Rorschach Comprehensive "International Norms": Cautionary Notes

Barry Ritzler and Anthony Sciara

Rorschach Training Programs

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"…and "creating a composite set of international norms at a level rarely achieved in personality assessment." (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 an important undertaking that may extend 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.

The conclusions of the editors of the Supplement were that these studies were legitimate as data to be combined to develop a set of international norms for the CS. As the editors say, "…the adult samples from around the world were generally quite similar and thus we encourage clinicians to integrate the composite international reference values into their clinical interpretation of protocols." (Meyer, Erdberg, & Shaffer, 2007, p. S201). This in a paper entitled "Toward international normative reference data for the CS."

The current authors 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, but we have a little to say about the child data.

The Supplement presents 21 studies from 17 countries (although the Supplement includes the study from France in the data, there was no presentation of that study in the Supplement for our review.) The final summary article in the Supplement presented only information and opinions that were supportive of the author's intention to develop an "international norm set." The current authors are attempting to introduce a cautionary evaluation of the information presented in the supplement. Those cautionary notes are generally summarized as follows:

1. Most of the studies (16 studies) utilized students in data collection (only four clearly indicate not using students (Argentina--Sanz, Japan, Portugal, Romania, and Exner--USA); the specific information regarding the French study was not available to determine if students were used as examiners.
2. The samples presented (with the exception of the Exner, Brazilian, and Finnish data) were not selected from a countrywide population, but from large urban populations and, therefore, are locale-specific data and not country-specific data
3. Many of the studies excluded individuals with <u>any</u> psychiatric history. Some of the studies included individuals with limited outpatient treatment (i.e., Argentina—Sanz, Israel—Tibon, Italy, Netherlands, Portugal, USA—Exner,

USA—Shaffer et al., and USA—older) and some of the studies did not mention outpatient treatment (i.e., Denmark, Finland, Greece, Israel—Berant, and Spain.) The CS norms included individuals with up to eight (8) psychiatric contacts; therefore, the exclusion criteria for the participants in some of these international data were quite different from the CS norms. The Netherlands sample actually included participants who were actively engaged in outpatient therapy.

4. The number of subjects in most of the studies was of inadequate sample size for the development of normative data. Only two international studies (Argentina [506 S's] and Spain [517 S's]) were greater than the Exner new non-patient sample (450 S's). All other county samples involved considerably fewer participants.
5. The "warm up" procedures used in the studies were not standardized and were usually only described as "brief". These procedures, if not appropriately done, can significantly affect subject participation and, therefore, prevent appropriate data collection
6. Administration procedures raised questions about the appropriateness of the practices used in presenting the Rorschach task to participants. There are concerns not only about students as data collectors (noted above), but also about site-specific "local" differences. There were concerns about both administration of the response proper and about inquiry. Administration concerns also included

a. concern for how such simple instructions as “What might this be?” were translated and the appropriateness of the meaning in the translated version.

b. as indicated by the Italian study included in the Supplement (Lis et al., 2007) inquiry is the most difficult part of administration and was not an area that was adequately controlled for. As the Japanese adult normative study indicated, the difficulty of dealing with non-patients was particularly troublesome and, therefore, more advanced examiners were required.

7. Child and adolescent data were interpreted with concerns about the identification of “pathology” as opposed to descriptions of behaviors.

1. Student Examiners

A majority of the studies used graduate students as examiners (16 out of 20 and we do not know about France). 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 THE NORMATIVE 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, et al, 2007) included in the Supplement concludes that CS administration is a difficult task that requires considerable experience and training. The Japanese adult normative 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, Fuchigami, & Tsugawa, 2007, p. S97). These are arguments used by Exner (personal communication) when he decided to exclude student examiners from his normative studies.

As an indication of the effect of student examiners on the non-patient results, Table 1 compares the weighted means (based on different Ns) from the studies using students with those of three studies that used trained examiners (Argentina—Sanz, Japan--Nakamura, and Portugal—minus those with less than 10 years of education) and Exner's new norms (N=450). The Romanian study, though not using students, was not included in the Table because the results were so unusual compared to the other studies (e.g., mean Lambda = 1.28 and 43% avoidant.)

Table 1 shows that the studies using students as examiners yielded different results than the studies not using students and the Exner study. The Table shows 20 of the 27 variables listed with the non-student studies falling between the student and Exner studies with eight of the variables showing little, if any, differences between the Exner and non-student studies (DQ+, Introversive %, Sum Shading, Experience Actual [EA], experience stimulation[es], Affective Ratio[Afr], and Intellectualization). The non-Exner studies not using students yielded results that were <u>somewhat</u> more simplistic than the Exner data, but the studies using students as examiners clearly were in an even more simplistic direction. 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. (This finding may be attributable to the selection procedures used in most of the studies in the Supplement—see the following section on generalizability.) While the differences between the Exner data and the data from studies not using student examiners suggests that the Exner 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 20) used 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 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 students) and at least some of the similarities across samples (the simplicity factor) can not be attributed to generalizability across countries.

2. 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, Finland, and Exner. The Brazil and Finland studies included participants from several different cities and even some participants from rural areas. Exner's study included participants from 22 of the 50 United States. 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 only include some of the areas outside of the cities where the study originated (e.g., Exner's sample includes protocols from on 22 of the 50 United States). yes

Furthermore, most of the studies in the Supplement indicate that they relied on "word of mouth" selection procedures to procure participants. Most psychologists and psychology students are introversive (or ambitent) in their Rorschach problem-solving styles—rarely extratensive I doubt if you are extratensive—you may have a lot of color on your Rorschach, but more M. If these individuals are using "word of mouth" for recruitment, then it is less likely that extratensive (i.e., color-dominant) individuals would be selected.

3. 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, Italy, Portugal, USA—Shaffer, and USA—older). One study included individuals who had some outpatient treatment more than five years from the time of the study (Argentina—Sanz). Exner 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, Finland, Greece, Israel—Berant, and Spain). Only one study clearly included participants in outpatient treatment at the time of the study (the Netherlands). Again, there is no information about the French study.

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 date as “an international normative reference group” (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.

4. Inadequate Sample Size

Sample size for any research is a key issue for ability to generalize the results. While the “international norms” presented 4704 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 with different urban, suburban and rural settings and varying educational backgrounds. Exner was in the process of a re-norming study that is represented by the Exner 450 sample presented in the Supplement. 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 and Spain) 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 inadequate samples into an overall "international norm" is not appropriate. In order to be considered "normative" sampling requires some type of stratification that is not represented in these findings.

5. "Warm-up" Procedures

Exner consistently talked about the necessity to establish a positive and cooperative relationship in the testing environment. In fact, this focus on establishment of the relationship was thought to affect in a very significant manner 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 is often thought to yield short, not very elaborate protocols. In effect, the subject is not engaged 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.

6. Procedures for Administration

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.

Our experience has demonstrated that even the simplest of instructions regarding administration, when used on an international basis must be evaluated. For example, the initiation of the Rorschach task should 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.

This issue is not isolated as the current authors have heard many different translations of the initiating question in Spanish. Consulting numerous sources including the authorized translation of the CS into Spanish (Pablo del Rio, 1974) and consulting a linguistic expert from the University of Mexico, the correct translation is, "Que podria ser esto?" This is only a minor example of the difficulties of assuring consistency in the administration of the Rorschach.

When we look at the area of Inquiry, matters get even more complicated. While the Inquiry is addressed by Exner (2001) it has not been elaborated on for many years. Recently, the current authors, in consultation with Exner 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 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 represented in the Supplement.

An article by Meyer, Viglione, Erdberg, 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 difference in whether a key word was inquired or not, or whether unnecessary inquiry questions were asked. Initial evaluation indicated

that "…across 129 variables, there were 36 scores that initially differed by d=.40 or larger. However, when all 80 protocols were blindly rescored by a third group of researchers, only three scores 92.3% still differed by this amount…indicating that much of the seeming variability was due to site specific scoring conventions." (Meyer et al., 2004). While this is identified as "scoring conventions" it actually addresses the issue of administration. 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.

The issue of site specific conventions could clarify issues of "outliers" in the current data set. For example, the data from the Romanian study demonstrated an exceedingly high Lambda and proportion of Avoidant records. All these records were collected and coded by the same individual. Likewise, the data from the Japanese child and adolescent study demonstrate significant variability from any norms. All these data were collected in the same locale in Japan and, therefore, raises the question of site specific conventions among the examiners.

7. Child-Adolescent Data

The Supplement clearly identifies there are significant problems with the child-adolescent data. This is not a particularly new finding as developing normative information regarding children and adolescents has always been a difficult undertaking.

Developmentally, children vary an extreme amount and, therefore, any attempts to develop norms that are age-based are flawed from the beginning. Exner often said (personal communication) that interpretation of a child protocol should include

information from one age group before the child's current age and one age group after the child's current age. The child and adolescent norms published by Exner clearly have difficulties in that they are dated and have too few participants in each age grouping. These data have always been used as guidelines for description of children as opposed to definitive statements about an individual child's functioning.

Since children do vary widely as they age, it is difficult to ascribe to a child certain "pathologic" findings. It is with this concern that is described in the Supplement that we take issue. Rorschach data should always be descriptive in nature, rather than focusing on "pathology". If we look at descriptive statements about children we can talk about how they manage different psychological functions at the time of testing. It is inappropriate to ascribe pathological functioning to a child based on one set of testing data at one point in time.

It is not, however, inappropriate to describe a child's functioning at a particular point in time and to describe how their functioning at that point in time based on Rorschach findings. For example, a 14 year old male student of superior intelligence who has as part of his Rorschach findings no M responses can appropriately be described as being quite overwhelmed by emotions. While this statement should always be correlated with clinical findings, the Rorschach findings are still descriptive.

The Supplement acknowledges difficulties with the child and adolescent findings, but a more parsimonious explanation of those findings may not be the children, but, as described above, may be a result of problems with examiners, selection criteria, etc.

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 and outlined in the CS. While the idea of international norms is enticing and important, the current research falls short of achieving that goal. It is our belief that, until better controlled studies are completed, it is inappropriate to describe personality and plan treatment from the "international norms."

While the data presented in the Supplement is impressive, it clearly does not reflect what has been consistently described as 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. It is our belief that if the CS had been strenuously adhered to, then use of the "international norms" would be appropriate.

One specific recommendation 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. Furthermore, samples should be selected to represent the entire country, not just a large urban area. The articles 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 large urban areas and do not include individuals from smaller

urban areas, other large urban areas, or rural areas—important geographical areas of any country.

Forensic psychologists must pay particular attention to the Supplement as it could bring about new challenges to using the Rorschach in court. The current research simply cannot stand up to a Daubert type challenge in court (Kumho Tire Co., Ltd. v. Carmichael, 1999). 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 again Wood, Nezworski, and Stejskal (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 data.

While the editors of the Supplement describe the examiners from the many different counties as "motivated", that is simply not justification to believe that their procedures are consistent with the CS. It has been our experience that using the Rorschach in personality assessment is not a project that should be undertaken lightly. We have seen 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 substance 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" 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. The old adage of "garbage in, garbage out" needs to be heeded. If the data that have been utilized are not consistent with CS guidelines then the "normative data" that are recorded by those data sources are not appropriately interpreted in CS values.

Next, we need to look at what we mean by norms. While we have a lot of sample data, it is not necessarily "normative" data. The samples included in the internationals studies are probably less normative and more locale samples. It is inappropriate to use these locale samples as normative data when they may inappropriately identify the demography which they purport to represent. We need to carefully stratify the samples before calling them normative.

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. A goal of revisiting the research which is represented in this Supplement would be well served by setting up specific guidelines for data collection, participant inclusion, examiner training, and administration rules.

Finally, statistical analyses should be revisited when the modifications to the overall research are accomplished. We have not based our concerns on a statistical critique of the Supplement because we lack the expertise for such an undertaking. It is our hope that someone with appropriate credentials will wade in with a statistical evaluation of the international studies. However, any analysis should focus on accessibility to the clinician. Statistical analyses that are outside the normal understanding of the average clinician do not encourage appropriate understanding of the system as a whole.

We do encourage continued research, 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 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

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

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.

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.

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

Weiner, I., Exner, J., & Sciara, A. (1996). Is the Rorschach welcome in the courtroom?, Vol 67, No2, pp 422-424.

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

International Studies (Adult) Covered in the Supplement

(All published in Journal of Personality Assessment, 89S1)

Berant, E. (Israel—N = 150) pp. S67-S74.

Campo, V. & Vilar, N. (Spain—N = 517) pp. S149-S153.

Daroglou, S. & Viglione, D. (Greece—N = 98) pp. S 61-S66.

deRuiter, C. & Smid, W. (The Netherlands—N = 233) pp. S113-S118.

Dumitrascu, N. (Romania—N = 111) pp. S142-S148.

Exner, J. (USA—N 450) pp. S154-S158.

Greenway, P. & Milne, L. (Australia—N = 128) pp. S20-S25.

Ivanouw, J. (Denmark—N = 141) pp. S42-S51.

Lis, A., Parolin, L., Salcuni, S., & Zennaro, A. (Italy—N = 249) pp. S80-S84.

Lunazzi, H., Urrutia, M., Fuente, M., Elias, D., Fernandez, F., & Fuente, S. (Argentina—N = 506) pp. S7-S12.

Mattlar, C., Forsander, C., Carlsson, A., Norrlund, L., Vesala, P., Leppanen, T., Oist, A., Maki, J., & Alanen, E. (Finland—N = 343).

Mormont, C., Thommessen, M., & Kever, C. (Belgium—N = 100) pp. S26-S34.

Nakamura, N., Fuchigami, Y., & Tsugawa, R. (Japan—N = 240) pp. S97-S102.

Nascimento, R. (Brazil—N = 409) pp S35-S41.

Pertchik, K., Shaffer, T., Erdberg, P., & Margolin, D. (USA—N = 52).

Pires, A. (Portugal—N = 309) pp. S124-S130.

Raez, M. (Peru—N = 233) pp. S119-S123.

Sanz, I. (Argentina—N = 90) pp. S13-S19.

Shaffer, T., Erdberg, P., & Haroian, J. (USA—N = 283) pp. S159-S165.

Tibon, S. (Israel—N = 41) pp. S74-S79.

.