WHAT FUTURE FOR THE COMPREHENSIVE SYSTEM? ANNE ANDRONIKOF

Emeritus Professor at the University Paris Nanterre

andronik@parisnanterre.fr

CONFERENCE AT THE IIIrd CONGRESS OF THE COMPREHENSIVE SYSTEM INTERNATIONAL RORSCHACH ASSOCIATION (CSIRA/ARISI)

Paris, September 6, 2019

Today, due to the recent surge of critics undermining the validity of the Comprehensive System, we, practitioners of this method, are bound to reflect on the scope and scientific basis of those critics. The questions are: Has the Comprehensive System (CS) lost its validity? As ethical professionals, must we discard it and move to another system? And in general, is the Rorschach still a valuable tool of assessment?

We will here offer our thoughts and conclusions on these important issues.

Analysis of the latest article discrediting the CS

In 2018, a fundamental critic of the Comprehensive System was published in the JPA by Mihura (2018), a paper which aims to discredit the scientific foundation of the CS. In essence, it represents a dramatic kill of John E. Exner's life work: revising very thoroughly every single evidence brought forward by John E. Exner in his book "The Rorschach: A Comprehensive System" (2003) which lays out the scientific and evidence-based arguments on which the System is built, these authors pinpoint a great number of errors, false research findings, mismatched references and other unscientific assertions.

In an apparent attempt to alleviate the opprobrium cast on the Comprehensive System and its creator, the authors underline that no other test than the Rorschach CS has ever been submitted to such scrutiny and call the assessment community to perform similar verification of all test manuals.

Strangely, Mihura, who is one of the promoters of the most recently created Rorschach system R-PAS, does not call for the same scrutiny of the R-PAS manual. Some fundamental critic has already been expressed about the purported evidence-base of the R-PAS, and curiously those critical manuscripts were constantly rejected by the JPA.

What is the R-PAS based on? As incredible as it might be, it is based on JE Exner's work, on the Comprehensive System data and concepts.

In a previous huge meta-analysis of the validating findings of the Comprehensive System's variables (2013), Mihura has shown strong validating findings for many of the CS variables, moderate for others and no validation for some.

So how can we explain the illogical position of the R-PAS group, who strive to validate the Comprehensive System and at the same time strive to demonstrate its fallacy, based on the purported dishonesty of its creator?

As early as 2011, Sciara & Ritzler (unpublished) analysed the situation created by the launching of a "new" system (the R-PAS) and predicted "When individuals want to initiate their own, new, approach to an established product it is common to attempt to discredit the old", and stated "Any new system must undergo the rigors of scrutiny by the professional community." Their accurate criticism of the foundations of the R-PAS is still valid today. They also announced that an international association (today called csira-arisi) "will continue to research the CS".

The Exner Research Council

The initiators of the R-PAS have sat in the Research Council installed by John Exner for the scientific development of the CS, some of them for many years. The Research Council was a brainstorming team of researchers and clinicians who met regularly to advance the system, critics and new ideas were welcome and thoroughly investigated.

I myself was invited by John E. Exner to join the Research Council and either in presence either through weekly telephone conversations with John E. Exner have been kept informed of the ideas and research projects currently debated by the group.

For instance, I remember that the group's suggestion to control the number of responses by asking a minimum of 2 and a maximum of 4 responses per card was duly considered and finally rejected by John E. Exner. The reason of the rejection stemmed from an asset-liability evaluation: on one hand it would indeed help avoiding "the problem of R" and permit stronger statistical data, on another hand it would restrict the subject's freedom and diminish the richness and clinical dimension of the test. John E. Exner, who had previously studied in depth "the problem of R" (Exner, 1992) deemed that the advantages of such a rule were not worth the loss it entailed.

To mention only the latest debates and projects, I recall John E. Exner's reaction to the numerous critics aimed at the normative data he had produced. Three things happened: first, he discovered that part of the original data had been duplicated (probably by an impatient or dishonest assistant). Therefore, a corrected and reduced non-patient sample was used for a new set of normative data.

Secondly, the findings of Shaffer and Erdberg ("Fresno" sample) on an independent non- patient sample were intriguing by the scope of differences with Exner's data. Suspecting a problem in scoring, John E. Exner obtained to have his team re-score part of the protocols and dramatic scoring discrepancies were found. When correctly scored, the data appeared to match Exner's normative data.

Finally, John E. Exner decided to update the normative data and recruited 450 new non- patients. Somme small differences were indeed found with the original normative data, interestingly showing an evolution in perception within the US population, but on the whole the results were the same (2007). John E. Exner was acutely aware of the scientific challenges posed by research with the Rorschach and called for sound research, both in application and validation studies (1995).

What do we learn from the preceding facts? We learn that John E. Exner was dedicated to sound and ethical science, that he welcomed critics and suggestions from colleagues and thoroughly pursued matters, weighing pros and cons before he adopted a suggestion, corrected an error or rejected a proposal.

In other words, we can outright reject any notion of dishonesty, misrepresentation or data falsification from the part of John E. Exner. And I believe that all the persons who have had the privilege to know him, work with him, including those who criticized him, remember him as a very honest, very open and extremely generous person.

Permit me to make here a self-citation (Andronikof, 2008) about John E. Exner's ethical stand:

"... As a researcher his motto might have been "learn from experience," meaning that he would not accept as true something that could not be proved or evidenced, and at the same time he was ready to change or even reverse his views in the face of new evidence. This motto also implies a constant quest for proving or disproving hypotheses."

So how can we explain the discrediting findings by Mihura?

I believe these authors have strived to perform an honest and meticulous analysis and therefore will review their main critics.

1/ Unpublished validity findings by the test developer (i.e. experiments performed by John E. Exner not submitted to peer review and publishing).

This criticism has been addressed to John E. Exner often, including by myself. To me, he responded that submission to a journal is a very long and time-consuming process and that, had he written articles for every single experiment, he would never have had the time to achieve his work plan. He also showed me where he kept the files pertaining to each experiment and told me that one day, we both might review them. Unfortunately, I was unable to obtain them from his family after his death in 2006. On another hand, he was very interested in replications of his studies and for instance shared with several researchers the LAD software for a replication of his EB study.

Conclusion: researchers are welcome to replicate Exner's studies (those which are described in the book and those which John E. Exner presented in the Alumni workshops). Only through replication studies the validity of the variables and indices can be challenged.

2/ Mihura (2018) found that "only 50 validity findings cited in the CS manual were published in peer-reviewed journal articles and used CS coding compared to the 1,156 CS finding in Mihura et al.'s meta-analyses." (p. 4)

That is an interesting point, to which answers can be found in the article by Smith, J. M., Gacono, C. B., Fontan, P., Taylor, E. E., Cunliffe, T. B., & Andronikof, A. (2018). These authors followed the recommendations written by John E. Exner in his book Issues and Methods in Rorschach Research (1995). In the introduction (p.3), Exner states "A huge number of published investigations (...) are clearly marked by errors in design, implementation, and/or analysis". He then proceeded to explain in detail the nature of these errors and how to avoid them.

Smith et al. (2018) also paid great attention to the difference between validation and application studies, as I. Weiner pointed out (1995): applied studies are those which use the Rorschach as a tool of investigation, validation studies are those which investigate the construct of Rorschach variables (researches *with* the Rorschach or *on* the Rorschach). Application studies are not meant to validate an instrument (Borsboom, Mellenbergh & van Heerden, 2004). Smith et al. (2018) then reviewed the quality of the 210 studies from the Mihura et al. (2013) meta-analyses in light of the criteria set by Exner (1995) and others (Cunliffe et al., 2012; Gacono et al., 2001). Results are rather sobering: of the 210 studies, only 104 could be classified as validation studies, and 58 of those (55%) presented 4 or 5 research issues out of the five criteria defined (reporting IQ/education level, number of responses, lambada/F%, interrater reliability, a sample size of minimum 20 participants). As Viglione and Exner (1995) stated: "All literature cannot be afforded equal weight" (p. 55).

Conclusion: meta-analyses which are performed on a maximum of literature without distinguishing between validation and applied studies and without discarding problematic researches cannot teach us anything about the validity of the constructs and variables of a test.

3/ "Practitioners who continue to use the Comprehensive System must consult the Comprehensive System meta-analyses (Mihura et al., 2013) – and use the international Comprehensive System norms (Meyer, Erdberg, & Shaffer, 2007; Meyer, Shaffer, Erdberg, & Horn, 2015; Viglione & Giromini, 2016) – to guide their interpretations." (Mihura et al., 2018, p. 13).

The so-called International norms were produced by collecting various existing normative data established in different countries and running a particular statistical analysis (T-scores). Two conclusions could be drawn from that hocus-pocus operation: A) the resulting data (descriptive statistics) are very different from Exner's norms (lower means), and B) there are no differences between countries.

1. *Differences between Exner's norms and the Composite norms.*
   The making of the Composite sample poses problems as it amounts to adding peas, carrots and potatoes (only good for obtaining a soup). Indeed, researchers have different conceptions of what

constitutes "normality" and will recruit for their norms either "anybody" consenting to respond to the Rorschach test, or "anybody non-patient", or "high functioning persons" (i.e. people well integrated in the society, job-wise and social- wise). The last option had been chosen by John E. Exner, and only replicated in the French recruitment for norms (Sultan 2004) and approximated in the Belgian sample. Mixing so heterogenous samples cannot, whatever statistical analysis is used, yield valid normative data. A simple example can be taken in pediatrics: to evaluate the height and proper growth of a child, physicians will naturally refer to the local growth curve and not to any international average.

2. *No cross-cultural differences.*

That notion is deeply flawed, as all test practitioners know and as is recognised by the International Test Commission (ITC) (2018): "Adaptation (of a test) needs to consider the whole cultural context within which a test is to be used (...) regardless of whether there is a need for translation". Four main categories are concerned:

1. a) cultural context
2. b) technicalities of instrument development and adaptation
3. c) test administration
4. d) documentation and interpretation.

Fortunately, the Rorschach does not require to be adapted in technicalities (the same inkblots can be used), nor in test administration (universal instructions). But it does need to be culturally adapted in "documentation and interpretation". Documentation covers the data obtained with the test in a culture and leads to adjustment of the interpretative value of variables and indexes. The ITC states that when a test is used in linguistically and culturally diverse populations, "Validity studies should be conducted to ensure that an adapted version of a test measures the intended construct(s), based upon its intended purpose." (Point 2.1.1. of the guidelines (International Test Commission, 2018). Cross-cultural differences are empirically evident, and we can cite many examples. It is sufficient to check the average of pure form responses (F% or Lambda) in various countries, or the average of good form responses (XA%); the frequency of positive depression indices (DEPI > 4, DEPI > 5 ...). Another example is the list of popular responses: some unlisted responses seem to be popular in a culture (for instance the Eiffel tower in the D11 of card X in the British population, two trolls as a global response in card II in Denmark, a mask as a global response on card I in France), some listed popular responses are completely absent in certain cultures (Humans in the D3 of card IX are very rare in France, where crabs in the D1 of card X often include the D5 extension ...). Cross-cultural differences have also been statistically demonstrated by Fontan and Smith (2018). It is important to note that ethnicity of participants is not at stake, it is a matter of culture and language and not ethnicity or race.

Conclusion, the so-called "international norms" are not adapted to the different cultures and each culture should strive to find out how their non-clinical population usually respond to the Rorschach.

We salute the very important work of our Argentinian colleagues (Lunazzi, 2019) who recently released new Form Quality Tables for Argentina

Another problematic aspect of most of the normative data produced in the world is that the proportion of F in protocols, expressed either by the Lambda or by F%, is not taken into account (Gacono, Loving, and Bodholdt, 2001) Cunliffe et al., 2012)). Consequently, the core CS notion of "expectancies" is discarded. The importance of both the number of responses (R) in a protocol and the proportion of pure F responses is well known since John E. Exner showed how different are the expected results (published in the Workbook), not mentioning the differences by EB style (Extratensive/Introversive). That importance has led the R-PAS developers to create the notion and coefficient of "complexity", in an attempt to overcome the difficulty. Unfortunately, it is not based on sound scientific reasoning (Fontan & Smith, 2018).

In conclusion, normative data for a population should, in addition, present high-Lambda and low- Lambda data. We do not expect the same configuration of data and same averages in these sub- groups, a fact that is fundamental for an accurate interpretation.

4/ John E. Exner's errors in citations. Yes, they do exist!

I offer two possible explanations:
1. Exner refers to the results obtained in a study before the publication of the article and

did not check the final published version. I have a personal example pertaining to an article cited as a flawed reference by Mihura (2018): Rorschach changes following brief and short-term therapy (Exner& Andronikof, 1992). Mihura (Open Science Framework at doi:10.17605/osf.io/ fxwat/) has found that "E&A reported two D results that are not in the article". As a matter of fact, the results for the D scores were presented in the first submission of our article and curiously disappeared in the printed version.

We note that Mihura herself mentions that possible explanation "Perhaps Exner ran these analyses, and they did not make it into the final published version of the article." (Mihura, 2018, p. 12).

In general, this type of errors should have been spotted during the peer-reviewing process.

Journals, not authors, are responsible of this type of errors.

2. Wrong reports by his staff
For this possible source of error, there is only one thing to say: *Errare Humanum est.* And Mihura herself prudently writes about her own article: "it would not be surprising to find a few errors in this article" (Mihura, 2018, p. 13).

Is the R-PAS an improvement?

Mihura et al. (2018) call for a systematic review of all psychological tests: "This article makes the case that all psychological tests should be subject to the same standards of systematic review" (p.2). So let's review the R-PAS.

1. All "evidence" that the R-PAS can today produce is based on data from the CS.
It is a known fact that a change of instructions modifies the results obtained with a test. No statistical manipulation can honestly transform data obtained with a particular instruction into data that would be potentially obtained with other instructions.
Comparison of instructions. Instructions in the R-PAS: "Try to give 2 responses ... or maybe three, to each card. That is, for each card try to see two different things; possibly three. [Hand Card I in the upright orientation to the respondent.] What might this be?" (Meyer, Viglione, Mihura, Erard, Erdberg, 2011). Instructions in the CS: "What might this be?" and nothing else. In addition, strict procedures are applied during the response phase of R-PAS ensuring that the number of responses on each card is no less than 2 and no more than 4 (prompt for 2, pull after 4). In the CS, the number of responses is not set in advance, leaving the testee free to produce more or less responses to each card, although a second response is gently encouraged on the first cards, and attempts to give a sixth response is firmly discouraged. Comparing the number of responses of 123 R- PAS R-Optimized protocols to 640 selected CS protocols, the authors determined that "fewer people give a third or a fourth response to a card following the R-Optimized guidelines" (p. 299 of the Manual). That seems to indicate that the R-PAS procedure tends to inhibit the production of responses.

Finally, the Normative Reference data of the R-PAS were produced through a 2-step statistical manipulation: first, the authors "selected" (the criteria for that selection are not indicated) "up to 100 records" from 15 international CS samples (N = 1396). Second, the authors "statistically modeled what an R-Optimized administration **would look like** (bold added) on the International 1396 Adult Reference Sample" (p. 299-300 of the Manual) through a complicated series of operations on the data.

Conclusion: R-PAS "normative reference data" is not based on actual protocols of real people.

2. The core factor of the R-PAS, Complexity, is akin to a tautology (i.e. flawed logic). Fontan (2018) convincingly showed that the Complexity factor, although a good idea, is not based on a sound scientific reasoning.

3. John E. Exner insisted on the necessity to keep calculations as simple as possible, to ensure that all practitioners would be able to establish the Structural Summary (protocol level) of their clients by themselves, either by hand or through a free computer program. He generously distributed the program he had devised, and nowadays anyone can download for free the CHESSSS program (Fontan, 2013). Conversely, the complexity of the R-PAS system does not easily allow the conversion from response-level to protocol level, therefore psychologists are required to send their scored protocols to the R-PAS website, which poses ethical problems in addition to costing money.

4. Automatic interpretation.
Always testing all possibilities, John E. Exner had tried to produce a computerized program for helping interpretation (RIAP) but, observing the resulting departure from the spirit of the CS, renounced the modality, and wrote (Exner, 2005a): "While computer programs can be of assistance to those using the Rorschach, the outputs from those programs can easily be abused if the user proceeds on the naive assumption that each statement generated by the interpretative segment of the program is valid and complete. [...] the computer cannot deviate from its assigned rounds. It cannot think, and it cannot integrate data at a level higher than that for which it has been programmed. The complexity and uniqueness of each human makes it essentially impossible for any program to be developed that would account for all of the idiosyncratic features that mark the individual, but the competent human interpreter can usually do this. [...] excessive reliance on interpretative programs is bad psychology and simply reflects a sort of naivety or carelessness by the program user and ultimately does a grave disservice to clients and the profession."

What future for the CS?

Considering that a) the validity of the CS has a solid evidence-base which has not been disproved, b) the attempt to discredit its creator is produced by promoters of a competing system, c) the R-PAS, although founded on Exner's work, deviates from and betrays Exner's interpretation principles and is therefore not a continuation or development of the CS, d) many clinicians and researchers in the world do not find the necessity to abandon the Rorschach CS, we can serenely contemplate the future of the CS.

We have two sources of information to outline the future of the CS: on one hand, an article by Exner which can be considered as his testament, on the other, the updating work performed by an international association (Csira-Arisi).

A. John E. Exner's testament.

In 2001, Exner wrote an article for a Portuguese journal, fortunately in English, which can nowadays be considered as a testament and action plan for the future.

Starting by reaffirming the foremost characteristic of the Rorschach as a test, i.e. its ability to capture the singularity of a person "one of its marvels is, indeed, the personal picture of the individual that can be derived from its yield (.. .)" (p. 10), Exner denounces the artificial search for "generalized truths":

"Mythical classes of humans are created that tend to ignore the individuality that marks each human being" (p. 9); "It is, after all, indisputable, that the human being is a very unique creation and, as such, behaves throughout his or her particular life span in a distinctive fashion that reflects that uniqueness." (p. 8).

Exner recognizes the difficulty of dealing scientifically with individual differences but deplores the current neglect of this dimension and the concrete use of statistics:

"An increasingly large number of researchers (. . .) embraced the fact that the issue of individual differences might be minimized or ignored by drawing conclusions based on laws of probability and the use of

sometimes questionable estimation levels of significance. Theorists naturally have tended to side-step the issue of individual differences by resorting to esoteric generalizations." (p. 9).

The largest part of the article is dedicated to the future of the Rorschach and Exner draws a research plan for future generations in four points:

1. Broadening information about the nature of the test: "It seems realistic to suggest that the future of the test is in jeopardy of some stagnation unless the matter of the blot characteristics and the response process is addressed more aggressively than has been the case." (p. 11).
2. Further researching the variables known to relate to various personality features or organizations. In this area, Exner stresses the importance of considering the variables "in light of response styles," and neglect of that necessity is "a costly error because the baseline frequencies, proportions, and so-called average ranges for many variables differ quite substantially across the three categories." (p. 11). Exner strongly disavows researchers' efforts to smooth out differences by broadly mixing heterogeneous samples: "(...) even when sophisticated procedures have been used, the findings may be very question-able because the data sets have not been subjected to partialing for response styles. [This] has led to overgeneralized conclusions (...) which in turn have been passed along for use by interpreters." (p. 13).
3. Searching "new directions with regard to features of personality and psychological functioning" (note the constant referral to the difference between personality and functioning), either known traits and features that are not yet identified in the Rorschach, or future conceptualizations of personality and functioning.
4. The fourth area of investigation is obviously, for Exner, the most important sector of future research, which, alone, could ensure the survival and success of the Rorschach. It is the relationship of the Rorschach to treatment and can be summarized in one question: Does pretreatment assessment contribute significantly to treatment selection or therapeutic outcome? Exner writes that research on the contribution of Rorschach to pretreatment assessment "can demonstrate its own credibility probably better than any other way."

B. Csira/Arisi
The Comprehensive System International Association (www.csira-arisi.org) is an active scientific society dedicated to the development of the CS. Clinical studies and researches from its members (from all continents) are presented in Csira congresses (every two years) and in many other meetings, congresses, workshops every year. Its membership is constantly growing, a testimony to the ever-living interest and use of the system as well as to the perceived need to upgrade and develop it.

In conclusion

The 2018 article by Mihura & al. is a shame: a shame for the authors whose purpose was to negate the scientific foundations of the Comprehensive System and cast a doubt on the intellectual honesty of its creator John E. Exner, meanwhile developing a system entirely based on the CS; a shame for the Journal which accepted to publish it, thus becoming accomplices in a destructive plot. And a shame for the scientific community which absorbs without discernment this kind of "fake truths".

The Comprehensive System is a sound and ever-promising system for coding and interpreting the Rorschach test, still developed by an international community of researchers and clinicians. Its future depends on our willingness and capacity to develop it according to John Exner's recommendations.