, (S1), S201–S216

ISSN: 0022-3891 print / 1532-7752 online
DOI: 10.1080/00223890701629342

N

GREGORY J. MEYER,[1] PHILIP ERDBERG,[2] AND THOMAS W. SHAFFER[3]

[1]
[2]

data on all CS variables and were a much better point of comparison for the CS norms than Wood et al.'s samples.

Meyer (2001) examined the 69 scores from the lower section of the CS structural summary that form the foundation for clinical interpretations. Cohen's *d* was computed to quantify deviations from both of the CS reference samples in a healthy or unhealthy direction, with the sign of each *d* determined empirically by comparing Exner's nonpatient means to his outpatient means, such that positive differences indicated greater health and negative differences indicated less health. Across all scores, the international sample was about 4 tenths of an less healthy than the old CS nonpatients (i.e., $Md = -.38$) and about equal to the CS outpatients ($Md = .03$), which was consistent with the supposition that Exner's nonpatients generally had positive evidence of social and/or vocational functioning, while the combined international sample reflected a broader and more general range of functioning in the population.

The largest differences were observed for the form quality (FQ) variables, with the combined international sample looking notably less healthy than Exner's nonpatient and outpatient reference samples. Meyer interpreted these differences as probably being due to changes that were made to the FQ tables after the reference samples were scored (Meyer & Richardson, 2001). Also, preliminary data from Exner's new normative sample (Exner, 2007/this issue; Exner & Erdberg, 2005) showed it differed from the older sample by about two tenths of a standard deviation, with the new CS norms being more similar to the international sample.

Despite the composite international sample being quite diverse with respect to selection procedures, examiner training, examination context, language, culture, and national boundaries, and despite the fact that the original CS norms had been collected 20–25 years earlier, Meyer (2001) concluded that the overall differences between the CS norms and the international composite sample were relatively small. At the same time, besides differences in form quality, relative to Exner's nonpatients, people in the combined international sample used more unusual location areas, incorporated more white space, used less color, had fewer blends, tended to see more partial than full human images, had less thematically elaborated movement (i.e., AG and COP), had more cognitive special scores, and gave fewer responses to the last three cards.

In this article we extend the previous analyses in several ways. First, we make use of the extensive data collected for this Supplement, which includes 20 adult samples and 19 child and adolescent samples. Second, we make use of data published elsewhere for adults (Sultan et al., 2004, 2006) and for children and adolescents (Exner, 2003). Third, and most important, rather than focusing on the extent to which Exner's CS reference values correspond to other samples, we use the available data to generate international normative reference values for the CS. The norms are based on adult protocols, with children and adolescents evaluated against the same standard as a way to highlight and quantify any developmental changes that may be present (Beizmann, 1970).

The samples in this Supplement differ in their quality (e.g., examiner training, scoring reliability, checks on administration

TABLE 1.—Composite adult international reference values for the Comprehensive System based on data from 17 countries: Average means and standard deviations for dimensional scores.

Variable			# of countries
Age	36.45	11.71	17
R	22.31	7.90	17
W	9.08	4.54	17
D	9.89	5.81	17
Dd	3.33	3.37	17
S	2.49	2.15	17
DQ+	6.24	3.54	17
DQo	14.68	6.74	17
DQv	1.09	1.50	17
DQv/+	0.29	0.67	17
FQx+	0.21	0.68	17
FQxo	11.11	3.74	17
FQxu	6.20	3.93	17
FQx–	4.43	3.23	17
FQxNone	0.33	0.71	17
MQ+	0.12	0.43	17
MQo	2.26	1.66	17
MQu	0.69	0.99	17
MQ–	0.63	1.05	17
MQNone	0.03	0.20	17
SQual–	0.87	1.15	17

the last seven variables in Table 1 were based on data from just 12 or 13 countries.

We encourage CS users to focus interpretation on psychometrically superior dimensional scores rather than categories formed by artificially dichotomized cut-off scores (see MacCallum, Zhang, Preacher, & Rucker, 2002). However, to facilitate clinical inferences regarding the presence or absence of certain CS scores, we also provide frequency data for the traditional classifications found in Exner's reference tables. Using the same procedures described above, we computed the average proportion of people across the adult reference samples in each classification category. We also computed the SD of these means to give an index of the variability across reference samples. Both sets of values are provided in Table 2.

From the descriptive data in Table 1, we generated Composite International T-scores. Although T-scores have never been used before with CS scores, they are a simple transformation of the reference data in Table 1, whereby the reference mean is set at a value of 50 and the reference is set to 10 points. For instance, Table 1 shows that *R* has $M = 22.31$ and

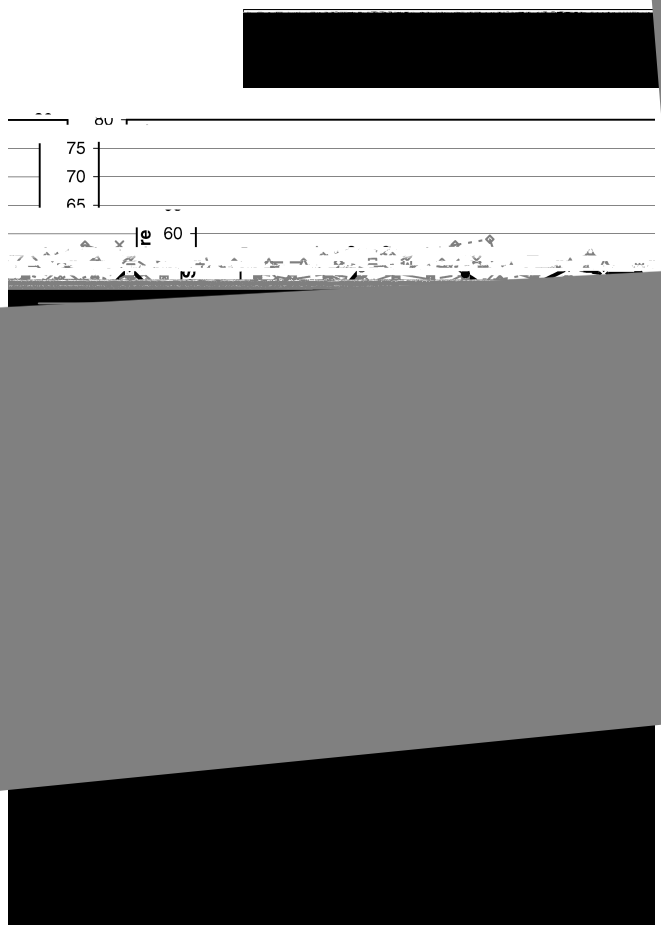
We first present and discuss graphs for the 20 adult samples that are part of this Supplement. For children and adolescents, we use the 19 samples published in this Supplement, as well as Exner's (2003) 12 samples that span the ages from 5 to 16 in yearly increments. The latter do not provide data for all the scores listed in Table 1, but to facilitate presentation we estimated their means for Form%, Blend%, NonPureH, An + Xy, and HRV by computing the sum, product, or difference using the reported mean values (e.g., Form% was estimated by dividing the mean for F by the mean for R).

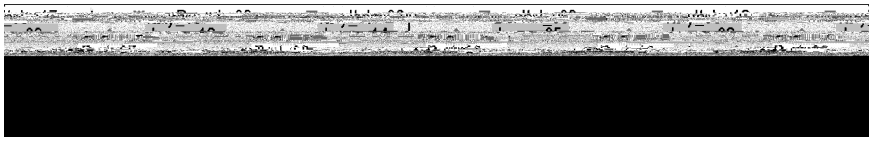
Results and Discussion

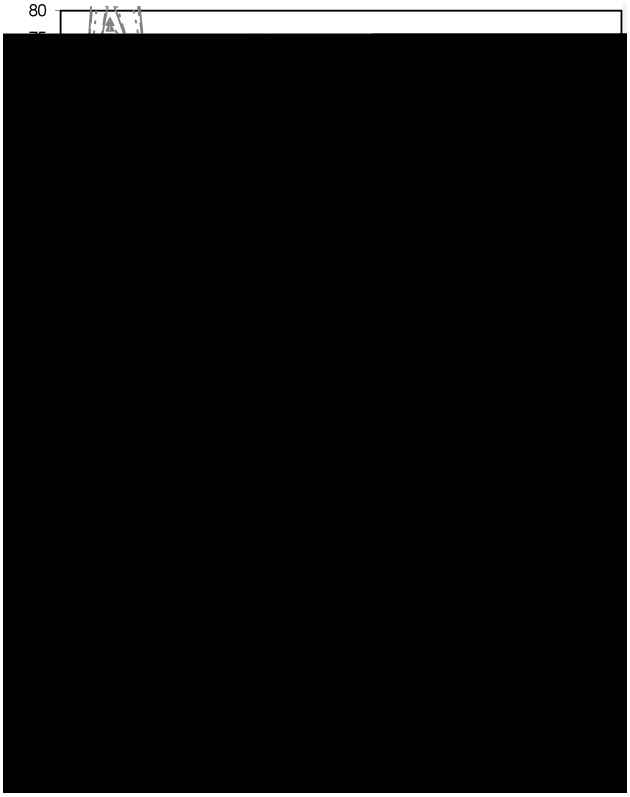
Before considering the substantive results, we computed reliability across samples from the data provided by each author. All but two samples (Exner, 2007/this issue; Mattlar et al., 2007/this issue) computed response-level percent agreement and iota (Janson & Olsson, 2004) values for response segments so we focused on these statistics. The 27 sets of reliability data were obtained from 997 protocols. Three reliability samples were notably larger than the others. Ivanouw (2007/this issue) used 191 protocols, Shaffer et al. (2007/this issue) used 92, and Nascimento (2007/this issue) used 80; the other samples ranged in size from 13 to 51, with a median across all samples of 25.

With results weighted by sample size, the average % Agreement was above .90 for all categories except FQ (.83). The average iota for coding complete responses was .84, which indicates excellent agreement. However, iota differed by segment: Location and Space = .92, DQ = .83, Determinants = .82, FQ = .72, Pairs = .91, Contents = .85, Popular = .90, Z-Scores = .87, and Special Scores = .67. Although showing adequate reliability, form quality and special scores clearly are the most challenging to code. In general, unweighted average iota values were slightly higher (e.g., complete responses = .86, FQ = .76, Special Scores = .71).

CS NORMATIVE REFERENCE DATA






80

the Composite International Adult normative data, and showed the largest within-country developmental trends. As such, in ways that were formerly unappreciated, these scores may be the most sensitive to the style or manner in which the test administration is conducted, the administration and inquiry skill of the examiner, across-site differences in administration and scoring conventions, developmental processes, and, perhaps most importantly, the interaction of all the forgoing factors with culture-specific conventions that may be present when a cue-sensitive child completes a rather unstructured and open-ended task with an unfamiliar adult.

CONCLUSIONS

Overall, if one embraces the goal of identifying normative reference values that transcend countries, cultures, languages, recruitment strategies, types of normative target populations, examiner training, and age, the data contained in this Supplement present a mixed picture for the CS. For adults, the findings reveal a reasonable degree of cross-sample and cross-national similarity. Relative to a composite international standard, adults from various countries around the world generally look similar.

Most instances when there were sample-specific divergences (e.g., DV1 in the older adults from the United States; FD in the Israeli sample of 41) did not appear tied to culture, as they either were not consistent across samples collected from the same country or the within-country differences were as large as the between-country differences. For instance, the two samples from Argentina differed by 10 T-score points on complexity markers (Zf, HVI Total); the two Israeli samples differed by this extent on determinant variables and their derivative scores (e.g., SumShading, FD, es, D-score, Blends); and the two general US samples differed by this degree on form quality (e.g., XA%, WDA%, X−

this research show a basic similarity in their CS scores. Although drawn from cultures that are largely Western in their orientation and level of education, there appears to be a basic consistency of human self-expression and perception across samples. Generally, people perceive and describe inkblot images similarly across cultures.

Although the findings in this Supplement strengthen our ability to use an international normative reference standard for the Rorschach with adults, the data in Figures 3–5 challenge our ability to do so for children and adolescents. For instance, it is clear that in a number of important ways Exner's (2003) reference samples for children are dated and atypical relative to the more recently collected samples from the United States, Denmark, Italy, Japan, and Portugal, as well as France (Andronikof, 1999). As such, they do not adequately serve as reference points

addition, we appreciate the insightful and critical comments on an earlier version of this article from Anne Andronikof, Jan Ivanouw, Bill Kinder, Carl-Erik Mattlar, Mariko Matsumoto, Joni Mihura, Regina Gattas do Nascimento, Barry Ritzler, Tony Sciara, Serge Sultan, Shira Tibon, Donald Viglione, and Irving Weiner.

REFERENCES

Allen, J., & Dana, R. H. (2004). Methodological issues in cross cultural and multicultural Rorschach research. , 189–206.

Andronikof-Sanglade, A. (1999, July). Data for French children. In P. Erdberg, & T. W. Shaffer, (Chairs) . Symposium presented at the XVIth Congress of the International Rorschach Society, Amsterdam, The Netherlands.

Andronikof-Sanglade, A. (2000). Use of the Rorschach Comprehensive System in Europe: State of the Art. In R. H. Dana (Ed.), (pp. 329–344). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Beizmann, C. (1970). . New York: Grune & Stratton.

Berant, E. (2007/this issue). Rorschach Comprehensive System data for a sample of 150 adult nonpatients from Israel. (Suppl. 1), S67–S73.

Campo, V., & Vilar, N. (2007/this issue). Rorschach Comprehensive System data for a sample of 517 adults from Spain (Barcelona). (Suppl. 1), S149–S153.

Daroglou, S., & Viglione, D. J. (2007/this issue). Rorschach Comprehensive System data for a sample of 98 adult nonpatients from Greece. (Suppl. 1), S61–S66.

de Ruiter, C., & Smid, W. (2007/this issue). Rorschach Comprehensive System data for a sample of 108 normative subjects from The Netherlands. (Suppl. 1), S113–S118.

Dumitrascu, N. (2007/this issue). Rorschach Comprehensive System data for a sample of 111 adult nonpatients from Romania. (Suppl. 1), S142–S148.

Erdberg, P., & Shaffer, T. W. (1999, July). J4th5.42s

French speaking community. , , , , (Suppl. 1), S26–S34.

Nakamura, N., Fuchigami, Y., & Tsugawa, R. (2007/this issue). Rorschach Comprehensive System data for a sample of 240 adult nonpatients from Japan. , , , ,