Assessing Depression in Russian Psychiatric Patients: Validity of MMPI and Rorschach

Jennifer Boyd Ritsher, Elena B. Slivko-Kolchik and Igor V. Oleichik

Assessment 2001; 8; 373
DOI: 10.1177/107319110100800403

ASSESSMENT
2001, Volume 8, Number 4, 373-389

ASSESSING DEPRESSION IN RUSSIAN PSYCHIATRIC PATIENTS: VALIDITY OF MMPI AND RORSCHACH

Jennifer Boyd Ritsher
University of Maryland and
Veterans Affairs Palo Alto Health Care System and
Stanford University School of Medicine

Elena B. Slivko-Kolchik
Igor V. Oleichik
Russian Academy of Medical Sciences, Moscow, Russia

We tested the criterion, concurrent, and content validity of depression indicators in 180 Russian psychiatric patients. Indicators from the Exner Rorschach (*DEPI*, *CDI*) and the Russian MMPI (Berezin Scale 2, Wiggins depression content) were compared to Hamilton (HRSD) scores and 3 types of diagnosis: traditional Russian, contemporary Western (ICD-10), and a mixed version. The MMPI scales had significant associations with each other and each criterion. The Rorschach indexes were unrelated to all other variables, even when their affective, cognitive, and interpersonal components were analyzed separately, response styles were taken into account, or the 2 indexes were used in combination. Nevertheless, sample means on 107 variables were roughly similar to Exner's norms. The study represents an initial step towards establishing the validity of instruments commonly used in Russia and North America for assessing depression among Russians.

Keywords: Cultural psychology, Rorschach, MMPI, Hamilton, psychological test validity, depression, Russia

Jennifer Boyd Ritsher, Center for Health Care Evaluation, Department of Veterans Affairs and Stanford University School of Medicine; Elena B. Slivko-Kolchik and Igor V. Oleichik, Department of Affective Disorders, Research Center for Mental Health, Russian Academy of Medical Sciences.

This research was supported by the International Research and Exchanges Board with funds provided by the U.S. Department of State (Title VII); the Academy for Educational Development, National Security Education Program; the Open Society Institute; the National Institute of Mental Health (MH-13043); and the Department of Veterans Affairs Health Services Research and Development Service and Mental Health Strategic Healthcare Group. None of these organizations are responsible for the views expressed. This article is based in part on Jennifer Ritsher's dissertation, which was under the joint supervision of Robert R. Dies and Robert D. Coursey at the University of Maryland, College Park. Versions of this article were presented at the annual meeting of the Society for Personality Assessment, Boston, MA, in February 1998, and at the XIVth Congress of the International Association for Cross-Cultural Psychology in August, 1998. The authors wish to thank those who participated in the study, including faculty, staff, and patients at the Research Center for Mental Health, the Alekseev Hospital Number 1, and the Mask Therapy Institute, as well as faculty and students from Moscow State University. We are also indebted to the assistance of colleagues at the Bekhterev Institute in St. Petersburg, Russia; the University of Maryland, College Park; the Psychiatric Epidemiology Training program at Columbia University, and St. Elizabeths Hospital in Washington, DC. We regret that space does not allow a full listing of their names.

Correspondence concerning this article and requests for offprints should be addressed to Jennifer Boyd Ritsher, Center for Health Care Evaluation, VA Palo Alto Health Care System (MPD-152) and Stanford University School of Medicine, 795 Willow Road, Menlo Park, CA 94025. E-mail: jennifer@ritsher.net

The explanatory power of psychological assessment is constrained by the fact that most psychological tests, most study participants, and indeed, most psychologists are of North American or Western European origin (Berry, Poortinga, Segall, & Dasen, 1992). Assessment tools are widely used with members of other cultural groups, such as Russians, with scant evidence regarding their cross-cultural validity. For example, they are used in forensic and school settings to evaluate immigrants, refugees, and ethnic minorities. Further, the tests are routinely used in Russia and many other countries worldwide, with varying degrees of modification from the original versions.

Cross-cultural investigations of assessment instruments are essential to determine whether it is appropriate to use a measure with members of a culture other than the one in which it was designed (Triandis & Berry 1980; van de Vijver & Leung, 1997). The small literature on the cross-cultural validity of psychopathology assessment techniques is better developed regarding their use with ethnic minorities than for their use abroad. Within-country differences by ethnicity are difficult to interpret because they may be due to cultural differences, acculturative stress, and/or differences in living conditions due to minority status (Frank, 1992; Ritsher, Ryder, Karasz, & Castille, in press). The cultural contrast is sharpened when the target culture sample is taken from the center of that culture (in this case Russia), rather than from immigrant groups found in the country that created the test (in this case, the USA).

The present study conducted in Russia sought to assess the validity of the Rorschach (Exner, 1991, 1993) and the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) in Russia, specifically to determine whether their depression indicators corresponded more closely to Western-style or traditional Russian clinical diagnoses of depression.

Both the prevalence and the symptom profile of clinical depression vary across societies (Kleinman & Good, 1985; Weissman et al., 1996), and are particularly likely to be different in Russia. Russians report low levels of "global well-being" (Balatsky & Diener, 1993), and the societal turmoil since the collapse of the Soviet Union has led to a sharp decline in life expectancy, widespread poverty, increased unemployment, and increased alcoholism, all of which are risk factors for depression (e.g., GosKomStat Rossii, 1995, 1999a, 1999b, 1999c; Ritsher, Warner, Johnson, & Dohrenwend, 2001). Moreover, it is likely that Russians are more willing than Americans to display depressive symptoms (Berezin, Miroshnikov, & Rozhanets, 1976). In Russia, psychological assessment and psychiatric diagnosis are becoming increasingly Westernized, amid considerable debate but little evidence about the cross-cultural validity of imported techniques (Calloway, 1993; Ritsher, 1997a, 1997b).

The Rorschach Inkblot test has long intrigued cross-cultural researchers because it seems to sidestep many of the language-related problems associated with questionnaire research without sacrificing the scientific virtues of standardization and replicability (Weiner, 1998). In contrast to the thousands of studies on the Comprehensive System (CS) in the U.S., there is a dearth of literature on its cross-cultural validity (Exner, 1994; Wood & Lilienfield, 1999). The cross-cultural literature that does exist on the Rorschach is very uneven in quality (Howes & DeBlassie, 1989). These issues are beginning to be addressed, as the International Rorschach Society provides a forum for cross-cultural dialogue, and large normative projects are underway in Europe and Japan (Exner, personal communication, August 19, 1996; Weiner, 1998).

Weiner (1998) argues that "Rorschach variables mean what they mean, regardless of a subject's socioeconomic status, ethnicity and national origin" (p. 46). In this view, the personality attributes measured by the Rorschach may be more or less common, or more or less adaptive, in a given cultural setting, but the attributes themselves are present as indicated by the Rorschach scores. Thus, the culture-related threats to validity would be limited to fairly manageable technical issues, such as the need for cultural and linguistic competence in order to avoid scoring errors (Weiner, 1999).

For example, the Russian Rorschach literature indicates that the most common answer on Card IV is an animal pelt, although it is a human or human-like figure in the U.S. (Bespalko, 1978, 1985;

Exner, 1993). Allowing an animal pelt to be scored "popular" in Russia would be a simple scoring modification, but would it fully address the cultural differences in Rorschach responses?

Exner and Weiner argue that such cultural differences are usually trivial, and that when cultural adaptation is necessary, relatively minor changes such as adapting the list of Popular responses or renorming the Form Quality tables should be sufficient (Exner & Weiner, 1995; Weiner, 1999). This theory is supported by some reviews of the literature (Butcher, Nezami, & Exner, 1998; Ritzler, 1996) but remains highly controversial. For example, Exner and Weiner report that in a sample of 900 "Non-American" records, "distributions in scores for each of the structural variables are similar to those for the American sample" (p. 49). Of course, specific cultural differences in opposite directions could cancel each other out in such a mixed-culture sample. The extent to which the CS is valid for specific uses in specific cultures remains an open question. Evidence showing cross-cultural or cross-ethnic differences in important Rorschach scores is mounting, as in the 15 studies cited by Wood and Lilienfeld (1999) and numerous other studies (e.g., Bylund, 1992; Hernandez-Guzman, Rey-Clericus, San Martin-Peterson, & Vinet-Reichhardt, 1989; Mattlar, Carlsson, & Forsander, 1993; Spigelman, Spigelman, & Engelesson, 1991; Takeuchi & Scott, 1986; Vizcarro, 1986). Given this state of affairs, Dana (1993) and Wood and Lilienfeld (1999) have urged clinicians to refrain from using the Rorschach cross-culturally until its cross-cultural validity has been more systematically evaluated.

If the Rorschach is not cross-culturally accurate, this poses a risk not just to the individuals who may be wrongly diagnosed, treated, involuntarily hospitalized, and so on, but also to the way that entire cultures are viewed by the dominant society. The Rorschach has long been popular in anthropological studies (e.g., DeVos & Boyer, 1989), and Weiner believes that its "utility" in "primitive societies" makes it "an ideal instrument for exploring cross-cultural differences" (1998, p. 46). Unfortunately, the Rorschach has been used to make sweeping generalizations about the character flaws typical of people in traditional cultures, such as "the life of the adult Indian is dominated by his primitive drives and early fixations" (Billig et al., 1948, in Lindzey, 1961). As recently as 1989, the Comprehensive System was used in Alaska to label the entire Tanaina culture as less psychologically sound than the nearby culture of the Upper Tanana, who were, perhaps not coincidentally, more acculturated to mainstream American culture (Boyer et al., 1989).

The task of cross-cultural Rorschach researchers is further complicated by the fact that it is not sensible to speak of the validity of "the" Rorschach, but rather only of specific indexes, because of the complexity of the CS (Wood, Nezworski, Stejskal, Garven, & West, 1999). The present study tests the Depression Index (*DEPI*), the Coping Deficit Index (*CDI*), and their component parts against other indicators of depression. Recent reviews of the *DEPI* have found it to have poor concurrent and criterion validity (Jørgensen, Andersen, & Dam, 2000; Wood et al., 1999), but it is generally a trusted component of the CS and widely used to assess depression.

Like the Rorschach, the MMPI has been used extensively abroad, as evidenced by the availability of 115 translations used in at least 65 countries (Berry, Poortinga, & Pandey, 1996; Butcher, 1996; see also Hall, Bansal, & Lopez, 1999). Some researchers have simply translated the items and used the American norms, others have created local norms, and still others have created new items and re-normed the test. To further complicate matters, there are often several variants within a given country. In Russia, there is disagreement about whether local norms are necessary. The version by Sobchik (1990) is a simple direct translation of the MMPI which uses American norms, whereas Berezin's version was adapted and re-normed (Berezin et al., 1976, 1994). The present study uses Berezin's version because it is the one most commonly used in clinical settings.

The present study examines the criterion, concurrent, and content validity of the Exner Rorschach and the Berezin MMPI for identifying Russian patients diagnosed with clinical depression, as defined by (a) the Hamilton scale (1960), (b) the traditional Russian/Soviet diagnostic system,

(c) the contemporary Western diagnostic system (the ICD-10; WHO, 1982), and (d) the mixed version (Moscow-ICD-9; 1982) commonly used in Russia. Each of the following hypotheses are formulated in the positive direction for the sake of clarity and because the relevant empirical literature was not clear enough to justify specific negative predictions.

Hypotheses

Criterion Validity

Each test will be a significant predictor of diagnosis and each will add a statistically significant increment of predictive power (incremental validity) when more than one test is used.

Concurrent Validity

Indicators from each test will correlate significantly with one another, which will indicate that they measure overlapping constructs. Due to method differences they are not expected to have perfect correlations.

Content Validity

Depression-related Rorschach scores will be significantly correlated with groups of MMPI items about topics which are similar to those that each Rorschach variable is thought to measure (affective, cognitive, or interpersonal aspects of depression).

Methods

Participants

The 180 participants included 108 adults from a psychiatric research hospital, the Research Center for Mental Health (RCMH) in Moscow; 50 adults from another traditional hospital, Alekseev; and 22 adults from a nontraditional private day treatment center, the Mask Therapy Institute. As the pattern of results was the same for each sample, they were combined to maximize statistical power. The three sites were chosen to represent a range of disorders and settings and to have a high concentration of cases with depression and a well-trained psychiatric staff. All patients at each site were invited to participate in the study, except those judged by the attending psychiatrist to be too acutely ill to be capable of participation. Each participant had a clinically significant and disabling degree of psychiatric impairment, and each gave informed consent.

Sample demographics are comparable to RCMH archival statistics and roughly comparable to nationwide demographic and diagnostic data from GosKomStat (1995, 1999a), the Russian government statistical agency. Like the general population, our sample was about half men (52%), four-fifths ethnic Russian (85%), and living in an average household size of about three (3.1). Since the study was in Moscow, participants were almost all urban (97%) and most came from households with a per capita monthly income over the equivalent of U.S.$100 (73% of sample versus 18% in general population). Our sample was younger than the general population (46% of our sample was under age 26 years), and therefore was less likely to be married (34% vs. 65%) or have children (35% of women in our sample vs. 83% in the general population).

Measures

Diagnosis

There are several systems of diagnosis in use at the present time in Russia: (a) the traditional Russian/Soviet "Moscow School" (Snezhnevsky, 1983); (b) the ICD-10, which is a direct and literal translation of the latest version of the international standard system (World Health Organization, 1992); and (c) a modified version of the International Classification of Diseases, 9th Edition (ICD-9), which was adapted during the Soviet period to be more compatible with the Moscow School (Ministerstvo Zdravokhraneniya SSSR, 1982). These three systems will subsequently be referred to as Snezhnevsky, ICD-10, and Moscow-ICD-9. Throughout Russia, the Moscow-ICD-9 is the most common diagnostic system in practice, whereas the ICD-10 is currently used only in research.

Staff psychiatrists with prior extensive training in all three systems diagnosed each participant using each of the systems. The psychiatrists were blind to the psychological test results until the end of the study, and the psychological assessment team was kept blind to the psychiatric diagnosis as well. For the study, psychiatrists received brief training on the three systems and the Hamilton scale (HSRD,

Hamilton, 1960) but financial constraints prevented the use of formal standardized psychiatric diagnostic interviews. Thus, the data may be considered representative of diagnoses given in routine clinical practice in elite research units, rather than strict research diagnoses or diagnoses typical of practice throughout Russia.

Depending on the diagnostic system used, "depression" was coded as present for a diagnosis of a depressive episode, syndrome, phase, disorder, or reaction versus "other" which included manic episode, schizophrenia, and others. A binary depression diagnosis variable was not coded for the traditional Snezhnevsky diagnoses, because it was not clear how to sort the disorders into mutually exclusive groups. For example, it would be difficult to classify a patient with "Progressive, Attack-like Schizophrenia, Depressive Attack." Snezhnevsky diagnoses were used in analyses of wider definitions of depression (any mention of any type of depression in the diagnosis) and in continuous variables (a 4-point scale indicating the degree to which depression is the central feature of the diagnosis).

According to the ICD-10 diagnoses, 23% of the sample had depression, 2% had a manic or mixed episode, 51% had a schizophrenia spectrum disorder, 14% had schizoaffective disorder, and 10% had other disorders such as epilepsy. According to the Moscow-ICD-9 diagnoses, these figures were 21%, 5%, 54%, 10%, and 10%, respectively. The percent of the sample with any mention of depression in the diagnosis was 50% for ICD-10, 72% for Moscow-ICD-9, and 63% for the Snezhnevsky diagnoses.

Hamilton Rating Scale for Depression (HRSD)

For participants at the main site (RCMH), psychiatrists completed the HRSD (Hamilton, 1960), which is a finely-graded indicator of the psychiatrist's assessment of the patient's level of depression, and is often used in cross-cultural research (Sartorius & Ban, 1986). It is a 26-item questionnaire assessing the presence and severity of 24 symptoms, and is commonly used as an outcome measure by Russian psychiatric researchers conducting clinical trials for pharmaceutical companies. We could find no published literature on the Russian HRSD. Scores from the HRSD in our sample were classified as "no depression" (a score from 0 to 8) or "some depression" (> 8).

Minnesota Multiphasic Personality Inventory (MMPI)

The version of the MMPI used most commonly in Russian clinical work was constructed by Berezin and colleagues (1976, 1994), who first altered the items to be more appropriate for the Russian language and culture, and then re-normed the scoring key. For example, the threshold for clinical significance was raised for Scales *2* and *4*, and the subject of the items was changed from "I" to "you." In terms of content, Berezin's version is more similar to the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) than to the original MMPI (Hathaway & McKinley, 1943), because most of the same changes were made during the American revision, such as the deletion of many items pertaining to bowel function and religion (Ritsher, 1997a). For example, on Scale *2*, 54 of the 57 MMPI-2 items are found in Berezin's version.

The present study uses two MMPI scales, the standard Russian clinical scale for depression (Scale *2*) and the American Wiggins depression content scale (Graham, 1977). We used Berezin's Scale *2* and his norms (Berezin et al., 1976, 1994). Using Berezin's Russian norms for the MMPI clinical scales, a *T* score of 65 is typically considered the clinically significant cutoff for what Russian clinicians refer to as a "character accentuation," with a score of 70 or above indicating the presence of a psychological disturbance (Berezin et al., 1994). We report analyses using the $T > 65$ cutoff score. (Aside from predictable changes in specificity and sensitivity, results for $T > 60$ and $T > 70$ were similar).

For the Wiggins depression content scale (which is not normally used in Russia), only 22 of the 33 items were found in the Berezin version. We added the remaining 11 items from Sobchik's verbatim Russian translation of the MMPI (Sobchik, 1990) to complete this American scale. In their book on MMPI special scales, Levitt and Gotts (1995) stated that "no study of the relationship between Rorschach factors and the special scales [including the Wiggins scales]...has yet been published" (p. 92). We chose to study the Wiggins depression scale because Levitt and Gotts concluded that

the existing (non-Rorschach) concurrent validity studies using the Wiggins scales "are almost unanimously positive" (p. 31). According to the American norms published by Graham (1977), a score of 16 or higher represents $T \geq 65$.

Items from Scale *2* and the Wiggins depression content scale were sorted into three categories—affective, interpersonal, and cognitive—by a team of three Russian clinician-researchers (two psychiatrists and one psychologist). These groups of MMPI items were used for construct validity analyses, comparing them to corresponding groups of Rorschach variables from the Depression Index (*DEPI*) and the Coping Deficit Index (*CDI*), following Exner (1993).

Rorschach–Comprehensive System.

The Comprehensive System for the Rorschach (Exner, 1993) is now the standard Rorschach scoring system in North America and has a large literature on its reliability and validity (e.g. Exner, 1995; Meyer, 1997a). However, it is based on American norms and its cross-cultural validity remains largely untested. Its accuracy for assessing depression in Russians is unknown.

We adhered strictly to American standards for administration and scoring (Exner, 1993). Training of the Russian research assistants (all advanced clinical psychology students at the elite Moscow State University) was conducted using materials co-written and supervised by an experienced Rorschach trainer (R. Dies), in consultation with J. Exner (personal communication with R. Dies and J. Ritsher, various dates, 1995-1996). The 10 Russian research assistants achieved 82% agreement in their coding of three protocols at the end of their training. (Kappa values were not calculated for these because there were only 3 protocols, making it difficult to establish expected rates of chance agreement.) The three most highly skilled research assistants, who also collected the majority of the data, achieved 92% agreement. However, since the first author re-scored each protocol and met with each research assistant to resolve discrepancies, it is more appropriate to measure scoring accuracy by comparing the final scores against an expert's scores. A set of 20 randomly chosen final protocols, 2 originating from each research assistant, was translated into English and scored by an expert (R. Dies). Correlations of the rater's CS index scores were high (*DEPI* $r = .85$, $p < .001$; *CDI* $r = .79$, $p < .001$). Overall inter-rater reliability was 87% agreement, or a kappa of .63 (see Table 1). When examining inter-rater reliability with non-translated protocols in English, a kappa over .60 is considered to be "good" (Meyer, 1999), and kappa over .40 is considered "fair." Given the changes in nuance brought about by the translation, and the fact that subsequent inspection of discrepancies revealed that most were attributable to ambiguities in translations of the protocols, we consider a kappa of .40 or higher to be evidence of acceptable inter-rater reliability using these methods. For this study, kappa is almost surely an underestimate of the quality of the scoring because of the errors introduced by the translation. By this standard, almost all of our kappas were acceptable, with the biggest discrepancies on variables most affected by nuances of language, such as the special scores (see Table 1). Following Shaffer, Erdberg, and Haroian (1999), we have not reported kappas for variables that had an extremely low level of occurrence (< 1%). English translations of protocols were not used in any other analyses reported in this paper, which all used data from the original Russian protocols scored by the research team directly.

The present study tests the validity in Russia of the two indicators of depression in Exner's Comprehensive System for the Rorschach—the Depression Index (*DEPI*) and the Coping Deficit Index (*CDI*). We used the standard (Exner, 1993) clinically significant cutting scores for the *DEPI* (*DEPI* > 4) and *CDI* (*CDI* > 3). Criterion validity was tested using hierarchical regression models as well as calculating the sensitivity, specificity, and hit rate. Sensitivity is the number of true positives divided by the number of true positives plus false negatives, or the percentage of the time that a test indicated depression when it was present. Specificity is the number of true negatives divided by the number of true negatives plus false positives, or the percentage of the time that a test indicated no depression when it had not been diagnosed. Hit rate is the percentage of accurate classification (true positive plus true negative) out of all the cases.

Table 1
Kappa Coefficients for 20 Rorschach Study Protocols Scored in Russian Compared to an Expert's Scores of the English Translations

Variable	Kappa	Variable	Kappa
R	1.0	*Ad*	.72
Location (*W, D, Dd*)	1.0	(*Ad*)	.44
Popular	.82	*Hx*	0.0
Z score	.73	*Art*	.53
DQ	1.0	*Ay*	.24
FQ	1.0	*Bl*	–
Pair	.86	*Bt*	.77
F	1.0	*Cg*	.67
FM	.74	*Cl*	.86
a/p for *FM*	.73	*Ex* or *Fi*	.94
M	.80	*Fd*	–
a/p for *M*	.76	*Ge*	–
m	.54	*Hh* or *Id* or *Sc*	.66
a/p for *m*	.49	*Ls*	.62
T	.52	*Na*	.66
Y	.37	*Sx*	–
C'	.68	*Xy* or *An*	.88
V	.59	*DV*	.28
FD	.57	*DR*	.39
C	.76	*INCOM*	.32
Col-Shd	.44	*FABCOM*	.57
Shd-Shd	.40	*CONTAM*	–
r	.92	*ALOG*	.19
H	.73	*Ab*	.56
(*H*)	.79	*COP*	.58
Hd	.58	*AG*	.19
(*Hd*)	.25	*MOR*	.59
A	.87	*PER*	.74
(*A*)	.49	*CP*	–

Note. Kappa coefficients not given for variables with very low occurrence (< 1%). Kappa coefficients are likely deflated by the error variance introduced by the Russian-English translation. No other analyses reported in this article use translated protocols.

There are no universally accepted thresholds for judging the adequacy of a test's hit rate, sensitivity, and specificity, because the relative importance of false positives and false negatives varies according to the purpose of the instrument. Thus, it is best to compare findings to published results using the same instrument. For identifying *DSM-III-R* (American Psychiatric Association, 1987) depression, studies reviewed by Dies (1994) found that the *DEPI* had a hit rate of over 80%, which Exner (1991) reported to improve to over 90% when the *CDI* is also considered. A recent review and meta-analysis by Jørgensen and colleagues (2000) found large variations in the diagnostic efficiency of the *DEPI* across studies, and did not support Exner's (1995) reported high sensitivity level of .75. They concluded that

"*DEPI* scores should be interpreted with a considerable amount of caution when applied for diagnostic purposes" (Jørgensen, Anderson, & Dam, 2000, p. 278). The present study contributes the issue of cross-cultural validity to this literature on the diagnostic criterion validity of the *DEPI*.

Rorschach-MMPI Interrelationship

Despite the vast literatures on the Rorschach and the MMPI, there is scant, contradictory, and typically imprecisely presented evidence about the nature of their interrelationships (Ganellen,1996a, 1996b). Recent studies have not found strong support for the convergent or incremental validity of the two instruments. For example, the relationship between the *DEPI* and MMPI Scale *2* varies across studies with reported correlations ranging from -.11 to .29 (Archer & Gordon, 1988; Archer & Krishnamurthy, 1993a, 1993b, 1997; Lipovsky, Finch, & Balter, 1989; Sells, 1990). Meyer (1997b) argues that the MMPI and Rorschach are such different approaches that "it should not be surprising if [they] obtain qualitatively different types of information" (p. 299). This lack of equivalence would be a useful feature of the tests if they both have incremental validity in predicting diagnosis. Even if their different characteristic strengths and weaknesses mean that these tests do not (and should not) agree in every way, they may still be expected to agree on the specific features that both tests measure. All three of these relationships (general agreement, incremental validity for predicting diagnosis, and agreement about specific symptoms) were tested in the present study.

Results

Criterion Validity: Predicting Diagnosis

The first set of analyses sought to determine each test's accuracy at predicting a diagnosis of depression, as defined in turn by the three diagnostic systems and the psychiatrists' HRSD rating.

Table 2 reveals that, in contrast to the Rorschach variables, the MMPI variables are almost all significantly related to the diagnostic category and to the presence of depression on the HRSD. Unlike the Wiggins depression content scale, the Russian MMPI clinical depression scale (Scale *2*) was not significantly correlated with Moscow-ICD-9 or ICD-10

Table 2
Criterion Validity: Correlations Coefficients for Rorschach and MMPI vs. Diagnosis and HRSD

Indicator	Moscow ICD-9 Depression vs. Other[a]	ICD-10 HRSD Depression vs. Other[b]	HRSD score > 8[c]	HRSD Continuous[d]
Rorschach variables				
DEPI > 4	-.10	-.05	.16	.13
DEPI continuous (0-7)	-.14	-.08	.03	.01
CDI > 3	.00	.00	-.01	.04
CDI continuous (0-5)	-.05	-.07	-.07	-.05
MMPI variables				
Scale *2* > 65	.16	.14	.42***	.44***
Scale *2* continuous	.08	.13	.39***	.41***
Wiggins > 65	.21*	.19*	.37***	.43***
Wiggins continuous	.19*	.23*	.36**	.41***

Note. Moscow-ICD-9 = Soviet/Russian modification of the International Classification of Diseases, 9th ed.; ICD-10 = International Classification of Diseases, 10th ed. (Russian translation, not modified); HRSD = Hamilton Rating Scale for Depression; *DEPI* = Rorschach Depression Index; *CDI* = Rorschach Coping Deficit Index; Scale *2* = Berezin's MMPI clinical scale for depression; Wiggins = American MMPI depression content scale.
[a]$N = 173$. [b]$N = 172$. [c]$N = 100$. [d]$N = 100$.
*$p < .05$. **$p < .01$. ***$p < .001$.

depression. Using the same set of variables, standard indexes of classification accuracy were calculated (Table 3).

Wider Definition of Depression

Using the wider definition of depression (described in the Diagnosis section), the pattern of findings was essentially the same as with the narrower definition, with the MMPI scales outperforming the Rorschach indexes (Table 4). The wider definition of depression produced somewhat higher hit rates and lower false positive rates than the narrower definition, but none of the Rorschach hit rates exceeded 50% (44%-48%). Furthermore, the hit rates were no better (42%-48%) for the subset of cases where the *DEPI* and *CDI* were either both positive or both negative. Widening the definition of depression still further to include manic disturbances also failed to improve the performance of the Rorschach variables (results not shown).

Incremental Validity

These analyses sought to determine whether each predictor variable (the Rorschach and MMPI variables) made an identifiable incremental contribution to the overall prediction of each criterion variable (diagnosis and HRSD), following the approach outlined by Wood and colleagues (1999) and Dawes (1999). Using hierarchical regression to predict the diagnosis variables, the MMPI indicator was entered on the first step, and the Rorschach indicator was introduced in the second step. In these analyses, the *DEPI* did not significantly add to the proportion of variance explained by Scale *2* alone, and the *DEPI* did not significantly predict depression on any of the classification systems: Moscow-ICD-9 (odds ratio [OR] for dichotomous *DEPI* variable = .50, 95% confidence interval [CI] 0.22-1.2, $p = .11$); ICD-10 (OR = .71, CI 0.32-1.6, $p = .40$), HRSD (OR = 1.6, CI 0.64-4.2, $p = .30$) using logistic regression, or predicting the continuous

Table 3
Criterion Validity: Accuracy of Categorizing Patients as Depressed or Non-depressed According to Psychiatrists' Diagnosis or Rating

Criterion and indicator variables	Sensitivity	Specificity	Hit Rate	False Positive	*N*
Moscow-ICD-9					
DEPI > 4	31%	57%	51%	34%	157
CDI > 3	31%	69%	61%	24%	157
Scale *2* > 65	60%	59%	59%	32%	149
Wiggins > 65	74%	49%	55%	39%	135
ICD-10					
DEPI > 4	37%	58%	53%	32%	157
CDI > 3	32%	69%	60%	24%	157
Scale *2* > 65	58%	59%	58%	31%	149
Wiggins > 65	73%	49%	55%	39%	135
HRSD					
DEPI > 4	52%	63%	57%	17%	89
CDI > 3	33%	66%	48%	16%	89
Scale *2* > 65	61%	80%	70%	9%	87
Wiggins > 65	69%	68%	69%	15%	86

Note. Moscow-ICD-9 = Soviet/Russian modification of the International Classification of Diseases, 9th ed.; ICD-10 = International Classification of Diseases, 10th ed. (Russian translation, not modified); HRSD = Hamilton Rating Scale for Depression; *DEPI* = Rorschach Depression Index; *CDI* = Rorschach Coping Deficit Index; Scale *2* = Berezin's MMPI clinical scale for depression; Wiggins = American MMPI depression content scale.

Table 4
Criterion Validity: Correlation Coefficients for the Wide Definition of Depression (Any vs. No Depression in Diagnosis)

Indicator	Depression diagnosis using Snezhnevsky criteria[a]	Moscow-ICD-9 criteria[b]	ICD-10 criteria[c]
Rorschach variables			
DEPI > 4	-.02	-.05	-.05
DEPI continuous (0-7)	-.08	-.11	-.05
CDI > 3	.05	.13	-.02
CDI continuous (0-5)	.03	.12	-.02
MMPI variables			
Scale *2* > 65	.32**	.18	.20*
Scale *2* continuous	.24*	.13	.24**
Wiggins > 65	.17	.13	.20*
Wiggins continuous	.22*	.23*	.24**

Note. Snezhnevsky = traditional Russian/Soviet "Moscow School" diagnostic system; Moscow ICD-9 = Soviet/Russian modification of the International Classification of Diseases, 9th ed.; ICD-10 = International Classification of Diseases, 10 ed. (Russian translation, not modified); *DEPI* = Rorschach Depression Index; *CDI* = Rorschach Coping Deficit Index; Scale *2* = Berezin's MMPI clinical scale for depression; Wiggins = American MMPI depression content scale.
[a]$N = 126$. [b]$N = 95$. [c]$N = 147$.
*$p < .05$. **$p < .01$.

HRSD score using linear regression ($\beta = .11$, $p = .28$). This was true whether Scale *2* or the Wiggins content scale was used in the first step, whether or not the *CDI* was introduced into the model along with the *DEPI*, or whether continuous or categorical predictor variables were used.

Hit Rates When DEPI and Scale 2 Agree

Another way of testing the incremental validity of the MMPI and Rorschach is to test whether the hit rate is higher when both tests are considered. For the group of 61 participants whose MMPI Scale *2* and Rorschach *DEPI* were either both elevated or both below the clinical threshold, the hit rates (the highest was 53% for ICD-10) are not improved over using the MMPI Scale *2* alone (which already had hit rates of 58%-70%, Table 3).

Concurrent Validity: Relationships Between Measures

To evaluate the concurrent validity (or convergent validity) of the psychological tests, they were correlated with one another. The Rorschach variables were unrelated to the MMPI variables (the largest r in the predicted direction was .08, $p = .35$, Table 5). In contrast, the MMPI variables were strongly related to each other ($r = .49$ and .67, $p < .001$, Table 5).

Content Validity: The Meaning of Measures

Theory-Based Comparisons

Following suggestions by Exner (1991, pp. 25-26), we classified the components of the *DEPI* and *CDI* into three dimensions of depression: affective, cognitive, and interpersonal/helpless (Table 6).

An MMPI scale was created for each of the three categories by a committee of Russian raters. The 21 affective items ($\alpha = .36$) included, for example, "In good weather your mood usually improves." The 29 cognitive items ($\alpha = .75$) included "You are quite indifferent to your fate." The 21 interpersonal/helpless items ($\alpha = .58$) included "Most people seem to be more pleased with their life than you." Using the general linear model, the relevant (dichotomous) Rorschach variables were included

Table 5
Concurrent Validity: Correlation Matrix of Rorschach and MMPI Scores

Indicator	*DEPI*		*CDI*		Scale *2*		Wiggins	
	Categorical	Continuous	Categorical	Continuous	Categorical	Continuous	Categorical	Continuous
Rorschach								
DEPI > 4	–							
DEPI continuous (0-7)	–	–						
CDI > 3	-.07	–	–					
CDI continuous (0-5)	–	.07	–	–				
MMPI								
Scale *2* > 65	.08	–	.03		–			
Scale *2* continuous	–	-.06	–	.00	–	–		
Wiggins > 65	.03	–	.05	–	.49***	–	–	
Wiggins continuous	–	-.10	–	-.01	–	.67***	–	–

Note. N = 173. *DEPI* = Rorschach Depression Index; *CDI* = Rorschach Coping Deficit Index; Scale *2* = Berezin's MMPI clinical scale for depression; Wiggins = American MMPI depression content scale.
***p < .001.

Table 6
Categorization of Rorschach Variables

Variable	Category
DEPI	
FV + *VF* + *V* > 0	Affective
FD > *2*	Cognitive
Color-Shading Blends > 0 *or* *S* > 2	Affective
(3*r* + (2)/*R* > .44 and *Fr* + *rF* = 0) *or* (3*r* + (2)/*R* < .33)	Cognitive
Afr < .46	Affective
Blends < 4	Cognitive
(SumShading > *FM* + *m*) *or* (*SumC'* > 2)	Affective
(*MOR* > 2) *or* (2 x *AB* + *Art* + *Ay* > 3)	Cognitive
(*COP* < 2) *or* (Isolate/*R* > .24)	Interpersonal/helpless
CDI	
EA < 6 *or* (*AdjD* < 0)	Interpersonal/helpless
(*COP* < 2) *and* (*AG* < 2)	Interpersonal/helpless
WSumC < 2.5 *or* (*Afr* < .46)	Affective
(*p* > *a* + 1) *or* (Pure *H* < 2)	Interpersonal/helpless
(*T* >1) *or* (Isolate/*R* > .24) *or* (*Fd* > 0)	Interpersonal/helpless

Note. Categories are based on Exner, 1991.

in an equation predicting each (continuous) MMPI score. The three models all failed to achieve statistical significance or to explain a meaningful amount of the variance in the MMPI score (Affective: R^2 = .024; Cognitive: R^2 = .005; Interpersonal/helpless: R^2 =.022). When the same three sets of Rorschach variables are made into scales (i.e., tally of the number of positive Rorschach criteria from Table 6, totaling up to 5 for the Affective scale, 4 for Cognitive, and 4 for Interpersonal/Helpless), and correlated with the 3 corresponding MMPI scales, the results do not improve (r between –.10 and .01).

Empirically Derived Comparisons

We hypothesized that the items that were statistically the most highly associated with a Rorschach variable would also be conceptually related to the putative meaning of the Rorschach variable. For each of the 14 Rorschach variables (the components of the *DEPI* and *CDI*), we listed the 10 MMPI items with the highest binary similarity coefficients (Gower's S2, Systat, 1992). Blind to the Rorschach variable involved, each list was read for shared themes. Next, the putative meaning of the corresponding Rorschach variable was compared to these themes. Most of the sets of MMPI items did not seem to be unified by any particular themes. Three out of the 14 sets of MMPI items did have clear themes. Of these three, one theme was very similar to the hypothesized meaning of the corresponding Rorschach variable, one was marginally similar, and one was quite dissimilar (detailed in Ritsher, 1997a).

Additional Analyses Investigating Possible Causes of the Rorschach Results

In order to probe the reasons for the poor performance of the Rorschach variables (*DEPI* and *CDI*) in the originally planned analyses, a series of additional analyses were performed, which focused on response styles on the MMPI and Rorschach, the effect of outliers, and the degree to which our sample frequencies of 107 Rorschach variables are similar to Exner's published norms.

Response Styles and Outliers

Following Archer (1996), we controlled for response style by constraining the sample to those 90 participants who had a consistent style on the Rorschach and MMPI—both tests indicating openness or constriction. Openness was defined as Lambda < 1.08 on the Rorschach and $F \geq 65$ on the MMPI. Constriction was defined as Lambda ≥ 1.08 on the Rorschach and either L or $K \geq 55$ on the MMPI. Correlations between the MMPI and *DEPI* remained weak (r = –.16, Scale *2*; r = –12, Wiggins; continuous versions) and nonsignificant, like the full sample results shown in Table 5. To test whether a few highly unusual Rorschach protocols were skewing the distributions of the variables and distorting the results, we removed protocols with $R > 37$ (following Meyer, 1993), which produced no change in the pattern of results (r = –.07, Scale *2*; r = –.02, Wiggins, $p > .05$, N = 143) .

Frequencies of Rorschach Variables in Russian Versus U.S. Samples

Cultural factors did affect the way some of the individual Rorschach responses were coded and interpreted. For example, a two-headed eagle is on the seal of Russia, which has recently become almost as ubiquitous as the hammer and sickle was during the Soviet period. The two-headed eagle was a common response to several of the Rorschach cards, but our strict adherence to the American scoring rules meant that it had to be coded as a deviant response unless the participant specified that it was a symbol rather than an actual bird with two heads. The form quality of these responses was usually quite good, and therefore scored as a "u" because it was not found on the American form quality tables. On average, our depressive sample had about one more *FQ*u response than Exner's depressive sample (mean of 6.1 vs. 5.2), two fewer *FQ*o responses (9.6 vs. 11.8), about the same number of *FQ*- responses (4.6 vs. 4.7), and about two fewer responses overall ($M R$ = 20.8 vs. 22.7). If culturally based differences like the two-headed eagle response were pervasive, we would expect the sample means for many Comprehensive System variables to be quite different from those in Exner's published norms for patients with depression.

We compared the means of 107 variables between our sample of 41 people with ICD-10 depression and Exner's sample of 315 "depressives" (Exner, 1993). As with our sample, most, but not all of Exner's

depressed sample were inpatients. Comparing demographics, our sample was similar, except somewhat younger and more urban. It is not clear what influence, if any, this would have on the data.

Of the 107 variables that we compared, 104 of our sample means were within 1 *SD* of Exner's reported mean and the remaining 3 (*Fr+rF*, *ALOG*, and Sum6) were within two *SD*s. (Exner [1993] lists 111 variables, but 4 were untestable, either because the scoring program could not export them [*Hx* and (*Ad*)] or because the *SD* was 0 [*CONTAM*, *CONFAB*].) Of the 61 variables that Exner (1993) reports to be normally distributed, 60 were within 1 *SD* in our sample and one was within 2 *SD*s. Of the variables that make up the *DEPI* and *CDI*, all were within 1 *SD* except for *Fr+rF*.

It is important to note that differences of less than 1 *SD* may represent clinically significant differences that would affect interpretation. For example, our sample had higher percentages of participants with at least one reflection response, a positive Schizophrenia Index, or an egocentricity index greater than .33. Regarding *EB* style, the Russian sample had a much lower proportion of Ambitents and much higher proportions of Introversives and Extratensives compared to Exner's depression sample. Similarly, the *D* score frequencies were different, with our sample having fewer $D = 0$, more $D > 0$, and about the same $D < 0$. More than a third of Exner's depression sample had *DEPI* scores of 6 or 7, but that was true of less than one-fifth of our sample (details are available upon request or in Ritsher, 1997a, pp. 169-179).

Although recent articles have compared sample norms to Exner's published norms (Frank, 1992; Glass, Bieber, & Tkachuk, 1996; Kaiser-Boyd, 1993), the main Rorschach research methodology text cautions against doing so, except to look for "extreme deviations from expected values" (Dies, 1995, p. 114). Therefore we present only descriptive statistics. At the level of group means for the raw variables underlying the Comprehensive System, our Russian responses were generally not "extremely deviant" from Exner's norms for people with depression.

Discussion

Overview

The study evaluated the criterion, concurrent, and content validity of MMPI and Rorschach depression indicators in a Russian clinical sample. The MMPI indicators (Berezin's Russian Scale *2* and the Wiggins depression content scale) were reliably associated with depression diagnosis (using traditional, Western, or mixed criteria) and psychiatrists' ratings (Hamilton). In contrast, the Rorschach variables (*DEPI*, *CDI*, and their components) were all quite poorly associated with all the criterion variables and MMPI scores. This lack of association did not improve when various data-sifting techniques were attempted as suggested by the recent literature. It is evident that the Rorschach did evoke a familiar pattern of responses from the Russian sample, but that the American Comprehensive System as it is currently configured has poor validity for detecting either clinically diagnosed or self-reported depression among Russian patients.

Criterion Measures

The traditional (Snezhnevsky), Western (ICD-10), and mixed (Moscow-ICD-9) diagnoses all performed similarly in terms of their overall agreement with the other depression measures. The associations were fairly weak but consistent and statistically significant. Among criterion variables, it was the Hamilton scale (HRSD) that had the strongest associations with the predictor variables. In general, the results do not favor one form of diagnosis over another in terms of agreement with the tests studied. Using formal structured clinical interviews as criterion variables in a future study would enhance precision and reliability.

MMPI

The main finding regarding the use of Berezin's Russian MMPI for depression was its criterion validity (significant associations with diagnosis and Hamilton scores, see Tables 2, 3, and 4). Berezin's version of the MMPI is the most commonly used in clinical practice in Russia, and is also commonly used to test Russians in North America. A new Russian version of the MMPI-2 is currently being

developed in the U.S. (Koscheyev & Leon, 1996; M. Atlis, University of Minnesota, personal communication, March 27, 2001). The fact that the results were more strongly supportive of the American Wiggins depression content scale than the Russian-normed Berezin clinical Scale *2* underscores the need for the re-norming effort. However, it is likely that Berezin's version will continue to be widely used in Russia for some time, due to its popularity and the difficulty of obtaining new instruments in Russia. Therefore our results are directly relevant to the use of the current standard (Berezin) in clinical practice, and are generally relevant to the cultural appropriateness of using this type of method (long true-false questionnaires) to assess Russian patients.

Rorschach

In contrast to the robust self-report findings, the Rorschach findings were weak at best and impervious to a wide range of analytic strategies. The lack of support for the criterion and incremental validity of the *DEPI* replicates many recent findings (reviewed in Jørgensen et al., 2000; Wood et al., 1999).

One explanation that we were not able to test is that the Rorschach is simply measuring something unique. Many believe that the Rorschach measures latent dynamics, personality structure, and styles of handling affect and cognition (e.g., Weiner, 1999). Therefore, a positive *DEPI* index in a person who does not report depression nor appear depressed to a diagnosing clinician might be attributable to the effect of unconscious defenses, compensatory resources, or problems with the criterion measures. The Rorschach may add clinical richness to the assessment enterprise regardless of whether it is able to increase the diagnostic hit rate. While the field awaits the results of further research, it is this line of thinking that accounts for the Rorschach's enduring popularity despite the ongoing debate about its validity (e.g., Dawes, 1994; Meyer, 1997a, 1997c; Weiner, 1996; Wood, Nezworski, & Stejskal, 1996, 1997).

The Russian and American sample means are strikingly similar—both in terms of structure (e.g., number of responses, percentage of responses using the whole inkblot) and in terms of content (e.g., number of responses containing humans), despite the culturally based error variance induced by culturally inappropriate coding of two-headed eagle responses and the like (for other examples, see Ritsher, 1997a). Furthermore, the frequency differences noted above in *EB* style and *D* scores—whatever the causes of those differences—may have affected the meaning of the *DEPI* scores in this sample. Comparing their data to Exner's published norms, researchers in Chile, Finland, Japan, Spain, and Sweden have also found substantial similarities on the various indexes studied, along with enough differences for them to uniformly stress the need for researchers in other countries to develop their own norms to use as the standard for comparison (Bylund, 1992; Hernandez-Guzman et al., 1989; Mattlar et al., 1993; Spigelman et al., 1991; Takeuchi & Scott, 1986; Vizcarro, 1986).

Exner and Weiner (1995) recommend re-norming the Form Quality tables if necessary but not making other culturally based adaptations to the Comprehensive System. Further research with Russian normative samples is necessary to determine whether this would be enough. Furthermore, there are indications that Exner's published norms may not be entirely representative of non-patient samples in the U.S. (Shaffer, Erdberg, & Haroian, 1999), which underscores the importance of using a control group in future research of this type (Dies, 1995).

Conclusion

In this Russian clinical sample, the MMPI functioned more accurately than the Rorschach in detecting depression, regardless of how it was defined. Although the Rorschach is one of the most often studied and most commonly used tests in clinical psychology, it is not yet clear which features of Exner's Comprehensive System are valid for which purposes under which conditions—particularly with regard to its use outside of North America. The fact that we achieved acceptable inter-rater-reliability and that the Russian and American variable means were relatively similar indicates that the Comprehensive System has the potential to become a valuable cross-cultural

research tool. Although our results do support the use of Berezin's Russian adaptation of the MMPI, it is clearly not currently appropriate to assess depression in Russian patients using the American norms inherent in the Comprehensive System for the Rorschach.

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Archer, R. P. (1996). MMPI–Rorschach interrelationships: Proposed criteria for evaluating explanatory models. *Journal of Personality Assessment, 67*(3), 504-515.

Archer, R. P., & Gordon, R. A. (1988). MMPI and Rorschach indexes of schizophrenic and depressive diagnoses among adolescent inpatients. *Journal of Personality Assessment, 52*, 276-287.

Archer, R. P., & Krishnamurthy, R. (1993a). A review of MMPI and Rorschach interrelationships in adult samples. *Journal of Personality Assessment, 61*, 277-293.

Archer, R. P., & Krishnamurthy, R. (1993b). Combining the Rorschach and the MMPI in the assessment of adolescents. *Journal of Personality Assessment, 60*, 132-140.

Archer, R. P., & Krishnamurthy, R. (1997). MMPI-A and Rorschach indexes of depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment, 69*, 517-533.

Balatsky, G., & Diener, E. (1993). Subjective well-being among Russians. *Social Indicators Research, 28*, 225-243.

Berezin, F. B., Miroshnikov, M. P., & Rozhanets, R. V. (1976). *Metodika mnogostoronnego issledovaniya lichnosti* [The multiscale personality test]. Moscow: Meditsina.

Berezin, F. B., Miroshnikov, M. P., & Sokolova, E. D. (1994). *Metodika mnogostoronnego issledovaniya lichnosti: Struktura, osnovy interpretatsii, nekotorye oblasti primeneniya* [Multiscale personality test: Structure, principles of interpretation, some areas of application]. Moscow: Folium.

Berry, J. W., Poortinga, Y. H., & Pandey, J. (1996). *Handbook of cross-cultural psychology, Vol. 1: Theory and method* (2nd ed.). Boston: Allyn & Bacon.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology: Research and applications*. Cambridge, United Kingdom: Cambridge University Press.

Bespalko, I. G. (1978). Normativnye dannye po metodike Rorshakha [Normative data for the Rorschach method]. In *Psikhologicheskie metody issledovaniya lichnosti v klinike* [Psychological methods of personality research in the clinic]. Leningrad, Russia: Meditsina.

Bespalko, I. G. (1985). O diagnosticheskikh vosmozhnostyakh populyarnykh otvetov v metode Rorshakha [On the diagnostic possibilities of popular answers in the Rorschach method]. *Psikhologicheskaya diagnostika pri nervno-psikhicheskikh i psikhosomaticheskikh zabolevaniyakh* [Psychological diagnosis in nervous-mental and psychosomatic illnesses]. Sbornik nauch. trudov, tom 112, Leningrad: Bekhterev Institute.

Boyer, L. B., Dithrich, C. W., Harned, H., Hippler, A. E., Stone, J. S., Walt, A., & Boyer, R. M. (1989). The relation between psychological states and acculturation among the Tanaina and Upper Tanana of Alaska: An ethnographic and Rorschach study. *Ethos, 17*, 450-479.

Butcher, J. N. (1996). *International adaptations of the MMPI-2*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the administration and scoring of the MMPI-2*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Nezami, E., & Exner, J. E. (1998). Psychological assessment of people in diverse cultures. In S. S. Kazarian & D. R. Evans (Eds.), *Cultural clinical psychology* (pp. 61-105). New York: Oxford University Press.

Bylund, M. (1992). Women in exile and their children. *Women and Therapy, 13*, 53-63.

Calloway, P. (1993). *Russian/Soviet and Western psychiatry: A contemporary comparative study*. New York: John Wiley & Sons.

Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology*. Boston: Allyn & Bacon.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Dawes, R. M. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment, 11*, 297-302.

DeVos, G. A., & Boyer, L. B. (1989). *Symbolic analysis cross-culturally: The Rorschach test*. Berkeley: The University of California Press.

Dies, R. R. (1994). The Rorschach Comprehensive System: Current status and clinical applications. *Independent Practitioner, 14*, 96-106.

Dies, R. R. (1995). Subject variables in Rorschach research. In J. E. Exner, Jr. (Ed.), *Issues and methods in Rorschach research* (pp. 99-121). Mahwah, NJ: Erlbaum.

Exner, J. E., Jr. (1991). *The Rorschach: A comprehensive system, Vol. 2: Interpretation* (2nd ed.). New York: John Wiley & Sons.

Exner, J. E., Jr. (1993). *The Rorschach: A comprehensive system, Vol. 1: Basic foundations* (3rd ed.). New York: John Wiley & Sons.

Exner, J. E., Jr. (1994). President's message. *Bulletin of the International Rorschach Society, 4*, 1.

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

Exner, J. E., Jr., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system, Vol. 3: Assessment of children and adolescents* (2nd ed.). New York: John Wiley & Sons.

Frank, G. (1992). The response of African Americans to the Rorschach: A review of the research. *Journal of Personality Assessment, 59*, 317-325.

Ganellen, R. J. (1996a). Exploring MMPI-Rorschach relationships. *Journal of Personality Assessment, 67*(3), 529-542.

Ganellen, R. J. (1996b). *Integrating the Rorschach and MMPI-2 in personality assessment*. Mahwah, NJ: Erlbaum.

Glass, M. H., Bieber, S. L., & Tkachuk, M. J. (1996). Personality styles and dynamics of Alaska native and nonnative incarcerated men. *Journal of Personality Assessment, 66*, 583-603.

GosKomStat Rossii. (1995). *Zdravookhranenie v rossiiskoi federatsii: Statisticheskii sbornik* [Health in the Russian Federation: Statistical handbook]. Moscow: Author.

GosKomStat Rossii. (1999a). *Handbook "Russia in figures,"* accessed via website www.gks.ru. Moscow: Author.

GosKomStat Rossii. (1999b). Poverty level change rate. *Current Statistical Survey, 1*(28), accessed via website www.gks.ru. Moscow: Author.

GosKomStat Rossii. (1999c). Mortality of juveniles (adolescents) by main classes of causes of death. *Current Statistical Survey, 3*(30), accessed via website www.gks.ru. Moscow: Author.

Graham, J. R. (1977). *The MMPI: A practical guide*. New York: Oxford University Press.

Hall, G. C., Bansal, A., & Lopez, I. R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment, 11*, 186-197.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56-62.

Hathaway, S. R., & McKinley, J. C. (1943). *Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.

Hernandez-Guzman, P. A., Rey-Clericus, R. J., San Martin-Petersen, C. L., & Vinet-Reichhardt, E. V. (1989). Diferencias entre las culturas angloamericana y latinoamericana en las respuestas dadas al psicodiagnostico de Rorschach [Differences between Anglo-American and Latin American cultures in the answers given to the Rorschach psychodiagnostic test]. *Terapia Psicologica, 8*, 62-66.

Howes, R. D., & DeBlassie, R. R. (1989). Modal errors in the cross-cultural use of the Rorschach. *Journal of Multicultural Counseling and Development, 17*, 79-84.

Jørgensen, K., Andersen, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review. *Assessment*, 7, 259-279.

Kaiser-Boyd, N. (1993). Rorschachs of women who commit homicide. *Journal of Personality Assessment, 60*, 458-470.

Kleinman, A., & Good, B. (1985). *Culture and depression: Studies in the anthropology and cross-cultural psychiatry of affect and disorder*. Berkeley: University of California Press.

Koscheyev, V. S., & Leon, G. R. (1996). The Russian translation and preliminary adaptation of the MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2*. Minneapolis: University of Minnesota Press.

Levitt, E. E., & Gotts, E. E. (1995). *The clinical application of MMPI special scales*. Hillsdale, NJ: Erlbaum.

Lindzey, G. (1961). *Projective techniques and cross-cultural research*. New York: Appleton-Century-Crofts.

Lipovsky, J. A., Finch, A. J., & Belter, R. W. (1989). Assessment of depression in adolescents: Objective and projective measures. *Journal of Personality Assessment, 53*(3), 449-458.

Mattlar, C. E., Carlsson, A., & Forsander, C. (1993). The issue of the popular response: Definition and universal vs. culture-specific popular responses. *British Journal of Projective Psychology, 38*, 53-62.

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

Meyer, G. J. (1997a). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment, 9*, 495-498.

Meyer, G. J. (1997b). On the integration of personality assessment methods: The Rorschach and MMPI. *Journal of Personality Assessment, 68*(2), 297-330.

Meyer, G. J. (1997c). Thinking clearly about reliability: More critical corrections regarding the Rorschach Comprehensive System. *Psychological Assessment, 9*, 480-489.

Meyer, G. J. (1999). Simple procedures to estimate chance agreement and kappa for the interrater reliability of response segments using the Rorschach Comprehensive System. *Journal of Personality Assessment, 72*, 230-255.

Ministerstvo Zdravokhraneniya SSSR. (1982). *Psikhicheskie Rasstroistva (razdel V "Mezhdunarodnoi statisticheskoi klassifikatsii boleznei, travm it prichini smerti 9-ogo peresmotra," adaptirovannii dlya ispolzovaniya v SSSR* [Mental disturbances (section V of "International classification of diseases, traumas, and reasons for death, 9th edition," adapted for use in the USSR]. Moscow: Author.

Ritsher, J. E. B. (1997a). Depression assessment of Russian psychiatric patients: Validity of MMPI and Rorschach scales. *Dissertation Abstracts International, 58*, 11B, 6246.

Ritsher, J. E. B. (1997b). Redefining Russian clinical psychology. *Journal of Russian and East European Psychology, 35*, 6-12.

Ritsher, J. B., Ryder, A. G., Karasz, A., & Castille, D. M. (in press). Integrating qualitative and quantitative approaches in the study of psychopathology across cultures. *New directions in cross-cultural psychology*. Warsaw: Polish Academy of Sciences.

Ritsher, J. E. B., Warner, V., Johnson, J., & Dohrenwend, B. (in press). Inter-generational longitudinal study of social class and depression: A test of social causation and social selection models. *British Journal of Psychiatry, 178*(Suppl. 40), s84-s90.

Ritzler, B. A. (1996). Projective methods for multicultural personality assessment. In L. A. Suzuki, P. J. Meller, & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment* (pp. 115-135). San Francisco: Jossey-Bass.

Sartorius, N., & Ban, T. A. (1986). *Assessment of depression*. New York: Springer-Verlag.

Sells, J. E. (1990). *A validity study of the DEPI index: The Rorschach comprehensive system*. Unpublished doctoral dissertation, University of Utah, Salt Lake City.

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

Snezhnevsky, A. V. (1983). *Rukovodstvo po psikhiatrii* [Handbook of psychiatry]. Moscow: Meditsina.

Sobchik, L. N. (1990). *Standartizirovannii mnogofaktornii metod issledovaniya lichnosti* [Standardized multifactorial method of studying personality]. Moscow: Moskva.

Spigelman, A., Spigelman, G., & Engelesson, I. L. (1991). Cross cultural differences between American and Swedish children regarding their Egocentricity Index. *Tidsskrift for Norsk Psykologforening, 28*, 316-319.

Systat. (1992). SYSTAT: Statistics (Version 5.2) [computer software]. Evanston, IL: Author.

Takeuchi, M., & Scott, R. (1986). Educational productivity and Rorschach location responses of preschool Japanese and American children. *Psychology in the Schools, 23*, 368-373.

Triandis, H. C., & Berry, J. W. (Eds.). (1980). *Handbook of cross-cultural psychology, Vol 2*. Boston: Allyn & Bacon.

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Vizcarro, C. (1986). Utilidad del Rorschach en el diagnostico de la esquizofrenia [Utility of the Rorschach in diagnosing schizophrenia]. *Evaluacion Psicologica, 2*, 79-88.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach inkblot method. *Psychological Assessment, 8*(2), 206-213.

Weiner, I. B. (1998). *Principles of Rorschach interpretation*. Mahwah, NJ: Erlbaum.

Weiner, I. B. (1999). What the Rorschach can do for you: Incremental validity in clinical applications. *Assessment, 6*, 327-340.

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H., Joyce, P. R., Karam, E. G., Lee, C., Lellouch, J., Lepine, J., Newman, S. C., Rubio-Stipec, M., Wells, J. E., Wickramaratne, P. J., Wittchen, H., & Yeh, E. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association, 276*, 293-299.

Wood, J. M., & Lilienfield, S. O. (1999). The Rorschach inkblot test: A case of overstatement? *Assessment, 6*, 341-349.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science, 7*(1), 3-17.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1997). The reliability of the Comprehensive System for the Rorschach: A comment on Meyer (1997). *Psychological Assessment, 9*, 490-494.

Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999). Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996). *Assessment, 6*, 115-129.

World Health Organization. (1992). *Mezhdunarodnaya klassifikatsiya boleznei, (10-ogo peresmotr): Klassifikatsiya psikhicheskikh i povedencheskikh rasstriostv* [International classification of diseases (10th ed.): Classification of mental and behavioral disturbances]. (Yu. L. Nuller & S. Yu. Tsirkin, Eds. of Russian language version). Moscow: Author.