

7-7-2014

Using the R-PAS' Aggressive Content Score For the Evaluation of Aggressive Behaviors in Children

Rebecca Dehass

Follow this and additional works at: https://digitalcommons.du.edu/capstone_masters



Part of the [Psychology Commons](#)

Recommended Citation

Dehass, Rebecca, "Using the R-PAS' Aggressive Content Score For the Evaluation of Aggressive Behaviors in Children" (2014).
Doctoral Papers and Masters Projects. 58.
https://digitalcommons.du.edu/capstone_masters/58

This Doctoral Capstone is brought to you for free and open access by the Graduate School of Professional Psychology at Digital Commons @ DU. It has been accepted for inclusion in Doctoral Papers and Masters Projects by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu.

Using the R-PAS' Aggressive Content Score

For the Evaluation of

Aggressive Behaviors in Children

A DOCTORAL PAPER

PRESENTED TO THE FACULTY OF THE

GRADUATE SCHOOL OF PROFESSIONAL PSYCHOLOGY

OFFICE OF GRADUATE STUDIES

UNIVERSITY OF DENVER

IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PSYCHOLOGY

BY

REBECCA DEHASS, M.A.

JULY 7, 2014

APPROVED: _____
E. Hale Martin, Ph.D., Chair

Courtney Welton-Mitchell, L.P.C., Ph.D.

Michele J. Forstot Nadel, Psy.D.

Abstract

The Rorschach Performance Assessment System (R-PAS) was developed in 2011 as an alternative to the previous Comprehensive System. The goal was to improve the psychometrics, and particularly the validity, of this assessment method. The norms for children were questionable in the Comprehensive system (e.g., outdated, low numbers of subjects) and validity studies for children were sparse. One of the indicators included in the R-PAS system, the aggressive content indicator (AgC), is intended to reflect aggressive behavior, but few studies have examined the validity of this indicator. This study examined the validity of AgC in a sample of 32 children and adolescents receiving services at a residential treatment center. Subjects' AgC scores were analyzed in relation to demographics and diagnosis, as well as ratings of aggression and conduct problems from the Behavioral Assessment System for Children-2 (BASC-2) Parent and Teacher Reports. Correlations between the AgC score and BASC-2 aggression and conduct problems scores were not statistically significant. None of the correlations between AgC score and a diagnosis of Conduct Disorder, Oppositional Defiant Disorder, Posttraumatic Stress Disorder, or Mood Disorders were significant either. Given the small sample size, null results may be a result of power concerns. The lack of significant correlations may however, indicate that operational definitions of aggression used in various forms of measurement reflect different constructs.

Keywords: Rorschach, R-PAS, BASC-2, aggression, conduct disorder, oppositional defiant disorder, posttraumatic stress disorder

The Rorschach Inkblot Method is one of the most common measures in the family of performance-based tests. Numerous studies have indicated that the Rorschach is a useful and valid measure (e.g., Hiller et al., 1999; Exner, 2003; Meloy and Gacono, 1992; Meyer, 2014; Meyer et al., 2014; Mihura et al., 2013; Erdberg, 1993). Over the years changes have been made to this test in the Rorschach Performance Assessment System (R-PAS). For instance, many changes were made to the administration and scoring guidelines in an effort to improve overall interpretability. Questions of reliability and validity remain an issue of considerable interest in the testing community. Without reliable norms and demonstrated validity, any test is limited in its use and, deservedly, subject to criticism.

The focus of this study was to examine the validity of an indicator of aggression included in the R-PAS system that has been the subject of little research in adults and virtually absent in child and adolescent samples. Aggression is an important behavior to monitor in children and adolescents as it can lead to problems with self-esteem, relationships, and future success (Meloy and Gacono, 1994; Reynolds and Kamphaus, 2004; Baity et al., 2000). These problems may lead to pervasive interpersonal dynamics in adulthood, possible legal consequences, and the victimization of others if not addressed. The aggressive content indicator (AgC) purports to reflect aggressive tendencies. Thus, it is potentially a useful indicator for both diagnosis and treatment planning and, therefore, merits considerable empirical investigation.

Some of the most prominent Rorschach researchers, Meloy and Gacono (1994), defined aggressive content (AgC) on the Rorschach as “any subject or object that would be qualified, by most people, as predatory, dangerous, malevolent, injurious or harmful” (p. 259). In their research, they were able to reach inter-rater agreements ranging from 92 to 100 percent for aggressive content scores. Such high agreement suggests strong reliability for this measure. They

further asserted that, during interpretation, it is extremely important to identify the quality, intensity and directionality of the aggressive content when choosing to code this score. There may be differences between aggression directed toward the self or others. Further, this creates a substantial spectrum of intensity in which answers fall. Both a lobster clenching something in his claw (the clenched claw being the AgC) and a tiger stalking its prey (the tiger being the AgC) both require identical content coding. The R-PAS manual (2011) also denotes common animals, weaponry, environmental forces, and creatures as falling under these criteria. In doing so, this again creates a rather broad category. Given these criteria, a hurricane, a grizzly bear, and gun all carry the same interpretive value. However, the writers were also quick to specify the use of clinical judgment. For example, a bull alone is not enough to code AgC; however, a “powerful bull” presumes a dangerous object (Meyer et al., 2011).

AgC differs from aggressive movement (AG), as it does not require actual movement or action, and only denotes an object that is representative of aggression or dangerousness. An example of AG would be one person hitting another. AgC also differs from past aggression (AgP), which includes any aggressive response that has occurred in the past. An example of AgP would be a response indicative of a person who *was* shot with a gun and is now dead. It should be noted that a response may be coded as both AG and AgC. An example of this would be a Viking attacking an enemy with a battle axe (with the axe required the AgC coding, whereas the action of attacking requires the AG coding).

Reliability and Validity Issues

Hiller et al. (1999) found that the Rorschach had an overall validity effect size of “substantial magnitude,” with an unweighted mean r of 0.29 and a weighted mean r of 0.22. This compared favorably to the MMPI-2, a highly utilized self-report personality inventory. They

noted further that these effect sizes even compare favorably to medical tests (though they did not specify which medical tests)(Hiller et al., 1999). They also mentioned that, “The Rorschach may have psychometric properties that are different from, and perhaps superior to, those of other ‘projective’ measures that are frequently mentioned in the same breath as the Rorschach” (Hiller et al., 1999, p. 289) such as the Thematic Apperception Test.

Still, concerns have been voiced over the years about the reliability and validity of indicators that emphasize content. Specifically, “with content data, issues of reliability and validity cannot be disregarded; nonetheless, such issues need not exclude or minimize the clinical richness of the material” (Erdberg, 1993, p. 139). Mihura et al. (2013) found that some of the indicators frequently used in interpretation lack adequate validity (e.g., Color Projection, Egocentricity Index, and Isolation Index). Some of the most recent research (Mihura et al., 2013) argues for dropping variables that lack support and adding variables that have greater interpretive value. As one example, Mihura et al’s work found little support for the AG score, which is a long-standing indicator in the Comprehensive System. As such, the R-PAS has added two additional aggressive scores (AgP and AgC) in order to increase the interpretive ability of aggressive answers. The current study addresses the validity of the AgC indicator.

Another challenging issue of interest for validation studies is that of inter-rater reliability. Janson (2008) presented numerous options for calculating and reporting this type of reliability. A noteworthy point that is largely unmentioned in other current research is that this agreement is contingent not only upon the skill of the coders, but it also hinges significantly upon the protocols themselves (i.e., the subject’s responses and the evaluator). The sample population and variance within the subjects (i.e., clinical versus nonclinical subjects, age, gender, etc.) can also greatly influence reliability (Janson, 2008). In order to observe both the agreement between

coders and the reliability (thus, accounting for potential random agreement), Janson encouraged the use of Cohen's *kappa*.

With specific regard to the latest scoring system, the Rorschach Performance Assessment System (R-PAS), Viglione et al. (2012) found an overall mean *kappa* of 0.88 and a median of 0.92, both of which indicate strong reliability of the measure. They also found that *kappa* for AgC was 0.79, which is indicative of strong reliability (Viglione et al., 2002). As a result, Cohen's *kappa* was used in this study.

R-PAS

In the early 1990s, leading psychologists in the Rorschach community posited that the copious amounts of data regarding various populations, diagnoses, and treatment collected over the years would lead to an even more impressive interpretation if implemented in a new publication of this measure's administration, scoring, and interpretation guidelines. Some researchers have noted that, at best, Rorschach interpretations using the Comprehensive System are correlated only moderately with self-report data (Meyer & Viglione, 2011). As such, the validity of scoring criteria and the subsequent interpretations have presented substantial concerns throughout the testing community.

Between 1997 and 2006, Exner's Research Council for the Comprehensive System worked to extend the research regarding Rorschach administration, scoring, and interpretation. Four of the council's members embarked on a path that led to the creation of a new system, the Rorschach Performance Assessment System (R-PAS). While normative data for the R-PAS exists for adults, developing adequate normative data for children has been a slower process (Meyer & Viglione, 2011).

Of course, issues of validity have remained a focal point of normative data for the Rorschach as the years have passed, and further research has demonstrated that there are specific needs for improvement in the interpretive value of many scores. This is potentially most important for child and adolescent norms. It appears that the Rorschach interpretation guidelines tend to conceptualize symptomology or behaviors in a static fashion, such that the fluidity of childhood development and behavior is not taken into account. Most recently, Meyer et al. (2014) acknowledged that the Comprehensive System's normative data for children tends to over-pathologize responses, which is one of the issues that the R-PAS was designed to rectify. Specifically, there is a distinct difference between profiles of children whose protocols were scored using the original normative data and those using the contemporary norms (both published by Exner; Meyer et al., 2014). The Comprehensive norms tended to make children appear more "disturbed," less resourceful, and less complex (Meyer et al., 2014). Further, they noted large variability across samples of children, even within the same age range (Meyer et al., 2014). This is not found to be the case with data collected for adults, despite samples from different age ranges, cultures and countries (Meyer et al., 2014).

In an effort to address this dilemma, Meyer et al. (2014) used data from an international sample to create three age ranges: five to eight, nine to twelve, and thirteen to eighteen. Further, the R-PAS online scoring program provides an interpretive overlay (based on specific age), which shows where a non-patient child of a given age would be expected to score (Meyer et al., 2014); from this, a percentile can be identified and then converted to a standard score (with a plus or minus fifteen point range; Meyer et al., 2014). It should be noted that the authors have referred to these norms as "transitional" or "provisional."

Of distinct interest to this paper is the introduction of several new aggressive scores (Aggressive Content and Aggressive Past) that are being used as part of the new system, in addition to the Aggressive Movement (AG) indicator, which has remained from the prior scoring system. Previously, AG was the only indicator of potential aggression; however, its interpretive meaning is more general, suggesting both possible aggression and poor interpersonal skills (Liebman et al., 2005). For this study, only the AgC score will be used given that past research has shown it is the only new aggressive score that has strong validity (Meloy & Gacono, 1992). Moreover, this study is specifically concerned with the potential for aggression or hostility, as opposed to more general poor capacity for relationships.

According to Gacono et al. (2005), early Rorschach research posited that both direct and covert AgC were indicative of aggressive impulses (2005). A direct AgC response would be “an angry grizzly bear,” suggesting an angry predator; a less obvious AgC would include “a lobster claw that is clenched,” suggesting tension and a sharp animal body part. Frequently, AgC may be coded with Aggressive Movement. Generally, however, studies have shown that increases in AgC scores were related more strongly to reports of highly aggressive behavior, as opposed to other variables currently used in Exner’s Comprehensive System (Gacono et al., 2005). Several studies have been conducted in order to assess the interpretive relevancy of the many aggression-related scores. As noted previously, the variable found to carry the most weight is the AgC score, perhaps because it is the variable that occurs most frequently in the aggressive indicators (Gacono et al., 2005).

Liberman et al. (2005) found that, of the various aggressive scores, only the AgC score predicted both aggressive potential and aggressive, violent behavior. Baity et al. (2000) found that aggressive movement and AgC were both suggestive of active aggression, verbal and

physical. Additionally, they noted that there was a significant overlap between the two variables. They concluded further that AgC is a reliable measure, even when untrained raters reviewed protocols (Baity et al., 2000).

Please see the Appendix (*Figure 1*) for the normative data currently available for AgC scores for children and adolescents (e-mail communication with Meyer, 2014).

Gender Differences in Aggressive Scores

One concern regarding aggressive scores (AgC, AG, and AgP) is the possibility that female and male children may demonstrate differences in their aggressive responses. Little research has examined gender differences in aggressive content using Exner's Comprehensive System, and no research has been conducted using the R-PAS with children and adolescents. Gacono and Meloy (1994) tested the hypothesis that gender differences exist. They asserted that even female psychopaths had fewer aggressive responses than males due to gender socialization regarding aggressive expression, despite how aggressive they behaved in "real life" (it should be noted that they only measured Aggressive Movement in the study). When comparing psychopathic and nonpsychopathic individuals, they found no significant differences in terms of aggressive responses (regardless of gender), despite real life differences in levels of physical violence (Gacono and Meloy, 1994, p. 263). Ultimately, they concluded that the "Various Rorschach indices of aggression appear less promising for the nomothetic comparison of different groups than for the idiographic understanding of the quality, intensity, and directions of aggressive impulses for a particular individual" (Gacono & Meloy, 1994, p. 273).

With specific regard to AgC, some of the factors that may need to be considered by the administrator include defensiveness, social awareness, and the face validity of AgC (Gacono & Meloy, 1994, 273.). When a client has a high Lambda score, the validity of the overall score is

qualified because the client answered in a defended (i.e., non-disclosing) manner. Rather, the interpretation likely underestimates the level of maladaptive behavior. It is not uncommon for individuals to be non-disclosing regarding all of what they see when being administered the Rorschach, particularly if there is a significant reason for the testing (i.e., forensic evaluations). For example, a truly aggressive, or even psychopathic, individual may know better than to offer answers with aggressive or violent content because this may reflect poorly upon his psychological functioning.

Assessing Problematic Behaviors in Children

Currently, one common way of measuring overt aggressive behavior in children and adolescents is the *Behavioral Assessment Symptom Checklist–Second Edition (BASC-2)*. This measure, which collects data from various informants (i.e., parents, teachers and self), has moderate to good reliability and validity (Kamphaus et al., 2011); as such, it is the additional measurement used in this study.

Specifically, the Aggression and Conduct Disorder scale scores were selected as potential covariates, as they capture the behaviors also suggested by the AgC indicator. The Aggression Scale of the *BASC-2* includes ratings on teasing, bullying, and hitting others. Examples of behaviors specified on the Conduct Problems scale include stealing, breaking rules, and disobeying orders. These, among the remaining symptom scales, measure the level of problematic behaviors as compared with same-age peers. T-scores between 41 and 59 fall in the Average range, while scores from 60-69 indicate the At-Risk range (i.e., potentially problematic behavior), and scores above 70 denote the Clinically Significant range (i.e., “maladaptive” behavior; Reynolds & Kamphaus, 2004).

Reynolds and Kamphaus (2004) noted that there is a strong inter-correlation between the Aggression and Conduct Problems ratings. For Teacher Ratings, the correlation is 0.88 for Aggression and Conduct Problems. For Parent Ratings, the correlation is somewhat lower, but at 0.74, still considered to be strong. The correlation between Teacher and Parent ratings of Aggression and Conduct Problems are both moderate, at 0.44 and 0.49, respectively. Reynolds and Kamphaus identified some differences in the context of these behaviors across environments, which may account for this inter-rater correlation being lower than intra-rater correlations (Reynolds & Kamphaus, 2004).

Goals of Current Research

This research will examine the validity of the AgC score. The AgC will be the only aggressive score studied here because AgC scores have been documented as occurring more frequently than other aggressive scores (Meloy and Gacono, 1992). Given the research available to date regarding aggressive scores, it is likely that AgC scores will show at least a moderate correlation with aggressive behavior, because it is based on the potential for physical aggression and hostility. Moreover, given that the clinical sample of protocols used in this study comes from children in residential treatment for behavioral problems (often aggressive and hostile during many interactions with others), AgC scores will likely be elevated. It is also hypothesized that AgC scores will correlate more highly with Conduct Disorder and Oppositional Defiant Disorder diagnoses than with other diagnoses because these diagnoses incorporate aggression and hostility. A final hypothesis is that, given socialization differences, males will offer more AgC responses than females as the AgC score is indicative of physical aggression which males produce more of than females. Though females are often hostile, this is usually more covert through social aggression.

Methods

Research Approval

This study was approved by the University of Denver Institutional Review Board (IRB), as well as the agency's IRB, from which the subject sample was drawn.

Participants

The subjects of the study were thirty-two adolescents (13 females, 19 males; age $M = 14.21$; age range 11-17 years) with previous psychological testing that included cognitive, self-report, and projective measures. Participants' parents or guardians signed a waiver of confidentiality upon permitting their child's participation in psychological testing. This waiver acknowledged that the child's psychological testing data could be used for the purposes of research at the discretion of the agency.

Subjects were identified by searching through all cases completed at the agency between 2001 and 2013 and selecting all cases that had a Rorschach (administered using Exner's Comprehensive System administration instructions) as well as both Teacher and Parent BASC-2 reports (to address the need for both observational data and projective data within the sample). Fifty subjects in total were available but 18 were discarded due to incomplete information regarding diagnosis, available BASC-2 results, or Rorschach protocols. After all available subjects were reviewed, only 32 cases that met these specific criteria.

Procedure

Statistics for the Rorschach protocols and the BASC-2 results were calculated using all cases due to the limited number of subjects. No outliers were removed from the sample, which may have impacted the overall descriptive statistics of this study. This decision was based upon

the small sample size and the decision to therefore use all available data meeting the above listed criteria. Protocol lengths for the Rorschach ranged from 14 to 41 responses. The mean number of responses was 20.89. The average number of aggressive scores was 3.48. The range of aggressive content scores per protocol was 0 to 11. BASC-2 T-scores were taken from previously scored protocols for these subjects. Of the 32 subjects, the mean T-score for Aggression per parent rating was 67.75, with a range of 43-94. The mean T score for Aggression per teacher rating was 62.18, with a range of 41-81. The mean T-score for Conduct Disorder per parent rating was 74.28, with a range from 39-104. The mean T-score for Conduct Disorder per teacher rating was 61.13, with a range from 46-83.

This study originally recruited five graduate students to code responses. Although all five were given the same directions for scoring, their coding was drastically different. The inter-rater reliability was so poor ($k = 0.00$) that the study was reconstructed two different times in order to reach inter-rater reliability that was adequate (ultimately resulting in the use of only two coders).

Two coders, advanced level doctoral students, were given unidentified Rorschach protocols for the 32 subjects. They were provided with no information about the subjects regarding demographics or diagnosis. The coders were asked to review the protocols and score responses for AgC only. Although these protocols had been scored previously as part of the subjects' psychological evaluations, the results were not made known to the coders. They were given directions regarding AgC scoring from the R-PAS manual; however, it is important to note that the coders had not previously received formal training regarding the R-PAS system. Once an AgC count was established for each subject, the score was then paired with the subject's BASC-2 scores and demographics. The overall raters' scoring only differed by one AgC response; therefore, neither coder was thought to over- or under-code responses.

Results

Inter-rater Agreement of AgC coding

Inter-rater agreement for AgC coding was calculated with the *kappa* statistic. In this case, a *kappa* of 0.89 was obtained, which is indicative of substantial agreement (*See Appendix: Figure 2*).

Inter-rater Agreement of BASC-2 ratings

With regard to the reliability of BASC- 2 Parent and Teacher raters for this study, the inter-correlation of Parent Ratings of Aggression and Conduct Problems was 0.82, higher than the inter-correlation reported for normative data (0.74) by Reynolds and Kamphaus (2004). The current study also showed an inter-correlation of Teacher Ratings of Aggression and Conduct Problems as 0.86, just slightly lower than that of the normative data (0.88) reported by Reynolds and Kamphaus.

With regard to Parent and Teacher ratings of Aggression and Conduct Problems, the normative sample showed 0.44 and 0.49 correlations, respectively (Reynolds & Kamphaus, 2004). In the current study, the correlation between Parent and Teachers ratings was 0.55 for Aggression and 0.42 for Conduct Problems. Thus, it appeared that the reliability of this measure matched or exceeded that of the normative sample reported by Reynolds and Kamphaus (See Table below).

		TEACHER AGGRESSION	PARENT AGGRESSION	TEACHER CONDUCT	PARENT CONDUCT
TEACHER AGGRESSION	Pearson Correlation	1	0.56	0.86	0.49
	Sig. (2-tailed)		0.001	0.00	0.004
	N	32	32	32	32
PARENT AGGRESSION	Pearson Correlation	0.56	1	0.45	0.82
	Sig. (2-tailed)	0.001		0.01	0.00

	N	32	33	32	33
TEACHER CONDUCT	Pearson Correlation	0.86	0.45	1	0.42
	Sig. (2-tailed)	0.00	0.01		0.02
	N	32	32	33	32
PARENT CONDUCT	Pearson Correlation	0.49	0.82	0.42	1
	Sig. (2-tailed)	0.004	0.00	0.02	
	N	32	33	32	33

Correlations

After the AgC was scored for all protocols by each rater independently, the mean of both raters' numbers of AgC was calculated. The average number of AgC scores was 3.48 (SD = 2.52), with the range of AgC scores falling between 0 to 11. The AgC average score was then compared to a two demographic variables (i.e., age and gender). Correlations were computed using Pearson's r to identify the relationship between AgC and each demographic variable. The results are as follows: AgC and Age ($r = -0.05, p > 0.05$), as well as Gender ($r = 0.15, p > 0.05$; see Appendix: Figure 3). There were no significant findings with regard to age or gender.

Of greatest interest to this study was the comparison of the AgC score and observed aggressive behavior (as evidenced by BASC-2 scores). Comparing these two variables demonstrated the relationship between projective answers suggestive of aggressive potential and observations of actual aggression. AgC scores were first compared with the BASC-2 Parent Report score for the Aggression scale. The correlation was not significant ($r = -0.03, p > 0.05$). The correlation between AgC scores and the Parent Report of Conduct Problems ($r = -0.07, p > 0.05$) was also not significant (see Figure 1). There were also no significant correlations between AgC and the Teacher Report of Aggression ($r = 0.09, p > 0.05$) or between AgC scores and the Teacher Report of Conduct Problems ($r = 0.12, p > 0.05$) (see Figure 2 below).

Figure 1: Correlations between AgC and Parent Ratings of Aggression and Conduct Problems

		AGG CONTENT	PARENT AGRESSION	PARENT CONDUCT
AGG CONTENT	Pearson Correlation	1	-0.03	-0.07
	Sig. (2-tailed)		0.86	0.70
	N	32	32	32
PARENT AGRESSION	Pearson Correlation	-0.03	1	0.82
	Sig. (2-tailed)	0.86		0.00
	N	32	33	33
PARENT CONDUCT	Pearson Correlation	-0.07	0.82	1
	Sig. (2-tailed)	0.71	0.00	
	N	32	33	33

Figure 2: Correlations between AgC and Teacher Ratings of Aggression and Conduct Problems

		AGG CONTENT	TEACHER AGGRESSION N	TEACHER CONDUCT
	Sig. (2-tailed)		0.61	0.53
	N	32	32	32
TEACHER AGGRESSION	Pearson Correlation	0.09	1	0.86
	Sig. (2-tailed)	0.61		0.00
	N	32	32	32
TEACHER CONDUCT	Pearson Correlation	0.12	0.86	1
	Sig. (2-tailed)	0.53	0.00	
	N	32	32	33

Additional correlations were computed between AgC and the following diagnoses (which were determined based on all clinical interviews with child and caregiver, as well as the measures used in this study): Posttraumatic Stress Disorder ($r = -0.08, p > 0.05$); Conduct

Disorder ($r = 0.17, p > 0.05$); Oppositional Defiant Disorder ($r = -0.002, p > 0.05$); and Mood Disorder ($r = -0.08, p > 0.05$). None of the correlations was significant (see Appendix: Figure 4).

Power and Effect Sizes

The effect size was computed for each correlation and the power observed for each correlation was also calculated (see Figure 3). In order to do so, the data were subjected to a univariate ANOVA. Of these, the relationship between AgC scores and Parent ratings of Conduct Problems had the highest effect size (as measured by *Cohen's d*) at 0.95 ($F_{26,5} = 3.36, p = 0.09$). The observed power of this correlation was 0.55, which is considered moderate.

Figure 3: Effect sizes for all variables compared to AgC

<i>Variables</i>	<i>F</i>	<i>Sig.</i>	<i>Partial Eta Squared</i>	<i>Observed Power</i>
<i>PRS Aggression</i>	0.42	0.95	0.42	0.16
<i>PRS Conduct</i>	3.36	0.09	0.95	0.55
<i>TRS Aggression</i>	0.51	0.91	0.45	0.17
<i>TRS Conduct</i>	0.94	0.57	0.63	0.30
<i>Age</i>	0.26	0.95	0.06	0.11
<i>Gender</i>	0.65	0.43	0.02	0.12
<i>PTSD</i>	0.21	0.65	0.01	0.07
<i>MOOD</i>	0.21	0.65	0.01	0.07
<i>ODD</i>	0.00	0.99	0.00	0.05
<i>Conduct Disorder</i>	0.90	0.35	0.03	0.15

Sixteen (50%) of the subjects were diagnosed with Conduct Disorder and an additional two (6%) were diagnosed with Oppositional Defiant Disorder. Per the DSM-IV-TR diagnostic criteria, a significant correlation should have existed because aggression and hostility are both

often characteristics of these two disorders. Of the eight subjects with an AgC score of five or higher, four had a diagnosis of Conduct Disorder or Oppositional Defiant Disorder. Further, two of the subjects with Conduct Disorder had AgC scores of zero.

The diagnosis of PTSD was included in this study as a point of comparison because such a high amount of the subjects both exhibited aggressive behavior and were given this diagnosis. However, it is important to note that this diagnosis is not characterized by aggression or hostility; rather, the impact of trauma is often manifested through anxious symptoms. Sixteen (50%) of the subjects had a diagnosis of PTSD. Of these, three of the subjects with elevated AgC scores had a PTSD diagnosis, with one being the subject with the highest AgC score (11). While the effect size and power were negligible, it is believed that this is due to the small sample size. It could also be the case that even with a more robust sample, the same pattern would be noted.

Of the 32 subjects, 24 presented with a mood disorder. Again, there was little evidence of any relationship between the AgC scores and mood disorder. This diagnostic category is included for discriminant validity—a correlation between mood and aggression would not be expected since aggression is not a diagnostic symptom of mood disorders in the Diagnostic and Statistical Manual—Fifth Edition (DSM-5). In future research, it may be beneficial to separate unipolar mood disorders from bipolar mood disorders to assess whether the presence of any AgC is related to one or the other. However, no hypothesis posited the presence of any significant correlation, regardless of type, and thus this study did not go to lengths to analyze them separately.

As previously noted, there were 13 female and 19 male subjects. While no significant correlation existed between gender and AgC, the data did provide an interesting spread. Notably, the subjects with the two highest AgC scores were female. Although the R-PAS does not suggest

different norms for female and male children, It was posited that there would be a gender difference. Specifically, it was thought that males would have higher AgC scores due to gender socialization. In future studies, it may be beneficial to assess whether any differences exist between the AgC responses given by females and males.

Discussion

Despite non-significant findings, the results of this study offer much to consider with multiple intriguing points. Meyer (2014) was gracious enough to provide the normative data for comparison to the data collected in this study. In the normative sample, the mean AgC for children and adolescents is between 3.10 and 3.13, depending upon age. In this study, the mean AgC for the adolescents was 3.48, although there were notable outliers. Thus, the mean score in this study did not present a meaningful difference from that of the normative data (i.e., the mean was not a standard deviation above the normative mean). This is surprising, given that at least half of the subjects were diagnosed with conduct disorder or oppositional defiant disorder. Even more interesting is that the agency from which these subjects' data were obtained includes a high proportion of children and adolescents who exhibit aggressive behaviors. Further, of the 32 subjects, only eight had AgC responses that exceeded the normative range (i.e., indicative of aggression or aggressive potential). Given the population, the AgC indicator does not appear to bear a strong predictive value to obvious aggressive behaviors. That is, given that it is meant to signify the presence of hostility, anger, and aggressiveness, the mean AgC for these children did not produce a correlation with BASC-2 reports. Rather, it appears that the BASC-2 may be a better indicator of aggressiveness than the AgC score because it addresses a variety of aggressive behaviors.

Reliability

It is important to recognize that the inter-rater reliability in this study was strong. Of the 32 protocols, the two coders differed only on three responses across all protocols. Of those three, differences in scoring occurred for responses where it was unclear whether the content exceeded the threshold to be coded as AgC. For example, coders disagreed about whether an answer of “an alligator” met the criteria. One coder noted that “alligator” was similar enough to “crocodile” to be qualified as an AgC response, whereas the other did not think that this was the case because nothing aggressive was explicitly stated about the “alligator.”

Overall, coders reported that there were very few responses (such as “alligator”) that were not listed in the R-PAS manual for AgC. While the inter-rater agreement was by and large acceptable, disagreement may still occur in the general testing community. Although many clear examples are given for AgC, as well as examples of what does and does not exceed the threshold, there is potential for clinical judgment and, thus, error may occur.

The original design for this study illustrates the issue of clinical judgment. This study initially recruited five graduate students to code responses. Although all five were given the same directions for scoring, their coding was drastically different. The inter-rater reliability was so poor ($k = 0.00$) that the study was reconstructed two different times in order to reach inter-rater reliability that was adequate (ultimately resulting in the use of only two coders). It is noteworthy that some coders significantly over-coded for AgC, and others severely under-coded. This haphazard pattern provided clear evidence that the AgC scoring process involves sufficient room for error.

In using the AgC, it will be important for clinicians to understand and adhere to the scoring criteria carefully. It is recommended that all professionals using the R-PAS attend a formal training session in order to best understand and use the measure. Moreover, the need for

supervision and consultation cannot be overemphasized in order to ensure that a high level of reliability is reached. With regard to training programs, it will be necessary for additional time to be spent establishing the bounds of clinical judgment and creating a threshold for answers that are not included in the R-PAS manual.

Further, the sample used in this study involved a population in which aggressive and problematic behaviors had already been identified by numerous professionals and caregivers, resulting in placement in a residential or day treatment program. Given that most of the subjects exhibited pre-existing problems with aggression, this may also explain why the BASC-2 reports suggested higher rates of both aggression and conduct problems. Of the current study BASC-2 ratings from parents and teachers, the reliability within and across raters matched or exceeded that of the normative data provided by Reynolds and Kamphaus (2004). This information, combined with the high *kappa* in this study, provides the most significant evidence that the BASC-2 is a better measure of aggressiveness for children and adolescents than the AgC indicator. Moreover, the results suggest that observational information is likely a more accurate measure of overt behaviors. It is important to note, however, that the AgC score is only indicative of aggressive potential, which is subtler and not as easily observed. Thus, the results of this study do not indicate that the AgC is an unreliable measure of aggressive behavior; rather, it highlights it is important for clinicians understand what the AgC is meant to measure and that it is not over-interpreted.

Validity

Despite the solid reliability obtained in this study, there is evidence that this indicator may fall short of its mark as a valid measure of aggression. As previously noted, it is possible that subjects moderated their Rorschach responses due to the face validity of AgC. Given this, it

seems reasonable that the interpretive value of one's AgC score should be compared with self-report measures and other observations. This relationship was not found in this study. Despite parent and teacher ratings that suggest frequent aggressive and conduct disordered behavior, AgC scores did not capture this same information. As such, it calls the validity of this measure into question.

Aggressiveness can also be a state as well as a trait. In general, the BASC-2 asks questions about specific behaviors more likely thought of as traits than states. Behaviors (e.g., aggression) may be context dependent, and thus, considered a state rather than a trait. It is unclear whether AgC captures aggression specifically as an enduring trait or as a current state. Currently, it does not appear to differentiate between state and trait. However, the results of this study suggest that, regardless of whether aggression is a state or trait characteristic, AgC did not capture the same information as that provided by the BASC-2.

Despite the limitations of this study, it is interesting that there may be a difference in the operational definition of aggression used for the BASC-2 and R-PAS. The R-PAS interpretation of AgC appears to be linked more to the *potential* for hostility, anger, and aggressiveness, whereas the BASC-2 scores identify aggressive behavior patterns that have already developed or been observed. Although a high correlation was expected between these separate measures, this was not the case. Most parsimoniously, the results suggest that the AgC score does not measure overt aggression, while the BASC-2 does.

Possible alternative explanations

The current sample is composed of children with highly problematic behaviors who were in need of a day treatment level of therapeutic intervention, at a minimum; most of the subjects were residential treatment patients. The major difference between the normative sample and the

sample used in this study was that this sample was under therapeutic scrutiny, while the normative sample was not. As such, the normative sample may have felt freer to report aggressive content without concern for impression management. However, if this were the case, it seems unlikely that this study sample would be highly skilled at impression management. Nevertheless, content scores undoubtedly have greater face validity than other Rorschach scores and thus individuals may be less willing to acknowledge content that may carry a negative connotation.

Perhaps, most importantly, it is necessary to consider that the normative sample was derived from an international population including 17 countries. This presents the potential for cultural differences that could affect the normative information. There may be international differences in AgC that could detract from the usefulness of these norms for American children and such differences may have skewed the results of this study. The applicability of diverse international norms to American children (and adults) is a significant concern among assessors (Meyer et al., 2014). Thus, gathering norms in an exclusively American sample would be important and is currently being pursued (Meyer et al., 2014). Because AgC is not an indicator on the Comprehensive System, there are no data to provide any guidance about this concern.

Quality of the AgC responses

Finally, an area of further study that may be useful involves assessing the quality of AgC scores. For example, there may be differences between a child and adolescent who see answers involving lobsters clenching something with their claws versus an angry monster attacking someone. Both exceed the AgC threshold (i.e., are worthy of an AgC coding), but they convey different levels of aggressiveness. While the R-PAS manual details the importance of using context when choosing whether to code an AgC, there are certain answers in which the context

itself would not indicate the necessity for coding it (e.g., the lobsters clenching something in their claws). Further, it might also be important to look at the differences between aggressive content that includes humans versus aggressive content related to animals. Per Exner, pure human content is indicative of interpersonal interest (2003). However, if the content includes a fantasy, fictional character or a partial human response (i.e., a person's hands), it suggests unrealistic perceptions of relationships (Exner, 2003). Similarly, partial or fictional animal content may indicate detachment from reality (Exner, 2003). As such, there may be a specific difference between fantastical and realistic aggressive content. For example, a fire-breathing dragon about to attack may demonstrate something different than a man pointing a gun at someone.

We must keep in mind that responsible Rorschach interpretation considers co-occurring indicators. Thus, answers with AgC that also include poor form quality, or special scores indicative of poor logical or human representations, may have different interpretive significance than AgC occurring without other indicators of concern. With AgC, as with any other indicator, it is important to compare and weave this information with other indicators in order to reach an overall conclusion regarding an individual's functioning.

Of further note is the fact that it may be important to acknowledge the child's individual interests and whether or not these may influence the content given. For example, it would be likely that a child with a fondness for *Harry Potter* might see a Dementor because he has simply read the book series recently. This may also be the case for individuals who take karate classes and, therefore, see ninjas with swords on a given card. As such, it is important that clinicians take care to understand the child's current interests, activities, and environment in order to avoid over-pathologizing answers.

Future Research

This study warrants replication with at least several improvements in design. The current study used two graduate students to score AgC in the study protocols independently. The methodology might be improved by using multiple raters who have had substantial experience scoring Rorschach protocols, including the use of AgC. Additionally, differences in coding should be examined closely to evaluate the root cause of discrepancies, as this might provide important information for potential revisions to the scoring criteria.

It is also recommended that five to ten coders be used in future studies. If strong agreement can be achieved between a large number of clinicians, this adds further evidence for the solid reliability of the AgC score. Moreover, it would be helpful to use licensed psychologists who have attending formal R-PAS training. Again, if high agreement is reached, this again provides merit to the reliability of the measure. Additionally, the number of subjects should be increased, potentially with a minimum of 100 protocols used. In the current study, the sample size was small and therefore may have skewed the results. To further assess validity, all aggressive scores should be coded (AG, AgC, and AgP) and then compared with BASC-2 Parent and Teacher Report scores. It may also be helpful to use the BASC-2 Self Report as well, to provide information regarding how the adolescent sees his or her own behavior.

Finally, a potential confound that might be controlled during future work involves the differences in the types of AgC responses given by female and male children. This study had 19 males and 13 females, and although no significant differences were found between these groups, a plethora of research suggests that differences exist. As previously noted, Gacono and Meloy (1994) asserted that males are generally more aggressive than females and may express their aggressiveness differently than females; thus, it may be helpful to consider males and females

separately. Whether differences are grounded in hormones or in different socializing experiences, there may be gender differences in aggressiveness and potential for hostility.

Conclusion

In summation, although the current study showed insignificant findings, the results do provide important implications for future research. This study provided merit to the strong reliability of scoring AgC answers, an important factor for any measure. In contrast, the validity of AgC was poor given the current findings. Based upon comparison with the BASC-2 scores, the AgC scores for this subject pool was thought to be higher. It is noteworthy that a significant charge for the Exner Comprehensive System to the R-PAS was to create norms that do not over-pathologize children; however, it may be that the current R-PAS norms under-pathologize children instead, at least with regard to aggressive potential. Of course, there are many other possible reasons, such as face validity, for no correlation between BASC-2 scores and AgC. Moreover, there may be a need to refine what constitutes an AgC score (i.e., direct versus indirect, fantastical versus realistic aggression). The R-PAS system is still in its infancy and will undergo substantial study in the upcoming years to continue proving its merits. Several recommendations have been made for further study of this system and specifically the AgC indicator. Future research will also provide evidence for ways in which the system may be improved, not only for use with children but adults as well.

References

- Baity, M., McDaniel, P.S., & Hilsenroth, M. (2000). Further exploration of the Rorschach aggressive content (AgC) variable. *Journal of Personality Assessment* 74(2), 231-241. Retrieved from <http://www.tandfonline.com/toc/hjpa>
- Baity, M. & Hilsenroth, M. (2002). Rorschach aggressive content (AgC) variable: A study of criterion validity. *Journal of Personality Assessment* 78(2), 275-287. Retrieved from <http://www.tandfonline.com/toc/hjpa>
- Crain, W. & Smoke, L. (1981). Rorschach aggressive content in normal and problematic children. *Journal of Personality Assessment* 45(1), 2-4. Retrieved from <http://www.tandfonline.com/toc/hjpa>
- Erdberg, P. (1993). The U.S. Rorschach scene: Integration and elaboration. *Rorschachiana* 18(1), 139-151. Retrieved from <http://www.hogrefe.com/periodicals/roorschachiana>
- Exner, J.E. (1993). *A Rorschach Workbook for the Comprehensive System*. Ashville, NC: Rorschach Workshops.
- Gacono, C.B. & Meloy, J.R. (1994) *The Rorschach Assessment of Aggressive and Psychopathic Personalities*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Gacono, C.B., Bannatyne-Gacono, L., Meloy, J.R., & Baity, M.R. (2005). The Rorschach Extended Aggression Scores. *Rorschachiana* 27(1), 164-190.
- Hiller, J., Bornstein, R., Rosenthan, R., Berry, D., & Brunell-Nuelieb, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment* 11(3), 278-296. Retrieved from <http://www.apa.org/pubs/journals/pas>

- Janson, H. (2008, March). *Calculating and reporting Rorschach intercoder agreement*. Paper presented at the annual meeting of the Society for Personality Assessment, New Orleans, LA.
- Kamphaus, R., Mays, K., Dowdy, E., Twyford, J., Chin, J., & DiStefano, C. (2011). Factor structure of the BASC-2 behavioral and emotional screening system student form. *Psychological Assessment* 22(2), 379-387. Retrieved from <http://www.apa.org/pubs/journals/pas>
- Kamphaus, R. & Reynolds, C. (2004). *Behavioral Assessment System for Children, Second Edition (BASC-2, 2nd ed.)*. Circle Pines, MN: AGS Publishing.
- Liebman, S., Porcerelli, J., & Abell, S. (2005) Reliability and validity of Rorschach aggression variables with a sample of adjudicated adolescents. *Journal of Personality Assessment* 85(1), 33-39. Retrieved from <http://www.tandfonline.com/toc/hjpa>
- Meloy, J.R. & Gacano, C. (1992). The Aggression Response and the Rorschach. *Journal of Clinical Psychology* 48(1), 104-114. Retrieved from <http://www.wiley.com/WileyCDA/WileyTitle/productCd-JCLP.html>
- Meyer, G. (personal communication, February 15, 2014).
- Meyer, G. (personal communication, February 16, 2014).
- Meyer, G. & Viglione, D.J. (2011, August). *New developments in Rorschach-Based behavioral assessment*. Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.
- Meyer, G., Viglione, D.J., Mihura, J.L., Erar, R.E., & Erdberg, P. (2011) *A Manual for the Rorschach Performance Assessment System*. Toledo, OH: R-PAS.

- Meyer, G.J., Viglione, D.J., & Girmini, L. (2014, January, 29). Current R-PAS transitional and adolescent norms [R-PAS website] Retrieved April 1, 2014 from <http://www.r-pas.org/currentchildnorms.aspx>
- Mihura, J.L., Meyer, G.J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin* 139, 548-605. Retrieved from <http://www.apa.org/pubs/journals/bul>
- Shapiro, D. (2012). Theoretical value of psychological testing. *Journal of Personality Assessment* 94(6), 558-562. Retrieved from <http://www.tandfonline.com/toc/hjpa>
- Viglione, D., Blume-Marcovici, A., Miller, H., Giromini, L., & Meyer, G. (2012). An inter-rater reliability study for the Rorschach Performance Assessment System. *Journal of Personality Assessment* 94(6), 607-612. Retrieved from <http://www.tandfonline.com/toc/hjpa>

APPENDIX

Figure 1: AgC normative data

<i>Age</i>	11	12	13	14	15	16	17
<i>Mean</i>	3.13	3.12	3.12	3.12	3.11	3.11	3.10
<i>St. Dev.</i>	2.28	2.25	2.22	2.19	2.17	2.14	2.12

Figure 2: Inter-rater Reliability

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	0.89	0.06	12.93	0.00
N of Valid Cases		32			

Figure 3: AgC with Age and Gender

		AGG CONTENT	AGE	GENDER
AGG CONTENT	Pearson Correlation	1	-0.05	0.15
	Sig. (2-tailed)		0.80	0.43
	N	32	32	32
AGE	Pearson Correlation	-0.05	1	-0.13
	Sig. (2-tailed)	0.80		0.48
	N	32	33	32
GENDER	Pearson Correlation	0.15	-0.13	1
	Sig. (2-tailed)	0.43	0.48	
	N	32	32	32

Figure 4: AgC and Diagnosis

		AGG CONTENT	PTSD	CONDU CT	ODD	MOOD
AGG CONTENT	Pearson Correlation	1	-0.08	0.17	-0.002	-0.08
	Sig. (2-tailed)		0.65	0.35	0.99	0.65
	N	32	32	32	32	32
PTSD	Pearson Correlation	-0.08	1	-0.07	-0.03	0.14
	Sig. (2-tailed)	0.65		0.70	0.87	0.44
	N	32	32	32	32	32
CONDUCT	Pearson Correlation	0.17	-0.07	1	-0.28	-0.06
	Sig. (2-tailed)	0.35	0.70		0.13	0.75
	N	32	32	32	32	32
ODD	Pearson Correlation	-0.002	-0.03	-0.28	1	-0.24
	Sig. (2-tailed)	0.99	0.86	0.13		0.18
	N	32	32	32	32	32
MOOD	Pearson Correlation	-0.08	0.14	-0.06	-0.24	1
	Sig. (2-tailed)	0.65	0.44	0.75	0.18	
	N	32	32	32	32	32