Journal of Personality Assessment, *89*(S1), S201–S216

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

Toward International Normative Reference Data for the Comprehensive System

Gregory J. Meyer,[1] Philip Erdberg,[2] and Thomas W. Shaffer[3]

[1]*Department of Psychology, University of Toledo*
[2]*Independent Practice, Corte Madera, California*
[3]*Independent Practice, Fresno, California*

We build on the work of all the authors contributing to this Special Supplement by summarizing findings across their samples of data, and we also draw on samples published elsewhere. Using 21 samples of adult data from 17 countries we create a composite set of internationally-based reference means and standard deviations from which we compute T-scores for each sample. Figures illustrate how the scores in each sample are distributed and how the samples compare across variables in eight Rorschach Comprehensive System (CS; Exner, 2003) clusters. The adult samples from around the world are generally quite similar, and thus we encourage clinicians to integrate the composite international reference values into their clinical interpretation of protocols. However, the 31 child and adolescent samples from 5 countries produce unstable and often quite extreme values on many scores. Until the factors contributing to the variability among these samples are more fully understood, we discourage clinicians from using many CS scores to make nomothetic, score-based inferences about psychopathology in children and adolescents.

Investigating normative reference values is a critical aspect of the science of psychological assessment. The authors contributing data to this Supplement (Shaffer, Erdberg, & Meyer, 2007) provide users of the Rorschach Comprehensive System (CS; Exner, 2003) with detailed descriptions of their 39 samples and a unique opportunity to review CS reference values from a diverse array of countries. When this Supplement was initially conceived, a primary goal was to provide CS users with a compendium of country-specific or locale-specific norms. However, these projects also introduce the possibility of creating a composite set of international norms at a level rarely achieved in personality assessment. In this article, we build upon the extensive work completed by the contributors to create such composite norms and explore their implications for clinical practice with adults and children.

The adequacy of the CS adult and child reference values (Exner, 2003) has been discussed and debated in the literature over the past decade, both with respect to samples from the US (e.g., Meyer, 2001; Shaffer, Erdberg, & Haroian, 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001a, 2001b) and from other countries (e.g., Andronikof-Sanglade, 2000; Mattlar, 2004; Sultan et al., 2004; Vinet, 2000). A study that sparked concern about the standard CS reference values was Shaffer et al.'s (1999) sample of 123 adults from Fresno, California. These participants were tested by graduate students, which Weiner (2001) questioned as a suitable level of training and experience to serve as a reference sample. Nonetheless, because both the Fresno sample and the traditional CS norms were obtained from nonpatients in the US, any disparities between them were notable. In particular, Shaffer et al. reported many shorter and more simplistic records than the existing CS norms. For instance, their sample had a mean $R = 20.8$ (versus 23.5 in the CS norms) and a mean Lambda = 1.22 (vs. .58), with 41% of their sample classified as having an avoidant style (i.e., Lambda > .99; vs. 14%).

Inspired by the Shaffer et al. (1999) study, Wood et al. (2001b) compared the old CS adult reference values to samples described in journal articles, book chapters, and dissertations. They examined 14 variables and, depending on the score, compared the CS reference values to those derived from 8 to 19 comparison samples. Mean differences were computed for 13 of the 14 variables and examined using Cohen's d as the effect size index, which indicates how far apart two means are in pooled standard deviation units. Wood et al. observed a wide range of differences; from what would be considered small effect sizes to very large effect sizes (Cohen's d values from .18 to 1.67). For nine of the variables, the differences were at least medium in size (i.e., $d = .50$, or half of a *SD*). The CS reference values were higher than the comparison samples for X+%, Afr, FC, P, WSumC, and Pure H, but lower for Fr+rF, X−%, and Y.

Wood et al.'s (2001) study had a number of difficulties (see Meyer, 2001, for details) so to investigate normative questions with a better comparison sample Meyer contrasted the existing CS adult normative group ($N = 700$) and Exner's (1993) sample of outpatients beginning therapy ($N = 440$) to a composite of 2,125 protocols from nine adult samples from Erdberg and Shaffer's (1999) symposium on internationally collected CS reference data. These samples were precursors to many of those in this Supplement and included data that are now part of the contributions by Shaffer, Erdberg, and Haroian (2007/this issue) from the US; Nakamura, Fuchigami, and Tsugawa (2007/this issue) from Japan; Campo and Vilar (2007/this issue) from Spain; Ivanouw (2007/this issue) from Denmark; Mattlar et al. (2007/this issue) from Finland; Mormont, Thommessen, and Kever (2007/this issue) from Belgium; Pires (2007/this issue) from Portugal; Raéz (2007/this issue) from Peru; and Sanz (2007/this issue) from Argentina.[1] These nine samples provided

Received July 19, 2007; Revised August 10, 2007
Address correspondence to Gregory J. Meyer, Department of Psychology, Mail Stop 948, University of Toledo, Toledo, OH 43606. Email: gmeyer@utnet.utoledo.edu

[1]The previous version of the adult sample from Italy that is presented in this Supplement (Lis, Parolin, Salcuni et al., 2007/this issue), was not included in

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 *SD* 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 *both* 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 quality), though all were collected by motivated and trained individuals seeking to advance the database of Rorschach assessment. CS users accustomed to Exner's reference values may have concerns about moving to a different type of standard. However, to the extent that the international samples converge on a common referent, the value in using these norms is that they should generalize across the diverse circumstances embodied in the international data collection efforts. In other words, because the CS international reference samples are quite diverse across a number of variables, to the extent that people look similar from one sample to the next, the composite norms have considerable generalizability across the same variables. These variables include strategies to recruit participants, training and experience of examiners, selection of participants based on background characteristics, context of the examination (e.g., office vs. other setting, method for recording responses, seating, warm-up procedures, administration of other measures), language, culture, and national boundaries.

METHOD

Samples and Procedures for Creating the Adult International Reference Values

To create international normative reference values for the CS, we used the 20 adult samples included in this supplement (Berant, 2007/this issue; Campo & Vilar, 2007/this issue; Daroglou & Viglione, 2007/this issue; de Ruiter & Smid, 2007/this issue; Nascimento, 2007/this issue; Dumitrascu, 2007/this issue; Exner, 2007/this issue; Greenway, & Milne, 2007/this issue; Ivanouw, 2007/this issue; Lis, Parolin, Salcuni, & Zennaro, 2007/this issue; Lunazzi et al., 2007/this issue; Mattlar et al., 2007/this issue; Mormont et al., 2007/this issue; Nakamura et al. 2007/this issue; Pertchik, Shaffer, Erdberg, & Margolin, 2007/this issue; Pires, 2007/this issue; Ráez, 2007/this issue; Sanz, 2007/this issue; Shaffer et al., 2007/this issue; Tibon, 2007/this issue), as well as one recently published normative sample from France (Sultan et al., 2004, 2006), for a combined sample based on 4,704 protocols. In those instances where there was more than one sample from a country, which was the case for Argentina (2 subsamples), Israel (2 subsamples), and the US (3 subsamples), we created a single, country-specific set of reference values by computing weighted average scores across the subsamples, giving more weight to larger samples. If one sample was missing information (e.g., the Argentinian sample of 90 for HRV), the composite was based on the remaining data.

Next we computed the international reference values for 143 scores that are included in the table of descriptive statistics accompanying articles in the Supplement. Because we wanted the norms to generalize across countries, we did not weight the samples by their size; rather, each of the 17 countries made an equal contribution to the average mean and average variance, which was then converted to a *SD*.[2] The resulting *M*s and *SD*s are provided in Table 1 (along with descriptive data for age). For most scores, all 17 countries contributed data. However, the French sample did not report results for 30 variables and

those analyses in part because of the reasons that are detailed by Lis, Parolin, Calvo et al., (2007/this issue).

[2] Although many CS variables are not normally distributed, according to the Central Limit Theorem, the distribution of the *M*s and *SD*s becomes increasingly normal as the size of the original samples increases. With the relatively large samples included here, computing an average *M* and *SD* should not be problematic.

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	*M*	*SD*	# 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
M	3.73	2.66	17
FM	3.37	2.18	17
m	1.50	1.54	17
FC	1.91	1.70	17
CF	1.65	1.55	17
C	0.34	0.66	17
Cn	0.02	0.14	17
Sum Color	3.91	2.53	17
WSumC	3.11	2.17	17
FC′	1.39	1.47	16
C′F	0.28	0.64	16
C′	0.06	0.28	16
FT	0.55	0.82	16
TF	0.08	0.30	16
T	0.01	0.11	16
FV	0.37	0.76	16
VF	0.12	0.39	16
V	0.01	0.13	16
FY	0.93	1.32	16
YF	0.36	0.73	16
Y	0.07	0.29	16
Fr	0.34	0.76	16
rF	0.07	0.33	16
Sum C’	1.75	1.71	17
Sum T	0.65	0.91	17
Sum V	0.52	0.92	17
Sum Y	1.34	1.63	17
Sum Shading	4.29	3.48	17
Fr+rF	0.41	0.88	17
FD	1.02	1.19	17
F	8.92	5.34	17
Pair	7.04	3.83	17
3r+(2)/R	0.38	0.16	17
Lambda	0.86	0.95	17
PureF%	0.39	0.17	16
FM+m	4.87	2.89	17
EA	6.84	3.76	17
es	9.09	5.04	17
D Score	−0.68	1.48	17
AdjD	−0.20	1.23	17
a (active)	4.96	3.08	17
p (passive)	3.73	2.65	17
Ma	2.09	1.83	17
Mp	1.67	1.61	17
Intellect	2.35	2.57	17
Zf	12.50	4.92	17
Zd	−0.67	4.72	17
Blends	4.01	2.97	17
Blends/R	0.18	0.13	17
Col-Shd Blends	0.60	0.92	17
Afr	0.53	0.20	17
Populars	5.36	1.84	17
XA%	0.79	0.11	17
WDA%	0.82	0.11	17
X+%	0.52	0.13	17
X–%	0.19	0.11	17
Xu%	0.27	0.11	17
Isolate/R	0.20	0.14	17
H	2.43	1.89	17
(H)	1.22	1.24	17
Hd	1.52	1.71	17
(Hd)	0.64	0.92	17
Hx	0.41	0.98	17
H+(H)+Hd+(Hd)	5.83	3.51	17
(H)+Hd+(Hd)	3.36	2.73	16
A	7.71	3.18	17
(A)	0.42	0.73	17
Ad	2.41	1.97	17
(Ad)	0.16	0.45	17
An	1.16	1.42	17
Art	1.22	1.45	17
Ay	0.52	0.87	17
Bl	0.25	0.55	17
Bt	1.41	1.44	17
Cg	1.89	1.77	17
Cl	0.18	0.46	17
Ex	0.19	0.48	17
Fi	0.50	0.80	17
Food	0.33	0.66	17
Ge	0.26	0.62	17
Hh	0.84	1.03	17
Ls	0.87	1.12	17
Na	0.75	1.11	17
Sc	1.11	1.35	17
Sx	0.47	0.94	17
Xy	0.19	0.52	17
Idiographic	0.89	1.21	17
An+Xy	1.34	1.54	16
DV	0.65	0.99	17
INCOM	0.73	0.97	17
DR	0.49	0.96	17
FABCOM	0.45	0.76	17
DV2	0.01	0.14	17
INC2	0.10	0.33	17
DR2	0.06	0.31	17
FAB2	0.08	0.31	17
ALOG	0.16	0.46	17
CONTAM	0.02	0.15	17
Sum 6 Sp Sc	2.75	2.39	17
Lvl 2 Sp Sc	0.25	0.62	17
WSum6	7.63	7.75	17
AB	0.32	0.82	17
AG	0.54	0.86	17
COP	1.07	1.18	17
CP	0.02	0.15	17
Good HR	3.70	2.18	17
Poor HR	2.86	2.52	17
MOR	1.26	1.43	17
PER	0.75	1.12	17
PSV	0.23	0.56	17
PTI Total	0.59	0.95	16
DEPI Total	3.80	1.33	16
CDI Total	2.90	1.24	16
SCon Total	4.67	1.62	16
HVI Total	2.79	1.64	16
OBS Total (1–5)	1.13	0.91	16
WD+	0.17	0.56	12
WDo	10.69	3.47	12
WDu	4.89	3.10	12
WD–	2.91	2.27	12
WDNone	0.34	0.69	12
EII–2	−0.15	0.95	13
HRV	0.94	2.98	13

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.

Samples and Procedures for the International T-Scores

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 *SD* is set to 10 points. For instance, Table 1 shows that *R* has $M = 22.31$ and $SD = 7.90$. For *R*, a T-score of 50 equates to a raw score of 22.31 and a T-score of 60 corresponds to a raw score one standard deviation higher, which would be a raw *R* value of 30.21. T-scores allow one to compare quickly individuals or samples to an expected reference value.

We computed T-scores for each sample in the Supplement by determining how far their sample mean was from the Table 1 reference mean in standard deviation units. For instance, Lunazzi et al.'s (2007/this issue) Argentinean sample had a mean $R = 18.71$. To compute the T-score for this sample, the International Reference Sample (IRS) mean (22.31) was subtracted from it and the difference (−3.60) was divided by the IRS *SD* (7.90), which indicated the *R* for this sample was .46 *SD* units lower than the IRS mean (i.e., $-3.60/7.90 = -0.456$). This value was then converted to a T-score by multiplying the *SD* difference by 10 and adding 50 (i.e., $[10 * -0.456] + 50 = 45.44$), which was then rounded to the nearest whole number, resulting in the Argentine sample having a Composite International T-score of 45 for *R*.

Not only do T-scores allow one to determine quickly how far a person or a sample is from expected norms, but the use of a common metric for expressing information allows one to plot all CS scores on a graph using a single axis for comparison. This makes it easy to see how typical or atypical values are for the person or sample that is being compared to the norms.

To facilitate cross-national comparisons, we will present two types of graphs. The first will be boxplots that present country-specific distributions for the first 136 scores listed in Table 1 (i.e., excluding the less often reported scores from WD+ to HRV). They show how the Composite International T-scores are distributed within each country. Given that positive and negative deviations from the mean cancel out, the most salient information in these graphs is the dispersion of scores. The second will be line graphs in which a separate line is plotted for each country across the scores in the 8 clusters of information found on the lower portion of the CS structural summary. These are dense graphs, though they allow one to see how each country compares to the others and identify any sample that produces unusual results.

TABLE 2.—Composite adult international reference values for the comprehensive system based on data from 17 countries: The average proportion of people in traditional score-based classifications and variability in those averages across samples.

Variables	Mean%	*SD* of M
Styles		
Introversive	26%	±9%
Pervasive Introversive	16%	±8%
Ambitent	31%	±7%
Extratensive	16%	±5%
Pervasive Extratensive	9%	±3%
Avoidant	28%	±9%
D-Scores		
D Score > 0	12%	±6%
D Score = 0	46%	±7%
D Score < 0	41%	±11%
D score < –1	23%	±7%
Adj D Score > 0	19%	±8%
Adj D Score = 0	52%	±4%
Adj D Score < 0	30%	±9%
Adj D score < –1	13%	±5%
Zd		
Zd > +3.0 (Overincorp)	19%	±6%
Zd < −3.0 (Underincorp)	29%	±5%
Form Quality		
XA% > .89	19%	±9%
XA% < .70	18%	±10%
WDA% < .85	49%	±12%
WDA% < .75	20%	±11%
X+% < .55	55%	±13%
Xu% > .20	68%	±13%
X−% > .20	41%	±14%
X−% > .30	14%	±9%
FC:CF+C Ratio		
FC > (CF+C)+2	13%	±7%
FC > (CF+C)+1	22%	±9%
(CF+C) > FC+1	24%	±9%
(CF+C) > FC+2	15%	±6%
Constellations		
S-Constellation Positive	4%	±2%
HVI Positive	12%	±7%
OBS Positive	0%	±1%
PTI = 5	0%	±0%
PTI = 4	2%	±1%
PTI = 3	6%	±4%
DEPI = 7	2%	±2%
DEPI = 6	10%	±5%
DEPI = 5	19%	±5%
CDI = 5	11%	±5%
CDI = 4	25%	±6%
Miscellaneous Variables		
R < 17	25%	±10%
R > 27	20%	±9%
DQv > 2	16%	±6%
S > 2	40%	±10%
Sum T = 0	57%	±12%
SumT > 1	15%	±7%
3r+(2)/R < .33	39%	±11%
3r+(2)/R > .44	30%	±8%
Fr+rF > 0	25%	±6%
Pure C > 0	25%	±9%
Pure C > 1	7%	±4%
Afr < .40	27%	±7%
Afr < .50	47%	±10%
(FM+m) < Sum Shading	35%	±13%
(2AB+Art+Ay) > 5	11%	±5%
Populars < 4	16%	±8%
Populars > 7	12%	±5%
COP = 0	42%	±12%
COP > 2	13%	±7%
AG = 0	64%	±8%
AG > 2	4%	±2%
MOR > 2	16%	±7%
Level 2 Sp.Sc. > 0	17%	±9%
GHR > PHR	57%	±12%
Pure H < 2	35%	±9%
Pure H = 0	11%	±4%
p > a+1	21%	±9%
Mp > Ma	32%	±8%

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

Scoring Reliability

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).

T-Scores in the Adult Samples

Figure 1 provides boxplots for the 20 adult samples across 136 CS scores. The samples are designated on the horizontal axis by short names for each country and the sample size. For each country, the box indicates where the central 50% of the Composite International T-scores fall, with the bottom of the box at the 25th percentile and the top of the box at the 75th percentile. The line in the box indicates the median T-score. For instance, in the Argentine sample of 506 participants, 50% of the 136 variables have T-scores that fall in a narrow range between 47 and 49, with the median at 48. For Exner's US sample of 450 participants, 50% of the variables have a T-score between 48 and 54 and the median is at a T-score = 49.5.

The lines or "whiskers" extending from the upper and lower end of each box indicate the range of scores, excluding outliers. The length of the whiskers is defined in relative terms; they extend no more than 1.5 times the length of the box from either its top or bottom. Values beyond this point are defined as outliers. The first panel shows the range of scores, excluding outliers. T-score values that fall 1.5 to 3.0 box lengths from the top or bottom of the box are considered outliers and are designated in the second panel of the graph with an "O." Extreme values are also defined in country-specific terms; these are T-score values that fall more than 3.0 box lengths from the top or bottom of the box

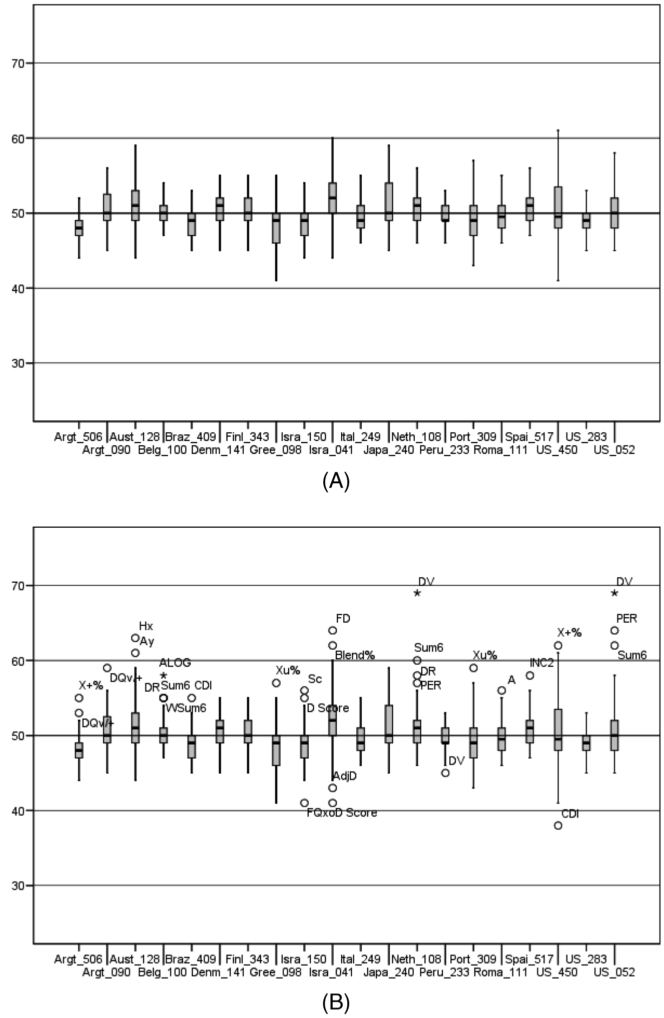

FIGURE 1.—Distribution of the adult international T-scores across 136 scores in 20 adult samples.

and they are designated with an asterisk (*). In the second panel, each of the outliers and extreme values are labeled. For instance, X+% is considered an outlier for the Argentinean sample of 506 and the US sample of 450. Note that because the T-scores for the Argentinean sample have such a tight distribution, the X+% T-score of 55 is considered an outlier in the plot, even though 55 is not a high T-score value and even though it would not be considered an outlier in most of the other samples.

The main message from Figure 1 is that with rare exception the T-scores across all 136 scores in all 20 adult samples fall in a fairly narrow range between 40 and 60, with half of the scores (as indicated by the length of the boxes) falling in an even narrower range between 47 and 53. In clinical practice, a difference of five T-score points (which is equal to 7.5 points on a traditional IQ scale) is considered to have potential clinical significance, and T-scores above 65 or below 35 are generally considered noteworthy because they are atypically elevated or suppressed relative to normative standards (e.g., Greene, 2000). Using the latter benchmark, the only noteworthy scores in Figure 1 are the Level 1 DV scores from The Netherlands and the sample of 52 older adults from the US. Because of the DV1 elevation, these samples are also elevated on Sum6 (but not WSum6). It is likely that the DV1 elevation in the older adult sample reflects some age-related word finding difficulties. We understand from our Dutch colleagues that the elevated rate of DV1 in that sample is largely related to a formally incorrect but fairly common substitution of the word for a human head (*hoofd*) to indicate a face versus an animal head (*kop*).

It is also noteworthy that Exner's (2007/this issue) reference sample of 450 produces the highest T-scores for X+% (T = 62) and the lowest T-scores for the CDI sum (T = 38). Thus, those reference values do not generalize well to other countries or to other samples from the US, which can be seen more clearly in the next set of graphs.

Figure 2 provides the line graphs for the eight CS clusters, with panels A through H, respectively, providing data on Resources and Controls, Affect, Interpersonal, Ideation, Mediation, Information Processing, Self-Perception, and the Constellations. With eight graphs to a page, the figures are small and are designed to give a gestalt perspective on cross-sample comparability. They are not in a form that allows results to be read for each country. However, each line represents the T-scores for a specific sample, with each differentiated by line type and symbol. For easy reference, Exner's (2007/this issue) sample of 450 is depicted with a heavier black line and filled circles to designate the T-score value for each variable.

Across graphs it can be seen that in almost all instances, each country has T-scores in the relatively narrow range between 45 and 55. These findings indicate that, by and large, adults look pretty much the same on the CS no matter what language they speak, what country they reside in, and what cultural background influences them. A question remains about whether the relatively small variation between countries is due to differences in culture, language, participant selection criteria, administration standards, coding benchmarks, and/or examiner skill. Despite all these potential influences, because the between-sample differences are modest, the findings support the transportability of the Rorschach across countries.

At the same time, however, some samples differentiate themselves from the rest. The Israeli sample of 41 has more Texture, FD, Blends%, Sum Shading, and es, and a correspondingly lower D and Adjusted D than other samples. As the author of this study has suggested, these findings may be the result of distress experienced from the Israeli Mideast conflict occurring at the time the protocols were collected (Tibon, 2007/this issue).

The most critical sample to consider in terms of relatively extreme scores is Exner's 450. Although the scoring applied to this sample sets the high mark for a number of variables, including D, DQ+, EA, FC, CF, Popular, and GHR, and the low mark for others, including Lambda, Dd, DQv, M–, Pure C, Xu%, and PHR, of greatest importance is that scores for certain variables are relative outliers. The most prominent are several of the constellations, including the PTI, DEPI, CDI, and S-CON, and the form quality variables X+%, WDA%, XA%, and X−% (though the Portuguese sample has similar values on the last two scores). The sample of 450 also is a relative outlier on COP, WSumC, and HRV scores. To generalize to samples collected around the world, the Composite International Reference values in Table 1 provide a more appropriate target than Exner's reference scores for these variables.

In summary, when plotted on the Composite International T-scores, these adult samples show a basic consistency that holds across cultures, languages, examiners, exclusion criteria, and recruitment strategies. Overall, adults from different countries and cultures look similar on the full range of CS scores. Exner's new sample of 450 is more similar to the others than his sample collected in the 1970s. However, the coding applied to the new sample still defines the healthy end of the form quality variables and gives notably lower PTI, DEPI, CDI, and S-CON scores.

If Exner's reference sample is set aside, the mean T-score across variables is 50 and the *SD* of the sample-specific mean T-scores (SD_{MT}) for each variable ranges from 1.1 to 6.9 (M = 2.6).[3] Examining the SD_{MT} across all variables, two variables are high outliers; DV1 and Sum6, with SD_{MT}s = 6.9 and 4.8, respectively. As discussed above, the high degree of mean T-score variability for these scores was a function of DV scores in the Dutch and older US sample. All other variables have more narrow T-score distributions, with SD_{MT}s ranging from 1.1 to 4.3.

T-Scores in the Child and Adolescent Samples

The available CS norms for children and adolescents were targeted to the 1970 US census and first published in 1982 (Exner & Weiner, 1982). At the time the authors questioned how representative their samples were, cautioning users that as a result of likely self- and parent-selection biases they were probably too healthy and well-functioning to generalize to typical participants. Particularly because of this, it is important to know what more recently collected samples look like when plotted on the Adult Composite International Norms. Table 3 provides a summary of the ages and sample sizes for the 31 child and adolescent samples included in these analyses (Exner, 2003; Hamel & Shaffer, 2007/this issue; Hansen, 2007/this issue; Lis, Salcuni, & Parolin, 2007/this issue; Matsumoto, Suzuki, Shirai, & Nakabayashi, 2007/this issue; Salcuni, Lis, Parolin, & Mazzeschi, 2007/this issue; Silva & Dias, 2007/this issue; Valentino, Shaffer, Erdberg, & Figueroa, 2007/this issue; Van Patten, Shaffer, Erdberg, & Canfield, 2007/this issue). As can be seen, these data draw on a total of 2,647 protocols but come from just five countries, with few covering the full developmental spectrum and many relying on fairly coarse age groupings. Samples in the first six rows are from this Supplement, while the 12 year-by-year samples in the last row are from Exner (2003).

Figure 3 provides boxplots for these 31 samples; each is designated on the horizontal axis by a short name and their average age. Exner's samples provided data for 116 scores, the three younger Italian samples for 135 scores (S-CON totals were not computed), and the remainder had data for 136 scores. As before, the first panel presents the distribution of adult-based T-scores in each sample excluding outliers and extreme values. The latter are shown and labeled in the second panel. What is most obvious from Figure 3 is the greater dispersion of scores relative to the adult distributions. For most samples, the range of T-scores, as indicated by the whiskers, extends from about 35 to 65, and the boxes and whiskers in these samples are notably longer than for adults. More evident, however, are the outliers and extreme values in each sample. In fact, the plotted T-score range from 20 to 80 is no longer adequate. To visualize all values, the T-score range on the vertical axis has to extend from

[3]Results are essentially identical with this sample included; the mean T-score across variables is 50 and the SD_{MT} for each variable ranges from 1.1 to 6.8 (M = 2.7).

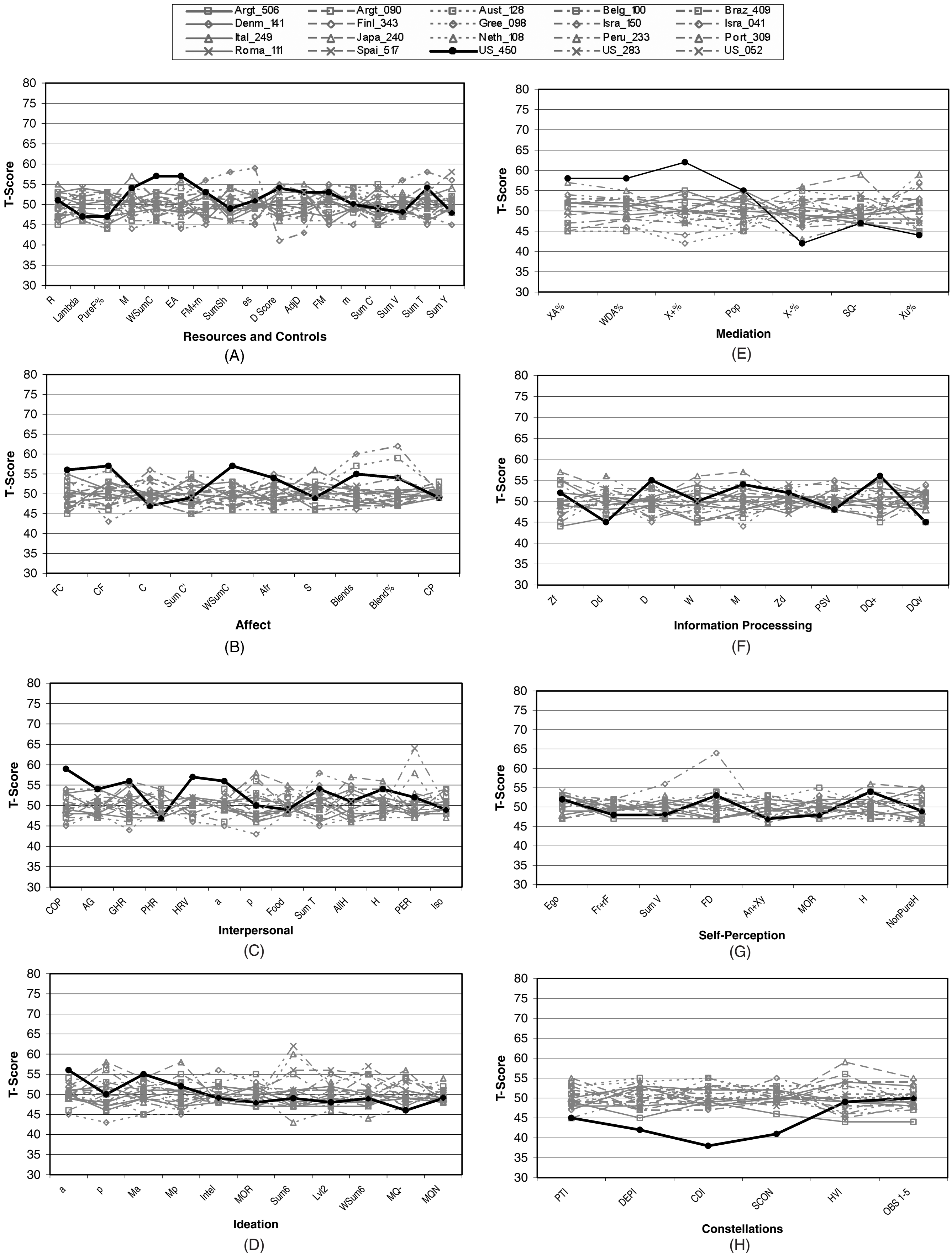

FIGURE 2.—Composite adult international T-scores for eight CS score clusters in 20 adult samples.

TABLE 3.—Age ranges and sample sizes for the 31 child and adolescent samples.

<table>
<tr><th>Country</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th></tr>
<tr><td>Denmark</td><td></td><td></td><td></td><td></td><td>75</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Italy</td><td colspan="3">75</td><td colspan="4">148</td><td colspan="3">116</td><td colspan="4">117</td></tr>
<tr><td>Japan</td><td colspan="2">24</td><td colspan="2">43</td><td colspan="2">42</td><td colspan="2">42</td><td colspan="2">39</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Portugal</td><td></td><td>86</td><td>69</td><td>75</td><td>66</td><td>61</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>US-Non-Exner</td><td></td><td colspan="4">50</td><td colspan="3">50</td><td></td><td></td><td colspan="3">37</td><td></td></tr>
<tr><td>US-Mexican American</td><td></td><td></td><td></td><td colspan="3">42</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>US-Exner</td><td>90</td><td>80</td><td>120</td><td>120</td><td>140</td><td>120</td><td>135</td><td>120</td><td>110</td><td>105</td><td>110</td><td>140</td><td></td><td></td></tr>
</table>

10 to 140, which is 4 *SD*s below and 9 *SD*s above the adult reference mean. This version of the figure is the third panel of Figure 3.

Two other features are noteworthy about these data. First, there are substantial disparities across samples and across countries. The data from Japanese children are the most unusual, with extremely high Lambda values and extremely low XA% and WDA% values. Second, Exner's samples produce high outlier values for X+% and often XA%; however, many of the other samples produce low outlier values for the same variables. Both of these issues are more evident with the line graphs, which are presented in Figure 4.

As before, the line graphs indicate the T-scores for each sample on specific variables. Samples from the same country share a symbol, with age groups differentiated by line type. As before, these graphs are small and are presented to give a visual gestalt of the data, not to read sample specific results. However, for easy reference, Exner's samples are shown with a heavier black line and the age groupings are differentiated by symbols. Exner's samples do not have dimensional total scores for the Constellation Indices so they are missing from the final panel. As noted previously, the scale on the vertical axis must be expanded to accommodate T-scores in the Resources and Mediation clusters. Both modified line graphs are presented in Figure 5.

There are often substantial disparities across the child and adolescent samples. Furthermore, relative to the adult international standard, these samples have particularly high Lambda scores and what would be considered unhealthy-looking form quality scores. The major exception, however, is Exner's reference data, where Lambda is a bit lower than in the adult reference norms and the coded form quality scores appear substantially healthier. The disparity of T-scores for FQ in the Mediation graph is particularly compelling and indicates the existing CS reference data for form quality cannot be used to evaluate the conventionality of perception in contemporary samples of children and adolescents.

As would be expected, the high Lambda and Form% values seen in the non-Exner samples are associated with lower frequencies of other scores related to the richness of verbalizations and visual perception, including determinants, contents, and non-cognitive special scores (e.g., COP, AG, MOR). T-scores for these variables are generally in the range of 35–55. Also, in the non-Exner samples FQ generally is less healthy than in adults, with T-scores ranging from 35 to 45 for X+%, XA%, WDA%, and Popular responses, and T-scores in the range from 55 to 65 for X−%.

The exception to both of these generalizations occurs for the Japanese samples. The markedly elevated Lambda values in these samples occurred after the authors had to discard 43% of their protocols because *R* was less than 14 even after two administrations of the test. It seems that these Japanese children had a difficult time engaging in the Rorschach task. They were reluctant, inhibited, or unable to offer many responses, and when a response was offered it was not elaborated with much depth or complexity, which also resulted in a higher number of PSV scores. This phenomenon may in part be the result of cultural factors (e.g., Matsumoto, 2005) associated with how to respond in situations of uncertainty or ambiguity, values emphasizing modesty (*tsutsumashii*) and considerate sensibility (*sassuru*) that can limit verbal self-expression, cultural constraints on what Japanese children are expected to do when evaluated by unfamiliar adults, simplification as a strategy to cope with extensive stimuli present in current Japanese society, and/or administration factors that can be associated with depth of responding, including quality of rapport and level of examiner training (e.g., Lis, Parolin, Calvo, Zenarro, & Meyer, 2007/this issue). However, the high Lambda values observed by this team of researchers are generally consistent with those Nakamura et al. (2007/this issue) described finding in another sample of Japanese adolescents.

The Japanese samples also had extreme values on form quality, with X−% markedly elevated at all ages (raw $M = .47$ to .66) and X+% markedly low ($M = .26$ to .41). In part, these findings are due to the authors' decision to generally code responses that could not be found in the FQ tables as minus responses, rather than extrapolating from other similarly shaped objects. This approach to coding also probably explains why Xu% was low in these samples ($M = .09$ to .12) relative to the other non-Exner samples ($M = .21$ to .46), though it should not have a notable impact on X+% or Populars, both of which are low but also more similar to other samples. If the FQ scores for the Japanese samples are set aside, it reduces but does not eliminate the notable disparities between Exner's reference data and the other contemporary samples.

Although the Japanese samples were notable for their Lambda and FQ scores, across the other child and adolescent samples there are a number of sample-specific atypical values. For instance, the Portuguese samples generally produced a high level of DQv responses. Hamel and Shaffer's (2007/this issue)

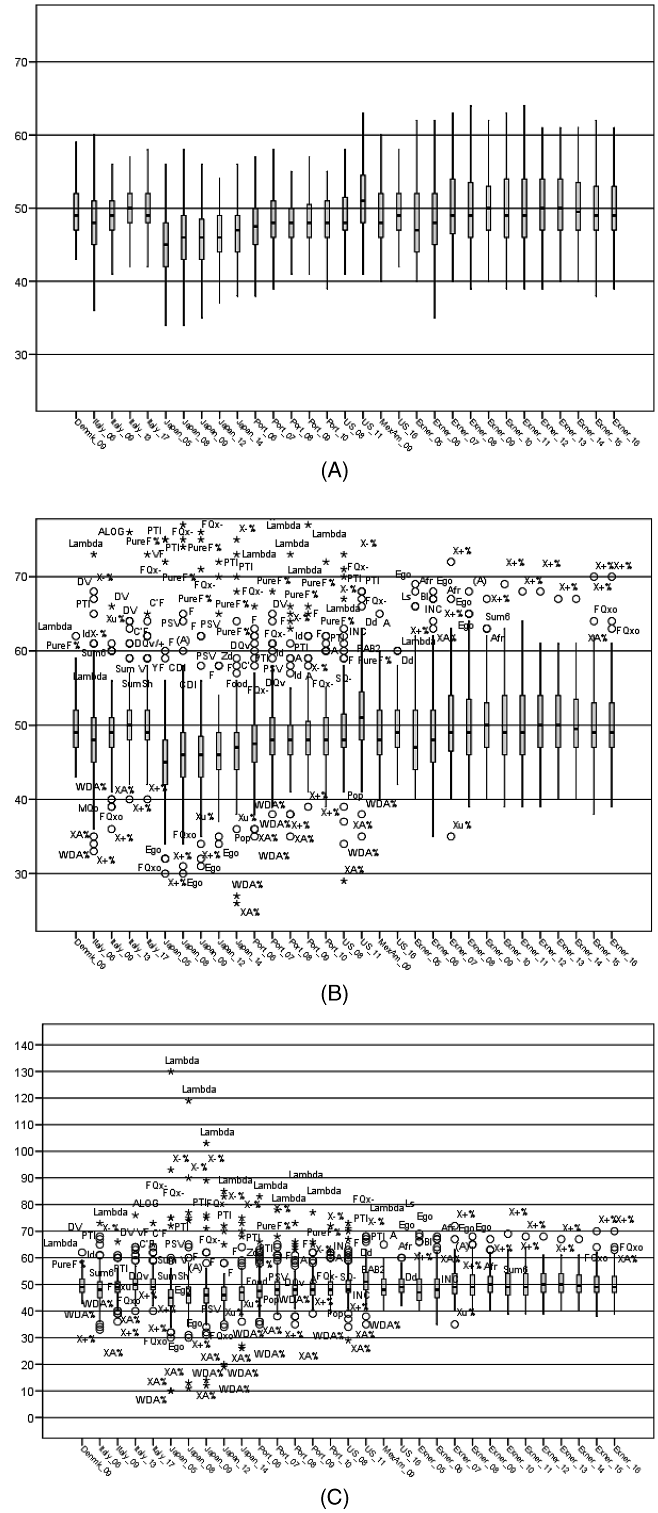

FIGURE 3.—Distribution of the adult international T-scores in 31 child and adolescent samples.

samples of US children produced atypically high Dd values, which is the consequence of the strict procedures the first author followed during administration and scoring (see Hamel & Shaffer, pp. S175–S176, S178–S179). In addition, their older sample had unusual elevations on m and S. Finally, the two Italian adolescent samples produced an markedly elevated number of dimensional responses (Vista raw $M = 1.4$–1.6; FD raw $M = 1.5$–1.6) relative to the other samples. When combined with their elevated frequency of diffuse shading and color responses, these two samples produced Lambda and Form% values that were notably lower than any of the other non-Exner samples. The complexity of these protocols carried over into an elevated number of blends and low values for D and Adjusted D.

Although the graphs do not make this clear, the data published in this supplement and also reported by Andronikof-Sanglade (1999) for French children provide some evidence for developmental changes. These changes were more visible within a country than across countries. In general, the most notable trends were for Lambda values to decrease with age and for form quality to become healthier with development. However, developmental trends on these two variables were not present in Exner's (2003) samples.

Although not as consistent across or within countries, other age-based developmental tendencies include increases in M, human content, and what can be considered more complex perceptions or the articulation of more subtle blot features (e.g., DQ+, Blends, Fr+rF, Vista, other shading). Table 4 provides a number of example scores by age for the samples included in this Supplement.

As noted above for the adult samples, with the exception of DV1 and Sum6, the *SD* of the mean T-scores (SD_{MT}) across samples for each variable in Table 1 was less than 4.5. If one disregards developmental differences and creates a combined reference sample using the 19 recently collected child and adolescent samples reported in this supplement (i.e., excluding those from Exner, 2003, which were obtained more than 25 years ago), applying the same benchmark reveals that 108 out of 143 variables have that degree of consistency across countries and samples. These are variables that show a fairly reasonable degree of consistency across samples, regardless of age, culture, language, examination context, examiner training, and potential differences in site-specific administration and scoring conventions. They are denoted with bold font in Table 5; the remaining 47 variables, all of which have SD_{MT}s > 4.5, are in standard font.

Table 5 also provides the average of the raw *M*s and *SD*s (computed from variances), as well as the average of the mean T-scores and the *SD* of these mean T-scores (SD_{MT}) across samples. As described above, the SD_{MT} indicates the average variability in mean T-scores across samples. About 70% of the sample-specific means fall in a range from 1 SD_{MT} below to 1 SD_{MT} above the average T-score; about 95% of the sample means fall in the range defined by $\pm$ 2 SD_{MT}s. For instance, Table 5 indicates that about 70% of the T-score means for *R* across the child and adolescent samples fall in a range from 47 to 53 (i.e., $50 \pm 1\ SD_{MT} = 47.0$ to 53.0).

A SD_{MT} cut-off of 4.5 indicates that about 70% of the sample means will fall in a 9-point range that is centered on the average T-score. For instance, if a mean T-score was 50, applying the $SD_{MT} < 4.5$ criterion would indicate that about 70% of the sample-specific T-score means would have to fall in a range that was narrower than 45.5–54.5. Although a cut-off of 4.5 is rather arbitrary, and although others may choose more stringent and conservative benchmarks, we believe that useful international reference values cannot be established for scores that vary this much or more across samples. However, to the extent that SD_{MT}s are more narrow than this (e.g., W = 2.9; SumT = 1.8), the descriptive information for the bolded variables in

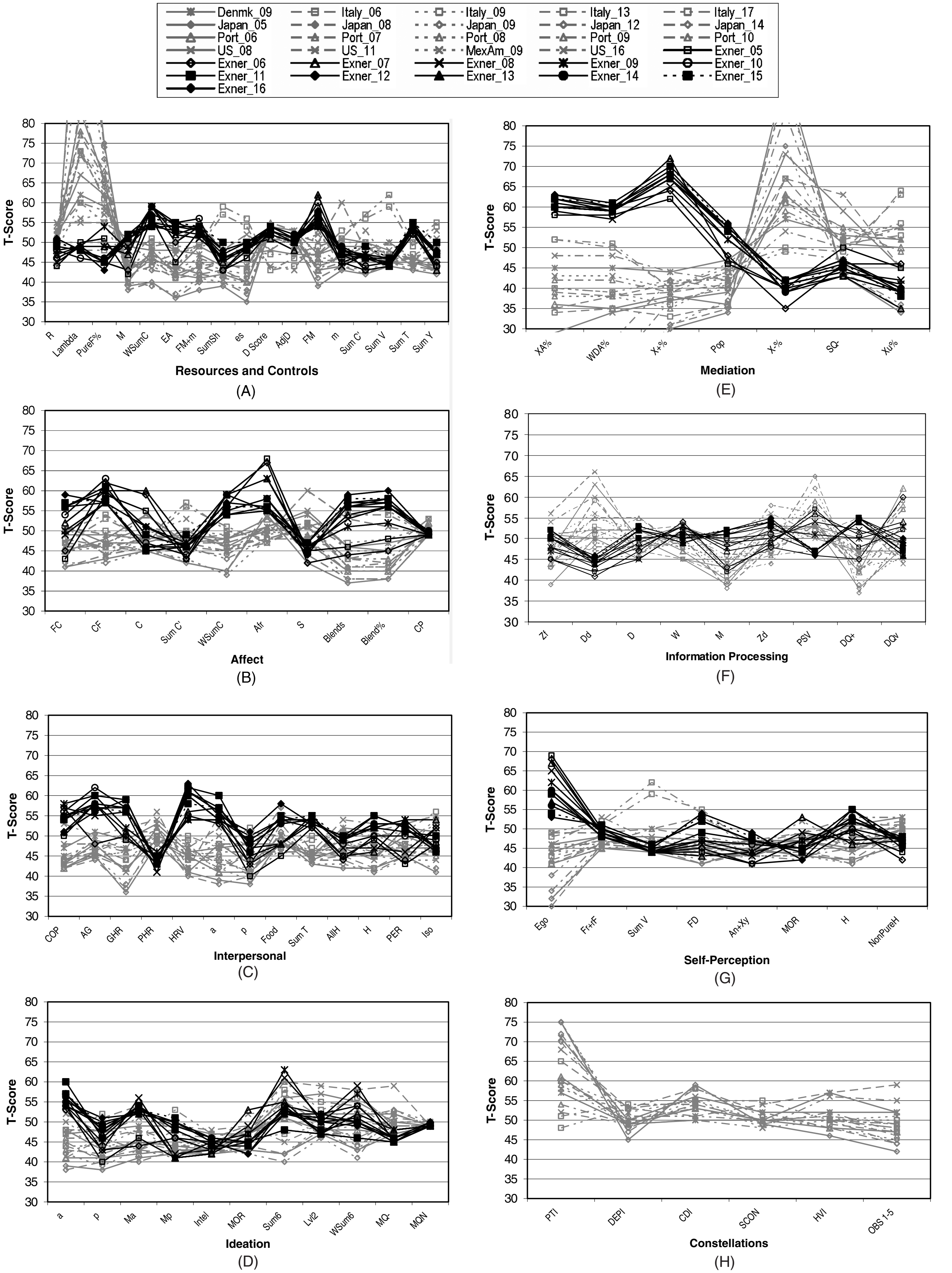

FIGURE 4.—Reference values for eight CS score clusters in 31 child and adolescent samples based using the composite adult international T-scores.

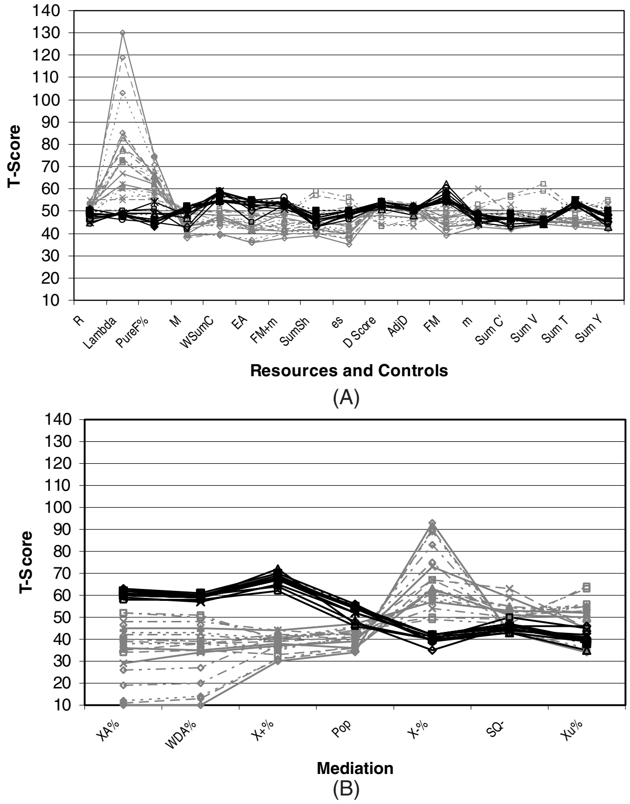

FIGURE 5.—Adjusted T-score ranges for the resources and controls and mediation clusters in 31 child and adolescent samples.

Table 5 should help guide clinical interpretation for children of all ages.

One should question our decision to combine data from children of all ages, as doing so disregards potentially important developmental progressions. However, we also examined two sets of data from more homogeneous age groups. The first considered three subsamples, aged 5–9, 8–12, and 12–18. The samples contributing to each group are indicated in Table 3 by different patterns of cross-hatching. Each group contained at least four samples from at least three countries. Table 5 provides the mean ages as well as the mean T-scores for each of these subsamples in columns 6–8. To facilitate developmental considerations, the final column gives the simple T-score difference between the oldest group and the youngest. T-scores in bold font indicate the mean T-scores across all samples in that age group were relatively homogeneous (i.e., $SD_{MT} < 4.5$); T-scores in regular font indicate means that were relatively heterogeneous. For instance, although the mean T-scores for S had fair variability when considering all 19 samples (column 5, $SD_{MT} = 4.0$), the regular font T-score of 49 in column 6 indicates that the 5–9-year-old age group had variability that exceeded our cut-off (although not shown in Table 3, the SD_{MT} in this group was 4.67).

Overall, of the 108 scores considered reasonably homogeneous when considered across all age groups, 17 no longer met this benchmark in at least one of the age-based subgroups (S, m, CF, YF, Sum C′, Sum Y, FA, FM+m, D Score, MQu, zd, Color-Shading Blends, (A), Food, FAB2, PSV, and OBS Total). No variable (GHR) became reasonably consistent in the age-specific subgroups when it was not initially.

Because the ages from 12–18 encompass many developmental changes, we also examined four groups with more narrowly defined ages for the teen years (5–8, 8–12, 11–14, and 15–18). In these analyses, 13 variables that were reasonably consistent for the three age subgroups no longer were (FY, a, p, Ma, Zf, Fi, Ls, Na, Isolate/R, INC2, LvI 2 Sp Sc, COP, and S-Con Total). In only one instance (Fd) did the reverse occur. Because the mean T-scores tended to be more variable in the more narrowly defined age groups, we decided to emphasize the overall combined sample in Table 5, while still presenting mean T-scores for the three age-defined subgroups.

The child and adolescent data are incomplete, emerging from just five countries that also varied in their sampling of the full spectrum of development and their grouping of samples across ages. This makes it difficult to disentangle what may be influences due to culture, age, or administration and scoring effects. Nonetheless, when considering the composite of child and adolescent data, the form quality variables and Lambda stand out. They were the most erratic across samples, the most divergent from Exner's data, the most divergent from

TABLE 4.—Selected mean scores by country and age.

Country	*N*	Age	R	Lambda	Dd	DQ+	DQv	Blends	M	H	GHR	EA	es	SumV	FrrF	Afr	X+%	X−%	WSum6	PTI
Denmark	75	9	23.6	2.0	3.2	6.1	0.9	2.9	3.4	2.6	3.6	6.0	6.2	0.1	.2	.55	.44	.27	9.4	1.1
Italy	75	5–7	21.2	3.0	6.2	3.5	1.5	2.0	1.3	0.9	1.9	3.5	5.6	0.2	.0	.48	.30	.38	10.4	2.0
Italy	148	8–11	20.7	1.8	3.8	5.1	1.3	3.2	1.9	1.5	2.4	4.7	8.2	0.2	.1	.53	.34	.31	12.1	1.5
Italy	116	12–14	22.3	0.9	4.1	5.2	1.6	5.3	2.8	2.4	3.5	6.2	11.4	1.4	.5	.46	.39	.19	12.8	0.7
Italy	117	15–18	21.8	0.7	4.2	5.0	2.5	5.4	3.5	2.4	3.9	6.6	12.2	1.6	.6	.48	.39	.18	4.8	0.4
Japan	24	5–6	17.2	8.5	3.2	1.8	0.8	0.3	0.7	0.9	0.8	1.6	1.7	0.0	.0	.61	.26	.66	2.6	3.0
Japan	43	7–8	20.7	7.4	2.7	1.5	1.1	0.4	0.7	0.8	0.9	1.7	2.7	0.1	.0	.63	.27	.63	2.5	3.0
Japan	42	9–10	20.2	5.9	3.1	2.5	0.6	0.4	1.2	0.8	1.1	1.9	3.1	0.1	.1	.48	.27	.62	2.9	3.0
Japan	42	11–12	20.0	4.2	1.6	3.3	1.1	1.1	1.7	1.5	1.9	3.5	3.8	0.1	.2	.53	.35	.55	2.5	2.7
Japan	39	13–15	21.8	3.1	2.4	4.5	0.9	1.4	2.1	2.0	2.3	3.8	4.7	0.1	.3	.49	.41	.47	1.0	2.5
Portugal	86	6	22.7	4.0	3.9	3.4	3.0	1.2	1.1	1.5	2.1	3.9	4.1	0.1	.0	.59	.37	.33	6.4	1.6
Portugal	69	7	25.3	3.5	4.0	4.7	2.3	1.4	2.0	1.9	2.9	4.4	4.2	0.1	.0	.53	.38	.31	7.4	1.5
Portugal	75	8	24.6	3.1	5.0	4.6	2.2	1.4	1.8	1.9	2.8	3.8	5.2	0.1	.0	.51	.33	.33	6.7	1.5
Portugal	66	9	25.7	3.4	5.0	4.0	2.2	1.9	1.9	1.7	2.9	4.3	5.4	0.2	.1	.52	.38	.29	5.7	1.0
Portugal	61	10	24.0	2.9	3.4	5.3	0.5	2.1	2.5	2.1	2.8	4.4	5.9	0.5	.1	.55	.42	.32	8.5	1.5
US	50	6–9	24.5	2.5	7.8	5.3	0.6	2.8	1.8	1.4	1.8	4.1	7.5	0.3	.2	.55	.35	.44	11.2	2.5
US (MA)	42	8–10	24.0	1.4	5.3	3.9	0.3	2.6	2.3	1.4	2.7	4.4	9.1	0.1	.1	.54	.39	.28	10.0	1.4
US	50	10–12	26.5	1.3	8.7	8.0	0.3	4.8	4.0	2.7	3.0	7.0	10.1	0.6	.3	.50	.40	.38	13.8	2.3
US	37	15–17	24.7	1.8	6.8	7.7	0.2	2.8	3.8	3.0	4.0	5.8	7.5	0.2	.7	.47	.44	.23	4.0	0.8

TABLE 5.—Combined child and adolescent reference samples: Descriptive data (*M*s and *SD*s for raw scores and T-scores) for a general sample averaged across all ages and T-scores for three age-based subgroups.

	Raw scores		T-scores		*M* Age-based T-scores			
Variable	*M*	*SD*	*M*	SD_{MT}	5–9	8–12	12–18	T Difference
Age	9.95				6.92	9.94	15.29	12/18–5/9
R	22.71	8.09	51	3.0	50	51	50	0
W	8.86	5.12	49	2.9	48	50	51	3
D	9.41	5.91	49	2.9	49	50	48	−2
Dd	4.44	4.29	53	5.5	54	53	53	−1
S	2.61	2.16	51	4.0	49	52	51	2
DQ+	4.49	3.80	45	4.7	42	46	48	6
DQo	16.70	7.14	53	2.6	53	54	51	−2
DQv	1.25	1.98	51	5.4	54	49	51	−2
DQv/+	.26	.67	50	4.9	49	48	54	5
FQx+	.02	.15	47	0.4	47	47	47	0
FQxo	7.99	3.28	42	4.7	39	43	44	5
FQxu	6.12	4.15	50	6.9	48	50	54	6
FQx–	8.36	4.45	62	8.3	66	63	55	−11
FQxNone	.22	.72	48	3.0	50	48	47	−3
MQ+	.01	.09	47	0.4	47	47	47	0
MQo	.95	1.16	42	3.0	40	43	45	4
MQu	.48	.91	48	3.9	46	47	53	7
MQ–	.67	1.07	50	2.8	48	52	51	3
MQNone	.01	.08	49	0.4	49	49	50	1
SQ–	.93	1.27	51	5.6	51	52	48	−3
M	2.12	2.22	44	3.7	41	45	47	6
FM	2.47	2.42	46	3.7	43	47	48	5
m	1.09	1.44	47	4.2	45	48	49	4
FC	1.40	1.53	47	3.3	45	48	49	3
CF	1.16	1.41	47	3.1	46	47	49	4
C	.21	.58	48	2.7	49	48	47	−2
Cn	.01	.10	49	1.2	50	49	49	−1
SumC	2.77	2.38	46	3.7	44	46	48	4
WSumC	2.17	2.00	46	3.4	45	46	47	3
FC′	1.04	1.39	47	3.3	46	48	49	4
C′F	.30	.68	50	5.1	49	49	55	5
C′	.07	.43	50	2.4	51	49	51	0
FT	.20	.51	46	1.7	45	45	47	2
TF	.03	.17	48	0.8	48	48	48	0
T	.00	.07	49	0.8	49	50	50	0
FV	.19	.58	48	2.4	46	47	51	5
VF	.13	.59	50	6.8	48	48	58	10
V	.00	.06	49	1.0	49	49	49	0
FY	.48	.94	47	3.9	44	47	50	6
YF	.22	.58	48	3.6	47	48	52	5
Y	.03	.20	49	1.0	48	49	49	1
Fr	.16	.53	48	2.4	46	47	52	5
rF	.02	.18	49	0.9	48	49	50	1
Sum C′	1.40	1.76	48	4.1	46	48	52	5
Sum T	.24	.57	46	1.8	45	45	48	3
Sum V	.31	.94	48	4.8	46	47	53	8
Sum Y	.73	1.20	46	4.2	44	47	50	6
SumSh	2.68	2.93	45	5.3	43	45	51	8
Fr+rF	.19	.59	47	2.4	46	47	51	5
FD	.56	.95	46	4.1	44	46	50	5
F	13.56	6.67	59	5.0	62	59	53	−9
Pair	5.20	4.06	45	5.8	44	46	45	1
Ego	.25	.16	42	5.9	39	42	45	6
Lambda	3.24	4.10	75	22.0	89	71	58	−31
PureF%	.60	.19	62	7.4	67	61	55	−12
FM+m	3.55	3.08	45	4.3	43	47	48	5
EA	4.30	3.26	43	4.1	41	44	47	6
es	6.23	5.00	44	5.8	41	45	50	9
D Score	–.59	1.36	51	3.9	52	51	48	−5
AdjD	–.28	1.05	49	2.8	50	50	48	−2
a	3.49	3.23	45	4.4	42	47	47	5
p	2.21	2.16	44	4.0	41	45	48	7
Ma	1.24	1.65	45	3.5	43	47	47	4
Mp	.89	1.14	45	3.1	43	45	50	6
Intel	.86	1.57	44	1.8	43	45	46	2
Zf	11.35	5.52	48	4.0	45	49	50	5
Zd	–.72	5.00	50	3.7	50	50	50	0
Blends	2.27	2.65	44	5.3	41	44	49	8
Blend%	.10	.11	44	5.1	41	44	49	8
C–Sh–Bl	.39	.79	48	3.5	46	48	50	4
Afr	.53	.20	50	2.2	51	50	48	−4
Pop	3.65	1.77	41	3.8	38	42	43	6
XA%	.62	.13	34	13.2	28	35	45	16
WDA%	.65	.14	35	12.4	29	35	44	15
X+%	.36	.13	38	4.3	35	39	42	7
X–%	.38	.13	67	13.3	73	67	57	−16
Xu%	.25	.11	49	9.7	46	48	55	8
Iso	.17	.14	48	4.2	47	47	52	5
H	1.75	1.68	46	3.6	44	47	50	6
(H)	1.08	1.25	49	2.3	49	49	48	−1
Hd	1.57	1.95	50	2.2	49	50	53	3
(Hd)	.63	.94	50	1.9	49	51	51	2
Hx	.15	.56	47	2.1	46	48	49	3
AllH	5.04	3.23	48	3.2	46	48	51	5
NonPureH	3.29	2.64	50	2.1	49	50	51	2
A	9.12	4.29	54	4.7	54	57	51	−4
(A)	.46	.88	51	3.5	51	50	51	−1
Ad	2.73	2.66	52	3.3	50	54	50	0
(Ad)	.15	.44	50	2.1	50	49	53	3
An	.52	.91	46	1.4	46	45	45	−1
Art	.52	1.11	45	2.2	45	46	44	−1
Ay	.17	.47	46	1.6	45	46	48	3
Bl	.17	.50	48	1.7	49	48	48	−1
Bt	1.27	1.44	49	2.4	48	49	50	2
Cg	1.57	1.62	48	2.9	48	48	49	1
Cl	.12	.40	49	1.5	49	49	49	1
Ex	.15	.40	49	2.4	47	50	51	4
Fi	.41	.79	49	3.1	49	49	49	0
Food	.25	.57	49	2.7	49	48	50	1
Ge	.07	.38	47	1.0	47	47	47	0
Hh	.62	.96	48	3.0	47	49	47	0
Ls	.80	1.26	49	3.7	48	49	52	4
Na	.82	1.23	51	3.8	50	50	53	3
Sc	1.18	1.53	51	3.1	50	49	54	4
Sx	.02	.14	45	0.3	45	45	45	0
Xy	.01	.12	46	0.6	46	46	47	1
Id	1.12	1.58	52	7.2	54	50	53	−1
An+Xy	.53	.92	45	1.3	45	44	45	−1
DV	1.04	1.43	54	8.8	56	53	52	−5
INC	.89	1.17	52	5.9	52	54	47	−5
DR	.17	.60	47	2.0	47	47	46	−1
FAB	.35	.75	49	2.5	48	51	47	−1
DV2	.02	.14	50	2.3	51	50	50	−2
INC2	.13	.45	51	2.8	51	51	51	0
DR2	.02	.20	49	1.1	49	49	48	−1
FAB2	.10	.41	51	3.8	51	52	49	−2
ALOG	.15	.53	50	6.6	49	49	55	7
CONTAM	.03	.18	51	2.7	50	52	50	0
Sum6	2.90	2.56	51	6.7	51	52	47	−4
Lvl2	.27	.76	50	3.5	50	51	49	−2
WSum6	7.09	7.82	49	5.3	49	51	47	−2
AB	.08	.45	47	1.7	46	47	49	3
AG	.27	.70	47	2.0	47	47	47	1
COP	.41	.75	44	2.6	43	45	46	4
CP	.02	.17	50	1.4	51	50	49	−2
GHR	2.48	1.85	44	4.5	42	45	49	7
PHR	3.01	2.59	50	2.2	50	51	51	1
MOR	.72	1.24	46	2.6	45	47	46	1
PER	.44	1.20	47	2.6	47	47	48	0
PSV	.47	.91	54	4.3	55	55	52	−4
PTI	1.79	1.18	63	8.6	67	63	55	−11
DEPI	3.82	1.08	50	2.4	49	50	51	2
CDI	3.45	1.04	54	2.7	56	54	53	−3
SCON	4.77	1.54	51	1.9	51	50	51	0
HVI	2.90	1.53	51	3.0	49	52	52	3
OBS 1–5	1.02	.81	49	4.0	47	49	50	3
WD+	.01	.12	47	0.4	47	47	48	0
WDo	7.49	3.08	41	4.2	38	42	43	5
WDu	4.46	2.91	49	5.9	47	49	51	4
WD–	6.15	3.48	64	10.1	67	65	57	−11
WDNone	.18	.60	48	2.5	49	47	47	−2
EII–2	.51	1.10	57	7.1	60	57	51	−9
HRV	–.53	2.93	45	2.8	44	45	48	4

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−%, Xu%), SumC, WSum C, and the CDI.

Some of the Israeli sample-to-sample differences appear due to intrinsic qualities of the participants (Tibon, 2007/this issue; Tibon, 2007). However, given emerging knowledge about how scores can differ as a function of potential across-site differences in scoring conventions or differences in administration, rapport, and inquiry (Lis, Parolin, Calvo et al., 2007/this issue; Meyer & Viglione, in press; Viglione & Meyer, 2008), almost all of the remaining sample-specific differences are likely to disappear with more thorough and detailed guidelines for administration and scoring, such as found in Viglione's (2002) *Rorschach Coding Solutions* and Sciara and Ritzler's (2006) book and DVD on administration and inquiry.

For instance, in addition to the findings described by Lis, Parolin, Calvo, et al. (2007/this issue), research by Meyer, Viglione, Erdberg, Exner, and Shaffer (2004) showed two important points when examining 40 protocols each from Exner's (2007/this issue) and Shaffer et al.'s (2007/this issue) samples. First, the Shaffer et al. sample had more instances when raters who were blind to the source of the protocols indicated a key word or phrase was never inquired ($d = .98$) or incorrectly inquired ($d = .98$), though they did not differ on the overall number of unnecessary inquiry questions ($d = .01$). Second, differences between the two US samples virtually disappeared when all protocols were rescored at a third site. Across 129 variables, there were 36 (27.9%) scores that initially differed by $d = .40$ or larger. However, when all 80 protocols were blindly rescored by a third group of researchers, only 3 scores (2.3%) still differed by this amount (FQNone, PER, DR1), indicating that much of the seeming variability was due to site-specific scoring conventions.

The value of the normative reference points presented in Tables 1 and 2 is that they indicate what can be expected from reasonably functioning adults across countries and cultures, while taking into account the limits of our current administration, inquiry, and scoring guidelines. Not only do they incorporate the seemingly small variability that may result from cultural differences, but more importantly they incorporate the kind of variability that can be expected within a country or region from different examiners administering, inquiring, and scoring the test. Using these values in clinical practice should help ensure that inferences about functioning generalize across examiners, levels of skill in administration and scoring, testing context, language, and culture. Inferences drawn from them also should help ensure that patients are being evaluated relative to a contemporary and broadly generalizable reference standard.

To date, the largest and most systematically organized effort to study personality around the world has been conducted with the Five-Factor Model of personality using the Revised NEO Personality Inventory (NEO–PI R; see McCrae & Terracciano, 2006; McCrae et al., 2005a, 2005b; Terracciano et al., 2005). These authors have compiled self- and observer-rating data on the NEO–PI R from 50 cultures and have documented both the transcultural similarity in adult personality and also how perceptions of so-called "national character" are based upon unfounded stereotypes that are not reflected in actual personality characteristics (McCrae & Terracciano, 2006; Terracciano et al., 2005). These authors based their conclusions on the results of the NEO–PI R. The data in Figures 1 and 2 are consistent with their conclusions, even though they are based upon the Rorschach, which is a very different method of assessment.

McCrae, Terracciano, and their colleagues (2005a, 2005b) used a single personality assessment instrument but typically compared findings from different cultures to the normative US reference sample for the NEO–PI R. As of yet, international norms for use with the NEO–PI R have not been published. Thus, the data in Tables 1 and 2 provide an important extension of the international study of personality characteristics by presenting for the first time normative reference values for a single test using adult participants tested from multiple countries around the world.

By emphasizing the consistency of findings across countries it may sound as if we think social and cultural factors are unimportant in shaping personality, attitudes, perceptions, and experiences. We do not hold this position. For instance, the Japanese sample has relatively higher T-scores (i.e., from 55–60) on W, S, DQ+, Zf, M, EA, D-score, HVI total, OBS total, human contents, Cg, Fi, Fd, p, X–%, and PTI. Although potential artifacts have not been ruled out, these results may reflect relatively unique qualities tied to Japanese culture. In addition, data show how education and socioeconomic status are correlated with many CS scores (e.g., Nascimento, 2004; Pires, 2007/this issue), and these variables are often tied to cultural or national differences. Thus, while we recognize the importance of understanding individuals in the context of their sociocultural background (see Allen & Dana, 2004; Ritzler, 2004), we also recognize how adults from the countries and cultures included in 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 for clinical applications and inferences about the contemporary functioning of children and adolescents.

Recently, Meyer and Viglione (in press; Viglione & Meyer, 2008) recommended that clinicians make a number of changes in their normative expectations for the CS based on the accumulating data from the internationally collected samples that are part of this Supplement. In general, their recommendations are consistent with the normative reference values provided in Tables 1 and 2. However, the authors concluded that the normative adjustments made for adults also probably would apply to children, assuming one took into account developmental progressions like those documented by Wenar and Curtis (1991). Although Meyer and Viglione noted that child reference data were unstable and cautioned clinicians about making inferences regarding psychopathology in children from CS data, given the findings in Figures 3–5, we take this caution further.

Figures 4 and 5 illustrate how examiners following current CS administration and scoring guidelines produce considerable variability within and across countries in interpretively important CS scores. Applying standard interpretive guidelines to these samples of normally functioning children and adolescents would incorrectly result in some very unhealthy inferences and attributions of psychopathology. Obviously, we need to avoid such unfounded inferences.

At present we do not understand the cultural, societal, examiner, and/or administration and scoring factors that are responsible for the erratic results seen with children and adolescents. It is unclear to what extent the differences reflect genuine cultural differences in personality and/or in culturally based child-rearing practices, or artifacts due to variability in the way the protocols were administered, inquired, or scored (Lis, Parolin, Calvo et al., 2007/this issue; Meyer & Viglione, in press; Viglione & Meyer, 2008).

Although it may seem that clinicians could rely on country-specific or "local norms" when assessing children, the findings in Figures 3–5 leave us concerned that normative information collected by one group in a particular locale may not generalize to the types of data obtained by all clinicians working in that locale. For instance, the *M*s differ markedly on many scores when comparing Exner's (2003) US samples to other US samples (e.g., Lambda ≈ 0.7 vs. 1.5; Dd ≈ 1.5 vs. 7.0). Even among the more recently collected US protocols, there is variability on these scores for children of the same age. The inconsistency also is not limited to the US. For instance, with Italian children aged 11 or 12, it is unclear whether one should anticipate a protocol with Lambda = 1.8, es = 8.0, SumV = 0.2, and X–% = .31, or a notably different standard of Lambda = 0.9, es = 11.4, SumV = 1.4, and X–% = .19.

An ambitious clinician could attempt to develop personal norms by obtaining CS protocols from healthy and normally functioning children at various ages. Doing so would control for any examiner effects associated with one's personal style of test administration and scoring and it would allow that clinician to understand what type of CS data he or she typically obtains from normal children, which is particularly important for the form quality variables, Lambda, and other scores related to the richness of verbalization in a protocol.

However, even if a clinician invests the time and energy necessary to have a personal normative base to draw upon, he or she should draw inferences about pathology quite cautiously—particularly with respect to Lambda and its associated scores (e.g., EA, es, M, WSumC) and form quality. The bolded scores in Table 5 should have some utility, though those in regular font probably should not be relied on for interpretation. Until we have a better understanding of the factors that influence CS data obtained from children and have stable within- or across-culture normative benchmarks for different levels of cognitive and emotional development, one could use much of the Rorschach data from children in an idiographic and exploratory manner, though not as a full suite of nomothetic CS scores tightly linked to inferences about psychopathology.

There are two exceptions to this guideline. One would be instances when a child or adolescent produces a healthy-looking record, akin to Exner's reference data, and comprised of a relatively high number of responses, low Lambda, high DQ+, healthy form quality, and low WSum6. For children like this, one could reasonably infer strengths and assets in functioning. This rationale suggests that clinicians working in the context of a diagnostic assessment with children and adolescents could expect that CS scores may provide better specificity for indicating health and effective functioning than they do sensitivity to pathology and disturbance.

The second exception would be instances when a child or adolescent obtains a score that is less healthy than the mean observed for any sample. Under these circumstances, one could begin to infer difficulties to the extent that the score deviated from the most extreme sample mean. Even the non-bolded scores in Table 5 could be considered from this perspective (e.g., one could begin to consider difficulties in functioning with T-scores >60 for WSum6, >70 for Dd, >75 for PureF%, or <35 for DQ+).

The data in this Supplement are less comprehensive and complete for children and adolescents than for adults and more systematic coverage is needed to adequately address questions about the relative influence of cultural, developmental, examiner, administration, or scoring factors. As such, it is clear the field needs additional carefully designed studies that examine developmental processes expressed on the Rorschach across cultures. We hope the data presented here and elsewhere in the Supplement facilitate those ends by inspiring others to collect and report reference data for children and adolescents and to design studies that help us understand the factors that may contribute to similarities and differences across samples.

Acknowledgments

We are grateful to all the authors who contributed data to this Supplement and to the analyses reported in this article. In 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. *Journal of Personality Assessment, 82*, 189–206.

Andronikof-Sanglade, A. (1999, July). Data for French children. In P. Erdberg, & T. W. Shaffer, (Chairs) *International symposium on Rorschach nonpatient data: Findings from around the world.* 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.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 329–344). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Beizmann, C. (1970). *Handbook for scoring Rorschach responses.* New York: Grune & Stratton.

Berant, E. (2007/this issue). Rorschach Comprehensive System data for a sample of 150 adult nonpatients from Israel. *Journal of Personality Assessment, 89*(Suppl. 1), S67–S73.

Campo, V., & Vilar, N. (2007/this issue). Rorschach Comprehensive System data for a sample of 517 adults from Spain (Barcelona). *Journal of Personality Assessment, 89*(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. *Journal of Personality Assessment, 89*(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. *Journal of Personality Assessment, 89*(Suppl. 1), S113–S118.

Dumitrascu, N. (2007/this issue). Rorschach Comprehensive System data for a sample of 111 adult nonpatients from Romania. *Journal of Personality Assessment, 89*(Suppl. 1), S142–S148.

Erdberg, P., & Shaffer, T. W. (1999, July). *International symposium on Rorschach nonpatient data: Findings from around the world.* Paper presented at the International Congress of Rorschach and Projective Methods, Amsterdam, The Netherlands.

Exner, J. E., Jr. (2003). *The Rorschach: A comprehensive system* (4th ed.). New York: Wiley.

Exner, J. E., Jr. (2007/this issue). A new US adult nonpatient sample. *Journal of Personality Assessment, 89*(Suppl. 1), S154–S158.

Exner, J. E., Jr., & Erdberg, P. (2005). *The Rorschach: A comprehensive system: Vol. 2. Interpretation* (3rd ed.). Oxford, England: Wiley.

Exner, J. E., Jr., & Weiner, I. B. (1982). *The Rorschach: A comprehensive system: vol. 3. Assessment of children and adolescents.* New York: John Wiley & Sons.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

Greenway, P., & Milne, L. (2007/this issue). Rorschach Comprehensive System data for a sample of 128 adult nonpatients from Australia. *Journal of Personality Assessment, 89*(Suppl. 1), S20–S25.

Hamel, M., & Shaffer, T. W. (2007/this issue). Rorschach Comprehensive System data for 100 nonpatient children from the United States in two age groups. *Journal of Personality Assessment, 89*(Suppl. 1), S174–S182.

Hansen, K. G. (2007/this issue). Rorschach Comprehensive System data for a sample of 75 Danish 9-year-old children: A pilot study. *Journal of Personality Assessment, 89*(Suppl. 1), S52–S56.

Ivanouw, J. (2007/this issue). Rorschach Comprehensive System data for a sample of 141 adult nonpatients from Denmark. *Journal of Personality Assessment, 89*(Suppl. 1), S42–S51.

Janson, H., & Olsson, U. (2004). A measure of agreement for interval or nominal multivariate observations by different sets of judges. *Educational and Psychological Measurement, 64*, 62–70.

Lis, A., Parolin, L., Calvo, V., Zenarro, A., & Meyer, G. J. (2007/this issue). The impact of administration and inquiry on Rorschach Comprehensive System protocols in a national reference sample. *Journal of Personality Assessment, 89*(Suppl. 1), S193–S200.

Lis, A., Parolin, L., Salcuni, S., & Zennaro, A. (2007/this issue). Rorschach Comprehensive System data for a sample of 249 adult nonpatients from Italy. *Journal of Personality Assessment, 89*(Suppl. 1), S80–S84.

Lis, A., Salcuni, S., & Parolin, L. (2007/this issue). Rorschach Comprehensive System data for a sample of 116 preadolescent and 117 adolescent nonpatients from Italy. *Journal of Personality Assessment, 89*(Suppl. 1), S91–S96.

Lunazzi, H., Urrutia, M. I., de La Fuente, M. G., Elias, D., Fernandez, F., & de La Fuente, S. (2007/this issue). Rorschach Comprehensive System data for a sample of 506 adult nonpatients from Argentina. *Journal of Personality Assessment, 89*(Suppl. 1), S7–S12.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Matsumoto, M. (2005, July). *Characteristics of Japanese children and related research as shown in Rorschach.* Paper presented at the XVIIIth Congress of the International Rorschach Society, Barcelona, Spain, July 28.

Matsumoto, M., Suzuki, N., Shirai, H., & Nakabayashi, M. (2007/this issue). Rorschach Comprehensive System data for a sample of 190 Japanese nonpatient children at five ages. *Journal of Personality Assessment, 89*(Suppl. 1), S103–S112.

Mattlar, C.-E. (2004). Are we entitled to use Rorschach Workshop's norms when interpreting the Comprehensive System in Finland? *Rorschachiana, 26*, 85–110.

Mattlar, C.-E., Forsander, C., Carlsson, A., Norrlund, L., Vesala, P., Leppänen, T., et al. (2007/this issue). Rorschach Comprehensive System data for a sample of 343 adults from Finland. *Journal of Personality Assessment, 89*(Suppl. 1), S57–S60.

McCrae, R. R., & Terracciano, A. (2006). National character and personality. *Current Directions in Psychological Science, 15*, 156–161.

McCrae, R. R., & Terracciano, A., & 78 members of the Personality Profiles of Cultures Project. (2005a). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547–561.

McCrae, R. R., & Terracciano, A., & 79 members of the Personality Profiles of Cultures Project. (2005b). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407–425.

Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science & Practice, 8*, 389–396.

Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment, 13*, 486–502.

Meyer, G. J., & Richardson, C. (2001, March). *An examination of changes in form quality codes in the Rorschach Comprehensive System from 1974 to 1995.* Paper presented at the annual meeting of the Society for Personality Assessment, Philadelphia, PA.

Meyer, G. J., & Viglione, D. J. (in press). An Introduction to Rorschach Assessment. In R. P. Archer & S. R. Smith (Eds.), *A guide to personality assessment: Evaluation, application, and integration.* Mawah, NJ: Lawrence Erlbaum Associates.

Meyer, G. J., Viglione, D. J., Erdberg, P., Exner, J. E., Jr., & Shaffer, T. (2004, March). *CS scoring differences in the Rorschach Workshop and Fresno nonpatient samples.* Paper presented at the annual meeting of the Society for Personality Assessment, Miami, FL, March 11.

Meyer G. J., Viglione, D. J., & Exner, J. E., Jr. (2001). Superiority of Form% over Lambda for research on the Rorschach. *Journal of Personality Assessment, 76*, 68–75.

Mormont, C., Thommessen, M., & Kever, C. (2007/this issue). Rorschach Comprehensive System data for a sample of 100 adult nonpatients from the Belgian

French speaking community. *Journal of Personality Assessment, 89*(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. *Journal of Personality Assessment, 89*(Suppl. 1), S97–S102.

Nascimento, R. S. G. F. (2004). The impact of education and/or socioeconomic conditions on Rorschach data in a Brazilian nonpatient sample. *Rorschachiana, 26*, 45–62.

Nascimento, R. S. G. F. (2007/this issue). Rorschach Comprehensive System data for a sample of 409 adult nonpatients from Brazil. *Journal of Personality Assessment, 89*(Suppl. 1), S35–S41.

Pertchik, K., Shaffer, T. W., Erdberg, P., & Margolin, D. I. (2007/this issue). Rorschach Comprehensive System data for a sample of 52 older adult nonpatients from the United States. *Journal of Personality Assessment, 89*(Suppl. 1), S166–S173.

Pires, A. A. (2007/this issue). Rorschach Comprehensive System data for a sample of 309 adult nonpatients from Portugal. *Journal of Personality Assessment, 89*(Suppl. 1), S124–S130.

Ráez, M. (2007/this issue). Rorschach Comprehensive System data for a sample of 233 adult nonpatients from Peru. *Journal of Personality Assessment, 89*(Suppl. 1), S119–S123.

Ritzler, B. (2004). Cultural applications of the Rorschach, apperception tests, and figure drawings. In M. Hilsenroth & D. Segal (eds.), Personality assessment. Volume 2 in M. Hersen (Ed.-in-Chief), *Comprehensive handbook of psychological assessment* (pp. 315–342). Hoboken, NJ: John Wiley & Sons.

Salcuni, S., Lis, A., Parolin, L., & Mazzeschi, C. (2007/this issue). Rorschach Comprehensive System data for two samples of nonpatient children from Italy: 75 aged 5–7 years and 148 aged 8–11 years. *Journal of Personality Assessment, 89*(Suppl. 1), S85–S90.

Sanz, I. A. E. (2007/this issue). Rorschach Comprehensive System data for a sample of 90 nonpatient adults from Argentina. *Journal of Personality Assessment, 89*(Suppl. 1), S13–S19.

Sciara, A. D., & Ritzler, B. A. (2006). *The little book on administration for the Rorschach Comprehensive System*. Asheville, NC: Rorschach Training Programs.

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

Shaffer, T. W., Erdberg, P., & Haroian, J. (2007/this issue). Rorschach Comprehensive System data for a sample of 283 adult nonpatients from the United States. *Journal of Personality Assessment, 89*(Suppl. 1), S159–S165.

Shaffer, T. W., Erdberg, P., & Meyer, G. J. (Eds.). (2007). International reference samples for the Rorschach Comprehensive System [Special issue]. *Journal of Personality Assessment, 89*(Suppl. 1).

Silva, D. R., & Dias, A. M. (2007/this issue). Rorschach Comprehensive System data for a sample of 357 Portuguese children at five ages. *Journal of Personality Assessment, 89*(Suppl. 1), S131–S141.

Sultan, S., Andronikof, A., Fouques, D., Lemmel, G., Mormont, C., Réveillère, C., et al. (2004). Vers des normes francophones pour le Rorschach en système intégré: Premiers résultats sur un échantillon de 146 adultes [Towards French language norms for the Rorschach Comprehensive System: Preliminary results in a sample of 146 adults.] *Psychologie Française, 49*, 7–24.

Sultan, S., Andronikof, A., Réveillère, C., & Lemmel, G. (2006). A Rorschach stability study in a nonpatient adult sample. *Journal of Personality Assessment, 87*, 330–348.

Terracciano, A., Abdel-Khalek, A. M., Ádám, N., Adamovová, L., Ahn, C.-k., Ahn, H.-n., et al. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science, 310*, 96–100.

Tibon, S. (2007/this issue). Rorschach Comprehensive System data for a sample of 41 adult nonpatients from Israel. *Journal of Personality Assessment, 89*(Suppl. 1), S74–S79.

Tibon, S. (2007). *Consistency in Rorschach Comprehensive System data for Israeli nonpatient adults across two samples (1996; 2006): Clinical and nonclinical implications*. Manuscript in preparation.

Valentino, M. A., Shaffer, T. W., Erdberg, P., & Figueroa, M. (2007/this issue). Rorschach Comprehensive System data for a sample of 42 nonpatient Mexican-American children from the United States. *Journal of Personality Assessment, 89*(Suppl. 1), S183–S187.

Van Patten, K., Shaffer, T. W., Erdberg, P., & Canfield, M. (2007/this issue). Rorschach Comprehensive System data for a sample of 37 nonpatient/nondelinquent adolescents from the United States. *Journal of Personality Assessment, 89*(Suppl. 1), S188–S192.

Viglione, D. J. (2002). *Rorschach coding solutions: A reference guide for the Comprehensive System*. San Diego: Author.

Viglione, D. J., & Meyer, G. J. (2008). An Overview of Rorschach psychometrics for forensic practice. In C. B. Gacono & F. B. Evans with N. Kaser-Boyd & L. A. Gacono (Eds.), *Handbook of forensic Rorschach psychology* (pp. 21–53). Mawah, NJ: Lawrence Erlbaum Associates.

Vinet, E.V. (2000). The Rorschach Comprehensive System in Iberoamerica. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 345–365). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Weiner, I. B. (2001). Considerations in collecting Rorschach reference data. *Journal of Personality Assessment, 77*, 122–127.

Wenar & Curtis (1991). The validity of the Rorschach for assessing cognitive and affective changes. *Journal of Personality Assessment, 57*, 291–308.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001a). The misperception of psychopathology: Problems with norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science & Practice, 8*, 350–373.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001b). Problems with the norms of the Comprehensive System for the Rorschach: Methodological and conceptual considerations. *Clinical Psychology: Science & Practice, 8*, 397–402.