Internal Reliability of Rorschach Oral Dependency Scale Scores

Robert F. Bornstein, Erica L. Hill, Kerry J. Robinson, Casey Calabrese and K. Shannon Bowers

Educational and Psychological Measurement 1996; 56; 130
DOI: 10.1177/0013164496056001009

INTERNAL RELIABILITY OF RORSCHACH ORAL DEPENDENCY SCALE SCORES

ROBERT F. BORNSTEIN, ERICA L. HILL, KERRY J. ROBINSON, CASEY CALABRESE, AND K. SHANNON BOWERS
Gettysburg College

The internal reliability of Masling, Rabie, and Blondheim's Rorschach Oral Dependency (ROD) scale was assessed in a mixed-sex sample of 200 undergraduate subjects (100 women and 100 men). ROD scale scores had adequate internal reliability when reliability was assessed using traditional procedures (i.e., odd-even, coefficient alpha). Internal reliability coefficients derived from recently developed "minimization" and "maximization" procedures differed from traditional internal reliability coefficients, with "minimized" coefficients being smaller than traditional indexes of internal reliability and "maximized" coefficients being larger than traditional internal reliability indexes. Implications of these results for the construct validity of ROD scale scores are discussed, and suggestions regarding the assessment of internal reliability for Rorschach-derived personality and psychopathology variables are offered.

Masling, Rabie, and Blondheim's (1967) Rorschach Oral Dependency (ROD) scale has been used in nearly 100 published studies since the late 1960s (Bornstein, 1992; Masling, 1986). Moreover, a comprehensive review of research on the etiology, correlates, and consequences of dependent personality traits revealed that the ROD scale has been the most widely used projective measure of dependency during the past 50 years (Bornstein, 1993), having been employed in approximately 20% of all published studies in this area.

The vast majority of investigations involving the ROD scale have used adolescent or adult subjects (see Masling, 1986; Masling & Schwartz, 1979), although a few investigations (e.g., Gordon & Tegtemeyer, 1983) used the ROD scale to assess dependency in children. Approximately two thirds of all

Correspondence regarding this paper should be sent to Robert F. Bornstein, PhD, Department of Psychology, Gettysburg College, Gettysburg, PA, 17325.

Educational and Psychological Measurement, Vol. 56 No. 1, February 1996 130-138

published studies involving the ROD scale have assessed dependency in college students (Bornstein, 1992, 1993). Except for a few investigations involving community subjects (e.g., Juni & Cohen, 1985), every other ROD study has used psychiatric inpatients (Bornstein & Greenberg, 1991) or outpatients (Masling, Schiffner, & Shenfeld, 1980). Regardless of subject age, socioeconomic background, educational level, or clinical status (i.e., patient vs. nonpatient), women and men typically obtain comparable ROD scores (Masling, 1986; O'Neill & Bornstein, 1990; Shilkret & Masling, 1981).

Evidence collected to date supports the construct validity of ROD scale scores as measuring dependency. High scores on the ROD scale are associated with a variety of dependent, help-seeking behaviors in college students and psychiatric patients (Bornstein, Krukonis, Manning, Mastrosimone, & Rossner, 1993; Greenberg & Bornstein, 1989; Masling, Weiss, & Rothschild, 1968; O'Neill & Bornstein, 1990; Shilkret & Masling, 1981; Weiss, 1969). In addition, ROD scores have positive correlations with scores on self-report measures of dependency (Bornstein, Manning, Krukonis, Rossner, & Mastrosimone, 1993; Bornstein, Poynton, & Masling, 1985) and are unrelated to scores on measures of social desirability, locus of control, and need for approval (Masling, 1986; Masling & Schwartz, 1979). ROD scores have adequate retest reliability over a 16-week period in both men and women (Bornstein, Rossner, & Hill, 1994) and are relatively immune from self-report and self-presentation biases that affect scores on other kinds of dependency measures (Bornstein, Rossner, Hill, & Stepanian, 1994). Finally, ROD scores have the predicted positive relationship with scores on measures of theoretically related psychopathologies such as depression (Bornstein et al., 1985), alcoholism (Weiss & Masling, 1970), and eating disorders (Bornstein & Greenberg, 1991).

Although studies have examined the convergent validity, discriminant validity, and retest reliability of ROD scores, no investigations have assessed the internal reliability of scores on the ROD scale. The paucity of evidence regarding the internal reliability of ROD scores has most likely resulted from two related methodological difficulties. First, a number of theoreticians and researchers have argued that because of differences in the types of responses evoked by different inkblots, Rorschach-derived scores measuring personality and psychopathology should not be expected to show high levels of internal reliability (Piotrowski, 1937; Stein, 1960, 1962). Second, the statistical procedures used to assess internal reliability on self-report tests are not always effective in assessing internal reliability on projective tests because these procedures yield internal reliability coefficients for projective tests that underestimate the actual internal consistency of the scores in question (Wagner, Alexander, Roos, & Adair, 1986; Wagner, Alexander, Roos, & Prospero, 1985).

Recently, Wagner et al. (1985, 1986) developed an approach to assessing internal reliability for Rorschach-derived variables that allows for a more

rigorous assessment of the internal consistency of the ROD scale than was possible when older approaches were used. Wagner et al. recommended computing all possible split-half correlations for the Rorschach variable in question and then retaining the maximum internal reliability coefficient obtained as an overall estimate of internal consistency. Wagner et al. argued that when such a procedure is used, the deleterious effects of Rorschach card heterogeneity on internal reliability estimates are minimized (see also Cronbach, 1943, 1951, for discussions of this issue).

In a recent investigation, Wagner et al. (1986) demonstrated that this "maximization procedure" yielded internal reliability estimates for Rorschach structural and content variables (e.g., D, F+, H, A) that were somewhat larger than those produced when the more conservative coefficient alpha (Cronbach, 1951) was used to assess internal reliability. When Wagner et al. employed a similar "minimization procedure" (wherein all possible split-half correlations were calculated and the lowest internal reliability estimate was retained), they found that reliability estimates derived using this procedure were substantially smaller than those obtained when traditional internal reliability computation formulas were used. Wagner et al. recommended reporting minimum and maximum internal reliability estimates for projective tests in addition to traditional indexes of internal consistency for these tests (e.g., odd-even reliability, coefficient alpha) to obtain as much information as possible regarding the upper and lower limits of the test's internal reliability.

The purpose of the present study was to assess the internal reliability of scores on Masling et al.'s (1967) ROD scale in a large, mixed-sex sample of college students. Assessing the internal reliability of the ROD scale fills an important gap in the literature on the construct validity of this measure. Moreover, use of Wagner et al.'s (1985, 1986) minimization and maximization procedures for deriving internal reliability estimates in addition to more traditional internal reliability estimates allows for a direct comparison of different procedures for assessing the internal consistency of scores on the ROD scale. We hypothesized that (a) ROD scores would show adequate internal reliability, regardless of the procedures used to derive internal reliability estimates; and (b) internal reliability estimates derived from Wagner et al.'s (1985, 1986) maximization procedure would be larger than those derived using traditional internal reliability formulas (i.e., odd-even and coefficient alpha), whereas internal reliability estimates calculated using Wagner et al.'s minimization procedure would be smaller than those yielded by traditional reliability formulas.

A secondary purpose of the present article was to examine the distribution of oral dependent percepts across the 10 Rorschach inkblots. A number of theorists have speculated that certain Rorschach inkblots should elicit oral dependent content more easily than do other inkblots, primarily because of variations in the form, texture, and (to a lesser extent) color of different inkblots (see Allison, Blatt, & Zimet, 1988; Aronson, 1952; Rapaport, Gill, &

Schafer, 1945; Schafer, 1954). To see whether certain Rorschach cards consistently elicit greater amounts of oral dependent content than do other cards, we calculated for each inkblot (a) the percentage of subjects who provided at least one oral dependent response to that blot, and (b) the overall percentage of responses to the blot that contained any scorable oral dependent content.

Method

Subjects

Two hundred undergraduate students (100 women and 100 men) participated in the study to fulfill a portion of their Psychology 101 research participation requirement. Subjects ranged in age from 17 to 21 years, with a mean age of 19.25 years ($SD = 1.34$).

Procedure

Groups of 10 to 12 subjects completed the ROD scale under standard conditions. As in other ROD studies (e.g., Bornstein, Rossner, & Hill, 1994), subjects were shown slides of Rorschach inkblots projected onto a screen and were asked to provide three written responses each to Cards 1, 2, 3, 8, and 10, and two written responses to each of the remaining cards. Subjects who provided fewer than 20 of the 25 required responses were dropped from the study. The ROD administration procedure took about 30 minutes per group to complete.

Scoring for oral-dependent content on the ROD scale followed the suggestions of Schafer (1954), as modified by Masling et al. (1967). One point was assigned for each oral-dependent Rorschach response. A detailed ROD scoring manual provided by Masling (1986) included the following categories: (a) foods and drinks; (b) food sources; (c) food objects; (d) food providers; (e) food receivers; (f) begging and praying; (g) food organs; (h) oral instruments; (i) nurturers; (j) gifts and gift givers; (k) good luck symbols; (l) oral activity; (m) passivity and helplessness; (n) pregnancy and reproductive anatomy; (o) "baby talk" responses; and (p) negations of oral percepts (i.e., "not pregnant," "man with no mouth").

All ROD protocols were scored for oral-dependent content by the first author. Reliability in ROD scoring was determined by having a second experimenter, blind to all information regarding individual subjects, rescore a random sample of 20 protocols containing a total of 500 responses. The two raters agreed regarding the scoring of 471 responses (94%). A Pearson correlation coefficient calculated between the two sets of scores was .93. These interrater reliability coefficients are comparable to those reported in

recent studies using the ROD scale (e.g., Bornstein & Greenberg, 1991; Bornstein, Krukonis, et al., 1993; Bornstein, Manning, et al., 1993; O'Neill & Bornstein, 1990).

To calculate the internal reliability of Masling et al.'s (1967) ROD scale, each Rorschach card was treated as a single test item (see Wagner et al., 1985, 1986). Thus, subjects' "scores" on Cards 1, 2, 3, 8, and 10 could range from 0 (no scorable oral dependent content) to 3 (three oral-dependent percepts), whereas subjects' scores on the remaining five Rorschach cards could range from 0 to 2. Whole-scale ROD scores therefore had a potential range of 0 to 25 (although in reality, few nonclinical subjects ever obtain ROD scores higher than 9) (see Bornstein, 1993).

Results

Mean ROD score in our college student sample was 3.52 (SD = 2.05, Range = 0 to 9). As expected, the ROD scores of men (X = 3.44, SD = 2.05) and women (X = 3.59, SD = 2.06) did not appreciably differ, $t(198) = 0.52$, n.s.; Cohen's $d = 0.07$. Consistent with previous studies in this area, subjects' ROD scores were normally distributed and ranged from 0 to 9, with most scores clustered between 1 and 5.

Table 1 summarizes the internal reliability estimates for ROD scores as a function of subject gender and computation method. Four reliability coefficients are included in this table: minimum, maximum, odd-even, and coefficient alpha. Although men produced slightly higher internal reliability coefficients than did women, focused comparisons of effect size (Rosenthal, 1984) indicated that the internal reliability coefficients produced by men and women in this sample did not differ significantly (all Zs < 1.00, all ps > .20). Consequently, the reliability coefficients produced by men and women were combined into overall (pooled) internal reliability estimates for further analysis.

A test for homogeneity of effect sizes (Rosenthal, 1984) revealed that there were statistically significant differences in the magnitude of internal reliability estimates as a function of computation method, $\chi^2(1) = 22.15$, $p < .001$. Follow-up focused comparisons of effect size indicated that the minimum internal reliability estimate was significantly smaller than the maximum, odd-even, and coefficient alpha reliability estimates (Zs were 4.69, 2.51, and 2.58, respectively, in these comparisons; all ps < .01). Although the odd-even and coefficient alpha internal reliability estimates did not differ (Z = 0.06, n.s.), focused comparisons of effect size confirmed that, as expected, maximum reliability estimates were significantly larger than reliability estimates calculated using odd-even and coefficient alpha procedures (Zs were 2.17 and 2.11, respectively, in these comparisons; both ps < .02).

The percentage of subjects who produced at least one oral-dependent Rorschach response on a particular card ranged from 14 (Card 8) to 29 (Cards

Table 1
Internal Reliability Estimates for ROD Scores

	Internal reliability estimate			
	Minimum	Maximum	Odd-even	Alpha
Women	.310	.761	.604	.610
Men	.359	.778	.614	.620

Note. *N* of subjects included in this table = 200 (100 women and 100 men). Minimum and maximum internal reliability estimates were computed using the procedures described by Wagner et al. (1985).

1 and 7). The percentage of Rorschach responses with any scorable oral-dependent content ranged from 8 (Cards 4 and 8) to 19 (Card 7). To examine the distribution of subjects' oral-dependent responses across the 10 Rorschach inkblots, separate analyses were performed comparing (a) the percentage of subjects producing at least one oral-dependent response on a particular card as a function of card number, and (b) the percentage of responses with any oral-dependent content as a function of card number. Both of these analyses yielded nonsignificant results, indicating that subjects' oral-dependent responses were randomly distributed among the 10 Rorschach inkblots, $\chi^2(9) = 5.48$, n.s. for Comparison 1, and $\chi^2(9) = 6.21$, n.s. for Comparison 2. When these analyses were performed separately for men and women, virtually identical results were obtained.

Discussion

Although researchers have long regarded projective tests as being psychometrically inadequate, recent findings suggest that the psychometric properties of Rorschach-derived personality and psychopathology indexes vary considerably (Parker, 1983; Parker, Hanson, & Hunsley, 1988). In general, these indexes show inadequate reliability and validity when (a) subjective scoring criteria are used, and (b) the projective test variables are not tied to an overarching theoretical framework that generates definitive predictions regarding the relationship of the variable in question to scores on other theoretically related personality and psychopathology measures (Parker et al., 1988). Conversely, when a Rorschach variable is scored using a highly elaborated, well-articulated set of decision rules and is explicitly linked to a theoretical framework that makes definitive predictions regarding critical intervariable relationships, then scores derived from this variable typically show adequate internal consistency, interrater reliability, and convergent and discriminant validity (Atkinson, 1986; Parker, 1983; Parker et al., 1988).

Clearly, Masling et al.'s (1967) ROD scale fits into the latter category: It is closely tied to an overarching theoretical framework that makes strong predictions regarding the relationship of ROD scores to scores on other

personality and psychopathology measures (Masling, 1986), and it uses an objective, lexical scoring system that yields excellent interrater reliability, even for relatively inexperienced raters (Bornstein, Krukonis, et al., 1993; Bornstein, Manning, et al., 1993; Bornstein, Rossner, & Hill, 1994). In this context, it is not surprising that ROD scores show adequate internal reliability: 25 years of research assessing the construct validity of the ROD score as a measure of dependent traits and tendencies have demonstrated that ROD scores have acceptable psychometric properties in most (if not all) areas.

In any case, the present results suggest that the internal reliability of the ROD scale is comparable to that of other commonly used Rorschach structural and content variables (see Wagner et al., 1986). These findings not only support the construct validity of ROD scale scores, but also support Parker et al.'s (1988) hypothesis that the psychometric properties of a projective test variable are largely a function of the clarity and specificity of the scoring criteria associated with the variable in question, as well as of the ability of the broader theoretical framework associated with that variable to generate definitive predictions regarding key intervariable relationships. It appears that greater attention to these issues during the delineation and development of Rorschach-derived variables has the potential to yield projective indexes of personality and psychopathology that have adequate construct validity.

One subsidiary finding from this investigation also warrants mention. Specifically, our results dovetail with those of Wagner et al. (1985, 1986), who found that internal reliability estimates of several Rorschach variables (e.g., D, F+, H, A) calculated using Wagner et al.'s maximization procedure were larger than internal reliability estimates calculated using traditional internal consistency indexes such as coefficient alpha (Cronbach, 1951), whereas internal reliability estimates for these variables calculated using the minimization procedure were smaller than those associated with traditional internal reliability formulas. Although Wagner et al. did not assess the statistical significance of differences in internal reliability estimates derived from different computation procedures, we did, and found that the maximization and minimization procedures did, in fact, yield internal consistency estimates for the ROD scale that differed significantly from those calculated in the traditional manner. Of course, statistical tests are largely driven by sample size, so these significance tests may be artifactual (Cohen, 1994). Without question, additional research assessing the internal consistency of variables derived from widely used projective measures is needed, as are studies comparing directly the internal consistency estimates calculated using different procedures and formulas.

In addition, continued examination of the psychometric properties of Masling et al.'s (1967) ROD scale is warranted. Although studies conducted to date support the construct validity of these scores as measuring dependent traits, attitudes, and behaviors (Bornstein, 1992, 1993; Masling, 1986),

additional data are needed in several areas (e.g., data regarding the long-term stability of ROD scores and the relationship of ROD scores to scores on other projective measures of dependency). It goes without saying that the degree to which researchers can draw firm conclusions regarding the correlates and consequences of dependent personality traits depends entirely upon the degree to which valid and reliable scores are available for empirical research in this area. In this respect, studies of the construct validity of widely used objective and projective dependency scales will not only contribute to the development and validation of the tests themselves, but also represent an important step in the continuing effort to increase our knowledge regarding the etiology, correlates, and consequences of dependent personality traits.

References

Allison, J., Blatt, S. J., & Zimet, C. N. (1988). *The interpretation of psychological tests*. New York: Hemisphere.

Aronson, M. L. (1952). A study of the Freudian theory of paranoia by means of the Rorschach test. *Journal of Projective Techniques, 16*, 397-411.

Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychology, 27*, 238-247.

Bornstein, R. F. (1992). The dependent personality: Developmental, social and clinical perspectives. *Psychological Bulletin, 112*, 3-23.

Bornstein, R. F. (1993). *The dependent personality*. New York: Guilford.

Bornstein, R. F., & Greenberg, R. P. (1991). Dependency and eating disorders in female psychiatric inpatients. *Journal of Nervous and Mental Disease, 179*, 248-252.

Bornstein, R. F., Krukonis, A. B., Manning, K. A., Mastrosimone, C. C., & Rossner, S. C. (1993). Interpersonal dependency and health service utilization in a college student sample. *Journal of Social and Clinical Psychology, 12*, 262-279.

Bornstein, R. F., Manning, K. A., Krukonis, A. B., Rossner, S. C., & Mastrosimone, C. C. (1993). Sex differences in dependency: A comparison of objective and projective measures. *Journal of Personality Assessment, 61*, 169-181.

Bornstein, R. F., Poynton, F. G., & Masling, J. M. (1985). Orality and depression: An empirical study. *Psychoanalytic Psychology, 2*, 241-249.

Bornstein, R. F., Rossner, S. C., & Hill, E. R. (1994). Retest reliability of scores on objective and projective measures of dependency: Relationship to life events and intertest interval. *Journal of Personality Assessment, 62*, 398-415.

Bornstein, R. F., Rossner, S. C., Hill, E. R., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63*, 363-386.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.

Cronbach, L. J. (1943). On estimates of test reliability. *Journal of Educational Psychology, 34*, 485-494.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Gordon, M., & Tegtemeyer, P. F. (1983). Oral dependent content in children's Rorschach protocols. *Perceptual and Motor Skills, 57*, 1163-1168.

Greenberg, R. P., & Bornstein, R. F. (1989). Length of psychiatric hospitalization and oral dependency. *Journal of Personality Disorders, 3*, 199-204.

Juni, S., & Cohen, P. (1985). Partial impulse erogeneity as a function of fixation and object relations. *Journal of Sex Research, 21*, 275-291.

Masling, J. M. (1986). Orality, pathology and interpersonal behavior. In J. M. Masling (Ed.), *Empirical studies of psychoanalytic theories* (Vol. 2, pp. 73-106). Hillsdale, NJ: Lawrence Erlbaum.

Masling, J. M., Rabie, L., & Blondheim, S. H. (1967). Obesity, level of aspiration, and Rorschach and TAT measures of oral dependence. *Journal of Consulting Psychology, 31*, 233-239.

Masling, J. M., Schiffner, J., & Shenfeld, M. (1980). Client perception of the therapist, orality, and sex of client and therapist. *Journal of Counseling Psychology, 27*, 294-298.

Masling, J. M., & Schwartz, M. A. (1979). A critique of research in psychoanalytic theory. *Genetic Psychology Monographs, 100*, 257-307.

Masling, J. M., Weiss, L., & Rothschild, B. (1968). Relationships of oral imagery to yielding behavior and birth order. *Journal of Consulting and Clinical Psychology, 32*, 89-91.

O'Neill, R. M., & Bornstein, R. F. (1990). Oral dependence and gender: Factors in help-seeking response set and self-reported psychopathology in psychiatric inpatients. *Journal of Personality Assessment, 55*, 28-40.

Parker, K.C.H. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment, 47*, 227-231.

Parker, K.C.H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach and WAIS: A meta-analytic comparison of reliability, stability and validity. *Psychological Bulletin, 103*, 367-373.

Piotrowski, Z. A. (1937). The reliability of Rorschach's Erlebnistypus. *Journal of Abnormal and Social Psychology, 32*, 439-445.

Rapaport, D., Gill, M., & Schafer, R. (1945). *Diagnostic psychological testing*. New York: Year Book Publishers.

Rosenthal, R. (1984). *Meta-analytic techniques for social research*. Beverly Hills, CA: Sage.

Schafer, R. (1954). Psychoanalytic interpretation in Rorschach testing. New York: Grune & Stratton.

Shilkret, C. J., & Masling, J. M. (1981). Oral dependence and dependent behavior. *Journal of Personality Assessment, 45*, 125-129.

Stein, H. (1960). Rotation and reliability of the Rorschach. *Journal of Projective Techniques, 24*, 171-181.

Stein, H. (1962). An analysis of two components entering into Rorschach reliability values. *Journal of Projective Techniques, 26*, 474-477.

Wagner, E. E., Alexander, R. A., Roos, G., & Adair, H. (1986). Optimum split-half reliabilities for the Rorschach: Projective techniques are more reliable than we think. *Journal of Personality Assessment, 50*, 107-112.

Wagner, E. E., Alexander, R. A., Roos, G., & Prospero, M. K. (1985). Maximizing split-half reliability estimates for projective techniques. *Journal of Personality Assessment, 49*, 579-581.

Weiss, L. R. (1969). Effects of subject, experimenter and task variables on compliance with the experimenter's expectation. *Journal of Projective Techniques and Personality Assessment, 33*, 247-256.

Weiss, L. R., & Masling, J. M. (1970). Further validation of a Rorschach measure of oral imagery: A study of six clinical groups. *Journal of Abnormal Psychology, 76*, 83-87.