

Rorschach Comprehensive System International Norms: Cautionary Notes

Barry Ritzler and Anthony Sciara

In December, 2007, a supplement to the *Journal of Personality Assessment* was published with the stated purpose of "...provid(ing) CS users with a compendium of country-specific or locale-specific norms..." (Meyer, Erdberg, & Shaffer, 2007, p S201). Even though this statement indicates the motivation behind the Supplement was to establish a set of country-specific or locale-specific norms, the editors, in their conclusions manuscript state that "...these projects also introduce the possibility of creating a composite set of international norms..." (Meyer, Erdberg, & Shaffer, 2007, p. S201)

It practically goes without saying that this is a legitimate project in that collecting norms for different countries and cultures is an important undertaking. The development of a set of norms that are country or culture-specific is meaningful in that they may increase the strength of the Comprehensive System (CS) in multiple sites. We commend the amount of effort that went into these projects and the concerns the investigators demonstrated regarding important issues of reliability and correct procedures for using the CS. A general result from these studies is well worth noting: i.e., there is remarkable consistency across most of the international studies. The editors of the Supplement emphasize this consistency and Ritzler (2004), the first author of these cautionary notes, has also suggested that this international consistency is an indication that the Rorschach method is relatively culture-free. Nevertheless, there are some difficult-to-explain differences, especially those between studies from the same country reported in the Supplement. Furthermore, we are concerned that some of the consistency across nations may be the result of methodological deficiencies that we will discuss in this manuscript.

We have concerns about the methodology of these studies. Due to those concerns we are presenting the following cautionary notes for clinicians to consider regarding the use of these international norms. Our concerns are expressed primarily in regard to the adult non-patient data contained in the studies included in the Supplement.

The Supplement presents 21 studies from 17 countries (the Supplement included a non-patient study from France, but did not present the data; we have located the study in an issue of the *Journal of Personality Assessment* and have included it in our remarks-Sultan, Andronikoff, Reveillere, & Lemmel, G., 2006). The final summary article in the Supplement presented only information and opinions that were supportive of the author's development of an international norm set. We are attempting to introduce a cautionary evaluation of the information presented in the Supplement.

We have two major concerns regarding the nature of nearly all the international studies reported in the Supplement. The first concern is the lower complexity levels reported compared to the complexity of the Comprehensive System non-patient sample of 450 United States adults. For instance, the mean Lambda of the international studies ranges from .48 to 1.28 with an overall mean of .83 compared to .58 in the CS non-patient sample reported in the Supplement (Exner, 2007)..

A second major concern we have is the noticeably lower amount of color responses in the international samples compared to the CS Supplement sample (Exner, 2007). For instance, the weighted sum of color responses ($1/2FC + CF + 1 \frac{1}{2}C$) in the international samples ranges from 2.19 to 3.85 with an overall mean of 3.15 compared to 4.54 in the CS Supplement sample. Also, the percentage of extratensive protocols (i.e., an $M/SumC$ ratio with a significant amount of $SumC$) ranges from 7% to 25% in the international samples compared to 31% in the CS Supplement sample.

The following sections summarize our concerns:

Graduate Student Examiners

A majority of the studies used graduate students as examiners (16 out of 21). Each of the authors of these studies tries to justify the use of students by documenting the amount of extra supervision given to the student examiners. If this supervision was delivered effectively by expert CS examiners, the examiners may be ready to serve as examiners in a normative study after their experience in their original study, but their ability to perform adequate CS administration was not likely to have been at an expert level during their involvement in the studies.

The study on inquiry by Lis et al. (Lis, Parolin, Calvo, Zennaro, & Meyer, 2007) included in the Supplement concludes that CS administration is a difficult task that requires

considerable experience and training. The Japanese (Nakamura, Fuchigami, & Tsugawa, 2007) study reports that "...we came to realize that full-fledged, experienced clinicians obtained data of better quality and made double-checking easier [than did students]. The problem was in our naïve assumption that non-patient data would be easier to gather than data in a clinical setting. However, that was not true. The non-patient spectrum was so wide that dealing with the variety of non-patients was not as easy as expected." (Nakamura et al., 2007, p. S97) These are arguments used by Exner (personal communication) when he decided to exclude most student examiners from his normative studies.

The non-Exner studies not using students yielded results that were somewhat more simplistic than the Exner data, but the studies using students as examiners clearly were in an even more simplistic direction (e.g., the Lambda variable). An exception to this finding is in the color responses where the non-student studies were more like the student studies in the amount of color in the Rorschach performances (e.g., the WSumC variable).. This finding may be attributable to the selection procedures used in most of the studies in the Supplement. While the differences between the Exner (2007) data and the data from studies not using student examiners suggest that the Exner (2007) data indicate a higher level of complexity (a conclusion drawn by the editors of the Supplement), at least some of the simplicity can be attributed to the lack of examiner experience.

Since a majority of the studies (at least 16 out of 21) used graduate student examiners, doubt is cast on the statement by the editors in their summary article that "...because the CS international reference samples are quite diverse across a number of variables...the composite norms have considerable generalizability across the same variables." (Meyer, Erdberg, & Shaffer, 2007, p. S202). The examiner variable does not show much diversity (mainly graduate students) and at least some of the similarities across samples (the simplicity factor) can not be attributed to generalizability across countries.

The results of the Romanian study (with protocols administered by an individual with less than graduate training) are quite far from the CS Supplement norms (Exner, 2007), particularly in regard to complexity (Lambda = 1.28) and color (WSumC = 3.28 and Extratensive % = 16).

Generalizability

A second cautionary note has to do with the generalizability of the non-patient samples. Most of the studies only sample from large urban areas. Exceptions are Brazil (Nascimento, 2007), Finland (Mattlar, Forsander, Carlsson, Norrlund, Vesala, Leppanen, Oist, Maki, & Alanen, 2007), Greece (Daroglou & Viglione, 2007), and Exner (2007). The Brazil, Finland, and Greece studies included participants from several different cities and even some participants from less populated areas. Exner's study included participants from 22 of the 50 United States including some areas with small urban populations. Most of the studies claimed that their samples were a good match for the population statistics of the country, but without a representative geographic sampling, these studies can only be said to be locale-specific, not country-specific. Even for the few studies with better geographic sampling (Brazil, Finland, and Exner), the samples are of questionable geographic generalizability because they involve a large proportion of participants from large urban areas.

Furthermore, most of the studies in the Supplement indicate that they relied on "word of mouth" selection procedures to procure participants. Such selection procedures increase the likelihood that the participants will be more like the examiners in demographic characteristics and maybe even psychological traits. For instance, there may not be many psychologists or acquaintances of psychologists who are extratensive. If this is true, then examiners using "word of mouth" for recruitment, are less likely to select extratensive (i.e., color-dominant) participants.

Exclusion Criteria

None of the studies included participants who had any history of inpatient psychiatric treatment. Several studies included individuals who had some outpatient treatment more than two years from the time of the study: Israel-Tibon (2007), Italy (Lis, Parolin, Salcuni, & Zennaro, 2007), Portugal (Pires, 2007), USA-Shaffer (2007), and USA-older (Pertchik, Shaffer, Erdberg, & Margolin, 2007). One study included individuals who had some outpatient treatment more than five years from the time of the study (Argentina-Sanz, 2007). Exner (2007) allowed up to eight outpatient contacts, but did not specify a duration time. Several studies did not refer to outpatient contacts in the exclusion criteria: Denmark (Ivanouw, 2007), Finland (Mattlar et al., 2007), Greece (Daroglou & Viglione, 2007), Israel (Berant, 2007), and Spain (Campo & Vilar, 2007). Only one study

clearly included participants in outpatient treatment at the time of the study (The Netherlands; de Ruiter & Smid, 2007)).

At the beginning of the Supplement, the editors make a distinction between non-patient and normative studies. They state that exclusion on the basis of any psychiatric treatment constituted a non-patient, but not necessarily a normative study. At the end of the Supplement, they refer to their combined data as “an international normative reference group.” (Meyer, Erdberg, & Shaffer, 2007, p. S201; emphasis added) Since the international studies show considerable variability in their exclusion criteria, it is recommended that a more consistent rule be established regarding a history of outpatient contact to better identify a reasonable normative rather than non-patient sample.

Inadequate Sample Size

Sample size for any research is a key issue for ability to generalize the results. While the international norms presented 5815 subjects collectively, it is important to look at the sample of each of the studies included in the international sample. The original sample from Exner’s norms (2001) included 600 individuals from 22 states, different urban, suburban and rural settings and varying educational backgrounds. Exner (Exner & Erdberg, 2005) was in the process of a re-norming study that is represented by the Exner 450 sample (Exner, 2007). Exner never intended that number as a final number and indicated that he wanted around 1000 for inclusion in the final study. (personal communication)

A review of the studies in the current sample indicate only two samples (Argentina; Lunazzi, Urrutia, de la Fuente, Elias, Fernandez, & de la Fuente, 2007 and Spain; Campo & Vilar, 2007) have more than Exner’s 450 with 506 and 517 respectively. Only 10 of the studies had between 100 and 450 subjects and seven of the studies had fewer than 100 subjects.

None of the studies, including the Exner 450, approached the original 600 in the CS norms and, clearly, none even came close to what Exner was trying to achieve with his ultimate re-norming study. It is difficult to believe that appropriate norms can be established with these small sample sizes, especially when many of the participants did not come from diverse parts of the country represented. In fact, then, the sampling

could only be representative of locale-specific findings. Summarizing these undersized samples into an overall international norm may not be appropriate. In order to be considered normative, sampling requires some type of stratification that is not represented in these findings.

Warm-up Procedures

Exner has warned that “many clients are not well prepared by those who have referred them, and examiners often must take some time to make sure that the client is not likely to be harboring negative or erroneous assumptions about the assessment process.” (Exner, 2001, p. 3) He suggested that “no special elaboration concerning the nature of the Rorschach should be required if the client has been properly prepared for the overall assessment process. In most cases this will be done after a relatively brief interview during which the examiner seeks to insure that the person has a reasonable understanding of the purpose of the assessment.” (Exner, 2001, p. 3) This focus on establishment of the relationship can affect the outcome of the testing based on the examiner’s ability to develop confidence and cooperation with the subject. Recalling that the Rorschach is a perceptual/cognitive problem solving task, cooperation by the subject in that endeavor is essential.

Non-cooperation often yields short, not very elaborate protocols. In effect, the subject is not as engaged in the task and produces a protocol that demonstrates that lack of engagement.

The studies in the Supplement give little specification of what the warm up procedure was and how it was presented. There are also a number of different warm up strategies that appeared to be engaged in with the subjects. It is difficult to determine from the studies the adequacy of these strategies and whether they were sufficient to generate appropriate engagement in the task.

Administration Procedures

The authors of the Supplement acknowledge that the composition of the “...international sample being quite different with respect to selection procedure, examiner training, examination context, language, culture, and national boundaries...” (Meyer, Erdberg, & Shaffer, 2007, p. S202), yet proceed with analysis of the data as normative because the

data were collected by "...motivated and trained individuals seeking to advance the database of Rorschach assessment." (Meyer, Erdberg, & Shaffer, 2007, p. S202) In terms of administration procedures, some basic issues seem to be ignored or at least minimized.

The success of the CS depends, to a large extent, on the simplicity of administration. For example, the administration of the Rorschach task should always begin with the question, "What might this be?" Inherent in that question is the basic task of the Rorschach of problem solving. In a recent communication to one of the authors (Sciara), Carl-Erik Mattlar questioned the difference between using "might" and "could" and how this was translated into Finnish. In order to clarify that issue, a distinction in English would be made as follows: "Might" is a prompt that expresses 'permission, liberty, probability, possibility that is consistent with the problem solving nature of the Rorschach. "Might" gives the individual permission to consider, compare, and give possibility to their responses. In effect, "might" does not communicate that the stimulus is actually something, but rather expresses probability or possibility of what the stimulus might be. This then allows for a generative type of problem solving. That is, the subject then can generate possibilities rather than just concretely try to figure out the right answer.

"Could" on the other hand is a word that presents less force or certainty. In effect, it is almost telling the individual that the stimulus being presented to him/her is actually something and we are asking them to figure out what that is.

As an example, we have heard many different translations of the initiating question into Spanish. When we reviewed the authorized translation of the CS into Spanish (del Rio, 1974) and consulted a linguistic expert from the University of Mexico, it became apparent that the correct translation is, "Que podria ser esto?" Any other Spanish translation of the initiating question may set a different task for the participant.

When we look at the area of Inquiry, matters get even more complicated. Exner (2003) indicates that "...the Inquiry has been one of the most misunderstood and abused features of the Rorschach. When done correctly, it completes the richness of the test data. When done incorrectly, it can muddle a protocol terribly and often generates data that may be of clinical interest but that represents something other than true Rorschach data." (Exner, 2003, p. 58) Recently, in consultation with Exner we developed The Little Book on Administration of the Rorschach Comprehensive System (Sciara & Ritzler,

2006) with an accompanying DVD which demonstrated correct administration and inquiry procedures. This is the first and most comprehensive guide to date.

If the inquiry phase is inappropriately and/or inadequately pursued by the examiner, then the resulting information can suffer tremendously. There was no real focus on evaluating the adequacy of inquiry on the majority of the samples collected in the countries represented by the studies in the Supplement.

An article by Meyer, Viglione, Erdberg, Exner, and Shaffer (2004) examined 40 protocols each from Exner's (n=450) and Shaffer et al.'s (n=283) samples to determine the adequacy of inquiry. They found differences in whether a key word was inquired or not, or whether unnecessary inquiry questions were asked. Initial evaluation indicated that "...across 129 variable, there were 36 scores that initially differed by $d=.40$ or larger." Clearly, when there is this type of variability in agreement among raters regarding whether or not inquiry was appropriate, then an evaluation of the appropriateness of any data to be used in developing an international norm should be scrutinized carefully.

Recommendations regarding the International Norms

Above, we have outlined our concerns for the development of the international norms and now we wish to express our recommendations for further use of those norms.

First, we believe there is insufficient data to support the use of the international norms over those developed by Exner. While the idea of international norms is enticing, and with appropriate methods is attainable, the current research falls short of achieving that goal. It is our belief that, until better studies are completed, it is inappropriate to describe personality and plan treatment from the international data.

While the data presented in the Supplement are impressive, they clearly do not constitute appropriate norms for the Comprehensive System. If professionals choose to utilize the data included in the international norms for evaluation and treatment, it is inappropriate to indicate that the CS has been used for the evaluation. The history of the Rorschach is replete with the development of differing systems and the current research is but a variation of this historical process.

One specific recommendation we have is that students should not be used to collect normative data. While the students used in many of the studies in the Supplement may now be ready to serve as examiners, they were not adequately experienced at the time of the studies.

A second recommendation is that samples should be selected to represent the entire country or culture, not just a large urban area and its associated culture. The studies in the Supplement mostly make the claim that their participants are characteristic of the country as a whole, but most of the samples involve only individuals living in a single large urban area and do not include individuals from smaller urban areas, other large urban areas, or rural areas.

Another recommendation is to not use word of mouth selection procedures. This method makes it more likely that the examiners will select individuals who are more like themselves than like the general population.

Forensic psychologists must pay particular attention to the current Supplement studies as they could bring about new challenges to using the Rorschach in court. The CS as developed by Exner, however has been shown to be a sturdy instrument that is well accepted in the courts in the United States (Weiner, Exner, & Sciara, 1996). Any use of the international norms to develop conclusions about individuals involved in the court system is likely to come under significant scrutiny. In fact, the criticisms leveled against Wood, Nezworski, and Stetsjal (1996) would in many ways be appropriately leveled against the current research. It is simply inappropriate to bring together data from varying perspectives and call them the same.

While the editors of the Supplement describe the examiners from the many different counties as motivated, that is simply not enough justification to believe that their procedures are consistent with the CS. There have been many instances in which individuals with minimal training and experience have attempted to use the CS when administering the Rorschach only to fall short of appropriate procedures and, therefore, appropriate conclusions. In some instances, individuals have modified their procedures, claiming they do not violate the constraints of the CS. Those modifications however, should be subjected to experimental validation. For instance, some individuals believe that taking the responses of the Rorschach utilizing a laptop computer is consistent with taking a Rorschach with the paper and pencil method. Unfortunately, there is no data to

support that contention. The use of a laptop in taking a Rorschach should be submitted to empirical validation before using that procedure.

It is our recommendation that assessment psychologists not "...integrate the composite international reference values into their clinical interpretation of protocols" (Meyer, Erdberg, & Shaffer, 2007, p. S201) as is recommended by the editors of the Supplement. As indicated in the current article, there is simply too much about the data that is questionable to make a shift to these reference values.

What do we recommend from here? First, there needs to be more emphasis on correct administration, including inquiry. If the studies in the Supplement have not used correct CS methods, then the data are not appropriate as CS norms.

Next, we need to look at what we mean by norms. The samples included in the international studies are probably less normative and more locale-specific samples. It is inappropriate to use these local samples as normative data if they are not consistent with the demography they are purported to represent.

In general, the idea of country-specific or culture-specific norms is a good goal to toward which the personality assessment community should strive. In these days of a true global economy and tremendous cultural interchange, the Rorschach can be a wonderful descriptor of individual strengths and weakness, but we need solid data on which to base our inferences. Establishing guidelines for normative studies is necessary for correct administration, data collection, participant selection, and examiner training.

We do encourage continued research, the development of novel approaches to understanding the CS, and expanding the utility of the System in varied cultural and community settings. We are encouraged by the current research in that it illuminates the need for better understanding of the procedures used in normative data collection in general and administration in particular. We hope the current research will be a stepping stone toward an international understanding of the Rorschach Comprehensive System that is consistent across cultures.

References

Berant, E. (2007) Rorschach Comprehensive System data for a sample of 150 adult nonpatients from Israel. Journal of Personality Assessment, 89 (S1), S67-S73.

Campo, V. & Vilar, N. (2007). Rorschach Comprehensive System data for a sample of 517 adults from Spain (Barcelona). *Journal of Personality Assessment*, 89 (S1), S149-S153.

Daroglou, S. & Viglione, D. (2007). Rorschach Comprehensive System data for a sample of 98 adult nonpatients from Greece. *Journal of Personality Assessment*, 89 (S1), S61-S66.

de Ruiter, C. & Smid, W. (2007). Rorschach Comprehensive System data for a sample of 108 normative subjects from The Netherlands. *Journal of Personality Assessment*, 89 (S1), S113-S118.

del Rio, P. (1978). *Sistema compensivodel Rorschach. Tomo I. Madrid: Impreson an espana por COSOL S.A., Poligono Industrial "El Balconcilio" (Guadalajara).*

Dumitrascu, N. (2007). Rorschach Comprehensive System data for a sample of 111 adult nonpatients from Romania. *Journal of Personality Assessment*, 89 (S1), S142-S148.

Exner, J. (2001). *A Rorschach Workbook for the Comprehensive System (5th ed.)*. Asheville, NC: Rorschach Workshops, Incorporated.

Exner, J. (2003). *The Rorschach: A Comprehensive System: Vol. 1. Basic Foundations (4th ed.)*. Hoboken, NJ: Wiley.

Exner, J., (2007). A new U.S. adult nonpatient sample. *Journal of Personality Assessment*, 89 (S1), S154-S158.

Exner, J. & Erdberg, P. (2005). *The Rorschach: A Comprehensive System: Vol. 2. Advanced Interpretation (3rd ed.)*. Hoboken, NJ: Wiley.

Ivanouw, J. (2007). Rorschach Comprehensive System data for a sample of 141 adult nonpatients from Denmark. *Journal of Personality Assessment*, 89 (S1), S42-S51.

Kumho Tire Co., Ltd. v. Carmichael, 119 Ct. 1167, (1999).

Lis, A., Parolin, L., Calvo, V., Zennaro, A., & Meyer, G. (2007). The impact of administration and inquiry on Rorschach comprehensive system protocols in a national reference sample. *Journal of Personality Assessment*, 89(S1), S193-S200.

Lis, A., Parolin, L., Salcuni, S. & Zennaro, A. (2007). Rorschach Comprehensive System data for a sample of 249 adult nonpatients from Italy. *Journal of Personality Assessment*, 89 (S1), S80-S84.

Lunazzi, H., Urrutia, M., de la Fuente, M., Elias, Fernandez, & de la Fuente, S. (2007). Rorschach Comprehensive System data for a sample of 506 adult nonpatients from Argentina. *Journal of Personality Assessment*, 89 (S1), S7-S12.

Mattlar, C., Forsander, C., Carlsson, A., Norrlund, L., Vesala, P., Leppanen, T., Oist, A., Maki, J., & Alanen, E. (2007). Rorschach Comprehensive System data for a sample of 343 adults from Finland. *Journal of Personality Assessment*, 89 (S1), S57-S60.

Meyer, G., Erdberg, P., & Shaffer, T. (2007). Toward international normative reference data for the comprehensive system. *Journal of Personality Assessment*, 89(S1), S201-S216.

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

Nakamura, N., Fuchigami, Y., & Tsugawa, R. (2007). Rorschach comprehensive system data for a sample of 240 adult nonpatients from Japan. *Journal of Personality Assessment*, 89(S1), S97-S102.

Nascimento, R. (2007). Rorschach Comprehensive System data for a sample of 409 adult nonpatients from Brazil. *Journal of Personality Assessment*, 89 (S1), S 35-S41.

Pertchik, K., Shaffer, T., Erdberg, P. & Margolin, D. (2007). Rorschach Comprehensive System data for a sample of 52 older adult nonpatients from the United States. *Journal of Personality Assessment*, 89 (S1), S166-S173.

Pires, A. (2007). Rorschach Comprehensive System data for a sample of 309 adult nonpatients from Portugal. *Journal of Personality Assessment*, 89 (S1), S124-S 130.

Ritzler, B. (2004). Cultural Applications of the Rorschach, Apperception Tests, and Figure Drawings. In Hilsenroth, M. & Segal, D. (eds.), *Comprehensive Handbook of Psychological Assessment: Vol. 2: Personality Assessment*. New York: Wiley & Sons.

Sanz, I. (2007). Rorschach Comprehensive System data for a sample of 90 adult nonpatients from Argentina. *Journal of Personality Assessment*, 89 (S1), S13-S19.

Sciara, A. & Ritzler, B. (2006). *The Little Book for Rorschach Comprehensive System Administration*. Asheville, NC: Grove Clinic.

Shaffer, T., Erdberg, P., & Haroian, J. (2007). Rorschach Comprehensive System data for a sample of 283 adult nonpatients from the United States. *Journal of Personality Assessment*, 89 (S1), S159-S165.

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

Tibon, S. (2007). Rorschach Comprehensive System data for a sample of 41 adult nonpatients from Israel. *Journal of Personality Assessment*, 89 (S1), S74-S79.

Weiner, I., Exner, J., & Sciara, A. (1996). Is the Rorschach welcome in the courtroom? *Journal of Personality Assessment*, 67, 422-424.

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

These online Articles do not represent the position of Rorschach Training Programs and are posted to share ideas and encourage discussion.