinking in children and adolescents: The Rorschach perceptual-thinking index. *Journal of Personality Assessment, 77*, 447–463.

Viglione, D. J. (1996). Data and issues to consider in reconciling self-report and the Rorschach. *Journal of Personality Assessment, 67*, 579–587.

Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 11*, 251–265.

Viglione, D. J., & Meyer, G. J. (1998). *Complexity Index scoring principles*. Unpublished manuscript.

Viglione, D. J., Perry, W., Jansak, D., Meyer, G., & Exner, J. E. (2003). Modifying the Rorschach Human Experience Variable to create the Human Representational Variable. *Journal of Personality Assessment, 81*, 64–73.

Viglione, D. J., Perry, W., & Meyer, G. J. (2003). Refinements in the Rorschach Ego Impairment Index incorporating the Human Representational Variable. *Journal of Personality Assessment, 81*, 149–156.

Wagner, E. E., Young, G. R., & Wagner, C. F. (1992). Rorschach blends, IQ, and the effect of R. *Journal of Personality Assessment, 59*, 185–188.

Walker, E. F., Harvey, P. D., & Perlman, D. (1988). The positive/negative symptom distinction in psychosis: A replication and extension of previous findings. *Journal of Nervous and Mental Disease, 176*, 359–363.

Weiner, I. B. (2003). *Principles of Rorschach Interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15*, 478–495.

APPENDIX

Complexity Composite Score Formula (Viglione & Meyer, 1998)

1. Location/Developmental Quality (DQ) Complexity: Any DQv = 0; Any DQv/+ or DQo = 1; D or Dd with DQ+ = 2; W, WS, DS, or DdS with DQ+ = 3.
2. Determinant Complexity: Pure F only = 0; one non-F determinant = 1, multiple determinant blend = Actual # of determinants in blend (e.g., $M^a.m^a.CF = 3$).
3. Content Complexity: Single content that is A, Ad, (A), or (Ad) = 0; Single non-A content = 1; Multiple contents = Actual # of content scores (e.g., A, Fd, Ad = 3).
4. Complexity Index: Sum of Location/DQ, Determinant, and Content Complexity scores.

Kristin L. Dean
University of Tennessee
Cherokee Health Systems Center of Excellence
2018 Western Avenue
Knoxville, TN 37921
Email: kdean@rand.org

Received May 2, 2006
Revised November 30, 2006