The Effects of Using the International Versus Comprehensive System Rorschach Norms For Children, Adolescents, and Adults

DOI:10.1080/00223891.2015.1136313

Viglione, Donald J.; Giromini, Luciano. The Effects of Using the International Versus Comprehensive System Rorschach Norms For Children, Adolescents, and Adults. JOURNAL OF PERSONALITY ASSESSMENT. 98 (4) pp: 391-397.
DOI: 10.1080/00223891.2015.1136313

The Effects of Using the International versus Comprehensive System – Rorschach Norms for Children, Adolescents, and Adults

Journal:	*Journal of Personality Assessment*
Manuscript ID	JPA-2015-150.R2
Manuscript Type:	General Submission
Keywords:	Rorschach < Measures, Norms, International, Comprehensive System

Using CS vs. International Norms

Abstract

Currently, there is some debate about whether to use Comprehensive System norms (CS, Exner, 2003) or the Composite International Reference Values (CIRV; Meyer, Erdberg, & Shaffer, 2007) when interpreting Rorschach Inkblot Method (RIM; Rorschach, 1921) protocols administered with the CS method. The goal of this study is to assist clinicians in making this decision by providing information about the effects of choosing one option or the other. Accordingly, the current research evaluates the effects of using the CS vs. CIRV norms with children, adolescents, and adults. First, we identified 43 variables for which the CS and the CIRV for children and adolescents differ from each other by at least a Cohen's *d* value of .5. Next, we evaluated whether these divergent variables are the same as those previously identified as divergent for the adult population. Results showed that for both children and adolescents, as well as for adults, relying on CS norms versus CIRV would result in interpretations that are more pathological in terms of: (a) perception and thinking, (b) psychological resources and cognitive and emotional abilities, and (c) representations of human relationships. A discussion on the clinical effects of utilizing one versus the other set of norms follows.

Keywords: Rorschach; Normative Data; Composite International Reference Values; Comprehensive System.

Using CS vs. International Norms

The Effects of Using the Composite International Reference Values versus Comprehensive System – Rorschach Norms for Children, Adolescents, and Adults

In recent years, some concern has been expressed about what norms should be referenced when interpreting Rorschach Inkblot Method (RIM; Rorschach, 1921) protocols administered using the Comprehensive System method (CS; Exner, 2003). The CS norms (Exner, 2003) have been criticized for being dated for they were mostly collected in the 1970s, and for making nonclinical adults appear maladjusted (Garb, Wood, Lillienfeld, & Nezworski, 2005; Shaffer, Erdberg, & Haroian 1999; Viglione & Hilsenroth, 2001; Viglione & Meyer, 2008; Wood et al., 2001a, 2001b). For example, in 2001 Wood and colleagues (Wood, Nezworski, Garb, & Lilienfeld, 2001a, 2001b) and Viglione and Hilsenroth (2001) reported that compared to CS norms, many American nonpatient samples described in the literature produced poorer Form Quality (FQ) values, lower values of color-related variables (e.g., WSumC and FC), and lower complexity (e.g., higher Lambda). A few years later, Meyer, Erdberg and Shaffer (2007) compared a sample of 450 American nonpatient adults (Exner, 2007) to large number of other nonpatient samples from other countries totaling approximately 4,500 individuals. This Exner sample differed from all other samples, especially in terms of FQ-related, color-related, and human representation-related variables. As pointed out by the authors, the CS norms (Exner, 2003, N = 600) are even more divergent than the Exner (2007, N=450) sample. For all variables, the direction of these differences indicate that using CS norms would make the typical or average nonpatient adult appear notably maladjusted (Viglione & Meyer, 2008). Accordingly, some authors have suggested that Meyer et al.'s (2007) composite, internationally based, nonpatient sample – often referred to as the Composite International Reference Values (CIRV) – should be

Comment [S1]: Need to have years by these authors.

Comment [LG2]: We have made the requested change used by practitioners, instead of standard CS norms, as the primary normative reference set for the CS (e.g., Viglione & Meyer, 2008).

The CIRV, however, have been criticized too. Although there are no published research studies on this matter, on the web page for Rorschach Training Programs, Ritzler and Sciara (2009) argued that: (a) the majority of the records making up the CIRV were collected by graduate students, which tends to generate protocols lacking of complexity (e.g., high Lambda); (b) data collection was mainly conducted in large urban areas, which poses questions regarding the representativeness of the CIRV; (c) inclusion and exclusion criteria varied across the multiple studies generating the CIRV, which potentially affects the variability of the reference values; (d) the individual studies making up the CIRV had relatively small *Ns*, which affects the accuracy of the stratification of the CIRV; (e) it is unclear whether all of the individual studies making up the CIRV followed the CS guidelines strictly, which might affect the overall quality of data collection.

Meyer, Shaffer, Erdberg, and Horn (2015) recently conducted additional analyses on Meyer et al.'s (2007) original dataset, to address some of the issues raised by Ritzler and Sciara (2009). The results of these additional analyses showed that the quality of data collection efforts did not affect the central tendency and dispersion values of the Rorschach scores included in the CIRV. Moreover, these additional analyses also showed that different sets of within-country local norms notably differ from each other, so that using a specific set of local norms for a given country, instead of CIRV, would be questionable anyways. In line with all these research findings, Meyer, Viglione, Mihura, Erard and Erdberg (2011) had previously argued that using international, rather than local norms would improve the applicability of the test. As such, the

Using CS vs. International Norms

Rorschach Performance Assessment System's (R-PAS; Meyer et al., 2011) reference values were derived from the CIRV.

To investigate which normative reference set might be more useful in clinical practice, Giromini, Viglione, and McCullaugh (2015) recently introduced a novel statistical approach. This method is designed to test whether a particular nonclinical sample more closely resembles one set of norms or another. Although the authors applied their statistical technique to a small, newly collected, nonclinical adult sample, the focus of their demonstration study was methodological. Thus, the authors offered no conclusion on the superiority of either set. However, in this same article the authors identified a large group of 28 "divergent variables" for which the standard CS norms for adults (Exner, 2003) and the CIRV differ from each other by at least a Cohen's *d* value of .50 (Cohen, 1988), that is one half of one standard deviation. For these 28 divergent variables, using one versus the other set of norms as a reference point for the clinical interpretation of an adult CS record will likely lead to notably different interpretive inferences. As a methodological study, any effects of using either reference sets was not addressed, so that the debate about using the RIM, CS adult data continues unabated.

There is perhaps even more uncertainty about what norms a clinician should use when interpreting a child or adolescent Rorschach record. Indeed, the available CS norms for children and adolescents may be outdated for they were first presented in 1976 (Exner, Weiner, & Schuyler, 1976), and to some extent atypical, too, compared to many other reference samples from various countries and cultures (Meyer et al., 2007). In fact, Exner's normative samples for children and adolescents produced by far more adequate FQ values than all other 29 samples from the five countries evaluated by Meyer et al. (2007). Different from other normative child samples, CS norms do not show more distorted FQ with younger ages (Ames et al., 1974; Ames et al., 1971; Meyer et al., 2007). In fact, Stanfill, Viglione, and Resende (2013) pointed out that the CS child and adolescent norms is the only normative sample in the literature or in their study in which children and adults produce similar FQ scores. In all other samples, as one would expect given the natural process of maturation and development, FQ, as a measure of perceptual accuracy, becomes more adequate with age. The reasons for these discrepancies are difficult to identify. Perhaps, they have to do with the fact that FQ coding guidelines have evolved over time (Meyer & Archer, 2001). In addition, it is possible that differences from one study to another in the inclusion or exclusion criteria, as well as in the data collection procedures also contributed to generating these discrepancies. Regardless, ~~Anyhow,~~ various empirical data concur to indicate that standard CS norms for children and adolescents may be non-optimal.

The international norms for children and adolescents introduced by Meyer et al. (2007), on the other hand, also are not without problems. First, the same concerns raised by Ritzler and Sciara (2009) for the adult data also apply to the child and adolescent data, i.e., the overall integrity and quality of the data collected is challenged by the fact that many records were collected by students, rather than experienced clinicians. In addition, unlike the adult samples, child and adolescent samples from different countries produce notable differences when compared to each other (Meyer et al., 2007). Thus, more research is needed to understand whether the internationally based, normative data introduced by Meyer et al. (2007) may be adequately representative across different cultural contexts.

Given the uncertainty about the superiority and the appropriateness of these two sets of norms, which norms should a clinician use? Currently, many clinicians rely on the CS norms (Exner, 2003). These norms are the reference set of data used by the CS computerized interpretation software, i.e., the Rorschach Interpretation Assistance Program (RIAP-5; Exner &

Weiner, 2005). Other authors (Viglione & Meyer, 2008), however, have asserted that using these norms might make the interpretations look more pathological than they actually are, and argue that the international norms introduced by Meyer et al. (2007) might serve as a better reference point.

To help the assessment psychologist choose the most appropriate set of norms when administering the Rorschach within the CS, the current research evaluates the effects of using the CS vs. CIRV Rorschach norms with children, adolescents, and adults. First, we focused on children and adolescents, and identified the Rorschach variables for which the CS (Exner, 2003) and the CIRV (Meyer et al., 2007) differ "notably" from each other. Secondly, we evaluated whether these divergent variables for children and adolescents are the same as those identified as "divergent" by Giromini, Viglione, and McCullaugh (2015), for the adult population. A discussion on the clinical effects of utilizing one versus the other set of norms follows.

Method

Rorschach Data

The reference adult sample (Exner, 2003) includes 600 U.S. records, mostly collected from a workplace setting. It is fairly well stratified in terms of gender, geographic distribution, and socioeconomic level, though the percentage of non-White respondents is somewhat lower than today's optimal standards. The mean age is ~~about 30~~31.7 (*SD* = 10.7), and the average level of education is ~~about 13~~13.4 (*SD* = 1.6) years. The reference child and adolescent sample described by Exner (2003) includes U.S. records form 1390 children and adolescents ranging in age from 5 to 16.

Meyer et al. (2007) summarized data from several countries, for a total sample size of 4,704 adult (21 samples from 17 countries) and 2,647 child and adolescent (31 samples from 5

Comment [S3]: I would prefer greater precision here….along with education.

Comment [LG4]: We have changed this passage accordingly

Using CS vs. International Norms

countries) protocols. In almost all of the cases, these data were derived from the 2007 supplement of the *Journal of Personality Assessment*, in which authors from all over the world contributed normative samples collected in their home countries, using the CS. For the combined, adult normative set, the mean age was 36; for the combined, child/adolescent normative set, the mean age was 10. Additional information concerning these samples can be found in Meyer et al. (2007).

Analyses

All Rorschach variables listed in the CS norms (Exner, 2003) and CIRV (Meyer et al., 2007) were included in this study, so that a total of 112 variables were examined. Giromini, Viglione, and McCullaugh (2015) previously identified the 28 variables for which the standard CS norms for adults (Exner, 2003) differ from the CIRV for adults (Meyer et al., 2007) by at least a Cohen's *d* value of .50 (Cohen, 1988). For the current study, we applied the same analysis to identify the variables for which the two sets of norms for children and adolescents (i.e., CS vs. CIRV) differ by at least a *medium* Cohen's *d* value of .50 (Cohen, 1988). That is, we compared the mean scores of CS and CIRV data, and used the effect size~~,~~ (Cohen's ~~*d* measure~~) to determine which variables produced at least a *medium-sized* difference (i.e., "divergent" variables, for which $d \geq .50$). The choice to focus on effect sizes greater than $d = .50$ is consistent with both Cohen's (1988) statistical recommendations, as well as with MMPI literature indicating that differences of 5 T points (i.e., $d = .50$) are to be considered as notable differences (e.g., Greene, 2000). Next, we evaluated whether these divergent variables for children and adolescents are the same as those identified as "divergent" by Giromini, Viglione, and McCullaugh (2015) for the adult population. To do so, we first used the *Phi* statistic to evaluate the strength of the association between type of variable (divergent vs. non-divergent) and age

Comment [S5]: I would like there to be a rationale somewhere in the paper for why this effect size? Is there any theory guiding this?

Comment [LG6]: We have added a brief explanation 3 lines below category (child/adolescent vs. adult). To supplement this comparison, we also calculated the Pearson correlation between the Cohen's *d* values observed within the child/adolescent versus adult populations.

Results

"Divergent Variables" for Children and Adolescents

The international norms for children and adolescents (*N* = 2647) combine available data from juveniles of all ages (5 – 18), for a grand mean age of ~~about~~ 10.0. Conversely, the CS norms present separate data for children of 5 (*n* = 90), 6 (*n* = 80), 7 (*n* = 120), 8 (*n* = 120), 9 (*n* = 140), 10 (*n* = 120), 11 (*n* = 135), 12 (*n* = 120), 13 (*n* = 110), 14 (*n* = 105), 15 (*n* = 110), and 16 (*n* = 140) years of age. ~~However, t~~The grand mean age of these twelve CS subsamples is ~~about 10, too (*M* =~~ 10.7~~; *SD* = 3.3)~~. Thus, wanting to compare the CIRV vs. CS norms, we first combined all year-by-year CS data in order to produce descriptive statistics for the entire, pooled CS reference sample (*N* = 1390). This first step produced the grand mean and grand standard deviation values of all 112 variables under investigation. Next, we used the Cohen's *d* statistic (Cohen, 1988) to compare these CS normative values against the CIRV.

Comment [S7]: See previous comment about level of accuracy

Comment [LG8]: We have changed it accordingly

Comment [S9]: Ibid.

Comment [LG10]: We have changed it accordingly

Across the 112 comparisons, the mean *d* (absolute values) was .57 (*SD* = .64), with values ranging from .00 (for DR2 and DV) to 3.30 (for X+%). Critically, 43 (or 38.4%) variables produced a Cohen's *d* greater than .50, thus indicating a "notable" or "medium" effect size (Cohen, 1988). These "divergent variables" are reported in Table 1.

As one may easily note, the most extreme differences between the two sets of norms are found in protocol-level FQ variables, namely X+% (*d* = 3.30), XA% (*d* = 2.58), WDA% (*d* = 2.33), X-% (*d* = 2.65), and FQxo (*d* = 2.20). In all these cases, compared to the CS norms, the CIRVproduced more pathological data, with fewer ordinary and more distorted perceptions.

Using CS vs. International Norms

Noteworthy, the CIRV also produced fewer emotional or cognitive resources, as demonstrated by the relatively lower number of color (see, for example, WSumC) and human movement (M) responses, as well as by the Experience Actual (EA) itself. Similar conclusions may also be made when looking at variables such as popular responses (Pop), human movement with ordinary FQ (MQo), or cooperative movement responses (COP).

Findings with Children and Adolescents versus Findings with Adults. To test whether the divergent variables that emerged from the analysis of child/adolescent normative samples were the same as those identified by Giromini, Viglione, and McCullaugh (2015) for the adult population, we ran three types of analyses. First, we produced a 2 by 2 contingency table and examined the number of divergent versus non-divergent variables in the child/adolescent versus adult samples. A *Phi* statistic was used to test the strength of the association between type of variable (divergent vs. non-divergent) and age category (child/adolescent vs. adult). As shown in Table 2, among the 112 variables under investigation, 58.0% were "non-divergent," and 21.4% were divergent for both the child/adolescent and the adult normative sets. Thus, for about 80% (i.e., 58.0% plus 21.4%) of the variables, findings with children and adolescents are similar to those with adults, reported by Giromini, Viglione, and McCullaugh (2015). From a different perspective, 24 out of 43 (or 55.8%) divergent variables for children and adolescents are also divergent for adults. Importantly, 65 out of 69 (or 94.2%) non-divergent variables for children and adolescents are also non-divergent for adults. Thus, the divergent and non-divergent variables in the child and adolescent population tend to be divergent and non-divergent also in the adult population, *Phi* = .56, $p < .001$.

To further evaluate how similar the distribution of the divergent variables was in the child/adolescent and adult samples, we also tested the Pearson correlation between the Cohen's *d* values observed within the child/adolescent versus adult populations. A high correlation would suggest that the same pattern of CS versus international differences characterize both the child/adolescent and adult samples. The results of this second analytic approach showed that the correlation between the Cohen's *d* values of the child/adolescent and the adult sets of norms was .75, $p < .001$ when the absolute *d* values were considered, and reached .84, $p < .001$ when the signed *d* values were used.

Lastly, a more detailed description of the divergent variables in the child/adolescent and/or in the adult norms is presented in Tables 3, 4, and 5. Again, what emerges from these analyses, is that the biggest differences between the two sets of norms are found in FQ related variables, with the CS norms (Exner, 2003) having more "healthy" values than the CIRV (Meyer et al., 2007). The results concerning variables such as, for example, EA, Afr, WSumC or COP also concur to indicate that the CIRV (Meyer et al., 2007) produced more pathological values than the CS (Exner, 2003) norms. Also noteworthy, these differences appear greater for the child and adolescent reference sets than they are for the adult norms. Indeed, the mean absolute Cohen's *d* values produced by the analyses of the child and adolescent sets of norms was .57 (*SD* = .64). For the adults, the corresponding mean value was .38 (*SD* = .33), a significantly lower value ($t(109) = 4.59$, $p < .0005$). Accordingly, even though the findings with children and adolescents are similar to those with adults, the differences in the child and adolescent sets are more extreme, compared to those with the adults norms.

Comment [S11]: Perhaps I am missing something, but it's not clear why this point is needed, as it repeats the sentence above with a few examples, which can be gleaned in the table.

Comment [LG12]: Thanks, we agree that this sentence is not needed

Discussion

Currently, there is a debate about whether to use the CS or CIRV norms when interpreting a Rorschach protocol administered with CS guidelines. It has been suggested that using the CS norms (e.g., Meyer et al., 2011; Viglione & Hilsenroth, 2001; Viglione & Meyer, 2008; Wood et al., 2001a, 2001b) makes respondents look more pathological than they actually are. To provide information about what norms to use, the current article investigated the effects of using one or another reference set on the nomothetic interpretation of the RIM.

First, we identified 43 variables for which the CS norms (Exner, 2003) and the CIRV (Meyer et al., 2007) for children and adolescents differ from each other by at least a Cohen's *d* value of .5. For these 43 variables, using one versus another set would have serious effects on the interpretation. More specifically, if one choses to use the CS norms as reference for interpretations, a given record will tend to look more pathological, in terms of form quality, color, and human representation variables, than it would be if one used the CIRV. This finding is consistent with previous literature suggesting that the CS norms might make nonclinical individuals look maladjusted and therefore warrant caution for clinical and forensic practice. It is also consistent with a previous, methodological work (Giromini, Viglione, & McCullaugh, 2015) for the nonpatient sample in that study appeared pathological when compared to the CS norms, but not to the CIRV. For these classes of variables (i.e., form quality, color, and human representation), thus, it is important to know that using the CS norms might accentuate weaknesses of the client, whereas applying the CIRV might lead one to overlook or de-emphasize his/her potential problems. Professionals should therefore 'weigh' or moderate their interpretations about perception or thinking problems, cognitive or emotional resources, and representation of human relationships, based on which set of norms they decided to use as a reference set.

When we tested whether these divergent variables found in the child and adolescent norms are the same as those identified as divergent by Giromini, Viglione, and McCullaugh (2015) for the adult populations, we found a striking similarity. Applying one versus the other normative set of values to children, adolescents, or adults would therefore produce very similar effects: In all cases, using the CS norms would suggest different inferences than the CIRV in respect to form quality, color, and human representation classes of variables. More specifically, relying on CS norms versus CIRV would result in interpretations that are more pathological in terms of (a) perception and thinking, (b) cognitive and emotional resources, and (c) representations of human relationships. From another perspective the community and non-patient samples collected in the international project (Meyer et al, 2007) would be judged as pathological if the CS norms were applied.

Looking at individual variables more closely with the child and adolescents reveals that the practical impact of the CS vs. CIRV differences are not small. For example, in Table 1 there are ~~19~~ 22, or more than half of the variables, in the child and adolescent samples with a Cohen's *d* exceeding one. We applied the CS norms and CS standard interpretive cutoffs (Exner, 2003) ~~to~~ for the means and likely associated medians of these ~~19~~ 22 variables. ~~Eight~~ Eleven of the values for these variables (X+%, X-%, XA%, WDA%, FQXo, FQx-, COP, Sum T, GoodHR, EA, F), although redundant to some degree, would be interpreted as pathological or highly problematic. ~~Four~~ Five (Popular, AG, Blends, a, MQo) would be interpreted as highly problematic and undesirable. Thus, the child and adolescent, community and non-patient samples collected in the international project (Meyer et al, 2007) would be judged globally as pathological if the CS norms were applied. More specifically, as a whole, the CS norms largely produce pathological and problematic attributions for the average child or adolescent in the international samples regarding (a) perception and thinking (X+%, XA%, WDA%, FQXo, FQx-, Popular) and (b) representations of human relationships (COP, Sum T, GoodHR, AG, MQo), and less so for (c) psychological and cognitive resources (EA, Blends, a, F, MQo). Applying this same procedure to

Comment [LG13]: In our last thorough review of the paper, we discovered a minor error in the counts for variables with d>1 (taken from Table 1). Namely, there were 3 additional variables that in our previous version we did not count. This does not change any of the findings but the relevant numbers have been changed here the adult comparisons would lead to similar observations, although the differences between CS and CIRV norms are not as great. Of course, non-patient status does not rule out pathological traits. However, the typical or average child or adolescent in the international sample would be considered to have key pathological or problematic traits if one applies CS norms, an observation that cannot be explained by within sample variability. Conversely, the average or typical child or adolescent in the CS normative sample would be considered to be function at an optimal level, with many adult like abilities if CIRV norms were applied.

When considering other variables, the impact is likely not pathological. Again selecting the child and adolescent group as an example, one observes that the color variables (CF, FC+CF+C+Cn, WSum C) are well-represented among the ~~19~~ 22 most divergent variables. According to CS guidelines, the differences would be associated with considerable less emotional responsiveness and less engagement in the international sample. However, these color variables are considered to represent a stylistic factor rather than a problematic one.

Also, the lower number of Populars in the international sample could be related to cultural factors rather than pathology. The CS Popular responses, defined as response occurring in more than 30% of American records, could be expressions of American culture so that non-American samples may report fewer. Thus, cultural and language factors could also account for CS vs. CIRV differences and should also be investigated in future research.

The reasons for these discrepancies – and for the similarity of such differences for both children/adolescents and for adults – are difficult to understand. To speculate, the selection of the target samples, the methods used to collect the data, the administration and coding procedures, as well as the motivation and engagement of the respondents might have played a role. If the CIRV data were improperly collected (Ritzler & Sciara, 2009), then they may invalid. However, as noted earlier the quality of the data collection within the CIRV sample was not associated with variability within that sample (Meyer, Shaffer, Erdberg, & Horn, 2015). Moreover, assuming that the CIRV methods and practices were faulty across 52 samples and 22 countries might seem presumptuous and would question the vary adaptability of the test across language, culture, and geography.

There is also a possibility that the samples themselves might differ in terms of psychological health, external stress, and pathology, as a result of the selection process. Or there might be examination context differences with associated differences in motivation. To address this issue, it is important to present information regarding the CS reference adult non-patient sample, since it has not been characterized in detail in the literature. The first records were collected before 1975, followed by various additions and revisions through 2003 (Exner, 1993, 2001, 2003). The CS normative reference sample is different from individuals that typically participate in research studies in that 409 of the 600 volunteers volunteered through their place of employment, were encouraged to participate by their supervisor or union leader, and were provided with time away from work for the testing. Of the remaining participants, 153 were solicited through social or special interest groups and 38 were accessed through social service agencies. A contribution to charity was made in the volunteer's name suggesting this group may have had altruistic motives. Furthermore, the non-patient CS reference sample is rather young, with 72% of the volunteers under 35 years of age, and well-educated, with an average grade level of over 13 years; only 85 individuals had less than 12 years of schooling. Therefore, the 2003 non-patient CS reference sample is predominately young, well-educated, employed and evidenced altruistic motivation characteristics which may be more representative of an optimally functioning subgroup of non-patients engaged in the examination process. Also, this examination context may have resulted in more motivation. Conversely, in other situations volunteer respondents might rush through the process.

In terms of administration and scoring, the CS Inquiry has been identified as a source of possible, troublesome, examiner variability (Exner, 2003; Meyer et al. 2011), and remains largely unexamined as a possible source for the difference. Scoring, or coding, is obviously quite critical to the scores on the divergent variables. Preliminary, but yet unpublished, findings (Meyer, Viglione, Erdberg, Exner, & Shaffer, 2004, March) suggest that coding differences largely accounted for the similar differences found between the sample of 450 American nonpatient adults (Exner, 2007) and the Shaffer, Erdberg, Haroian (1999) non-patient American sample.

Obviously, future research should address ~~about~~ the possible impact of these factors on Rorschach data. Regardless of what does cause the differences between the CS and CIRV reference data, it is important that practitioners be aware of the effects of using either set of norms. Particular caution, in our opinion, should be paid to interpretation of those classes of variables for which the two sets of norms produce the most notably different reference values, i.e., variables related to form quality, color, and human representation. Also noteworthy, Table 5 shows that the discrepancies between the two sets of norms become more extreme if one considers the CS reference values and the CIRV for children and adolescents. When testing young individuals, using one versus the other set of norms would largely impact interpretations, especially for those classes of variables described above.

~~As a rule of thumb, one might consider to be 'extreme' only those values that would be classified as 'extreme' by both sets of norms. Alternatively, whenever a value would result to be 'extreme' only for one set of norms, but not for the other, one might want to keep in mind that~~

Comment [S14]: I think it is inappropriate to say "false positives" and "false negatives', since variables are not diagnostic. Furthermore, there is no data for the CIRV that provides a comprehensive discussion of what high or low scores mean. Thus, I recommend to remove this paragraph or reword.

Comment [LG15]: We decided to remove this paragraph, as we fully appreciate your concern about it. Thanks.

Using CS vs. International Norms

~~the CIRV are more likely to produce 'false negatives,' while the CS norms are more likely to produce 'false positives.' Said differently, if one uses the CIRV, non-pathological scores on form quality-, color-, or human representation-related variables do not necessary reflect presence of strengths (nor absence of weaknesses). Vice versa, if one uses the CS norms, pathological scores on those classes of variables do not necessary reflect psychopathology. Importantly, this is particularly true for evaluations with children and adolescents, where the discrepancies between the two sets of norms are more marked.~~

Nonetheless, two limitations of our study deserve mentioning. First, to compare the CS norms versus the CIRV for children and adolescents, we combined data for children and adolescents ranging in age from 5 to 18. Given that some Rorschach scores vary with age (Giromini, Viglione, Brusadelli, et al., 2015; Meyer et al., 2015; Meyer, Viglione & Giromini, 2014; Stanfill, Viglione & Resende, 2013), combining data from different ages may be not an optimal choice. On the other hand, the mean age of both the normative samples was ~~about~~ ~~10~~very similar ~~(~~, i.e., 10.7 for the CS norms and 10.0 for the CIRV~~)~~, and therefore it is very unlikely that combining data from different ages really impacted our findings. Or, at least, it is very unlikely that it had any impact on the *mean values* of the Rorschach scores taken into consideration.

Second, the choice of using $d = 0.50$ as criterion to classify a variable as divergent versus non-divergent is somewhat arbitrary. However, this is the same criterion used by Giromini, Viglione, and McCullaugh (2015) in their methodological study on the Rorschach norms, and it is also consistent with statistical authorities such as Cohen (1988), as well as with the literature on the MMPI, ~~where~~ which suggests that deviations of 5 T scores ~~are~~ can be understood to

Comment [S16]: See earlier comments

Comment [LG17]: We have changed it accordingly

Using CS vs. International Norms

represent usually conceived as notable differences, from an interpretative point of view (Greene, 2000).

Despite these limitations, our study demonstrates that using CIRV versus CS norms has notable effects on the interpretation. Future research should therefore try to more directly help practitioners by showing which set of norms would be more appropriate for use with CS data. For example, one might assemble contrast well-defined samples of impaired versus well-functioning individuals, generate Standard Scores based on both CIRV and CS norms, and then use efficiency diagnostic statistics to determine which set of norms produces more accurately predicts group. classifications. Alternatively, one might collect additional non-patient samples, and use Giromini, Viglione, and McCullaugh (2015) procedures to determine which of the available sets of norms (i.e., CS vs. CIRV) more closely resemble the Rorschach profiles produced by these new nonpatient samples.

Comment [S18]: See my earlier comment about your justification of d=.50

Comment [DJVJ19]: As noted this was addressed above. Here we made the language more suggestive rather than definitive.

Comment [DJVJ20]: Re the next comment.. are referring here only to group contrasts in which the IV is a dichotomous categorical variable. We would think that well-defined high functioning versus impaired individuals would be an understandable example of such a contrast. We changed a few words here and there to convey that.

Comment [S21]: Again, RIM scores in and of themselves are not diagnostic. However, it looking toward behaviorally-referenced criteria would be a good alternative.

References

Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. B. (1974). *Child Rorschach responses: Developmental trends from two to ten years*. New York, NY: Brunner/Mazel.

Ames, L. B., Metraux, R. W., & Walker, R. B. (1971). *Adolescent Rorschach responses: Developmental trends from ten to sixteen years*. New York, NY: Brunner/Mazel

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Exner, J. E. (2003). *The Rorschach: A comprehensive system, Volume 1* (4th ed.). New York: Wiley.

Exner, J. E., Jr. (2007). A new US adult nonpatient sample. *Journal of Personality Assessment, 89(S1),* S154–S158.

Exner, J. E., & Weiner, I. B. (2003). *RIAP 5 – Rorschach Interpretation Assistance Program: Version 5 for Windows*. Odessa, FL: PAR Psychological Assesment.

Exner, J. E., Weiner, I. D., & Schuyler, W. (1976). *A Rorschach Workbook for the Comprehensive System*. Bayville, NY: John E. Exner.

Garb, H. N., Wood, J. M., Lilienfeld, S. O., & Nezworski, M. T. (2005). Roots of the Rorschach controversy. *Clinical Psychology Review, 25*, 97–118. doi:10.1016/j.cpr.2004.09.002

Giromini, L., Viglione, D. J., Brusadelli, E., Lang, M., Reese, J. B., & Zennaro, A. (2015). Cross-cultural validation of the Rorschach Developmental Index. *Journal of Personality Assessment, 97*, 348-353. doi: 10.1080/00223891.2014.960927

Giromini, L., Viglione, D. J., & McCullaugh, J. (2015). Introducing a Bayesian Approach to Determining Degree of Fit with Existing Rorschach Norms. *Journal of Personality*

Assessment, Advance Online Publication, 97, 354-363. doi: 10.1080/00223891.2014.959127

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

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., Erdberg, P., Shaffer, T. W. (2007). Toward international normative reference data for the Comprehensive System. *Journal of Personality Assessment, 89(S1)*, S201-S216.

Meyer, G. J., Giromini, L., Viglione, D. J., Reese, J. B., & Mihura, J. L. (2014). The association of gender, ethnicity, age, and education with Rorschach scores. *Assessment, Advance Online Publication, July 24*. DOI: 10.1177/1073191114544358

Meyer, G. J., Shaffer, T. W., Erdberg, P., & Horn, S. L. (2014). Addressing issues in the development and use of the Composite International Reference Values as Rorschach norms for adults. *Journal of Personality Assessment, Advance Online Publication, Oct 8*. DOI: 10.1080/00223891.2014.961603

Meyer, G. J., Viglione, D. J., & Giromini, L. (2014). An introduction to Rorschach assessment. In R. P. Archer and S. R. Smith (Eds.), *Personality Assessment, 2nd Ed*. (pp. 301-369). New York, NY: Routledge.

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation and technical manual*. Toledo, OH: Rorschach Performance Assessment System.

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.

Ritzler, B., & Sciara, A. (2009). Rorschach Comprehensive System international norms: Cautionary notes. Unpublished manuscript. Retrieved from http://www.rorschachtraining.com/wp-content/uploads/2011/10/Rorschach-Comprehensive-System-International-Norms.pdf

Rorschach, H., 1921. *Psychodiagnostik*. Bern, Bircher.

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

Stanfill, M. L., Viglione, D. J., & Resende, A. C. (2013). Measuring psychological development with the Rorschach. *Journal of Personality Assessment, 95*, 174-186. doi:10.1080/00223891.2012.740538. Stanfill, M. L., Viglione, D. J., & Resende, A. C. (2013). Correction to: Measuring psychological development with the Rorschach. *Journal of Personality Assessment, 95*, 435. DOI:10.1080/00223891.2013.779563

Viglione, D. J., & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment, 13*(4), 452–471.

Viglione, D. J., & Meyer G. J (2008). An overview of Rorschach psychometrics for forensic practice. In C. Gacono, F. Evans, N. Kaser-Boyd, & L. Gacono (Eds.) *The Handbook of Forensic Rorschach Assessment* (pp. 21-53). New York, NY: Routledge/Taylor & Francis Group.

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*(3), 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(3)*, 397–402.

Using CS vs. International Norms

Table 1

Comparison between CIRV (Meyer et al., 2007) and CS (Exner, 2003) Norms for Children & Adolescents: Divergent Variables.

Formatted: Font: Italic

	CIRV (N = 2647)		Combined CS Norms (N = 1390)		*d*
	M	*SD*	*M*	*SD*	
Higher in CS					
X+%	0.36	0.13	0.75	0.09	3.30
XA%	0.62	0.13	0.91	0.06	2.58
WDA%	0.65	0.14	0.92	0.05	2.33
FQXo	7.99	3.28	14.86	2.79	2.20
3r(2)/R	0.25	0.16	0.55	0.12	2.01
MQo	0.95	1.16	3.02	1.78	1.47
COP	0.41	0.75	1.59	0.97	1.42
P~~opular~~	3.65	1.77	5.82	1.19	1.36
CF	1.16	1.41	3.03	1.35	1.35
Sum T	0.24	0.57	0.96	0.50	1.32
FC+CF+C+Cn	2.77	2.38	5.66	2.04	1.28
WSum C	2.17	2.00	4.54	1.58	1.27
GoodHR	2.48	1.85	4.86	1.93	1.26
EA	4.30	3.26	8.01	2.27	1.25
AG	0.27	0.70	1.18	0.81	1.24
Blends	2.27	2.65	5.29	2.22	1.20
a (active)	3.49	3.23	6.72	1.67	1.15
Pair	5.20	4.06	9.11	1.84	1.13
FM	2.47	2.42	4.79	1.68	1.06
Afr	0.53	0.20	0.71	0.15	0.96
FM+m	3.55	3.08	5.81	1.74	0.84
Ma	1.24	1.65	2.53	1.35	0.83
DQ+	4.49	3.80	7.17	2.12	0.81
FC	1.40	1.53	2.44	1.43	0.70
M	2.12	2.22	3.58	2.14	0.67
H	1.75	1.68	2.74	1.62	0.60
es	6.23	5.00	8.57	2.69	0.54
FQX+	0.02	0.15	0.21	0.56	0.54
Sc	1.89	1.16	1.18	1.53	0.50
Higher in Int'l					
X-%	0.38	0.13	0.08	0.06	2.65
FQx-	8.36	4.45	1.78	1.29	1.79

Using CS vs. International Norms

F	13.56	6.67	6.95	2.43	1.18
Xu%	0.25	0.11	0.16	0.07	0.92
Dd	4.44	4.29	1.40	1.32	0.85
FQxu	6.12	4.15	3.25	1.38	0.83
PoorHR	3.01	2.59	1.23	1.03	0.82
DQo	16.70	7.14	11.98	2.33	0.79
Lambda	3.24	4.10	0.70	0.41	0.76
Hd	1.57	1.95	0.50	0.69	0.65
S	2.61	2.16	1.49	0.94	0.61
Ad	2.73	2.66	1.37	1.14	0.60
S-	0.93	1.27	0.34	0.59	0.54
MQ-	0.67	1.07	0.20	0.40	0.52

Note. This table only reports data for divergent variables, i.e., variables for which the CIRV (Meyer et al., 2007) and CS (Exner, 2003) differ of at least a Cohen's *d* of .5.

Using CS vs. International Norms

Table 2

Number of divergent and non-divergent Variables in the Adult vs. Child and Adolescent Norms

Formatted: Font: Italic

Formatted: Font: Italic

	Adult Norms		
	Non-Divergent Variables	Divergent Variables	Total
Child and Adolescent Norms			
Non-Divergent Variables	65 (58.0%)	4 (3.6%)	69 (61.6%)
Divergent Variables	19 (17.0%)	24 (21.4%)	43 (38.4%)
Total	84 (75.0%)	28 (25.0%)	112 (100.0%)

Note. Percentage values in parenthesis indicate the proportion of the frequency of a cell divided by the total.

Using CS vs. International Norms

Table 3.

List of Divergent Variables (n = 24) in both the Child/Adolescent and Adult Norms

Formatted: Font: Italic

Variable	Child and & Adolescent Norms (Cohen's *d*)	Adult Norms (Cohen's *d*)
X+%	3.30	1.98
XA%	2.58	1.23
WDA%	2.33	1.14
FQxo	2.20	1.44
MQo	1.47	0.78
COP	1.42	0.77
Pop	1.36	0.68
CF	1.35	0.50
SumC	1.28	0.87
WSumC	1.27	0.59
GHR	1.26	0.58
EA	1.25	0.50
AG	1.24	0.64
Afr	0.96	0.71
FC	0.70	0.96
FQx+	0.54	0.71
MQ–	-0.52	-0.56
SQ–	-0.54	-0.56
PHR	-0.82	-0.55
FQxu	-0.83	-0.72
Dd	-0.85	-0.67
Xu%	-0.92	-1.13
FQx–	-1.79	-0.94
X–%	-2.65	-1.14

Note. Cohen's *d* values greater than zero indicate that the mean value of the variable is bigger in the CS norms (Exner, 2003); Cohen's *d* values lower than zero indicate that the mean value of the variable value is bigger in the CIRV (Meyer et al., 2007).

Using CS vs. International Norms

Table 4.

List of Variables Divergent in the Adult Norms but Non-Divergent in the Child/Adolescent Norms (n = 4)

Variable	Child & Adolescent Norms (Cohen's d)	Adult Norms (Cohen's d)
MQ+	0.36	0.69
Bt	0.31	0.67
D	0.08	0.53
MQu	-0.13	-0.51

Note. Cohen's *d* values greater than zero indicate that the mean value of the variable is bigger in the CS norms (Exner, 2003); Cohen's *d* values lower than zero indicate that the mean value of the variable value is bigger in the CIRV (Meyer et al., 2007).

Using CS vs. International Norms

Table 5.

List of Variables Divergent in the Child/Adolescent Norms but Non-Divergent in the Adult Norms (n = 19)

Formatted: Font: Italic

Variable	Child & Adolescent (Cohen's d)	Adult (Cohen's d)
F	-1.18	-0.19
DQo	-0.79	-0.17
LAMBDA	-0.76	-0.29
HD	-0.65	-0.41
S	-0.61	-0.44
Ad	-0.60	-0.06
Sc	0.50	0.01
es	0.54	-0.15
H	0.60	0.42
M	0.67	0.22
DQ+	0.81	0.33
Ma	0.83	0.45
FM+m	0.84	0.05
FM	1.06	0.18
PAIR	1.13	0.40
a (active)	1.15	0.49
Blends	1.20	0.40
Sum T	1.32	0.34
3r(2)/R	2.01	0.13

Note. Cohen's *d* values greater than zero indicate that the mean value of the variable is bigger in the CS norms (Exner, 2003); Cohen's *d* values lower than zero indicate that the mean value of the variable value is bigger in the CIRV (Meyer et al., 2007).