The Diagnostic Efficiency of the Rorschach Depression Index and the Schizophrenia Index: A Review

Kasper Jorgensen, Tom J. Andersen and Henrik Dam

Assessment 2000; 7; 259
DOI: 10.1177/107319110000700306

ASSESSMENT
2000, Volume 7, Number 3, 259-280

THE DIAGNOSTIC EFFICIENCY OF THE RORSCHACH DEPRESSION INDEX AND THE SCHIZOPHRENIA INDEX: A REVIEW

Kasper Jørgensen
Department of Neurology, Bispebjerg Hospital, Copenhagen, Denmark

Tom J. Andersen
Henrik Dam
Department of Psychiatry, Rigshospitalet, Copenhagen, Denmark

This review focuses on the diagnostic efficiency of the new versions of the Rorschach Comprehensive System Depression Index (*DEPI*) and the Schizophrenia Index (*SCZI*). Clinical diagnosis according to the *Diagnostic and Statistical Manual of Mental Disorders* was chosen as the external validation criteria. The sensitivity, specificity, and overall classification rates for the indices were presented from the studies or computed from the data when possible. The positive and negative predictive validity was estimated at three different base rates. As regards the *DEPI* the results showed a large variation in diagnostic performance as the index seemed to have relatively more success in identifying nonpsychotic and unipolar depression than psychotic and bipolar depression. The *DEPI* did not successfully identify depression among adolescent patients. As regards the *SCZI* the results more consistently indicated that the index effectively discriminates between psychotic and nonpsychotic patients and the predictive validity of both a positive and negative *SCZI* was found to be high.

Keywords: Rorschach, depression, schizophrenia, DEPI, SCZI

The *Rorschach Comprehensive System* (CS; Exner, 1978, 1986, 1991) includes a number of empirically derived constellation indices among which are the *SCZI*, aimed at facilitating the identification of schizophrenia, and the *DEPI*, aimed at the identification of depression (Exner, 1991, 1993, 1995). The indices were developed and validated by Exner and his coworkers, but subsequent independent studies on the diagnostic efficiency of the indices have produced diverse and sometimes dissimilar results, in particular regarding the *DEPI*. In a review of recent research addressing the utility of the Rorschach, Viglione (1999) concluded that on the basis of the available data, one cannot recommend routine application of the *DEPI* for diagnostic purposes. Similarly, Wood, Nezworski, Stejskal, Garven, and West (1999) concluded that independent peer-reviewed studies of the *DEPI* have nearly all found that it is unrelated to diagnoses of depression in either adolescents or adults.

We wish to thank Gregory J. Meyer for sharing information regarding his clinical database with us. We wish to thank both Gregory J. Meyer and James M. Wood for their comments on the manuscript and help with literature search. We also wish to thank Thomas W. Teasdale for help on the language.

Correspondence concerning this article and requests for offprints should be addressed to Kasper Jørgensen, MA, Department of Neurology, Bispebjerg Hospital, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark. E-mail kj03@bbh.hosp.dk

The diagnostic performance of an index or a test can be described in terms of *sensitivity*, *specificity*, and diagnostic *hit rate* or *overall correct classification* (Kessel & Zimmerman, 1993; Sackett, 1992). In order to calculate these statistics an external criterion or "gold standard measure" for the target disorder must be chosen, such as clinical diagnosis according to the criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM;* 3rd ed., 1980; 3rd ed. rev., 1987; 4th ed., 1994). Diagnostic efficiency statistics thus relate to the concept of criterion validity. Sensitivity, also known as detection rate, measures the proportion of those with the target disorder, according to external criteria, who are correctly identified by the index [true positives / (true positives + false negatives)]. Specificity measures the proportion of those without the target disorder who are correctly identified by the index as not having the target disorder [true negatives / (true negatives + false positives)]. The term diagnostic hit rate or overall correct classification is a general measure of the diagnostic performance of the test, as it indicates which proportion of the total number of individuals (both with and without the target disorder) are correctly classified [(true positives + true negatives) /(sum of individuals in the target group and the control group)]. The diagnostic efficiency statistics are based on the implicit assumption that the external criterion is objective, that is that the diagnostic classification is accurate in the first place. If the clinical diagnosis that constitutes the external criterion is less than accurate, the relevance of the diagnostic efficiency statistics is reduced. The reliability of the *DSM-III* and *DSM-III-R* criteria for schizophrenia has been examined by Flaum et al. (1998). The interrater reliability of the *DSM-III* criteria and the *DSM-III-R* criteria for schizophrenia were both reported to have kappa values of .77 (.05 standard error). The test-retest reliability for the *DSM-III* criteria for schizophrenia was reported as a kappa value of .79 (.05 standard error) whereas the test-retest reliability for the *DSM-III-R* criteria was reported as a kappa value of .74 (.05 standard error). Although all the reported kappa values are high, the data indicate that clinical diagnosis according to *DSM* criteria should not be considered as an absolutely objective criterion.

Sensitivity and specificity provide important information about the psychometric properties of a test. But a diagnostician in a clinical context would also like to know (a) the probability that an individual has the target disorder given the index is positive, and (b) the probability that an individual does not have the target disorder given the index is negative. The predictive value of a positive test is known as the *positive predictive value* (PPV) and is calculated as [true positives / (true positives + false positives)]. The predictive value of a negative test is known as the *negative predictive value* (NPV) and is computed as [true negatives / (true negatives + false negatives)]. The PPV and the NPV depend upon both the sensitivity and the specificity of the index but they are also determined by the *prevalence* or the base rate of the disease of interest. When the base rate is increased, so is the PPV (it is more likely to draw a patient with the target disorder) whereas the NPV is decreased, and vice versa. In many studies the PPV and the NPV are estimated directly from the study sample, but this is inappropriate if the base rate in the study is markedly different from the typical clinical population. It is also unrealistic to compare PPV and NPV estimated from different samples without considering differences in base rates. Alternatively, if the sensitivity and the specificity is known, the PPV and NPV can be estimated for individually selected base rates of the disease of interest (Foldspang, Juul, Olsen, & Sabroe, 1986; Sackett, 1992).

Literature search

Searches in MEDLINE and PsycINFO were used to identify potentially relevant studies published between January 1990 and December 1999. The start of this period was chosen because we wanted to include only studies using the most recently revised versions of the *DEPI* and the *SCZI*. We excluded studies involving only children but included studies involving adolescent samples. We were primarily interested in studies including clinical samples diagnosed by *DSM* criteria but because of the limited number of such studies available, we decided also to include studies including nonclinical samples where no systematic diagnostic procedure was applied. Search terms

were *Rorschach, DEPI, SCZI, depression, and schizophrenia*. The number of references found when searching MEDLINE (1990-1999) was: Rorschach: 386; *DEPI*: 6; *SCZI*: 3; Rorschach and depression: 54; Rorschach and schizophrenia: 47. The number of references found when searching PsycINFO (1990-1999) was: Rorschach: 790; *DEPI*: 15; *SCZI*: 13; Rorschach and depression: 170; Rorschach and schizophrenia: 113. A total of 62 references were retrieved for review. Of these 35 were excluded for various reasons (limited relevance, child samples, previous versions of the indices).

The Depression Index (*DEPI*)

An experimental depression index constituted by five variables was introduced in 1982 (Exner & Weiner, 1982), but according to subsequent independent studies the index had unsatisfactory diagnostic performance as the sensitivity in particular was too low (Archer & Gordon, 1988; Ball, Archer, Gordon, & French, 1991; Lipovsky, Finch, & Belter, 1989; Viglione, Brager, & Haller, 1988). In 1990 the revised *DEPI* was published, constituted by 15 structural variables related to either cognitive or affective depressive features (Exner, 1990). These 15 variables are combined into 7 constellation criteria or tests, 5 of which must be met for the index to be considered positive (see Table 1). A *DEPI* value of 5 indicates that the presence of a depression is likely but not necessarily definitive, whereas values of 6 or 7 should provide more certainty (Exner, 1991). The revised *DEPI* was derived from statistical analysis of protocols from patients with diagnosed depression. In 1986, Exner described a protocol pool including 812 depressed cases, 680 of which were of the dysthymic or unipolar variety, and 132 were bipolar and schizoaffective cases. In 1991 the pool had grown to include 1,421 protocols from individuals with "*DSM-III-SADS* diagnosed first-admission major affective disorders" (Exner, 1991, p. 23). Using non-test data this sample was subdivided into a target sample of depressed patients (n = 471) as distinct from a sample of patients characterized as "helpless in the face of contending with a complex society" (n = 213). The revised *DEPI* correctly identified 85% of the cases from the target sample of depressives as opposed to only 17% of the "helpless" sample. In Exner's psychiatric reference sample (n = 315) of inpatient depressives (Exner, 1995, Table 23), 75% have *DEPI* values equal to or greater than 5 as compared with only 4% of a sample (n = 700) of adult nonpatients (Exner, 1995, Table 11). Wood, Nezworski, and Stejskal (1996) raised doubts about the diagnostic criteria for the selection of depressive cases in Exner's data mentioning the risk of *criterion contamination* which means that Rorschach data may have entered into the diagnostic procedure and inflated the diagnostic properties of the *DEPI*. In a reply to this Exner (1996) stated that the patients were diagnosed independently of Rorschach data with diagnoses derived by means of the Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robbins, 1985), the Schedule for Affective Disorders and Schizophrenia structured interview (SADS; Spitzer & Endicott, 1978), or both. Using descriptive data reported by Exner (1991), Ganellen (1996) calculated diagnostic efficiency statistics for the revised *DEPI*. The sensitivity was reported as .75 and the specificity was .81 when calculated from a schizophrenic control group (n = 320) alone and .92 when calculated from a mixed clinical and nonclinical control group (n = 1,020). Ganellen recognized that his evaluation of the *DEPI*'s diagnostic efficiency must be considered preliminary as it was based exclusively on Exner's data with no attempt of cross-validation through independent studies. Using a very similar method we computed the diagnostic performance of the *DEPI* in two ways. The classification rates for *DEPI* reported by Exner (1995, Table 23) was compared first to the classification rates reported for

Table 1
The Rorschach Depression Index (DEPI)

1. (FV+VF+V > 0) **or** (FD > 2)
2. (Color-Shading Blends > 0) **or** (S > 2)
3. (3r+(2)/R > .44 *and* Fr+rF = 0) **or** (3r+(2)/R < .33)
4. (Afr < .46) **or** (Blends < 4)
5. (SumShading > FM+m) **or** (Sum C'> 2)
6. (MOR > 2) **or** (2AB+Art+Ay > 3)
7. (COP < 2) **or** (Isolate/R > .24)

Note. The Index is positive if 5 or more conditions are true.

700 nonpatients (Exner, 1995, Table 11), and second to the rates reported for a mixed clinical sample comprised of 320 inpatient schizophrenics (Exner, 1995, Table 21) and 180 patients with character disorders (Exner, 1995, Table 25). As expected the diagnostic efficiency statistics are similar to the statistics reported by Ganellen (1996) but computing the statistics based on comparisons with a pure nonclinical and a pure clinical sample respectively seems more appropriate for two reasons: (a) it illustrates the impact of the control comparison on the performance of the index, and (b) it allows for comparison with subsequent independent studies, as the majority of these use either clinical or nonclinical control groups (see Table 2). The impact of the control condition is illustrated by the fact that the *DEPI* performs relatively better when comparing depressed patients with nonclinical controls than when comparing with another clinical sample.

Vincent and Harman (1991) analyzed the clinical validity of the CS by comparing the descriptive data for three of Exner's psychiatric reference groups (inpatient schizophrenics, inpatient depressives, and character disorders) with Exner's adult nonpatient sample. They examined how many of the CS structural variables in each of the three reference groups differed significantly (defined as ± 2 *SD*) from the nonpatient mean. According to Vincent and Harman these variables can be said to meet the criteria for "clinical significance" as they have adequate discriminative properties in a clinical assessment situation (when $n = 1$). In the depression sample, only 12 of 111 variables met the criteria for clinical significance and only one variable (*pure C*) was identified as specific for the depression sample. They concluded that the CS demonstrated little differential utility for detecting depression. In our view the relevance of Vincent and Harman's conclusion is limited by the fact that their criterion for clinical significance is probably too restrictive to be clinically meaningful.

Clinical Studies Including Adult Samples

In a study on the validity of the *DEPI*, Sells (1990/1991) reported Rorschach data on a group of 29 adult inpatients with *DSM-III-R* discharge diagnoses of depression (major depression = 28; dysthymia = 1), a group of nondepressed inpatients with personality disorder and schizophrenia ($n = 25$), and a diagnostically mixed group of inpatients with both affective disorder and concurrent personality disorder ($n = 55$). Judging from the description of the study design there was a possibility of criterion contamination as the majority of the patients were tested as part of a clinical procedure prior to their discharge. Based on the classification rates reported by Sells when comparing the depressed and the nondepressed group and excluding the diagnostically mixed group we calculated diagnostic efficiency statistics for the *DEPI* finding a sensitivity of .62 and a specificity of .56 (see Table 2). The low specificity should be seen in the light of the fact that Sells used a clinical control group. Sells also compared the performances of the original and the revised *DEPI* concluding that the revised *DEPI* improved the hit rate considerably in his study sample.

Singer and Brabender (1993) examined the effectiveness of the Rorschach in differentiating subtypes of affective disorders in a sample of 62 inpatient depressives diagnosed by means of both the SADS structured diagnostic interview according to the RDC criteria and *DSM-III-R* admitting diagnoses. The diagnoses were determined prior to the administration of the Rorschach. Patients were excluded if they showed evidence of psychosis, if they were found to be organic or below average intelligence, or if they had an Axis II diagnosis. The sample was subdivided into three groups: unipolar depressed ($n = 29$), bipolar depressed ($n = 15$), and bipolar manic ($n = 18$). Singer and Brabender reported that the *DEPI* identified 59% of the patients in the unipolar group, 37% of the bipolar manic group, and only 26% of the bipolar depressed group. We adjusted the percentage of the bipolar manic group identified by the *DEPI* to 39% (7 of 18 patients). Based on these classification rates we computed the sensitivity of the *DEPI* for the combined sample of depressives as .45. Singer and Brabender concluded that despite the limitations of their study design (small group sizes, no nonclinical or nonaffective control groups) the results suggest that the *DEPI* more successfully identifies depression in a unipolar than in a bipolar population. They also

Table 2
Diagnostic Efficiency Statistics for the Rorshach DEPI Applied to Adults

Author	*N*	Age	Diagnostic criteria	Blind	SN	SP	OCC	Base rate	PPV	NPV
Exner (1995)	D = 315 NC = 700	Adults	SADS/RDC	yes	.75	.96	.90	31%	.90	.90
Exner (1995)	D = 315 CC[a] = 500	Adults	SADS/RDC	yes	.75	.83	.80	39%	.74	.84
Sells (1990/1991)	D = 29 CC[b] = 25	Adults	*DSM-III-R*	no	.62	.56	.59	54%	.62	.56
Caine et al. (1995)	D = 20 NC = 20	Adults	*DSM-III-R*	no	.10	.80	.45	50%	.33	.47
Jansak (1996/1997)	D[c] = 60 NC = 30	Adults	*DSM-IV*/SCID-IV	yes	.42	.93	.59	67%	.93	.44
Ilonen et al. (1999)	D = 70 NC = 60	Adults	*DSM-IV*	yes	.51	.88	.68	54%	.84	.61
Ilonen et al. (1999)	D = 70 CC[d] = 27	Adults	*DSM-IV*	yes	.51	.59	.54	72%	.77	.32
Meyer (personal communication, 2000)	D = 267 CC[e] = 65	Adults	*DSM-III-R*	yes	.49	.63	.52	80%	.85	.23
Ball et al. (1991)	D[c] = 45 CC[e] = 54	Adolescents	*DSM-III-R/DSM-IV*	no	.18	.76	.40	45%	.35	.49
Archer & Krishnamurthy (1997)	D = 56 CC[f] = 96	Adolescents	*DSM-III-R*	yes	.36	.71	.58	37%	.42	.65

Note. Blind = diagnostic clinicians were blind to Rorschach data; SN = sensitivity; SP = specificity; OCC = overall correct classification or hit rate; Base rate = percentage of depressed participants in study sample; PPV = positive predictive validity; NPV = negative predictive validity; D = number of depressed participants; NC = number of nonclinical controls; CC = number of clinical controls; SADS = Schedule for Affective Disorders and Schizophrenia; RDC = Research Diagnostic Criteria.
[a]Inpatient schizophrenics (n = 320) and character disorders (n = 180). [b]Personality disorder and schizophrenia. [c]Major depression and/or dysthymia. [d]First-episode schizophrenia. [e]Diagnostically mixed. [f]Conduct disorder and adjustment disorder.

noted that the *DEPI* may not be attuned to the ideational disturbances seen in bipolar disorders.

In a study concerning the impact of response frequency (*R*) on the CS constellation indices, Meyer (1993) reported a lack of diagnostic discrimination for the *DEPI* since total scores on the index did not differ significantly between patients with and without a diagnosis of depression. Meyer's sample consisted of psychiatric patients (n = 90) diagnosed by *DSM-III-R* criteria and selected from a larger pool of patients on the basis of response productivity criteria. When examining the *DEPI* the sample was subdivided into a group of patients with various diagnoses of depression (major depression, n = 33; depressive disorders, including major depression, bipolar disorder with depressed mood or mixed affective features, depressive disorder NOS, and dysthymia, n = 48) and a clinical control group of patients with other psychiatric diagnoses. Only 16 patients in the sample were reported to be without any depressive symptoms. The *DEPI*'s lack of diagnostic discrimination was seen when both narrow and broad diagnostic classifications were utilized and in both high-*R* ($R \geq 29$) and low- *R* protocols ($R \leq 17$). A moderating impact of *R* on the *DEPI* was found as high-*R* protocols were positive on the *DEPI* significantly more often than low-*R* protocols.

Based upon an extended data set of 332 patients split into a subgroup (n = 267) with depressive disorder (major depression, bipolar disorder with mixed or depressed features for the most recent episode, dysthymic disorder, depressive disorder NOS) and a clinical control group (n = 65) with other psychiatric diagnoses, Meyer (personal communication, February 2000) reported a sensitivity for the *DEPI* of .49 and a specificity of .63 (see Table 2). About 88% of the sample was diagnosed according to *DSM-III-R* criteria and the rest was diagnosed according to *DSM-IV* criteria. The diagnoses were assigned independently of Rorschach data, but as they were assigned for billing purposes as part of usual clinical practice they cannot be considered a "gold standard." A small, but significant difference was found when comparing mean *DEPI* scores in the depressive disorder group and in the clinical control group (t = 2.19, p = .029; see Table 6). Furthermore, based on a section of this extended data set (n = 262) Meyer reported a small but significant correlation between the *DEPI* and diagnosis of depressive disorder (r = .111, p = .037, one-tailed).

As part of a study focused on the possibility of malingering depression on the Rorschach and MMPI-2, Caine, Kinder, and Frueh (1995) compared Rorschach data from a group of depressed inpatient adult females diagnosed by the means of *DSM-III-R* (n = 20) with a nonclinical control group of 20 female university students recruited from undergraduate psychology courses (an experimental group of 20 students instructed to simulate depression was also included). In the inpatient group, 15 patients met *DSM*-criteria for major depression and 5 patients had other types of depressive diagnoses (adjustment disorder with depressed mood, depressive disorder NOS, etc.). The Rorschach data for this group were drawn from patient files, meaning that diagnosis might have been influenced by Rorschach data. In the nonclinical control group Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was used as a screening measure to exclude patients with depressive symptoms. Contrary to expectations, two patients in the depressed group and four patients in the control group obtained *DEPI* scores equal to or greater than 5. Based on these classification rates we calculated diagnostic efficiency statistics for the *DEPI* and found very low sensitivity (see Table 2). However, the inpatient depressives were found to produce significantly more responses with morbid content (*MOR*) than the control group. Discussing the *DEPI*'s seeming lack of sensitivity in this patient group, Caine et al. stated that the patients' length of hospital stay was relatively brief (average 18.6 days) suggesting the presence of reactive (or mild) depression rather than an enduring affective disturbance.

In a study on depression heterogeneity Jansak (1996/1997) tested a group of adults with depressive disorder (major depression and/or dysthymia; n = 60) and a nonclinical control group of volunteers (n = 30) matched according to gender and age. The depressive disorder group was diagnosed

prior to Rorschach testing according to *DSM-IV* criteria by means of the Structured Clinical Interview for *DSM-IV* Axis I Disorders (SCID-IV; First, Spitzer, Gibbon, & Williams, 1995). The volunteers were screened for depressive symptoms and those meeting the criteria for past depressive disorder were excluded. Jansak found the *DEPI* to be a highly specific (.93), but not sensitive (.42), measure of depression in this sample (see Table 2). She also examined the contributions of each of the 7 *DEPI* constellation criteria and the 15 single variables in the *DEPI* to the classification of depressive disorder. Of the seven constellation criteria, only one (Criterion 6) contributed significantly to the classification but two (Criteria 2 and 7) approached significance at the $p < .05$ level (Fisher's Exact Test, two-tailed). Of the 15 single variables in the *DEPI* none contributed significantly to the classification, but two variables (Sum C' > 2) and (2AB+Art+Ay > 3) approached significance. In our view, a two-tailed significance test is probably too conservative here as realistic hypotheses regarding the direction of the difference between the two groups can be formed on the basis of previous research. Actually, for all of the seven *DEPI* constellation criteria and for 14 of the 15 single variables in the *DEPI* the group differences went in the direction that could be predicted from Exner (1991). Jansak found one constellation criterion (Criterion 3) that did not seem to contribute to the diagnostic accuracy of the index commenting that the *DEPI* might become more precise by excluding this criterion. Erlebnistypus (*EB*) was found to have a moderating impact on the *DEPI* as 12 of the 15 (80%) depressed individuals with an extratensive *EB* (sum of human movement less than weighted sum of color responses by 1.5 or more) had a positive *DEPI* compared to only 3 of 12 (25%) of the introversives and 10 of 33 (30%) of the ambitents (see also Viglione, 1999).

Carlson, Kula, and St. Laurent (1997) examined the ability of the *DEPI* to identify depression in inpatients with *DSM-III-R* diagnosed major depression ($n = 20$) and inpatients with major depression and concurrent borderline personality disorder ($n = 20$). The Rorschach was administered as part of a diagnostic assessment at the time of the intake and the psychologist providing the assessment took part in diagnosing the patients. Carlson et al. reported that the *DEPI* identified 7 of 20 (35%) in the major depression group and only 4 of 20 (20%) in the major depression/borderline personality disorder group. Using the Coping Deficit Index (*CDI*) as a "second depression index" they reported classification rates for the *DEPI* and the *CDI* in combination finding that 11 of 20 (55%) in the major depression group were identified by either one or both indices.

Ilonen et al. (1999) examined the diagnostic efficiency of the *DEPI* for detecting severe depression with and without psychotic features diagnosed by *DSM-IV* criteria independently of Rorschach data. They tested a sample of 70 adult Finnish patients with depression (nonpsychotic depression, $n = 29$; psychotic depression, $n = 28$; and bipolar I disorder, $n = 13$). Two control groups were included: a clinical control group consisting of patients with first-episode schizophrenia ($n = 27$) and a nonclinical control group of volunteers ($n = 60$). The three groups were comparable in age and education. Ilonen et al. reported a moderate sensitivity (.51) for the *DEPI* applied on the total depression sample (see Table 2). However, when examining specific subtypes of depression, the sensitivity was reported as .79 for the identification of nonpsychotic depression, .46 for the identification of bipolar depression, and only .25 for the identification of psychotic depression. The specificity varied from .59 for the clinical control group to .88 for the nonclinical control group. The poor sensitivity of the *DEPI* regarding psychotic depression was discussed by the authors noting that this subgroup produced significantly fewer responses than the other groups (and high *Lambda* as well) thus reducing the likelihood of a positive *DEPI* (cf. Meyer, 1993). They also found that their results may support the notion that psychotic depression should be considered a distinct disorder rather than a subtype of severe major depression.

Clinical Studies Involving Adolescent Samples

In a study including 99 adolescent psychiatric patients (12 through 18 years old; mean age = 15 years) Ball et al. (1991) examined the relationship

between the *DEPI* and *DSM-III* or *DSM-III-R* discharge diagnoses of depression. The adolescent sample was divided into a target group of patients with dysthymia or major depression (n = 45) and a clinical control group with other diagnoses (conduct disorder, n = 26; personality disorder, n = 13; adjustment disorder or developmental disability, n = 12; and schizophrenia or schizophreniform psychosis, n = 3). A sample of outpatient children was also included in the study. The Rorschach was administered at admission to inpatient service and the discharge diagnoses represented the consensual judgement of the inpatient treatment meaning that Rorschach data might have influenced diagnosis. Ball et al. reported a very low sensitivity for the *DEPI* (.18) but a moderate specificity (.76) (see Table 2). When the target sample was limited to include only cases with major depression, a small and almost negligible increase of sensitivity (.24) was found. In Exner's sample of nonpatient 15-year olds (n = 110) there are no cases with *DEPI* scores equal to or greater than 5 equivalent to a specificity of 1.00 (Exner, 1995, Table 19). Ball et al. also examined the diagnostic performance of the original *DEPI* which identified none of the depressed adolescents (.00 sensitivity) and only one of the adolescents in the clinical control group (.98 specificity).

Archer and Krishnamurthy (1997) compared the accuracy of the *DEPI* and MMPI-A (Butcher et al., 1992) scales related to diagnoses of depression in a clinical sample of 152 adolescents (13 through 18 years old; mean age = 15.13 years) diagnosed according to *DSM-III-R* criteria. The diagnoses were determined by a licensed clinical psychologist independent of psychological test data. The sample was divided into a target group of adolescent patients (n = 56) with primary diagnoses of depression (major depression, dysthymia, bipolar disorder with most recent episode depressed, or depressive disorder NOS) and a clinical control group (n = 96) comprised of other diagnoses, the majority of which were conduct and adjustment disorders. Archer and Krishnamurthy reported a low sensitivity (.36) and a moderate specificity (.71) for the *DEPI* (see Table 2). The frequency of 12 other Rorschach variables and indices was examined but a significant difference was found only for *Vista* responses which occurred more frequently in the depressed group. The single variable test *Vista* > 0 was found to produce better classification accuracy (sensitivity = .52; specificity = .68) than the *DEPI*.

Meta-Analysis

Inspection of Table 2 shows a large variation in diagnostic efficiency statistics for the *DEPI* reported or calculated from various studies. The large differences in sample sizes make direct comparison of the statistics across studies difficult. In order to compensate for this we calculated sample-weighted average values of sensitivity and specificity for the *DEPI* using the relative proportion of participants from each study as weights. Because Exner's (1995) sample of inpatient depressives was the largest depression sample included in this review and because the diagnostic performance statistics calculated from Exner's data are different from most other studies, we calculated sample-weighted average sensitivity both with and without this sample (see Table 3). Omitting Exner's (1995) sample reduces the sample-weighted average sensitivity for studies with adults from .57 to .47. For the two studies including adolescent samples, the sample-weighted average sensitivity of the *DEPI* is even lower, .28. As the control condition is known to have an impact on specificity, we calculated sample-weighted average values of specificity for studies including clinical controls and studies including nonclinical controls separately (see Table 4). Both sets of calculations were done with and without Exner's (1995) control groups and the two studies including adolescents were treated separately. As expected, the specificity is markedly higher in the nonclinical control condition. Inclusion of Exner's (1995) samples increases the sample-weighted average specificity from .61 to .79 in the clinical control condition and from .88 to .95 in the nonclinical control condition. The weighted average value of specificity for the two studies including adolescents is not markedly different from the weighted average values for the studies including adults.

Table 5 displays the predictive value of a positive and a negative *DEPI* respectively at three different base rates and at three different levels of sensitivity

Table 3
Sample-Weighted Average Values of Sensitivity for the Rorschach DEPI

Condition	Number of studies	*N*	Weighted SN
Adults; Exner's (1995) sample included	8	843	.57
Adults; Exner's (1995) sample excluded	7	528	.47
Adolescents	2	101	.28

Note. *N* = total number of participants. Weighted SN = sample-weighted average sensitivity.

Table 4
Sample-Weighted Average Values of Specificity for the Rorschach DEPI

Condition	Number of studies	*N*	Weighted SP
		Clinical controls	
Adults; Exner's (1995) sample included	4	617	.79
Adults; Exner's (1995) sample excluded	3	117	.61
Adolescents	2	150	.73
		Nonclinical controls	
Adults; Exner's (1995) sample included	4	810	.95
Adults; Exner's (1995) sample excluded	3	110	.88

Note. *N* = total number of participants. Weighted SP = specificity sample-weighted average.

and specificity. As a basis for calculation of the PPV and the NPV of the *DEPI* we chose three different conditions: (a) a high level of sensitivity and specificity calculated from Exner's (1995) data comparing his inpatient depression sample with nonclinical controls; (b) a moderate level of sample-weighted average sensitivity and specificity calculated from three independent studies including adult depression samples compared with clinical controls (Sells, 1990/1991; Ilonen et al., 1999; Meyer, personal communication, January 2000); and (c) a low level of sample-weighted average sensitivity and a moderate level of sample-weighted average specificity calculated from the two studies including adolescents with depression compared with clinical controls. Base rates of 5%, 10%, and 25% were chosen for depressive disorder. As predicted, the prevalence has a major impact on the probability that an individual who has a positive *DEPI* actually has a diagnosis of depressive disorder. With the relatively high sensitivity of .75 computed from Exner (1995) there is an estimated .53 probability (53% chance) that a positive *DEPI* is associated with a diagnosis of depressive disorder, when the base rate is 5%. This may not seem very high, but as the chosen probability of a depressive disorder was only .05, using the *DEPI* increases the estimated probability of identifying depression by a factor of 10. The estimated NPVs are very high at all three base rates when based upon the high specificity (.96) calculated from Exner's data. At a moderate sample-weighted level of sensitivity (.50) calculated from three independent studies, the estimated PPVs are reduced to levels close to the base rates. The NPV estimates are also close to base rates when based on a moderate sample-weighted level of specificity (.61). At a base rate of 5% (.95 probability that a given individual does not have a depressive disorder) the NPV is .96, at a base rate of 10% (.90 probability of no depression) the NPV is .92, and at a base rate of 25% (.75 probability of no depression) the NPV is .79. At a low sample-weighted level of sensitivity (.28) calculated from the two studies including adolescents, the estimated PPVs are reduced to levels almost identical to base rates and the index does not seem to contribute anything to diagnostic classification.

Table 5
Estimated PPV and NPV for the Rorschach DEPI at Different Base Rates

Condition	*N*	SN	SP	PPV at different base rates			NPV at different base rates		
				5%	10%	25%	5%	10%	25%
Adult depressives vs. nonclinical controls[a]	D = 315 NC = 700	.7524[d]	.9643[d]	.53	.70	.88	.99	.97	.92
Adult depressives vs. clinical controls[b]	D = 366 CC = 117	.5041[e]	.6116[f]	.06	.13	.30	.96	.92	.79
Adolescent depressives vs. clinical controls[c]	D = 101 CC = 150	.2798[e]	.7280[f]	.05	.10	.26	.95	.90	.75

Note. SN = sensitivity; SP = specificity; PPV = positive predictive validity; NPV = negative predictive validity; D = number of depressed participants; NC = number of nonclinical controls; CC = number of clinical controls.
[a]Calculated from Exner (1995). [b]Calculated from the combined data of Sells (1990/1991), Ilonen et al. (1999), and Meyer (personal communication, 2000). [c]Calculated from the combined data of Ball et al. (1991) and Archer and Krishnamurthy (1997). [d]Sensitivity and specificity is reported to four decimals as even small variations in these statistics may affect the PPV and NPV estimates. [e]Sample-weighted average sensitivity is reported for groups of studies. [f]Sample-weighted average specificity is reported for groups of studies.

Table 6
Rorschach DEPI Scores in Clinical and Nonclinical Samples

Author	Age	Diagnostic criteria	Group									
			Depression		D & BPD	Mixed		CD		NC		
			M	*SD*	*M*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ball et al. (1991)	Adolescents	*DSM-III/DSM-III-R*				3.32[a]	1.17				
Caine et al. (1995)	Adults	*DSM-III-R*	3.20[b]	1.01						3.45[c]	1.23
Carlson et al. (1997)	Adults	*DSM-III-R*	4.0[d]		3.75[e]						
Archer & Krishnamurthy (1997)	Adolescents	*DSM-III-R*	3.9[f]	1.3		3.7[g]	1.6	3.7[h]	1.3		
Greenwald (1997)	Adults	—								4.20[i]	1.13
Franklin & Cornell (1997)	Adolescents	—								3.63[j]	1.54
										4.40[k]	1.26
Meyer (personal communication, 2000)	Adults	*DSM-III-R/DSM-IV*	4.38[l]	1.33		3.97[m]	1.48				

Note. D & BPD = depression and concurrent borderline personality disorder; CD = conduct disorder; Mixed = mixed diagnoses; NC = nonclinical. Dashes indicate that participants were not diagnosed.
[a]$n = 99$. [b]$n = 20$. [c]$n = 20$. [d]$n = 20$. [e]$n = 20$; [f]$n = 56$. [g]$n = 50$. [h]$n = 46$. [i]$n = 41$. [j]$n = 19$. [k]$n = 43$. [l]$n = 267$. [m]$n = 65$.

Studies Involving Nonclinical Samples

Information on the specificity of the *DEPI* can be inferred from studies regarding Rorschach testing of nonclinical groups even when no independent diagnostic evaluation of the individuals is available. Obviously some occurrence of depression must be expected in randomly selected nonclinical samples, but a very high prevalence of nonpatients with positive *DEPI* is unlikely and might indicate a problem with the specificity of the index. Mattlar and colleagues (1993) presented data from testing a group of adult Finnish nonpatients (n = 70) reporting that no less than 30 of these (43%) had a positive *DEPI*. The individuals were participating in a follow-up study as part of normative and epidemiological study and randomly drawn from the Finnish population register (Mattlar, 1986). In a study on the correlation between the *DEPI* and self-report measures of affect and related personality constructs Greenwald (1997) tested a group of undergraduate college students (n = 41; mean age = 19 years) reporting that 6 students (15%) had *DEPI* scores of 6 (the number of cases with *DEPI* ≥ 5 was not reported). The mean *DEPI* score was relatively high, 4.2 (see Table 6).

As part of a presentation of current nonpatient data for the Rorschach, WAIS-R (Wechsler, 1981), and MMPI-2 (Hathaway & McKinley, 1989) tests, Shaffer, Erdberg, and Haroian (1999) reported Rorschach data from a sample of adults (n = 123) residing in central California and found 18 adults (15%) with a positive *DEPI*. Potential participants were excluded if they had a prior psychiatric hospitalization or if they had been in psychological treatment within the past 2 years. Jørgensen and Olsen (1999) reported Rorschach data from a small sample of adult Danish nonpatients (n = 27) finding 8 adults (30%) with *DEPI* equal to or greater than 5. Potential participants reporting prior psychiatric or neurological disease during a preliminary interview were excluded. The relatively high incidence of nonpatients with a positive *DEPI* contrasts with the normative data presented by Exner (1995) where only 25 of 700 nonpatients (4%) had a positive *DEPI*. Franklin and Cornell (1997) reported mean *DEPI* scores for high-ability and creative adolescent females subdivided into a group of *accelerants* (i.e., students enrolled in an early college entrance program, n = 43; mean age = 14 years) and *nonaccelerants* (n = 19; mean age = 14 years; see Table 6). The number of students with *DEPI* equal to or greater than 5 was not reported but the mean *DEPI* score for each group was unexpectedly high when compared to adolescent psychiatric patients. When comparing the data from clinical and nonclinical samples no clear relation between mean *DEPI* scores and a diagnosis of depression is seen (see Table 6). The highest mean *DEPI* value of 4.4 was reported in a nonclinical sample whereas the lowest mean *DEPI* value of 3.2 was reported in a sample with diagnosed depression.

The Schizophrenia Index (*SCZI*)

An early experimental version of the *SCZI* was published in 1978 (Exner, 1978), revised in 1984 (Exner, 1986), with a subsequent revision in 1990 (Exner, 1990). The 1990 version of the *SCZI* is constituted by 10 structural variables combined into 6 constellation criteria or tests with a cutoff score equal to or greater than 4 (see Table 7). The variables included in the *SCZI* refer partly to problems concerning perceptual accuracy and partly to ideational slippage and deviant verbalizations. According to Exner a *SCZI* value of 4 indicates a significant probability that schizophrenia is present whereas *SCZI* values of 5 or 6 are more definitive as they indicate a strong likelihood of schizophrenia and a very low probability of a false positive (Exner, 1991). In Exner's psychiatric reference group (n = 320) of schizophrenic patients (Exner, 1995, Table 21) 82% have *SCZI* equal to or greater than 4 compared to only 0.3% of the 700 nonpatient normative adults (Exner, 1995, Table 11). Exner (1991, 1995) did not specify whether diagnoses of schizophrenia in the psychiatric reference group were determined independently of Rorschach data. Ganellen (1996) calculated diagnostic efficiency statistics for the revised *SCZI* using classification rates reported by Exner (1991) for two samples of schizophrenics (n = 320 and n = 500) compared with a sample of inpatient depressives (n = 315), a sample of nonpatients (n = 700), and a heterogeneous group of outpatients and nonpatients (n = 1,500). The sensitivity of the *SCZI* was reported as .81 and .83 respectively

Table 7
The Rorschach Schizophrenia Index (SCZI)

1. (X+% < .61) *and* (S-% < .41) **or** (X+% < .50)
2. X-% > .29
3. (FQ- >= FQu) **or** (FQ- > FQo + FQ+)
4. (Sum Level 2 Special Scores > 1) *and* (FAB2 > 0)
5. (Raw Sum of 6 Special Scores > 6) **or** (Weighted Sum of 6 Special Scores > 17)
6. (M- > 1) **or** (X-% > .40)

Note. The Index is positive if 4 or more conditions are true.

whereas the specificity was reported as .90 regardless of the control condition. Using a similar method we computed the diagnostic performance of *SCZI* in two ways (see Table 8). The classification rates for *SCZI* reported by Exner (1995, Table 21) were compared first to the classification rates reported for 700 nonpatients (Exner, 1995, Table 11), and second to the classification rates reported for a combined clinical sample of inpatient depressives (n = 315; Exner, 1995, Table 23) and character disorders (n = 180; Exner, 1995, Table 25). When applied to Exner's nonpatient sample, the specificity of the *SCZI* is almost a perfect 1.00 (the exact value is .997) due to the fact that only 2 out of 700 nonpatients had a positive *SCZI*. The impact of the control comparison is illustrated once more as the *SCZI* performs relatively better when contrasting schizophrenic patients with nonclinical controls.

Vincent and Harman (1991) found that 21 of 111 structural variables in Exner's descriptive data for the psychiatric reference group of inpatient schizophrenics met the criteria for clinical significance (i.e., ± 2 *SD* from the nonpatient mean). Through comparison with Exner's psychiatric reference samples regarding inpatient depressives and character disorders 9 of these 21 variables were identified as specific for the schizophrenic sample. The majority of these 9 variables were related to deviant or disordered verbalization and conceptualization (e.g., *CONTAM, Sum 6 Special Scores, CONFAB*). The variables relating to perceptual inaccuracy demonstrated limited differential utility, as they deviated significantly from the nonpatient mean in all three psychiatric reference samples.

Clinical Studies

Netter and Viglione (1994) reported *SCZI* classification rates for a sample of 20 inpatient schizophrenics diagnosed by the means of *DSM-III-R* criteria based on clinical interviews and histories and a nonclinical control group of 20 volunteers (an experimental group of nonpatients instructed to malinger schizophrenia was included). As a further validation of this diagnosis they all exceeded the cutoff score on a test designed to assess thinking disturbances (Gorham Proverbs Test; Gorham, 1956). Based on the classification rates reported by Netter and Viglione we calculated diagnostic efficiency statistics for the *SCZI* finding a sensitivity of .70 and a specificity of .85 (see Table 8). Netter and Viglione also reported means, medians, and modes for the *SCZI*, finding markedly higher values in the schizophrenic group (see Table 9). A significant difference regarding the number of positive *SCZI*s was found between the schizophrenics and the controls using chi-square analysis, $(1, N = 40) = 10.23$, $p < .005$.

Meyer (1993) reported that *SCZI* scores were significantly higher in a group of patients with *DSM-III-R* diagnosed psychotic disorders (n = 39) than in a group of patients with nonpsychotic diagnoses (n = 42; see Table 9). The psychotic sample included schizophrenia, schizoaffective disorder, bipolar disorder with psychotic features, major depression with psychotic features, delusional disorder, and psychotic disorder NOS (the exact number of patients in each diagnostic category was not specified). There was a significant correlation between *SCZI* scores and a psychotic diagnosis in average-length protocols (Spearman r = .48, p = .004, one-tailed), but the correlation was nonsignificant in both high-R protocols ($R \geq 29$) and low-R protocols ($R \leq 17$). R was reported to have a significant moderating impact on constellation criteria 4, 5, and 6 of the *SCZI*, meaning that as R increases so does the probability of these three criteria to become positive. Meyer (personal communication, February 2000) split the abovementioned extended data set of 332 patients into a subgroup (n = 158) with *DSM* diagnosed psychotic disorder (schizophrenia, schizophreniform, or schizoaffective disorder, depressive or bipolar disorder with psychotic features, delusional disorder,

Table 8
Diagnostic Efficiency Statistics for the Rorschach SCZI

Author	*N*	Diagnostic criteria	Blind	SN	SP	OCC	Base rate	PPV	NPV
Exner (1995)	P[a] = 320 NC = 700	*DSM-III*	–	.82	1.00	.94	31%	.99	.92
Exner (1995)	P[a] = 320 CC[b] = 495	*DSM-III*	–	.82	.93	.88	39%	.88	.89
Netter & Viglione (1994)	P[a] = 20 NC = 20	*DSM-III-R*	yes	.70	.85	.78	50%	.82	.74
Hilsenroth et al. (1998)	P = 33 NC = 50	*DSM-IV*	yes	.73	1.00	.89	40%	1.00	.85
Hilsenroth et al. (1998)	P = 33 CC[c] = 45	*DSM-IV*	yes	.73	.60	.65	42%	.57	.75
Ilonen et al. (1999)	P[a] = 27 NC = 60	*DSM-IV*	yes	.70	1.00	.91	31%	1.00	.88
Ilonen et al. (1999)	P[a] = 27 CC[d] = 70	*DSM-IV*	yes	.70	.87	.82	28%	.68	.88
Meyer (personal communication, 2000)	P = 158 CC[e] = 174	*DSM-III-R*	yes	.47	.76	.63	48%	.65	.62

Note. Blind = diagnostic clinicians were blind to Rorschach data; SN = sensitivity; SP = specificity; OCC = overall correct classification/hit rate; Base rate = percentage of psychotic participants in study sample; PPV = positive predictive validity; NPV = negative predictive validity; P = number of psychotic participants; NC = number of nonclinical controls; CC = number of clinical controls. Dashes indicate that the information was not reported.
[a]Schizophrenic patients. [b]Inpatient depressives (n = 315) and character disorders (n = 180). [c]Borderline personality disorder and personality disorders. [d]Depression. [e]Diagnostically mixed.

Table 9
Rorschach SCZI Scores in Clinical and Nonclinical Samples

Author	Age	Diagnostic criteria	Group									
			P		NP		BPD	CAPD	CCPD	NC		
			M	*SD*	*M*	*SD*	*M*	*M*	*M*	*M*	*SD*	
Meyer (1993)	Adults	*DSM-III-R*	3.18[a]		2.40[b]							
Netter & Viglione (1994)	Adults	*DSM-III-R*	4.10[c]	1.92						1.40[d]	1.79	
Hilsenroth et al. (1998)	Adults	*DSM-IV*	4.5[e]				3.0[f]	3.3[g]	1.8[h]	1.1[i]		
Franklin & Cornell (1997)	Adolescents	–								1.90[j]	1.10	
										2.49[k]	1.47	
Bannatyne et al. (1999)	Adults	*DSM-IV*	2.47[l]	1.54								
Meyer (personal communication, 2000)	Adults	*DSM-III-R/ DSM-IV*	3.32[m]	1.77	2.25[n]	1.60						

Note. P = psychotic disorder; NP = nonpsychotic disorder; BPD = borderline personality disorder; CAPD = cluster A personality disorder; CCPD = cluster C personality disorder; NC = nonclinical. Dashes indicate that participants were not diagnosed.
[a]$n = 39$. [b]$n = 42$. [c]$n = 20$. [d]$n = 20$. [e]$n = 33$. [f]$n = 23$. [g]$n = 9$. [h]$n = 13$. [i]$n = 50$. [j]$n = 19$. [k]$n = 43$. [l]$n = 180$. [m]$n = 158$. [n]$n = 174$.

shared psychotic disorder; brief psychotic disorder, psychotic disorder NOS, schizotypal or borderline personality disorder) and a clinical control group (n = 174) with nonpsychotic diagnoses. He reported a sensitivity for the *SCZI* of .47 and a specificity of .76 (see Table 8). A highly significant difference was found between *SCZI* mean scores in the two groups, t = 5.75, p = < .001 (see Table 9). Based on a section of this extended data set (n = 262), Meyer furthermore reported significant correlations between the *SCZI* and schizophrenia by independently assigned billing diagnosis (r = .175, p = .002, one-tailed) and between the *SCZI* and a diagnosis of psychotic disorder (r = .332, p = .000, one-tailed).

Hilsenroth, Fowler, and Padawer (1998) found that the *SCZI* is empirically related to the presence of a diagnosis of psychotic disorder as it was effective in differentiating psychotic patients (n = 33) from patients with a borderline personality disorder (n = 23), cluster A personality disorder (n = 9), cluster C personality disorder (n = 13), and nonclinical controls recruited from undergraduate psychology classes (n = 50). The clinical samples were diagnosed prior to Rorschach testing using *DSM-IV* criteria. The psychotic disordered sample consisted of paranoid schizophrenia (n = 19), undifferentiated schizophrenia (n = 7), schizoaffective (n = 6), and psychotic disorder NOS (n = 1). When comparing the *SCZI* scores across the five groups Hilsenroth et al. found significant main effect differences, $F(4, 123) = 29.5$, p < .0001 (see Table 9). By comparing the frequencies of the six individual *SCZI* criteria for each of the five groups, it was found that Criteria 4 and 5 were the most specific for the psychotic disordered sample. Both criteria refer to the presence of ideational slippage and problems in judgment and/or conceptualization (cf. Vincent & Harman, 1991). Hilsenroth et al. (1998) presented diagnostic efficiency statistics for the *SCZI* in one nonclinical group comparison (n = 50) and four clinical group comparisons at three different cutoff scores. In order to compare the diagnostic efficiency statistics with the statistics calculated from Exner (1995) we combined the three nonpsychotic clinical groups (n = 45) in the Hilsenroth et al. study before computing diagnostic efficiency statistics (see Table 8). As expected, the index performed better in the nonclinical control comparison, with perfect specificity (1.00) and PPV (1.00). In the clinical comparison the performance of the index is lower than the statistics computed from Exner (1995) but considering the Hilsenroth et al. clinical control group had a significant proportion of borderline patients, this is not surprising. Hilsenroth et al. argued that in a clinical context it may be more appropriate to employ the *SCZI* as a dimensional measure of psychosis—a *Psychosis Index*—as the index is sensitive to phenomena such as impaired reality testing and disordered thinking which are not specific to schizophrenia but may be present in other kinds of psychotic disorders as well.

Ilonen et al. (1999) examined the diagnostic efficiency of the *SCZI* for detecting first-episode schizophrenia independently diagnosed by *DSM-IV* criteria. A sample of 27 adult Finnish patients meeting the criteria for schizophrenia were compared first to a nonclinical control group (n = 60) and secondly to a clinical control group comprised of patients with depression (n = 70; nonpsychotic depression, psychotic depression, and bipolar I disorder). They reported a sensitivity of .70 for the *SCZI* and high levels of specificity ranging from 1.00 in the nonclinical comparison to .87 when comparing with the clinical control group (see Table 8). As the majority of the false positives in the clinical control group were from the subgroup with bipolar disorder the specificity actually varied from .54 when comparing with the bipolar group to .97 when comparing with the nonpsychotic depressives. Ilonen et al. interpreted the occurrence of a positive *SCZI* among bipolar disorder patients with the most recent manic episode as a sign of the presence a common psychopathological structure consisting of psychotic, negative, and disorganized dimensions. In accordance with Hilsenroth et al. (1998) they proposed that clinically it may be more useful to employ the *SCZI* as a dimensional measure of psychosis rather than an index specific for schizophrenia.

Bannatyne, Gacono, and Greene (1999) reported *SCZI* classification rates for three groups of chronic, psychotic forensic patients: paranoid schizophrenics (n = 89), undifferentiated schizophrenics

(n = 38), and schizoaffective patients (n = 53). All patients were in outpatient treatment and on neuroleptic medication at the time of psychological testing. The diagnoses according to *DSM-IV* criteria were determined by consensus among several evaluators including psychologists, and it was not specified whether Rorschach data might influence the diagnosis. In all three groups of psychotic patients the occurrence of a positive *SCZI* was relatively low: 36%, 21% and 26% respectively, equivalent to sensitivity values from .21 to .36. For the total sample (n = 180) the mean *SCZI* score was 2.47, which is also relatively low compared to other psychotic groups (see Table 9). Bannatyne et al. note that factors such as defensiveness, denial, chronicity, concurrent character pathology, and neuroleptic medication may contribute to the low incidence of positive *SCZI*s in this sample.

Singer and Brabender (1993) reported *SCZI* classification rates for unipolar depressed inpatients (n = 29), bipolar depressed patients (n = 15), and bipolar manic patients (n = 18). From these data we estimated the specificity of the *SCZI* for the three subgroups finding very high levels of specificity in the unipolar depressed group (.97) and the bipolar depressed group (1.00) but only a moderate level of specificity in the bipolar manic group (.67). The relatively high proportion of false positive scores on the *SCZI* among bipolar/manic patients found by both Ilonen et al. (1999) and Singer and Brabender (1993) seems to indicate that the index reacts to psychotic or psychotic-like phenomena in the manic phase of bipolar disorder (disorganized thinking, loose associations) thereby creating difficulties in differentiating schizophrenia from mania.

Meta-Analysis

On the basis of the data on the diagnostic performance of the *SCZI* presented in Table 8 we calculated sample-weighted average values of sensitivity for the *SCZI* (see Table 10). The sample-weighted averages of sensitivity were calculated both with and without Exner's (1995) sample of inpatient schizophrenics. Excluding Exner's (1995) data the sample-weighted average sensitivity is reduced from .71 to .55. This is partly due to the impact of Meyer's data set (Meyer, personal communication, February 2000) which is relatively large and with a markedly low sensitivity. The low sensitivity reported by Meyer might be explained by the fact that a number of patients in his data set with a psychotic disorder by history were without psychotic symptoms at the time of testing due to medication. Furthermore, the psychotic disorder group in his sample was more heterogeneous than other psychotic disorder groups (containing also depressive or bipolar disorder with psychotic features). We therefore also calculated the sample-weighted average sensitivity based on three studies (Hilsenroth et al., 1998; Ilonen et al., 1999; Netter & Viglione, 1994) including mostly schizophrenic patients finding a sensitivity (.71) identical to the sensitivity calculated from Exner's (1995) data for inpatient schizophrenics. The sample-weighted average values of specificity were calculated for clinical controls and nonclinical controls separately and both with and without Exner's 1995 data (see Table 11). Again, the control condition can be seen to have a significant impact on specificity. Table 12 displays the estimated predictive values of a positive and a negative *SCZI* respectively at three different base rates and

Table 10
Sample-Weighted Average Values of Sensitivity for the Rorschach SCZI

Condition	Number of studies	N	Weighted SN
Exner's (1995) sample included	5	558	.71
Exner's (1995) sample excluded	4	238	.55
Netter & Viglione (1994); Hilsenroth et al. (1998); Ilonen et al. (1999)	3	80	.71

Note. N = total number of participants. Weighted SN = sample-weighted average sensitivity.

at two levels of sensitivity and specificity. As a basis for the estimation of the PPV and the NPV of the *SCZI* we chose two different conditions: (a) a high level of sensitivity and specificity calculated from Exner's (1995) data comparing his inpatient schizophrenic sample (n = 320) with nonclinical controls (n = 700) and (b) a moderate level of sample-weighted average sensitivity and specificity calculated from two independent studies including mostly schizophrenic patients compared with clinical controls (Hilsenroth et al., 1998; Ilonen et al., 1999). Although the base rate can be seen to have some impact on the probability that an individual who has a positive *SCZI* actually has a diagnosis of psychotic disorder, all the estimated PPVs are higher than the base rates. The lowest PPV of .14 estimated from two independent data sets is still almost three times higher than the base rate of 5% (.05 probability). Also the NPVs are considerably above probability levels at all three base rates.

Studies Involving Nonclinical Samples

As mentioned earlier, information on the specificity of an index might be inferred from Rorschach data on nonclinical samples. Even when no independent diagnostic evaluation of the participants is available we would not expect to find a high incidence of cases with a positive *SCZI* in nonclinical samples. However, this was the case in the aforementioned study by Franklin and Cornell (1997) on the personality of high-ability and creative adolescent females. In the target group of *accelerants* (n = 43) no less than 14 students (33%) had *SCZI* scores equal to or greater than 4. In the control group of *nonaccelerants* (n = 19) two students (11%) had *SCZI* = 4. In Exner's (1995) normative data on 14-year-olds (n = 105) the number of individuals with a positive *SCZI* is 0. The *SCZI* elevations in Franklin and Cornell's samples were caused by a low level of form quality rather than by special scores referring to ideational slippage. In the nonpatient sample (n = 123) presented by Shaffer et al. (1999) 20 participants (16%) had a positive *SCZI*. The *SCZI* elevations seemed to be based on both form quality distortions and elevated cognitive special scores (see Shaffer et al., Tables 5 and 6). In the study on Danish nonpatients (n = 27) by Jørgensen and Olsen (1999) only one participant had *SCZI* equal to or greater than 4. As can be seen from the data presented in Table 9, mean *SCZI* values are generally at a high level in groups with a psychotic disorder, at an intermediate level in groups with other psychiatric diagnoses, and at a low level in nonclinical groups–with the exception of the data reported by Franklin and Cornell (1997).

Discussion

Although we chose clinical diagnosis as the external criterion for the diagnostic efficiency of the *DEPI* and the *SCZI*, we recognize the fact that clinical diagnosis does not represent a perfectly objective classification of psychiatric disorders (cf. Flaum et al., 1998). The clinical studies on the *DEPI* including adults show a large variation in

Table 11
Sample-Weighted Average Value of Specificity for the Rorschach SCZI

Condition	Number of studies	*N*	Weighted SP
		Clinical controls	
Exner's (1995) sample included	4	784	.87
Exner's (1995) sample excluded	3	291	.76
		Nonclinical controls	
Exner's (1995) sample included	4	830	1.00
Exner's (1995) sample excluded	3	130	.98

Note. *N* = total number of participants. Weighted SP = sample-weighted average specificity.

Table 12
Estimated PPV and NPV for the Rorschach SCZI at Different Base Rates

Condition	*N*	SN	SP	PPV at different base rates			NPV at different base rates		
				5%	10%	25%	5%	10%	25%
Schizophrenia vs. nonclinical controls[a]	P = 320 NC = 700	.8156[c]	.9971[c]	.94	.97	.99	.99	.98	.94
Schizophrenia vs. clinical controls[b]	P = 60 CC = 115	.7165[d]	.7643[e]	.14	.25	.50	.98	.96	.89

Note. SN = sensitivity; SP = specificity; PPV = positive predictive validity; NPV = negative predictive validity; P = number of psychotic disorder patients; NC = number of nonclinical controls; CC = number of clinical controls.
[a]Calculated from Exner (1995). [b]Calculated from Hilsenroth et al. (1998) and Ilonen et al. (1999). [c]Sensitivity and specificity is reported to four decimals as even small variations in these statistics affect the PPV and NPV estimates. [d]Sample-weighted average sensitivity is reported for groups of studies. [e]Sample-weighted average specificity is reported for groups of studies.

diagnostic performance, in particular regarding the sensitivity of the index. None of the independent studies so far support the relatively high sensitivity (.75) calculated from Exner's (1995) data. Some of the variation in the reported diagnostic performance can probably be explained by methodological difficulties. Several studies make group comparisons based on very small samples and many studies did not include nonclinical control groups. Most of the studies were characterized by diagnostically heterogeneous depression samples including a mixture of diagnoses such as major depression, dysthymia, bipolar affective disorder, and depressive disorder NOS. When the *DEPI* was applied to separate subtypes of affective disorders some interesting differences emerged. The highest values of sensitivity were reported for nonpsychotic depression (.79; Ilonen et al., 1999) and for unipolar depression (.59; Singer & Brabender, 1993) whereas the lowest values of sensitivity were reported for borderline personality disorder and concurrent major depression combined (.20), psychotic depression (.25), and bipolar depression (.26). One interpretation of these results might be that nonpsychotic depression and unipolar depression represent depression in a more unambiguous form—and thus easier to identify for the *DEPI*—than psychotic and bipolar depression where psychotic phenomena might be a confounding factor. Studies including adolescent samples accordingly indicate that the *DEPI* has an insufficient ability to identify adolescent depression. The specificity of the *DEPI* is clearly influenced by the control condition, in the sense that it is relatively lower when applied to other psychiatric groups than when applied to nonclinical groups. This does not in itself indicate a problem with the specificity of the *DEPI* as it might correctly identify secondary depressive phenomena in various psychiatric groups. However, the large variation in the proportion of cases with positive *DEPI* in nonpatient samples raises doubts about the specificity of the index. Regarding the *SCZI* the majority of the clinical studies indicate that the index effectively discriminates between individuals with and without a psychotic disorder. The relatively low sensitivity reported by Meyer (personal communication, February 2000) may be explained partly by the fact that some of the patients were free of psychotic symptoms at the time of testing and partly by the heterogeneity of Meyer's psychotic group that included individuals with depressive or bipolar disorder. In general, the diagnostic efficiency statistics should probably not be interpreted as an indication of the *SCZI*'s ability to identify schizophrenia per se, but rather as an indication of the ability of the index to identify psychotic features in a broader sense. Data from nonclinical samples show an unexpectedly large variation in the proportion of cases with positive *SCZI*.

It could be argued that the estimates made here regarding the predictive validity of the constellation indices at different base rates are too pessimistic. The estimates simulate a situation where the indices are used indiscriminately or "blindly" as screening instruments. In a clinical setting, several other sources of information (anamnestic and interview data, results from other tests or self-rating instruments, etc.) are usually involved in the diagnostic procedure, thereby greatly improving the probability of an accurate diagnosis. Among patients referred for assessment to address a question about a specific disorder (e.g., psychotic disorder) the base rate is likely to be quite high, as psychological testing is often used only when the clinical symptoms already point to a particular diagnosis. But if the *DEPI* or the *SCZI* are to be used as a diagnostic aids or instruments, we consider it relevant to focus on how much predictive power these indices will yield. This makes comparison with other diagnostic measures possible.

One question raised by this review is why the diagnostic performance of the *DEPI* is lower than the performance of the *SCZI*. In a discussion on the validity of the *DEPI*, Wood et al. (1996) referred to the statistical concept of *shrinkage during cross-validation*: the fact that empirically derived variables showing adequate discriminative power when applied with the original sample sometimes show poor predictive power when applied to a new sample with a different composition of individuals. In a reply to Wood et al., Exner (1996) expressed remarkable pessimism regarding the possibilities of identifying *DSM* diagnosed depression by means of Rorschach data:

> Depression is a complex subject, and I believe that there are other measures such as the Minnesota Multiphasic Personality Inventory or the Beck Depression Inventory that might identify the presence of reported depression much more accurately than the Rorschach. As Wiener (1989) has noted, "Depression is used indiscriminately as a label for a state, trait, sign, syndrome, disease, as a category name and, at the same time as an explanatory concept" (p. 296). When the checklists from the Diagnostic and Statistical Manual of Mental Disorders are applied to arrive at diagnosis of depression, a huge number of combinations of symptoms and behaviors may yield the same conclusion, and it is unrealistic to assume that any Rorschach measure of depressive features can be that broad. (Exner, 1996, p. 12).

Another question raised by this review is whether the diagnostic performance of the *DEPI* can be improved through an adjustment or revision of the index. As suggested by Meyer (1993) controlling for the effects of *R* may improve the external validity of the *DEPI* in relation to diagnostic criteria as *DEPI* was more often positive in longer records. This was found also to be true for four of the seven constellation criteria of the *DEPI*. Thus, controlling Criteria 1, 2, 6, and possibly Criterion 5 for *R* might enhance the external validity of the overall index. Other potential moderators of diagnostic efficiency are *Lambda* and *EB*. As several of the constellation criteria in the *DEPI* refer to the use of determinants other than pure form, a high *Lambda* is likely to reduce the probability of a positive *DEPI*. The results of Jansak (1996/1997) indicated that depressed individuals with an extratensive *EB* had a positive *DEPI* more often than individuals with other *EB* styles. In the present review, it was not possible to examine the potential moderating effects of *R*, *Lambda*, and *EB* as exact information on these variables is absent in the majority of the studies. Another approach to optimize the *DEPI* might be to assign different weights to the constellation criteria or single variable tests constituting the index. Jansak (1996/1997) found that some of the constellation criteria in the *DEPI* (Criteria 2, 6, and 7) as well as some of the single variable tests (SumC' > 2; 2AB+Art+Ay > 3) contribute relatively more to the diagnostic accuracy than other criteria and variables. Caine et al. (1995) reported that the single variable test *MOR* > *2* had a relatively high discriminative power, whereas Archer and Krishnamurthy (1997) reported that *Vista* > 0 had better classification accuracy than the complete *DEPI*. It has not been possible to check the validity of these findings across the studies, as the classification accuracy of individual *DEPI* constellation criteria and single variables is generally not reported. However, the potential moderating effects of *R*, *Lambda*, and *EB* as well as the contribution of the individual constellation criteria to the diagnostic performance of the *DEPI* should be relevant focus points for future research. Until more documentation is available we believe that *DEPI* scores should be interpreted with considerable caution when applied for diagnostic purposes. By contrast, a positive *SCZI* should probably be seen as a valid indicator of the presence of psychotic or disordered thinking.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.) Washington, DC: Author.

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

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

Archer, R. P., & Gordon, R. A. (1988). MMPI and Rorschach indices of schizophrenic and depressive diagnoses among adolescent inpatients. *Journal of Personality Assessment*, *52*(2), 276-287.

Archer, R. P., & Krishnamurthy, R. (1997). MMPI-A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment*, *69*(3), 517-533.

Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach depression indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment*, *57*(3), 465-476.

Bannatyne, L., Gacono, C. B., & Greene, R. L. (1998). Differential pattern of responding among three groups of chronic, psychotic forensic outpatients. *Journal of Clinical Psychology*, *55*(12), 1553-1565.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent): Manual for administration scoring and interpretation.* Minneapolis: University of Minnesota Press.

Caine, S. L., Kinder, B. N., & Frueh, B. C. (1995). Rorschach susceptibility to malingered depressive disorders in adult females. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (pp. 165-174). Hillsdale, NJ: Erlbaum.

Carlson, C. F., Kula, M. L., & St. Laurent, C. M. (1997). Rorschach revised DEPI and CDI with inpatient major depressives and borderline personality disorder with major depression: Validity issues. *Journal of Clinical Psychology, 53*(1), 51-58.

Exner, J. E. (1978). *The Rorschach: A Comprehensive System. Volume 2: Current research and advanced interpretation* (1st ed.). New York: Wiley.

Exner, J. E. (1986). *The Rorschach: A Comprehensive System. Volume 1: Basic foundations* (2nd ed.). New York: Wiley.

Exner, J. E. (1990). *A Rorschach Workbook for the Comprehensive System* (3rd ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E. (1991). *The Rorschach: A Comprehensive System. Volume 2: Interpretation* (2nd ed.). New York: Wiley.

Exner J. E. (1993). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (3rd ed.). New York: Wiley.

Exner, J. E. (1995). *A Rorschach Workbook for the Comprehensive System* (4th ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E. (1996). A comment on "The Comprehensive System for the Rorschach: A critical examination." *Psychological Science, 7*(1), 11-13.

Exner J. E., & Weiner, I. B. (1982). *The Rorschach: A comprehensive system. Volume 3: Assessment of Children and adolescents.* (1st ed.). New York: Wiley.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). *Structured clinical interview for the DSM-IV Axis I disorders–patient edition.* New York: New York State Psychiatric Institute, Biometrics Research Department.

Flaum, M., Amador, X., Gorman, J., Bracha, H. S., Edell, W., McGlashan, T., Pandurangi, A., Kendler, K. S., Robinson, D., Lieberman, J., Ontiverros, A., Tohen, M., McGorry, P., Tyrell, G., Arndt, S., & Andreasen, N. C. (1998). *DSM-IV* field trial for schizophrenia and other psychotic disorders. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, W. Davis, & M. Kline (Eds.), *DSM-IV sourcebook* (pp. 687-716). Washington, DC: American Psychiatric Association.

Foldspang, A., Juul, S., Olsen, J., & Sabroe, S. (1986). *Epidemiologi* (2nd ed.). Copenhagen, Denmark: Munksgaard.

Franklin, K. W., & Cornell, D. G. (1997). Rorschach interpretation with high-ability adolescent females: Psychopathology or creative thinking? *Journal of Personality Assessment, 68*(1), 184-196.

Ganellen, R. J. (1996). Comparing the diagnostic efficiency of the MMPI, MCMI-II, and Rorschach: A review. *Journal of Personality Assessment, 67*(2), 219-243.

Gorham, D. R. (1956). A proverbs test for clinical and experimental use. *Psychological Reports, 2,* 1-12.

Greenwald, D. F. (1997). Comparisons between the Rorschach Depression Index and depression-related measures in a nonpatient sample. *Psychological Reports, 80*(3), 1151-1154.

Hathaway, S. R., & McKinley, J. C. (1989). *MMPI-2 manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Hilsenroth, M. J., Fowler, J. C., & Padawer, J. R. (1998). The Rorschach Schizophrenia Index (SCZI): An examination of reliability, validity and diagnostic efficiency. *Journal of Personality Assessment, 70*(3), 514-534.

Ilonen, T., Taiminen, T., Karlsson, H., Lauerma, H., Leinonen, K.-M., Wallenius, E., Tuimala, P., & Salokangas, R. (1999). Diagnostic efficiency of the Rorschach schizophrenia and depression indices in identifying first-episode schizophrenia and severe depression. *Psychiatry Research, 87,* 183-192.

Jansak, D. M. (1997). The Rorschach Comprehensive System Depression Index. Depression heterogeneity, and the role of self-schema (Doctoral dissertation, California School of Professional Psychology, San Diego, 1996). *Dissertation Abstracts International, 57,* 6576B.

Jørgensen, K., & Olsen, H. (1999). Rorschachtesting af danske kontrolpersoner i tværkulturelt perspektiv - et pilotprojekt (Rorschach data on Danish nonpatients in cross-cultural comparison: A pilot study). *Nordisk Psykologi, 51*(3), 203-218.

Kessel, J., & Zimmerman, M. (1993). Reporting errors in studies of diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results and implications for the peer review process. *Psychological Assessment, 53,* 395-399.

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. (1986). Finnish Rorschach responses in cross-cultural context: A normative study. *Jyvaskyla Studies in Education, Psychology and Social Research, 58.*

Mattlar, C. E., Forsander, C., Norrlund, L., Öist, A.-S., Mäki, J., & Alanen, E. (1993). Rasgos del contenido y del resumen estructural Rorschach caracteristicos de los finlandeses: Una comparacion transcultural. *Revista Sociedad Espanola Rorschach Met Y Proy,* 21-27.

Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment, 60*(1), 153-180.

Netter, B. E., & Viglione, D. J., Jr. (1994). An empirical study of malingering schizophrenia on the Rorschach. *Journal of Personality Assessment, 62*(1), 45-57.

Sackett, D. L. (1992). A primer on the precision and accuracy of the clinical examination. *Journal of the American Medical Association, 267,* 2638-2644.

Sells, J. E. (1991). A validity study of the DEPI index: The Rorschach Comprehensive System (Doctoral dissertation, University of Utah, 1990). *Dissertation Abstracts International, 51,* 5590B.

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

Singer, H. K., & Brabender, V. (1993). The use of the Rorschach to differentiate unipolar and bipolar disorders. *Journal of Personality Assessment, 60*(2), 333-345.

Spitzer, R. L., & Endicott, J. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 35*, 837-844.

Spitzer, R. L., Endicott, J., & Robins, E. (1985). *Research Diagnostic Criteria (RDC) for a selected group of functional disorders* (3rd ed.). New York: New York State Psychiatric Institute, Biometrics Research.

Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 11*(3), 251-265.

Viglione, D. J., Brager, R. C., & Haller, N. (1988). Usefulness of structural Rorschach data in identifying inpatients with depressive symptoms: A preliminary study. *Journal of Personality Assessment, 52*(3), 524-529.

Vincent, K. R., & Harman, M. J. (1991). The Exner Rorschach: An analysis of its clinical validity. *Journal of Clinical Psychology, 47*(4), 596-599.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised*. New York: Psychological Corporation.

Wiener, M. (1989). Psychopathology reconsidered: Depressions interpreted as psychosocial transactions. *Clinical Psychology Review, 9*, 295-321.Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*(1), 3-10.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*(1), 3-10.

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*(2), 115-129.