CHAPTER 2

STASTICAL FOUNDATIONS OF THE CRISI WARTEGG SYSTEM

Jacob A. Palm, Southern California Center for Collaborative Assessment
Alessandro Crisi, Istituto Italiano Wartegg, Sapienza University Rome

As reviewed in Chapter 1, the Crisi Wartegg System developed over the course of thirty years, through a process of continuous refinement based upon clinical experience and empirical research. This chapter summarizes major research studies concerning the CWS. It should be noted that while many of these research studies have been published previously, this marks the first time they are presented for English publication. Additionally, much of the previous research has been presented at conferences and symposia during the previous two decades, including annual conventions of the Society for Personality Assessment (SPA) and the International Rorschach Society (ISR). As such, the collected data presented here is based upon translations of Italian publications, journal articles, and academic texts; summaries of data presented at conferences and symposia; review of recently published English-language peer-reviewed journal articles; and presentation of recent reliability data not previously published. In some cases, when possible, primary authors were contacted for further information about research design, clarification of statistical data, or broader explanation of findings. In some cases, when contact was not possible, the data summarized here is limited to that which was previously published or presented. It should be noted that all tables from previous publications are reproduced with the permission of the respective author or publisher.

Prior to research summarizing the specific validity and reliability of the Crisi Wartegg System (CWS), the 2012 meta-analytic study conducted by Soilevuo Grønnerød and Grønnerød will be reviewed. Following this review, published and presented data related to the interrater and test-retest reliability of the CWS will be presented. Lastly, relevant convergent validity data will be summarized, beginning with early exploratory studies, and continuing chronologically, grouped by study focus, over the past several decades.

Meta-Analytic Study of the WDCT (Soilevuo Grønnerød & Grønnerød, 2012)

The varied interpretive systems and approaches for the Wartegg Drawing Completion Test that developed in isolation (described in Chapter 1), have resulted in limited comprehensive research. To address this limited integration of study data, a result of challenges in accessing previously conducted research due to geographic distance and publication language, Soilevuo Grønnerød and Grønnerød (2012) undertook analysis of available empirical studies on the reliability and validity of the Wartegg Test. In completing their meta-analytic review, the authors conducted a comprehensive literature search noting that advancements in online text access and databases yielded greater numbers of Wartegg-specific articles than previously conducted similar searches by several researchers. It was noted that some articles were unable to be retrieved, including dissertations from the United States, and unpublished works including Master's theses and conference presentations. Lastly, despite a long history of Wartegg use in Japan, studies published in Japanese were excluded due to translation challenges. The authors located 507 references of scholarly work from 31 countries. Following exclusion of non-empirical publications, 37 studies (containing 38 data sets) were determined to meet inclusion criteria for analysis. Full exclusion criteria, and case-by-case exclusion determination are well described in the authors' published article, and will not be summarized here (see Soilevuo Grønnerød & Grønnerød, 2012, for full details).

In terms of reliability, results of the 2012 meta-analysis appear quite favorable. Inter-rater reliability coefficients averaged in the excellent range (r_w=.79; 15 results from 12 samples). Similarly, internal consistency coefficients averaged in the satisfactory range (r_w=.74; 3 results from 2 samples). The authors noted that test-retest reliability coefficients were "disappointingly low" (page 14, r_w=.53*;* 3 results from 2 samples), although suggested that this weighted average is difficult to understand due to lack of clarity on the state versus trait aspects of included variables.

In terms of validity, the authors found similarly positive results from the analyzed research. Analysis of studies with a clearly stated research hypothesis yielded a large effect size (r_w=.33; 290 results from 14 samples). This effect size was noted to be slightly greater than meta-analytic results of both the Rorschach and the MMPI-2 (Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999; as cited in Soilevuo Grønnerød & Grønnerød, 2012). Overall, a lower magnitude effect size was found for all results (r_w=.19*; 95% CI* =.14-.26); however, various factors (including scoring system, number of examined criteria, and scorer blindness) were determined to significantly impact results. For example, in considering scorer blindness during regression analysis, difference was noted between no scorer blindness (r=.12, predicted effect size) versus full scorer blinding (r=.35, predicted effect size). Given the small number of studies for each scoring system, including the CWS, effect sizes for each system were unable to be analyzed.

In their discussion of results, Soilevuo Grønnerød and Grønnerød (2012) noted "surprise" regarding the large effect size of studies involving specific clinical hypotheses. Based upon this result, and the comparability of this effect size with other performance-based and objective personality measures, the authors concluded, "the research on the WZT may reach levels comparable to other assessment methods, given sufficient focus on study quality" (p. 482). The authors further noted that based upon review, WDCT results appeared well correlated with other free-response methods and clinical observation, whereas were not well-associated with self-report measures of personality. This lack of relationship was noted to occur with other commonly used free-response methods as well, as discussed in the assessment research literature (see Bornstein, 2009). In summation, the authors asserted, "…based on our meta-analysis, we argue that there is no reason to dismiss the Wartegg method altogether as a method for personality evaluation. However, it is necessary to build a solid, cumulative research tradition to produce knowledge and create a basis for the use of the Wartegg method in psychological practice… We strongly encourage, however, more research built on previous studies that will cultivate the strongest part of the method" (p. 483).

RELIABILITY AND VALIDITY OF THE CRISI WARTEGG SYSTEM

While the meta-analytic findings of Soilevuo Grønnerød and Grønnerød (2012) included research regarding several scoring and interpretation systems of the Wartegg Test, among them the Crisi Wartegg System (CWS), significant research has been conducted specifically on the CWS. In the last 20 years, studies have investigated the interrater and test-retest reliability, as well as the convergent validity of the CWS.

Reliability Studies

Projective and performance-based personality tests may be thought of as potentially reactive or subjective techniques, as their administration and scoring may be significantly impacted by factors related to the evaluator. These factors may include competence, training, relationship to the client, setting of test administration, and familiarity with the test being administered. While the WDCT according to the CWS is considered standardized, based upon an objective scoring and administration system, given the potentially subjective nature of the method it is crucial to evaluate agreement (interrater reliability) between scorers.

Once interrater reliability is established between professionals, it is then possible to examine consistency (test-retest reliability) between deferred administrations (Balboni & Cubelli, 2004). With this in mind, both interrater and test-retest reliability of the CWS have been researched, with consistently positive results. Eight studies on interrater reliability and 1 study on test-retest reliability are presented below.

Interrater Reliability

Many previously conducted studies involving the Wartegg Drawing Completion Test have demonstrated high levels of interrater agreement, including Kappa coefficients of 0.94 (Roivainen & Ruuska, 2004) to those ranging between 0.66 to 1.0 (Alves, Dias, Sardinha, & Conti, 2010). These studies have evaluated varied scoring methodologies, as discussed above. The interrater reliability of the Crisi Wartegg System is equally established, as described below in multiple studies.

Preliminary Interrater Reliability Analysis (Crisi, 1998; Crisi, 2007)

Examination of the interrater reliability of CWS scoring was first undertaken in 1999, following basic standardization and formalization of the administration and scoring process. Initial research, published in the first and second editions of the Italian-language CWS Manual (Crisi, 1998, 2007), reviewed interrater reliabilities between three pairs of judges, each of whom independently reviewed and scored 18 CWS protocols randomly selected from the archives of the *Istituto Italiano Wartegg*. Data was collected under two conditions, to evaluate the effectiveness and impact of written CWS scoring and administration materials. First, judges were instructed to score the selected protocols without referencing published scoring guidelines; that is, scoring was based upon previous training and experience. In the second condition, the same judges were instructed to score selected protocols with the assistance of the instructional manual.

Interrater reliability was further evaluated based upon the experience level of the raters. Raters were divided into three categories: “Expert,” indicating psychologists who had been practicing assessment with the CWS for at least five years; “Practical,” indicating psychologists who had two years of experience; and “Novice,” indicating psychologists who had only recently completed training on the use of the WDCT according to the CWS. Each dyad was examined using paired raters of similar or different experience levels, resulting in comparisons between Expert-Expert, Expert-Practical, and Expert-Novice.

In considering this initial interrater reliability research, it should be noted that indexes of agreement were calculated for only the most important scoring categories of the CWS. These categories included Evocative Character (EC), Affective Quality (AQ), Form Quality (FQ), and Special Scores (SS). Some scoring categories were not initially evaluated (Popular Responses, Content, Movement), given the objective scoring criteria for these scoring categories assures interrater agreement and might artificially inflate overall correlations of agreement. For example, Popular responses for each box are provided to users in list format, yielding near-perfect interrater agreement. Lastly, the theoretically derived category of Impulse Responses was not studied, given that depending upon the theoretical orientation of the examiner, this category can be scored or not scored without detracting from the overall quantitative or interpretive power of the test.

The degree of agreement between raters was calculated using Cohen's measure of agreement corrected for chance agreement (κ), between pairs of examiners for all possible combinations. That is, for each analyzed variable (i.e., EC, AQ, FQ), scores in each of the eight boxes, over a total of 18 rated protocols, were compared yielding a total of 432 points of comparison. These comparisons were analyzed for both condition #1 (scoring without referencing scoring manual) and condition #2 (scoring using formal scoring manual). For example, in comparing Expert-Expert ratings, identical ratings increased from the first condition (363 out of 432) to the second condition (380 out of 432), as further described below. Whereas the above-mentioned variables (i.e., EC, AQ, FQ) are mandatory in scoring (requiring a score in each of the 8 boxes of every protocol), Special Scores are not mandatory and are only scored when warranted by the client's drawings or verbalizations. As such, the number of cases by which interrater agreement was calculated was determined independently by the presence of an assigned Special Score by any rater. That is, if one rater assigned a special score to Box 1 of a protocol, but another rater did not, the presence of a special score was considered possible (regardless of the degree of agreement or disagreement); therefore, interrater agreement was confirmed if both raters agreed on the presence or absence of a Special Score, and was not confirmed if either assigned a Special Score while the other did not.

Considering results from the first condition, in which the protocols were scored without reference materials, the highest degree of interrater reliability was obtained in the Expert-Expert comparison (κ=0.84 to 0.88, "Almost Perfect"). As expected, agreement was lower between Expert-Practical (κ= 0.68-0.78, "Substantial") and lowest between Expert-Novice (κ= 0.55-0.59, "Moderate").

In the second condition, during which raters referenced written standardized scoring guidelines, interrater reliability was noted to significantly increase. Agreement between raters, on average, was high between Expert-Expert (κ=0.91, "Almost Perfect") and Expert-Practical (κ=0.84, "Almost Perfect") raters, and substantial between Expert-Novice (κ=0.69, "Substantial").

In considering the results of this preliminary study, it is important to note that while a high level of agreement between expert examiners was discovered, the lower (yet still substantial) levels of interrater reliability found in this study would not significantly impact interpretation of the test results. Similar to other tests of this nature, including the Rorschach and MMPI-2, many of the indices obtained, as well as the resulting calculations and computations derived from these indices (i.e., in the CWS: EC+%, AQ+%, etc.; in the MMPI-2, the clinical scales), are interpreted via normative ranges of values. For example, in considering Evocative Character of the CWS, the EC+% (which is calculated by summing the individual EC scores for each box, dividing by 8, and converting into a percentage), has a "normal" range from 56% to 81%. Translated into raw scores derived from scoring, this equals 4.5 to 6.5 points. Considering that in each of the 8 boxes of the Wartegg, a response can be assigned an EC raw score of 0, 0.5, or 1, let us imagine that a rater assigned 6 points (75%) to a CWS protocol. If a second rater assigns 5 points (62%), for interpretive

purposes, the understanding of the client remains the same given that both calculated indices fall within the "normal" range (between 56% and 81%). While this is an important consideration, it should be noted that among experienced examiners, differences of more than one half a point are extremely rare.

Crisi (2011b)

CWS interrater reliability was further evaluated in a clinical sample randomly selected from the archives of the *Istituto Italiano Wartegg*. In this study, the protocols of 30 subjects were blindly evaluated by 3 independent judges certified in the Crisi Wartegg System. Intraclass correlation coefficients (ICC) were computed between the three judges on major CWS scoring categories. ICCs were calculated for all formal indices but not for specific Content, Movement and Special Score categories of scoring whose frequencies were too low for meaningful comparisons (i.e., those clinical phenomena that are captured by formal scoring, but occur in less that 2% of cases).

The majority of evaluated indexes exhibited excellent levels of ICC ranging from .77 to .97 (p<.01), as presented in Table 2.1. Several scoring categories, including Form Quality (ICC=.60), Original Responses (ICC=.68) and Popular Responses (ICC=.68) demonstrated substantial inter-rater agreement. Results are presented in Table 2.1.

Table 2.1
Intraclass Correlation Coefficient (ICC) Values for Major CWS Indices

Index	ICC*
Evocative Character (EC+%)	0.77
Affective Quality (AQ+%)	0.90
Form Quality (FQ+%)	0.60
Popular Responses (P%)	0.68
Original Responses (O%)	0.68
Anxiety Index (AI)	0.87
Impulsivity Index (IM)	0.82
Primary Movement (M)	0.90
Secondary Movement (m)	0.79
Human Content (H %)	0.98
Object Content (OBJ %)	0.92
Animal Content (A%)	0.93
Symbol Content (SIG%)	0.89
Botanical Content (BOT%)	0.95
Astronomical Content (AST%)	0.93
Architecture Content (ARC%)	0.94
Human Simulacrum Content (HS%)	0.90
Nature Content (NAT%)	0.88
Note. *p<0.01 for all variables. ICC calculated between 3 independent raters.	

Crisi (2011b)

Given that the Crisi Wartegg System introduced two new scoring categories to the evaluation of the WDCT—Evocative Character (EC) and Affective Quality (AQ)—interrater reliability analyses were conducted as specifically related to these categories to ensure consistent scoring. The AQ and EC of 30 protocols were reviewed. Both Cohen's kappa (κ) and Fleiss' kappa (K) were calculated, with the former examining each possible correlation between pairs of three (blind) independent raters, and the latter examining correlations between three independent evaluators concurrently.

In considering agreement at the scoring level (raw data level of agreement), independent raters can arrive at three levels of agreement: total agreement, partial agreement, or total disagreement. In both of the scored categories (Evocative Character and Affective Quality), possible scores are limited to 0, 0.5, or 1 in each box, for each rater. Therefore, in considering levels of agreement between three raters, the following possibilities exist:

1. Total Agreement: All raters assign the same score to a box (for example: 1, 1, 1);
2. Partial Agreement: One rater's score differs from the other two raters' scores (for example: 0,1,1);
3. Total Disagreement: All raters assign a different score to a box (for example: 1, 0.5, 0).

Preliminary analysis of the raw data found significant levels of total agreement. For Evocative Character, for example, 76.6% of boxes demonstrated total agreement (*n*=191) between raters, 17.5% indicated partial agreement (*n*=42), and only 2.9% demonstrated total disagreement (*n*=7). Results for Affective Quality exceeded those of Evocative Character: 85.8% of boxes demonstrated total agreement (*n*=206), 14.2% indicated partial agreement (*n*=34), and 0% demonstrated total disagreement (*n*=0).

Results of interrater reliability analyses for both Evocative Character and Affective Quality (by Box) are summarized in Table 2.2.

Table 2.2
Interrater Reliability by WDCT Box:
Evocative Character (EC) and Affective Quality (AQ)

Box	**Evocative Character**		**Affective Quality**	
	Fleiss' K	***Cohen's κ***	***Fleiss' K***	***Cohen's κ***
Box 1	.84	.85	.84	.87
Box 2	.44	.50	.56	.66
Box 3	.79	.85	.95	.96
Box 4	.59	.60	.66	.70
Box 5	.69	.69	.81	.79
Box 6	.24*	.18*	.79	.81
Box 7	.75	.76	.87	.88
Box 8	.80	.80	.96	.96
Mean	*.64*	*.64*	*.81*	*.83*

Note. *As initial analysis suggested poor results, scoring rules for Box 6 were specifically clarified, resulting in higher levels of agreement in subsequent studies. Despite this questionable result in Box 6, overall interrater reliability fell in the substantial range of classification (.61-.80).

Crisi & Dentale (2016)

Crisi and Dentale investigated the interrater reliability of "three new scoring categories" introduced by the Crisi Wartegg System: Evocative Character (EC), Affective Quality (AQ), and Form Quality (FQ). Two independent scorers blindly reviewed 30 randomly selected clinical protocols. Intraclass correlation coefficients were calculated, with each variable demonstrating significant levels of agreement: EC (ICC=.74, F=7.81, df=29, p<0.001), AQ (ICC=.92, F=28.17, df=29, p<0.001), and FQ (ICC=.71, F=5.77, df=29, p<0.001).

Crisi, Vari, Velotti, Carlesimo, Guzzi, & Zavattini (2014)

As part of a larger comparative study investigating depression and negative affect states in dermatology patients, interrater reliability analyses were conducted. Two independent raters scored 84 protocols (42 experimental, 42 control), blind to group membership. Intraclass correlation coefficients (ICC) were calculated for three CWS variables to determine mean consistency. On average, interrater reliability for the three investigated CWS variables (Evocative Character, EC; Affective Quality, AQ; and Form Quality, FQ) fell within the *almost perfect* range (*ICC =.826*).

More specifically, the ICC for each studied variable is presented in Table 2.3.

Table 2.3
Interrater Reliability of Major CWS Indices: Evocative Character (EC), Affective Quality (AQ), Form Quality (FQ)

Variable	*ICC*	*F*	*df*	*p*
EC	.887	8.862	83	.000
AQ	.793	4.829	83	.000
FQ	.802	5.502	83	.000

Note. ICC calculated between 2 independent raters.

Overall, consistent interrater reliability (ranging from *substantial* to *almost perfect)* was demonstrated on three crucial CWS indices, including two indices original to the Crisi Wartegg System-- Evocative Character and Affective Quality.

Daini, Lai, Festa, Maiorini, Pertosa, & De Risio (2006)

As part of a larger comparative study investigating affect states and impulsivity in patients being treated for eating disorders, interrater reliability analyses were conducted. Two independent raters scored 40 randomly selected protocols (from 181 protocols included in the research study). Interrater agreement on five specific scoring categories was studied: Evocative Character (EC), Affective Quality (AQ), Form Quality (FQ), Anxiety Index (AI), and the Index of Impulsivity (IM). Results for each comparison fell in the excellent range. For EC, AQ, and FQ, mean interrater reliability was reported by the authors (*K minimum*=.79; *K maximum*=.82; *K mean*= 0.80). Reliability coefficients for both the Anxiety Index (r =.80, p<.0001) and the Index of Impulsivity (r_w=.88, p<.0001) were found to be similarly adequate.

Daini, Petrongolo, Manzo, & Bernardino (2012)

As part of a larger comparative study investigating personality factors impacted by work as a professional nurse, interrater reliability analyses were conducted. Two independent evaluators scored 321 protocols (111 experimental, 210 control), blind to group membership. Interrater

reliability coefficients were calculated for relevant indices, including Evocative Character, Affective Quality, Form Quality, Frequency (Popular Responses), Anxiety Index, Index of Impulsivity, Index of Inner Tension, and some Content category percentages. Overall mean interrater reliability was reported in the excellent range (r_w=.*95*).

Crisi, Palm, & Lops (2016)[i]

Interrater reliability was further evaluated in an American sample of protocols, with clinicians trained in the United States serving as independent raters. In this study, 30 randomly selected protocols from the American normative sample were independently scored by 5 clinicians trained and certified in the CWS. Given training, certification, and experience, these clinicians were deemed experienced in the CWS, reflecting an advanced level of competence in both scoring and interpretation. A 6th judge, considered expert in Italy, additionally participated in the study. Interrater agreement was calculated at the individual scoring-decision level, the aggregate protocol-level of calculated indices and percentages, and overall evaluation of consistency between pairs of judges. Each will be discussed in turn below.

Agreement at the Scoring Level

Interrater agreement was calculated for the most common scoring elements of the CWS, including Evocative Character (EC), Affective Quality (AQ), Form Quality (FQ), Content (CONT), Popular Responses (P), Anxiety Stroke (AS), Crossed Border (CB), and Movement (M). Given that scores are assigned for each of 8 boxes in a protocol, and 30 protocols were included in this study, interrater comparisons were calculated on a total of 240 data points (i.e., 30 protocols x 8 boxes = 240) for each scoring category evaluated.

In evaluating agreement between the 6 raters, Fleiss' kappa (*K)* was utilized. This statistic, typically used as an alternative to the Cohen's kappa in cases where more than three raters are evaluated, provides statistics relevant to level of agreement between judges while taking into account the probability of agreement due to chance. Moreover, given that for some CWS variables (i.e., FQ, AQ, EC) scoring values are limited to three alternatives on an ordinal scale (0, 0.5. and 1) for which the differences are weighted (i.e., the difference between 0 and 0.5 is different from that between 0 and 1), Fleiss' *K* is considered a better choice (Chiorri, 2011).

In considering overall agreement, a further consideration is relevant related to the paradox of high agreement concurrent with a low kappa (Randolph, 2005). Moreover, most statistical evaluations of interrater reliability assume that raters are limited in the manner in which cases may be distributed across categories, which in the case of the CWS, is not true. Given that there are no limits placed on CWS raters regarding the frequency of assigned specific scores (i.e., raters can assign 0, 0.5, or 1 without limitation or expectation), the results of the Free Marginal Kappa are also presented because this coefficient may more accurately reflect interrater agreement under these circumstances. Brennan and Prediger (1981) recommend using free marginal kappa (K_{free})when raters are not forced to assign a certain number of cases to each category (as in the case of the CWS), and fixed marginal kappa (K_{fixed}) in cases where they are. Given this recommendation, two variations of kappa are presented in Table 2.4: Siegel and Castellan's (1988) fixed-marginal multirater kappa and Randolph's (2005) free marginal multirater kappa (Randolph, 2005; Warrens, 2010).

Table 2.4 Interrater Reliability of Major CWS Scoring Categories					
Variable	**% Overall Agreement**	K_{fixed}	K_{free}	**Z** (K_{fixed})	***p*** (K_{fixed})
EC	82.2	.679	.733	31.09	.000
AQ	88.7	.809	.831	44.88	.000
FQ	87.1	.292	.807	12.37	.000
Content	88.7	.867	.882	44.91	.000
Popular Response (P)	91.5	.827	.873	33.31	.000
Anxiety Stroke **(Special Score)**	87.5	.617	.749	14.79	.000
Crossed Border **(Special Score)**	93.9	.875	.879	45.20	.000
Movement (M)	91.4	.770	.896	16.93	.000
Note. K_{fixed}: fixed-marginal multirater kappa (Siegel and Castellan, 1988); K_{free}: free marginal multirater kappa (Randolph, 2005).					

Given the occurrence of a paradoxically high level of overall agreement coupled with a low kappa, the scoring category Form Quality (FQ) must be examined. In this case, as presented in Table 2.4, the overall percentage of agreement appears high (87.1%), whereas the *kappa* appears low (K_{fixed}=.292). In understanding this, we must consider the FQ variable. FQ scoring, as compared to the other indices evaluated, demonstrates less variability. That is, while raters have the opportunity (as with other variables) to assign scores of 0, 0.5, or 1, the majority of ratings typically represent a score of 1, with relatively few scores of 0 assigned. Clinically, scores of 1 are expected and normative, whereas scores less than 1 are typically representative of difficulties, or in extreme cases, pathology. Given this distribution of scores, it is not surprising that there would be a high percentage of agreement between raters, but a corresponding low fixed marginal kappa (due to limited prevalence of scores based upon variable distribution). As such, the free marginal kappa appears a more reliable estimate of agreement corrected for chance (K_{free} =.807).

In considering the results presented in Table 2.4, specifically as related to free marginal kappa values, agreement for all variables falls above 0.7, with the majority (6 of 8 studied scoring variables) falling higher than 0.8. Relating these values to the interpretive scale devised by Landis and Koch (1977), three variables (EC, AS, FQ) demonstrate *substantial* agreement, and five variables demonstrate *almost perfect* agreement (AQ, Content, Popular, CB, Movement). These results provide strong support for consistent scoring between clinicians who have been trained and certified in the CWS. Considering that the CWS is a performance-based or projective personality test, requiring complex scoring decisions based upon clinician judgment, these numbers are more compelling.

Agreement at the Protocol Level

Following analysis of individual scoring decisions made by the 6 judges, protocol-level calculations and indices were derived according to CWS guidelines. In general, the individual scoring data points spread over the 8 boxes of the CWS are mathematically converted into approximately 60 calculated scores and indices on which interpretation is subsequently based. For the purposes of this study, the 17 most relevant scoring elements were studied with other calculated indices excluded due to low frequency or statistical rarity.

To evaluate the agreement between the six raters, intraclass correlation coefficients (ICC) were calculated, given that this technique is suitable for situations with more than two judges and rating scales with multiple points (Chiorri, 2011). A two-way random effects model was utilized taking into account that each of 6 raters evaluated each case, and these 6 raters are regarded as a sample of the entire population of clinicians certified in the CWS. Each studied index, presented in Table 2.5, was evaluated using the same parameters (random two-way effects model, absolute agreement definition, consistent estimator).

Table 2.5
Interrater Reliability (ICC):
Major Calculated Indices of the CWS

Index	***Single Measures***	***p***	***Average Measures***	***p***
EC+%	.722	.000	.940	0.000
AQ+%	.800	.000	.940	0.000
FQ+%	.350	.000	.763	0.000
P%	.680	.000	.860	0.000
AI	.750	.000	.947	0.000
IM	.928	.000	.987	0.000
M	.881	.000	.978	0.000
m	.656	.000	.920	0.000
H %	.916	.000	.985	0.000
OBJ %	.846	.000	.971	.000
A%	.987	.000	.998	.000
ARC%	.712	.000	.937	.000
SIG%	.863	.000	.974	.000
NAT%	.615	.000	.906	.000
BOT%	.932	.000	.988	.000
AST%	.839	.000	.969	.000
HS%	.823	.000	.965	.000

Note. ICC calculated based upon 6 independent raters.

In evaluating results according to Cicchetti's guidelines (1994), as with previous findings, most indices on average demonstrate *good* to *excellent* levels of agreement.

Consistency Between Pairs of Judges

Lastly, aggregate correlations between each possible pair of judges were calculated for two scoring categories unique to the CWS: Evocative Character (EC) and Affective Quality (AQ). Calculated using Spearman rank order correlation (r_s), all resulting correlations ranged from *strong* to *very strong*. See Results of correlations between raters for Evocative Character (EC) in Table 2.6.

Table 2.6
Correlations between Rater Pairs: Evocative Character (EC)

	N	r_s	*t(n-2)*	*p*
Rater 1 & Rater 2	30	.9274	13.12187	.0000000
Rater 1 & Rater 3	30	.6836	4.957344	.0000312
Rater 1 & Rater 4	30	.7682	6.351255	.0000007
Rater 1 & Rater 5	30	.8119	7.359818	.0000001
Rater 1 & Rater 6	30	.7358	5.750079	.0000036
Rater 2 & Rater 3	30	.6249	4.235486	.0002228
Rater 2 & Rater 4	30	.7805	6.606388	.0000004
Rater 2 & Rater 5	30	.8342	8.004766	.0000000
Rater 2 & Rater 6	30	.7267	5.598184	.0000054
Rater 3 & Rater 4	30	.6075	4.047235	.0003702
Rater 3 & Rater 5	30	.7006	5.195174	.0000163
Rater 3 & Rater 6	30	.7302	5.655459	.0000046
Rater 4 & Rater 5	30	.7704	6.394437	.0000006
Rater 4 & Rater 6	30	.7529	6.053823	.0000016
Rater 5 & Rater 6	30	.8104	7.319362	.0000001

Note. r_s*=Spearman rank order correlation.*

Results of correlations between raters for Affective Quality (AQ) are presented in Table 2.7.

Table 2.7
Correlations between Rater Pairs: Affective Quality (AQ)

	N	r_s	*t(n-2)*	*p*
Rater 1 & Rater 2	30	.8501	8.542847	.0000000
Rater 1 & Rater 3	30	.7516	6.029415	.0000017
Rater 1 & Rater 4	30	.7803	6.602315	.0000004
Rater 1 & Rater 5	30	.7915	6.852993	.0000002
Rater 1 & Rater 6	30	.8306	7.893332	.0000000
Rater 2 & Rater 3	30	.7731	6.450924	.0000005
Rater 2 & Rater 4	30	.8616	8.981222	.0000000
Rater 2 & Rater 5	30	.7419	5.854662	.0000027
Rater 2 & Rater 6	30	.8451	8.365695	.0000000
Rater 3 & Rater 4	30	.7920	6.863811	.0000002
Rater 3 & Rater 5	30	.8935	10.52798	.0000000
Rater 3 & Rater 6	30	.7723	6.432695	.0000006
Rater 4 & Rater 5	30	.8078	7.251393	.0000001
Rater 4 & Rater 6	30	.8547	8.711901	.0000000
Rater 5 & Rater 6	30	.8344	8.009299	.0000000

Note. r_s*=Spearman rank order correlation.*

It should be noted that these scoring categories, novel to the CWS, require judgment on the part of the clinician while scoring. Therefore, this strong agreement between raters underscores the effectiveness of CWS training, written guidelines, and scoring methodology. Overall, all components of this study demonstrated significant interrater reliability of the CWS at both the scoring level and the computational/interpretive level.

Test-Retest Reliability

While many studies have evaluated the interrater reliability of the CWS, less research has investigated the stability of CWS indices over time via formal analysis of test-retest reliability. Given the supposition that some variables of the CWS measure *state* personality characteristics and others personality *traits,* clinical experience suggests that not all CWS variables will remain consistent over time (similar to other personality measures). Further studies regarding the test-retest reliability of the CWS, both with normative and clinical populations (to assess the impact of psychotherapeutic intervention on functioning), are necessary (with several currently in process).

Crisi, Janson, & Palm (2016)[ii]

Initial exploratory research into test consistency over time followed 38 normative study participants (non-clinical) recruited as volunteers from the general population. These individuals were administered the WDCT according to CWS standardized administration guidelines on two occasions, approximately 4-5 weeks apart ($\bar{x}$=31.2 days, *SD*=2.7 days; minimum 28 days, maximum 45 days). The participants were all adults (26 women, 12 men) with a mean age of 37.7 years (minimum 18 years, maximum 71 years). The participants were tested solely for research, gave informed consent to participate in this study, and were not compensated for their participation. None of the participants had a known psychiatric diagnosis or disorder, and none had a history of psychiatric or psychotherapeutic treatment. Assessment took place individually at the offices of the *Istituto Italiano Wartegg*. The examiners included 5 psychologists proficient in the CWS. All protocols were coded by one expert coder.

14 CWS variables were selected for analysis: EC+%, AQ+%, FQ+%, the Affective Stability Index (A/F Ratio), P%, Impulsivity Index (IM), Anxiety Index (AI), WIP Quadrant and Area (QUAD/AREA), Global Assessment classification, Index of Suicidal Tendencies (IST), H%, OBJ%, ARC%, and SIG%. These variables included the most central protocol-level aggregate variables as well as content code percentages with a frequency of >10% across the two administrations. Due to the small sample and the limited variation of some variables in this sample, test-retest reliability for low-frequency variables (including O%, less frequent Content codes, and Special Scores) or variables based on a subset of drawings (including P+% and O+%), were not investigated.

Two sets of analyses were performed, treating the variables as continuous and categorical, respectively. The analyses of the variables as continuous addresses the agreement between Times 1 and 2 for variable values. However, in clinical interpretation, interpretive ranges of variables are of primary importance, rather than absolute numeric values of variables. Thus, the agreement for the selected CWS variables was also analyzed in terms of time-to-time categorical agreement with categories representing interpretive ranges. The analyses of the categorized variables address the agreement between Times 1 and 2 in terms of stability within one interpretive range.

In regards to the Affective Stability Index (A/F Ratio), for which the interpretive ranges depend on both terms, the separate terms of the ratio (i.e., A and F) were included in the continuous analyses rather than the ratio. For the continuous evaluation of the WIP Quadrant and Area (QUAD/AREA),

the separate terms that make up the basis for the evaluation (i.e., IIT-1 and the first term of the IIT-2 ratio) were included in the continuous analyses, as well as the Area. In the categorical analyses, the A/F ratio was analyzed as one categorical variable (but not the separate terms A and F). Likewise, for the categorical analyses, the WIP Quadrant and Area (QUAD/AREA) were analyzed as one categorical variable (but not the separate terms IIT-1, IIT-2, and Area).

Table 2.8 presents results for the selected variables analyzed as continuous. The table denotes means and standard deviations for variables at Times 1 and 2, with *t*-test *p*, and Pearson test-retest-correlations (*r*). Results suggest overall small systematic shifts in variable means between Time 1 and Time 2 (paired *t*-test, $p<.05$ for 6 out of 18 variables, which is more than what would have been expected by chance, but the absolute size of systematic shifts were small and seemingly not important). The time-to-time-correlations were on the whole sizeable, with a mean Pearson *r* of .61. More than half of the correlations investigated found correlations >.60, five out of seventeen reaching >.70, and three exceeding >.80. The lowest test-retest correlation was found for P% ($r=.32$).

Table 2.8
Test-Retest Reliability: CWS Continuous Variables

Variable	Time 1		Time 2		*r*
	$\bar{x}$	*SD*	$\bar{x}$	*SD*	
EC+%	73	9	76	12	.63
AQ+%	63	13	60	13	.47
FQ+%	95	6	98*	4	.72
A	2.7	0.8	2.4	0.8	.52
F	3.7	0.4	3.9*	0.3	.78
P%	24	12	24	13	.32
IM	.30	.21	.23*	.21	.81
AI	.80	.22	.83	.20	.83
IIT-1	1.2	0.4	1.1*	0.3	.80
IIT-2 (Term 1)	4.7	1.3	4.8	1.5	.66
Area	1.6	0.6	1.5	0.5	.68
Global Evaluation	1.8	1.3	1.3*	0.9	.50
IST	2.4	1.8	1.8*	1.5	.45
H%	14	13	13	12	.55
OBJ%	32	19	37	20	.55
ARC%	16	10	17	9	.38
SIG%	10	13	10	13	.69

Note. *$p<.05$ for mean Time 2 – Time 1; difference by paired *t-test*.

Table 2.9 presents time-to-time agreement results for selected CWS variables recoded into interpretive ranges and treated as categorical. Test-retest agreement is presented as percent agreement (i.e., the proportion of participants who stayed within the same interpretive range from Time 1 to Time 2) and Cohen's (1968) kappa for chance-corrected test-retest agreement (κ). Results are presented both at the variable level (i.e., the agreement for classification in the same category out of all those possible at Times 1 and 2) and for each category (i.e., the agreement for having the same status - Present or Absent - for that specific category at both times). The overall proportion of agreement across times was generally high, averaging .78. With the exception of two variables (AQ+% and A/F Ratio), all test-retest agreement fell above .65, with results as high as .97 for two variables (FQ+%, ARC%). Chance-corrected test-retest agreement was generally lower, averaging .41. Six variables fell between .50 and .84, whereas eight variables were lower than .50. An uneven

distribution of participants over categories, with most participants classified in one single category--a condition which is known to be problematic for the estimation of chance-corrected agreement--was present for several variables, thereby making the interpretation of chance-corrected agreement results (i.e., kappa) more difficult.

Table 2.9
Test-Retest Reliability
CWS Categorical Variables

Variable and Category	Overall Proportion of Agreement[a]	Cohen's kappa (κ)[b]	Frequency	
			Time 1	Time 2
EC+%	.76	.38		
1. 0 to 50%			2	3
2. 51 to 81%			33	26
3. 82 to 100%			3	9
AQ+%	.45	.10		
1. 0 to 49%			3	5
2. 50 to 63%			21	21
3. 64 to 75%			9	10
4. 76 to 100%			5	2
FQ+%	.97	.00		
1. 0 to 74%			0	0
2. 75%			1	0
3. 76 to 100%			37	38
A/F Ratio	.50	.17		
1. A<F			21	23
2. A<<F			2	7
3. A>F			10	7
4. A>>F			5	1
P%	.66	.35		
1. 0 to 12%			5	4
2. 13 to 31%			24	24
3. 32 to 100%			9	10
IM	.79	.57		
1. 0 to .25			20	26
2. .26 to 1.0			18	12
AI	.89	.79		
1. 0 to .37			1	1
2. .38 to .85			16	14
3. .76 to 1.0			21	23
Quadrant/ Area	.92	.83		
1. Positive[c]			23	26
2. Negative[d]			15	12
Global Evaluation	.74	.31		
1. NOP + LSI			26	33
2. MSI			8	4
3. PTL + SSI			4	1
IST	.87	.23		
1. 0 to 4			33	36
2. 5 to 7			5	2

3. >7			0	0
H%	.74	.53		
1. 0 to 12%			11	15
2. 13 to 25%			23	20
3. >25%			4	3
OBJ%	.74	.26		
1. 0 to 12%			4	3
2. 13 to 50%			31	29
3. 51 to 100%			3	6
ARC%	.97	.65		
1. 0 to 25%			37	36
2. 26 to 100%			1	2
SIG%	.92	.62		
1. 0 to 25%			34	33
2. 26 to 100%			4	5

Note. [a]The overall proportion of agreement for a variable describes the proportion of participants who were classified in the same category at Times 1 and 2.
[b]*kappa* (κ) for a variable describes the chance-corrected agreement for the multi-category classification across Times 1 and 2.
[c] Positive Quadrant/Area= Area alpha of all quadrants + Area beta of Quadrant A.
[d] Negative Quadrant/Area= Area beta of Quadrants B, C, and D + Area gamma of all quadrants.

Overall, the results of the current analyses suggest substantial time-to-time correlations for continuous variables, and high overall percent agreement (but moderate chance-corrected agreement) across time for variables when treated as categorical in interpretive ranges. Some Wartegg variables that are known to be more influenced by day-to-day fluctuations, including AQ+% and A/F Ratio (Crisi, 2007), were among those variables that showed less stability across the four- to five-week retest period. It was noted that given the small sample size, and the relatively normative population studied, limited variation in some variables may have further artificially decreased the chance-corrected consistency data (kappa). Given the exploratory nature of this study, and these limitations, further research on test-retest reliability is needed.

Validity Studies

With reliability established, it is important to evaluate the validity of test measures prior to clinical use. In evaluating validity, we address several important questions: Does the CWS measure what it purports to measure? Does the CWS "agree" with other established personality measures (both self-report and performance-based)? Can the CWS discriminate between various clinical presentations, personality traits, or symptoms? Over the past twenty years, numerous published studies have addressed the validity of the Crisi Wartegg System, providing an empirical foundation with which we can begin to answer these questions.

In reviewing clinical research related to the validity of the CWS, we will first examine studies addressing construct validity, in particular the convergent validity (agreement with established personality tests) of the CWS. As the CWS is considered a projective test, performance-based measure, free-response task, or stimulus-attribution test, it is not designed to be (nor expected to be) face valid. Similarly, while informed by the theories of personality, the Wartegg Drawing Completion Test is not considered to demonstrate content validity. This lack of face validity and content validity results in reduction of clients' anxiety, skirting of defenses, and increased participation in assessment tasks. Following a review of convergent validity, published and presented case studies will be discussed, demonstrating the clinical utility of the CWS at the individual level.

CONSTRUCT VALIDITY

Construct validity is "the degree to which a test measures what it claims, or purports, to be measuring" (Cronbach & Meehl, 1955). Convergent and discriminant validity are the two subtypes of validity that make up construct validity. Convergent validity refers to the degree to which two measures of constructs that theoretically should be related are, in fact, related. In contrast, discriminant validity tests whether concepts or measurements that are supposed to be unrelated are, in fact, unrelated (Campbell and Fiske, 1959).

Similarities between the Wartegg and other personality measures, including the Rorschach Inkblot Test (Rorschach, 1921, 1942) and the MMPI-2 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 1989), have been frequently described by numerous Wartegg scholars over the past two decades (see Crisi, 1999, 2008; Fontana, 1984, 2005; Gardziella, 1985; Mattlar, 2005; Scarpellini, 1962, 1964; for a review). These authors have generally affirmed the complementarity of the Wartegg with other measures of personality, even acknowledging surprise at the diagnostic capacity of the seemingly simple Wartegg stimuli (Fontana, 2013).

Concerning the CWS specifically, convergent validity has been studied as related the MMPI-2, the Guilford-Zimmerman Temperament Survey (GZTS; Guilford & Zimmerman, 1949), the Rorschach Inkblot Test, the Separation Anxiety Test (SAT; Klagsbrun & Bowlby,1976; Italian revision: Attili, 2001), the Adult Attachment Projective (AAP; George & West, 2012), and others. Convergent validity has further been evaluated as compared to independent psychiatric ratings, genetic testing, and expectations based upon client diagnosis, economic advantage, learning style, and psychiatric symptoms and impairments.

Early Exploratory Convergent Validity Studies

During the development of the CWS (see Chapter 1), several early exploratory validity studies were completed with children diagnosed with learning disabilities, children who exhibited articulation difficulties including stuttering, and Deaf children. For the purposes of these initial exploratory research studies, the term *Deaf* includes children who were deaf and Hard of Hearing. This term is

capitalized in respect and reference to the Deaf population that is culturally Deaf and uses sign language. As explained by Padden and Humphries (1988):

> We use the lowercase deaf when referring to the audiological condition of not hearing, and the uppercase Deaf when referring to a particular group of deaf people who share a language – American Sign Language (ASL) – and a culture. The members of this group have inherited their sign language, use it as a primary means of communication among themselves, and hold a set of beliefs about themselves and their connection to the larger society. We distinguish them from, for example, those who find themselves losing their hearing because of illness, trauma or age; although these people share the condition of not hearing, they do not have access to the knowledge, beliefs, and practices that make up the culture of Deaf people. (p. 2)

In total, four studies were completed by Crisi and colleagues, spanning over a decade, investigating differences between control group children and those who were Deaf, demonstrated articulation difficulties, or were diagnosed with learning disorders. The purpose of these studies was to gather exploratory data regarding differences between groups, as well as determine the effectiveness and validity of the WDCT with populations that had previously been viewed as challenging to assess in a valid and fair manner. Some researchers, for example, have noted, that the dearth of appropriate psychological instruments to assess the Deaf population may contribute to previously reported findings of differences between groups (Iqbal & Hall, 1991). Moreover, authors have asserted that given potentially immature language abilities of Deaf children, rapport is difficult to establish and therefore, frustration and anxiety may be more common in assessment situations (Bailly, Dechoulydelenclave, & Lauwerier, 2003). Given these findings, the WDCT was identified as a potentially fair and inclusive test measure to use with Deaf, learning disabled, and stuttering children, given the reduced expressive language requirements.

The first set of studies, conducted as part of normative data collection at the *Istituto di Ortofonologia* of Rome (an institution that provides assessment and intervention services to children with speech and language disorders), focused on differences between all groups mentioned above. As the remainder of exploratory research summarized here focused solely on the Deaf population, a brief literature review is provided below summarizing previous research related to this topic.

Brief Literature Review of Mental Health Symptoms and Distress in the Deaf Population

In reviewing relevant research, authors have noted that Deaf and Hard of Hearing populations have often been associated with both psychological and neuropsychological disorders, but little agreement regarding etiology of these differences has existed between researchers. The World Health Organization cites negative functional, social, and emotional impact as prevalent within the Deaf community (WHO, 2017). Historically, Deaf individuals have been described as emotionally and socially immature, with characteristics of lability and behavioral explosiveness (Denmark, 1966). Hindley, Hill, McGuigan, and Kitson (1994) found the presence of psychological disorders in approximately 50% of the Deaf population studied. Further authors have consistently noted significant levels of psychological symptoms among Deaf individuals (Checinski, 1991, as cited in Fellinger et al, 2007; de Graaf & Bijl, 2000). These symptoms have included depression (Werngren-Elgström, Dehlin, & Iwarsson, 2003), suicide risk (Turner, Windfuhr, & Kapur, 2007), anxiety (Kvam, Loeb, & Tambs, 2006), impulsivity and externalizing behaviors (Barker, Quittner, Fink, Eisenberg, Tobey, & Niparko, 2009), obsessive and compulsive behaviors (Kitson & Fry, 1990; Marschark, Mouradian, & Halas, 1994), and social challenges (Ridgeway, 1997).

While these findings are assumed to be the multi-determined, including biological, environmental, and subjective causal factors, several authors have also offered specific theories as to the nature and etiology of these mental health symptoms. Hindley, Kitson, and Leach (2000) highlighted the difficulty of Deaf children in establishing emotional relationships with their mothers, whereas Palermo and Rhodes (2007) asserted that later psychopathology may stem from difficulties in recognizing the emotional facial expressions of others. Others (Vaccari & Marschark, 1997) have

asserted that parent-child communication styles, and resulting frustration, may increase psychiatric symptoms, whereas Gallagher (1999) posits both interpersonal and intrapersonal language development as crucial to developmental of psychiatric symptoms. Further, authors suggest considerable heterogeneity within the Deaf population, including significant differences between Deaf children born into Deaf families as opposed to those born into hearing families (Rainer, 1963). As noted above, some of these etiological theories and research findings may stem from the lack of appropriate and effective psychological test measures to assess the Deaf population, historical biases, communication-based administration challenges, and cultural differences between assessors and individuals being assessed.

More recent literature review (Bailly, Dechoulydelenclave, & Lauwerier, 2003) suggests that deafness is a multifaceted condition with medical and social aspects, including developmental communication challenges and social-cultural-familial impact. Despite the assessment difficulties and potential biases described above, the authors listed several symptoms prevalent in the population of Deaf children studied, including delays in social maturity; however, the authors asserted that the prevalence of psychopathological disorders generally appears no higher in Deaf children when compared to their hearing counterparts. Exceptions to this include autism and attentional disorders, which appear more prevalent in the Deaf population.

Citing difficulties in assessment, Fellinger and colleagues (2005) assessed mental distress and quality of life in 233 Deaf research participants, using written measures paired simultaneously with computer-based sign language presentation. Results demonstrated significantly elevated levels of social distress in the Deaf population, as compared to the general population. Quality of life was found to be negatively impacted in the domains of physical and psychological functioning, whereas social relationships were not impacted (with results similar to those of the general population). Primary identified symptoms included emotional and conduct problems, "interpersonal sensitivity," insecurity, and feelings of inferiority. The authors suggested several potential reasons for higher levels of mental distress in the Deaf population, including lack of developmental communication opportunities, increased likelihood of victimization or trauma experienced by members of the Deaf community, on-going cultural ostracism by those in a predominantly hearing culture, and associated cerebral pathology. The strong ties within the Deaf community, including sign language as a common communication and cultural system, were noted as robust protective factors for Deaf individuals.

Bianchi di Castelbianco, Crisi, & Di Renzo (1993, 1996)

Given previous research findings, a series of studies examining the ability of the CWS to distinguish between normal and historically challenging-to-assess experimental groups of children were conducted at the *Istituto di Ortofonologia* of Rome, beginning with the collection of the first childhood normative sample. In the current study, the authors reported significant differences between the childhood normative sample and three clinical groups of children.

The clinical groups were comprised of children who stutter (Stutter: N=98; 29 female, age: $\bar{x}$=9.4138, SD=2.822; 69 male, age: $\bar{x}$=8.9565, SD=2.825; range: 7-14), children with diagnosed learning disabilities (LD: N=72; 14 female, age: $\bar{x}$=9.25, SD=2.418; 48 male, age: $\bar{x}$=9.125, SD=2.208; range: 7-14), and Deaf children, including Hard of Hearing children (Deaf: N=102; 48 female, age: $\bar{x}$=8.833, SD=2.562; 54 male, age: $\bar{x}$=9.277, SD=2.149; range: 7-14). Each experimental group was divided into three age bands: 6-7 years (Stutter: N=37; LD: N=22; Deaf: N=35), 8-10 years (Stutter: N=31; LD: N=29; Deaf: N=42), and 11-14 years (Stutter: N=30; LD: N=21; Deaf: N=25). These clinical groups were compared to a control group comprised of children from the original childhood normative sample (N=282; 113 female, age: $\bar{x}$=9.3602, SD=2.6043; 169 male, age: $\bar{x}$=9.0953, SD=2.549; range: 7-14; See Appendix A for further information). Three trained psychologists administered all control group evaluations, with two different clinicians administering experimental group protocols. All protocols were scored by an expert scorer, blind to

group membership.

Initial investigation examined differences between the control group and the Deaf group. Specifically, mean differences between groups in 46 CWS interpretive data points were examined, including 22 Content categories, 17 Special Score frequencies, 4 movement categories, and 3 Impulse Response ratings. Differences were quantitatively tallied by the boxes in which they occurred. Statistically significant results ($p<.05$) were discovered for each age group studied. In the 6-7 year age group, the most statistically significant differences (4) were discovered in Box 6; that is, in Box 6, the Deaf group differed significantly from the control group on four investigated indices. Similar results were noted for the 8-10 year age group, with 7 statistically significant differences noted in Box 6, and a high number of differences also noted in Box 1 (5) and Box 2 (5). The 11-14 year age group demonstrated the most statistically significant differences in Box 1 (4). The authors concluded that in the first two developmental age bands, differences in the Deaf group are related to interaction with the environment and perceptual challenges (evidenced by difficulties in Box 6). As they mature, Deaf students, consistent with research literature, may begin to experience feelings of inadequacy, inferiority, insecurity, and limited self-concept (as evidenced by differences in Box 1).

Similar analysis was subsequently undertaken with the sample of children diagnosed with learning disabilities. In the 6-7 year age group, the most statistically significant differences (9) were discovered in Box 5. In the 8-10 year age group, the most statistically significant differences were noted in Boxes 4 (6) and 3 (5). Similarly, the 11-14 year age group demonstrated the most statistically significant differences in Box 4 (5). The authors concluded that students with learning disabilities initially react to the environment with frustration and anger (as evidenced by differences in Box 5), with this anger later converting into consistent difficulties interacting with authority figures and meeting environmental expectations (as evidenced by differences in Box 4). Some degree of inconsistency in performance and goal-achievement is likely (as evidenced by differences in Box 3).

Statistical analyses evaluated the differences between each clinical group and the control group, both overall and by corresponding age band. Specifically, mean differences between groups were investigated. See Results of overall ANOVA analyses for major CWS indices in Table 2.10.

Table 2.10
Differences between Control Group and Clinical Groups (All Ages): Major CWS Indices

	AQ+% (*p*)	**FQ+%** (*p*)	**A/F*** (*p*)	**P%** (*p*)	**P+%** (*p*)	**O%** (*p*)	**O+%** (*p*)	**IIT-1** (*p*)	**AS** (*p*)	**CB** (*p*)
Stuttering	62	55	1.25	23	65	20	38	.36	.21	.15
	---	.001	.01	.05	.05	.001	---	.001	.05	.001
Learning Disabled	65	50	1.65	24	53	23	35	.42	.22	.20
	---	.001	.01	---	.01	.001	---	.001	.05	.001
Deaf	58	51	1.3	17	55	20	35	.32	.22	.09
	.05	.001	.001	.001	.01	.001	---	.001	.01	.01
Control	62	65	1	26	72	12	34	.19	.15	.03
ANOVA**	.05	.0001	.0001	.0001	.0001	.0001	---	.0001	.0001	.001

Note. *A/F calculated by converting ratio into decimal. **p values for ANOVA.

Reprinted with permission from: Bianchi, F., Crisi, A., & Di Renzo, M. (1996). *Il test di Wartegg nell'età evolutiva. Un contributo psicodiagnostico allo studio clinico della balbuzie, della sordità e dei disturbi dell'apprendimento* [The Wartegg Test in childhood. A clinical contribution to the study of stammering, deafness, and learning disorders]. Rome, Italy: Edizioni Magi.

As expected, control group subjects demonstrated significantly higher levels of Form Quality (FQ+%) and more balanced Affective Stability Indexes (A/F), suggesting a higher level of cognitive development and affective control. Control group subjects further demonstrated higher levels of Popular (P%) responses and lower Original (O%) responses, suggesting more conventional thinking. Lastly, the clinical groups demonstrated a significantly higher Index of Inner Tension (IIT-1) than control group subjects, indicating greater levels of internal tension, distress, state anxiety and impulsivity. When investigating by age band, the highest number of significant differences was noted in the 8-10 year group, with lowest number of significant differences noted in the 11-14 year age band. This suggests that with continued maturation, clinical group members may experience a decrease in symptoms and appear more similar to the control group.

Further analysis was conducted for each age band of each clinical group, as compared to the control group. These analyses were compounded by "missing data" in the clinical groups, due to Global Rejections, meaningless drawings, or incomprehensible explanations/verbalizations of depicted content. While these phenomena are rare in the normative population, the clinical groups demonstrated significantly higher percentages of missing data, which in and of itself is an important clinical finding. Table 2.11 notes missing data percentages and statistically significant differences from the control group for each age band of each clinical group.

Table 2.11
Comparisons between Clinical and Control Groups:
Missing Data Percentages and Statistically Significant[1] Differences in Content, Special Scores, Movement (M/m), and Impulse Responses (IR)

Group	Age Band	Missing Data%[3]	Statistically Significant Differences			
			Contents[2]	Special Scores	M/m[2]	I.R.
Stuttering	6-7	3.7%	ARC-, BOT+, FD-, HS+, SIG-	AS, CB, PP, AP, GR, PR	---	---
	8-10	0%	ARC+, AST+, BOT-, FD-, OBJ-, SIG+	AS, CB, RB, II, PP, AP, SR, TR	M+ MI+	---
	11-14	2.9%	A-, ARC+, AST+, BOT-, HS-, NAT+, OBJ-, SIG+	AS, CB, GR, SR	M+	---
Learning Disability	6-7	2.3%	A-, ARC-, AST+, BOT+, CLD+, NAT+, SIG-	AS, CB, II, PP, AP, PR, SR	MI+	---
	8-10	1.3%	A+, ARC+, BOT-, H+, NAT+, CLD+, OBJ-	AS, CB, II, PP, AP, GR, PR, SR	M+	---
	11-14	0.6%	A-, FD+, SIG+	AS, CB, PP, GR, PR, SR	M-	---
Deaf	6-7	16%	ARC-, BOT+, OBJ-, SIG-	AS, CB, II, SR	M- MI+	---
	8-10	10%	A+, ARC-, AST+, H-, OBJ-, SIG-	PA, AS, CB, II, AP, GR, PR, SR, TR	M-	---
	11-14	10.5%	ARC+, AST+, BOT-, H-, OBJ-, NAT+, SIG-	AS, CB, II, AP, GR, PR, SR	M- MI+	---

Note. [1]all differences at $p<0.5$ level.

[2]+= clinical group demonstrates higher level than control, -= clinical group demonstrates lower level than control group.
[3]Missing Data % for Control Group: 6-7: 0.2%; 8-10: 0.2%; 11-14: 0.3%.

Results suggested significant differences between control and clinical groups, primarily in content categories, presence of special scores, and movement responses. Overall, clinical groups generally demonstrated lower percentages of adaptive Content categories (i.e., Human [H], Object [OBJ], Symbol [SIG]), while simultaneously demonstrating higher levels of Content categories suggesting detachment and internal conflict (i.e., Astronomical [AST], Cloud [CLD]). Some tendencies toward distanced compensatory sensitivity to others (i.e., Nature [NAT] and Botany [BOT]) were noted in the clinical groups. Similarly, without exception, the clinical group at all age bands demonstrated higher frequency of Special Scores denoting cognitive difficulties (i.e., Perseveration [PP], Stimulus Repetition [SR], Inadequate Integration [II],), internal conflict (i.e., Global Rejection [GR], Partial Rejection [PR], Anxiety Stroke [AS], Crossed Border [CB], Reversed Box [RB]) and thinking challenges (i.e., Arbitrary Performance [AP], Personalized Answer [PA], Transparency [TR]). In terms of movement, clinical groups, in general, demonstrated higher presence of Secondary Movement (m), suggesting the presence of internal conflict and distress. The Deaf group, in particular, also demonstrated lower levels of Primary Movement (M) as compared to controls, suggesting potentially limited cognitive control, fewer internal cognitive resources, and a reduction in coping skills. No statistically significant differences were noted between groups in regards to Impulse Responses.

Overall, the authors concluded that the CWS identifies significant differences between groups, both in terms of box-by-box differences, as well as via aggregate differences in CWS indices, Content categories, Special Scores, and Movement responses.

Crisi (2002)

Extending the work of Bianchi di Castelbianco and colleagues (1993, 1996), this study further investigated differences in levels of psychopathology between Deaf and control groups. The goal of the study was to investigate the effectiveness of the CWS in detecting group differences, with the aim to determine the utility of the CWS in working effectively with the Deaf population in a valid and non-pathologizing manner.

One hundred participants were originally included in the clinical group, although 36 were subsequently excluded due to communication difficulties between the examinee and examiner or difficulties in comprehending the instructions of the CWS. The remaining 64 participants (39 male, 25 female) were divided into three age groups: 7-13 (*N=19*, 9 female; 10 male, age: $\bar{x}$=9.895, *SD*=1.9971), 14-19 (*N*=23, 8 female, 15 male; age: $\bar{x}$=16.478, *SD*=1.9038), 20 or above (*N*=22, 8 female, 14 male; age: $\bar{x}$=25.273, *SD*=4.4953). Level of hearing loss was determined as Severe in 23 participants, and Moderately-Severe in 41 participants. In further analyses, no statistically significant differences were discovered between groups, related to either age or gender.

The clinical group described above was compared to a control group comprised of children, adolescents, and young adults (aged 7-19) from the original CWS normative sample (*N*=282; 113 female, age: $\bar{x}$=9.3602, *SD*=2.6043; 169 male, age: $\bar{x}$=9.0953, *SD*=2.549; range: 7-19; See Appendix A for further information*).* Initial analysis investigated the ability of the CWS to differentiate between groups based upon the Global Assessment classification assigned to each participant's protocol. In completing the CWS scoring mechanics, each protocol is assigned an overall Global Assessment classification, ranging from "No Pathology" (NOP) to "Pathological"

(PTL). This classification rubric is divided into five levels of increasing severity (1) NOP, 2) LSI, 3) MSI, 4) SSI, 5) PTL), reflecting the degree of symptoms and functional impairment likely to be experienced by the examinee. Two of these five levels are considered positive classifications ("No Pathology" [NOP] + "Low Symptoms" [LSI]) reflecting limited pathology and functional impairment; one is considered indicative of moderate impairment ("Moderate Symptoms" [MSI]); and two are considered negative classifications ("Severe Symptoms" [SSI] and "Pathological" [PTL]) representing severe impairment in functioning. Clinical and control groups were compared based on their classification within this Global Assessment rubric. It was hypothesized that the experimental group would produce higher rates of negative classifications (i.e., SSI + PTL) as compared to control subjects. Results of chi-square analyses are presented in 2.12.

Table 2.12
Differences in Positive, Neutral, and Negative Global Assessment Classifications: Deaf Group vs. Control Group

Global Assessment	Group	Present		Absent		χ^2	*p*	Φ^1
		N (%)	χ^2	N (%)	χ^2			
Positive	Deaf	19 (29.7%)	62.56	45 (35%)	20.03	99.42	.00001	.58
	Control	209 (89.3%)	5.22	25 (10.7%)	16.78			
Neutral	Deaf	17 (26.6%)	8.32	47 (73.4%)	1.28	12.21	.001	.20
	Control	23 (9.8%)	2.26	212 (90.2%)	0.35			
Negative	Deaf	28 (43.8%)	72.13	36 (56.2%)	8.07	102.13	.00001	.59
	Control	2 (0.9%)	19.73	232 (99.1%)	2.21			

Note. Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Results of chi-square analyses indicate significant differences between clinical and control groups, with the control group generally demonstrating a more positive global assessment, and the clinical group demonstrating a more negative global assessment as predicted. The author concluded from this initial analysis that a condition of impairment is present in the Deaf group, although asserted that this does not necessarily indicate that Deaf individuals demonstrate clinical psychopathology; rather, a level of inner tension or distress is present, which likely impacts personality development and everyday functioning. This heightened internal level of distress may limit a more functional and balanced development, as suggested by other research.

As follow-up to the initial investigation, differences between clinical and control groups were similarly investigated for each box of the WDCT. In the Analysis of Sequence 1, an interpretive strategy for the CWS, each of the eight test boxes is assigned a positive or negative Code computed mathematically by summing the Evocative Character (EC) and Affective Quality (AQ) of each box. Six Codes are possible in each box, 2 considered positive (C, PC) and 4 considered negative (NC, AC, AD, D). Frequencies of positive versus negative Codes for both groups were analyzed by chi-square for each box of the WDCT. It was hypothesized that the control group would produce significantly more positive codes per box than the clinical group. Results of these analyses are presented in Table 2.13.

Table 2.13
Frequency of Negative and Positive Codes by WDCT Box:
Differences between Deaf (*N=39*) and Control Groups (*N=234*)

Box	Group	Frequency [Negative] (χ^2)	% [Negative]	Frequency [Positive] (χ^2)	% [Positive]	χ^2	*p*	Φ^1
Box 1	Deaf	24 (10.77)	61.5	15 (5.04)	39.5	18.44	.0001	.26
	Control	63 (1.80)	26.9	171 (0.84)	73.1			
Box 2	Deaf	18 (10.87)	46.2	21 (3.00)	53.8	16.17	.0001	.24
	Control	41 (1.81)	17.5	193 (0.50)	82.5			
Box 3	Deaf	16 (2.44)	41	23 (0.94)	59	3.93	.05	.12
	Control	60 (0.41)	25.6	174 (0.16)	74.4			
Box 4	Deaf	18 (0.13)	46.2	21 (0.10)	53.8	0.27	.6023	.03
	Control	98 (0.02)	41.8	137 (0.02)	58.2			
Box 5	Deaf	19 (1.35)	48.7	20 (0.80)	51.3	2.50	.1133	.1
	Control	83 (0.22)	35.4	151 (0.13)	64.6			
Box 6	Deaf	9 (9.94)	23.1	30 (0.91)	66.9	12.66	.001	.22
	Control	14 (1.66)	6.1	220 (0.15)	93.9			
Box 7	Deaf	25 (0.54)	64.1	14 (0.67)	35.9	1.42	.2329	.07
	Control	126 (0.09)	53.7	108 (0.11)	46.3			
Box 8	Deaf	14 (7.91)	35.9	25 (1.64)	64.1	11.14	.001	.20
	Control	33 (1.32)	14.1	201 (0.27)	85.9			
	Note. Positive Codes: C+PC; Negative Codes: NC+AC+AD+D. Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).							

Overall, significant differences were noted between groups, with the Deaf population generally demonstrating more negative codes in all boxes. Boxes 1, 2, 3, 6, and 8 demonstrated statistically significant differences between groups in the expected direction. Boxes 4, 5, and 7 demonstrated this trend, albeit without significance. The author concluded that experimental hypotheses were confirmed.

As related to Box 1, 61.5% of Deaf participants produced negative codes, as compared to only 26.9% of control group subjects. This suggests that for the clinical group, self-evaluation is more

likely to be based in insecurity, lower self-esteem, and lack of integration of ego functions. Significant differences in Box 2 indicated more negative codes in the clinical population (46.2%) as compared to the control group (17.5%). This, consistent with research conducted by Hindley, Kitson, and Leach (2000), suggests that the clinical group demonstrates greater challenge in entering into contact with the environment, which may stem (in part) from difficulties communicating with the maternal figure during development. Given the lack of significance in Box 4 (the box of the paternal figure), lack of verbal and non-verbal communication with the mother in particular appeared related to observed psychopathological indicators.

Box 3 illustrates the amount of energy available to individuals in service to the ego, facilitating adaptation to the environment; that is, the internal energy and drives available to the individual for productive use. Consistent with other results, the Deaf population produced significantly more negative codes (41.0%) as compared to the control group (25.6%). Considering the Jungian psychoanalytic understanding of affective complexes (see Jung, 1960 for a review), the authors hypothesized post-hoc that this lower score in Box 3 was likely related to the presence of affective difficulties draining available energy to ego functioning.

One of the most sensitive boxes on the WDCT, Box 6 relates to judgment, reality testing, one's capacity to structure and organize the environment, the ability to manage and resolve challenges, and general cognitive/rational processes. Similar to other studies, significant differences were noted, with the Deaf group demonstrating a higher frequency of negative codes (23.1%) as compared to the control group (6.1%). Confirming previous research, Box 6 appears one of the most sensitive differentiating measures between the Deaf and normative populations.

Lastly, Box 8, which measures areas of socialization, social skills, interactions with others, and the ability to enter into mature, socially flexible, and genuine relationships, differed significantly between groups. Again, the Deaf group produced a higher frequency of negative codes (35.9%) as compared to the control group (14.1%). On the basis of previous clinical literature citing interpersonal and social reciprocity challenges in the Deaf population, this finding was both expected and confirmed.

Overall, data from previous research studies was confirmed by the present research. Box 6 (rationality and cognitive processes) was the strongest differentiator between groups. Boxes 3, 4, and 8 indicate interruptions in interests and energy, as well as potential challenges in interpersonal relationships. Lastly, consistent with past studies, significant difficulties were noted in Box 2, suggesting sensory deficits may negatively impact both the relationship to mother and the sensitivity to others.

Given that these studies were conducted with children and adolescents, it is likely that these observed discrepancies may reduce or normalize during development (as suggested by previously reviewed CWS research). Citing the protective factors of the Deaf community, these differences are likely more visible during the childhood developmental period prior to full immersion in the Deaf culture and full acquisition of sign language abilities.

Bianchi di Castelbianco, Crisi, & Palermo (2002)

In continuing their research investigating the Deaf and Hard of Hearing population, Bianchi di Castelbianco and colleagues investigated differences within the Deaf community as compared to heterogeneous and homogenous genotypes of the Serotonin Transporter Gene, linked to anxiety and internal distress via previous research.

In reviewing previous research, the authors noted a main objective of the study to include providing a description of Deaf individuals' personality via non-language based assessment instruments, rather than potentially over-pathologizing the Deaf community through use of instruments designed for a hearing population. As such, the CWS was utilized, because of its reduced verbal language demands. Further, genetic studies were utilized, specifically focusing on polymorphism in genes that transmit serotonin, a neurotransmitter linked to mood disorders and stress, and believed to play an important role in brain development. While no definitive proof has been established between polymorphic gene-influenced serotonin-transmission and neuropsychological disease, authors (Melke, Landén, Baghei, Rosmond, Holm, Björntorp, Westberg, Hellstrand, & Eriksson, 2001) have described an association between S alleles (in the gene in question) and traits of anxiety. Therefore, the specific aim of the current study was to demonstrate that the presence of anxiety in some Deaf individuals is not due solely to the inability to hear, but rather also represents a genetic vulnerability at the polymorphic level.

More recent research highlights the link between the S allele of the S/S genotype and affective disorders, including major depressive disorder, unipolor or bipolar depression, and seasonal affective disorder (as described in Margoob & Mushtaq, 2011). In some of these cases, heterogeneous genotype (i.e., L/S) appeared to constitute a protective factor as related to environmental stress. Conversely, individuals with homogenous genotype L/L who were exposed to environmental adversity displayed significantly higher rates of depressive and anxiety disorders (Laucht, Truetlein, Blomeyer, Buchmann, Schmid, Becker, et al., 2009), replicated by Zhang, Xu, Xu, Yang, Luo, Sun, Sun, Wang, and Shen, 2009). For comprehensive summary of current research on genetic sensitivity and the Serotonin Transporter Gene, see Caspi, Hariri, Holmes, Uher, and Moffit (2010).

54 Deaf participants (36 male, 18 female) were included in the current study, representing a subset of the sample included in a larger study. Degree of hearing loss was distributed from mild to severe (Mild=2, Moderate=12, Severe=24). Genetic testing related to the polymorphic Serotonin Transport Gene in question identified three types of alleles: L/S (heterogeneous; *N=32*; 22 male, 10 female; age: $\bar{x}$=17.5), S/S (homogenous; *N=4*; 3 male, 1 female), and L/L (homogeneous; *N=18;* 11 male, 7 female; age: $\bar{x}$=14).

To ascertain differences between allele groups, the authors evaluated categorical differences related to overall Global Assessment assigned to each participant's CWS protocol. As described above, in completing the CWS scoring mechanics, each protocol is assigned an overall Global Assessment classification, ranging from "No Pathology" (NOP) to "Pathological" (PTL). This classification rubric is divided into five categories, two of which are considered positive (NOP + "Low Symptoms" [LSI]) and three of which can be considered negative (PTL + "Moderate Symptoms" [MSI] and "Severe Symptoms" [SSI]). Clinical and control groups were compared based on their classification within this Global Assessment rubric. Given previous research, individuals with the L/L homogenous genotype were expected to produce significantly more negative codes, as compared to heterogeneous L/S counterparts. For the purposes of analysis, the S/S group was not included due to small sample size. As expected, significant differences between the L/S and L/L groups were identified, presented in Table 2.14.

Table 2.14
Frequency of Distribution of Positive and Negative Global Assessment Classifications: Differences with the Deaf Group* between Genotype L/S and Genotype L/L Groups

Group	Frequency [Negative] (χ^2)	% [Negative]	Frequency [Negative] (χ^2)	% [Negative]	χ^2	*p*	Φ^1
L/S	18 (0.46)	56	14 (0.89)	44	3.7656	.05	.27
L/L	15 (0.82)	83	3 (1.59)	17			

Note. Positive Global Assessment classification= NOP + LSI; Negative Global Assessment classification= MSI + SSI + PTL.
*When Genotype groups (L/S and L/L) were compared to matched adolescent control sample, statistically significant differences were noted (χ^2=8.2305, *p*=.016) with clinical groups demonstrating greater frequency of negative Global Assessment.
Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

In examining the results, 83% of individuals with the homogenous L/L genotype produced a negative global assessment, as compared to 56% of the heterogeneous L/S genotype. The authors concluded that this confirms the experimental hypotheses, and suggests the potentially protective factor of allele heterogeneity of the Serotonin Transport Gene. The authors further asserted that these findings support previous research identifying anxious characteristics within the Deaf community, specifically higher in the homogenous allele group (Jorm, Prior, Sanson, Smart, Zhang & Easteal, 2000). Additionally, while a small sample size, these findings give biological support to the utility of the CWS in diagnostic assessment and follow-up.

Bianchi di Castelbianco, Crisi, Palermo, Palladino, Sommaruga, Sgueglia, & Vichi (2002)

In further investigating differences between the Deaf and normative population, Bianchi di Castelbianco and colleagues conducted a pilot study with Deaf children. Citing literature indicating the presence of psychological discomfort and difficulty in the Deaf population (Hindley, Kitson, & Leach, 2000), the authors hypothesized that deprivation of hearing during preverbal development likely limits exposure to emotional content and information, which may impact relationships with primary attachment figures. Given this expectation, specific hypotheses of this study predicted that Deaf children, as compared to their matched control counterparts, would demonstrate more negative codes in WDCT Boxes 2 and 4 (representing relationship to the maternal and paternal figure, respectively). Additional expectations, based upon previous literature and research, predicted more negative values in Boxes 1 and 6 for Deaf children, reflecting higher levels of self-doubt and recognition of greater difficulties, and less cognitive flexibility. The authors asserted that early intervention during this developmental period must focus not only on communication, but also include simulation of emotional and relational skills and abilities.

The pilot study included 33 participants, divided into 3 groups. The first group (*N*=11) was comprised of children participating in language-based therapeutic intervention. The second group (*N*=13), was comprised of children participating in gestural rehabilitation therapy, including therapeutic intervention provided via sign language. The third group (*N*=9) was receiving a combination of the two above-mentioned therapeutic interventions. The three groups were compared to a matched control sample with no diagnosed hearing loss. Comparisons between the Code of each WDCT box were made between the four groups (3 clinical, 1 control). Given the

small number of participants, the study was considered exploratory, and while significant findings were reported, no statistical data was provided by the authors in publication.

Overall results confirmed the experimental hypothesis, with clinical groups (specifically the first two groups, each of which received one form of intervention in isolation) demonstrating less positive codes in the boxes of male (Box 4) and female (Box 2) relationships, related to parents when assessing children. Additionally, all three groups demonstrated significant differences in the predicted direction as compared to the control group, suggesting less positive self-concept in the clinical groups (Box 1), and a strong tendency toward rationality and rigidity of thought (Box 6). Additional findings, not hypothesized by the authors, indicated less positive codes in Box 3 for Deaf children, suggesting potential difficulty in planning and maintaining a productive direction of psychic energies. Overall, these findings were noted to be consistent with literature and previous research.

CONVERGENT VALIDITY

More recent convergent validity studies have examined the CWS in relationship to a variety of symptoms, attachment styles, and diagnoses and conditions, in both the clinical and selection fields. Organized by focus, CWS convergent validity research will be reviewed as related to: learning disorders, general psychiatric conditions (including psychosis), depression and suicide, trauma and adjustment conditions (including work abuse), eating disorders, psychiatric symptoms associated with medical conditions, attachment, and test use within special situations (including infertility). CWS convergent validity in the selection field will be reviewed. Lastly, the ability of specific CWS content scores to indicate pathology and personality traits will be discussed.

Learning Disorders

Crisi (2006)

Following earlier research related to children with learning disabilities, Crisi hypothesized that children with learning disorders would produce more movement responses on the Wartegg, but also would produce a higher number of Global Rejections as compared to control group children. These hypotheses were based in the guiding beliefs of the *Istituto di Ortofonologia*, where the research was conducted, that observed learning disabilities often reflect emotional or psychoaffective symptoms, rather than organic cognitive deficits. Given this theory, it was hypothesized that children with learning disabilities would demonstrate average to above average cognitive abilities, as evidenced by average or above average primary movement scores (M) and positive codes in Box 6, the box of rational thought. At the same time, it was hypothesized that children with learning disabilities would demonstrate greater levels of internal tension and distress, indicated by above average secondary movement scores (FM, MI, FE) and higher levels of internal conflict, indicated by the presence of the special score Global Rejection (GR).

To investigate these hypotheses, 195 children (LD: 125 male, 70 female) independently diagnosed with learning disabilities were compared to 971 control group subjects (504 male, 467 female). The experimental group was divided into two age bands: 6-7 years old (CN: N=118; 81 male, 37 female) and 8-10 years old (N=77; 44 male, 33 female). No significant differences between groups were noted concerning age or gender.

In terms of overall cognitive abilities, the experimental group (LD) produced a higher number of primary movement (M) responses than the control group (CN), as predicted by the author (LD: 2.82; CN: 2.05, p<0.05). While not statistically significant, the experimental group also produced

higher levels of secondary movement, as predicted (FM: LD=0.73, CN=0.44; MI: LD=5.06, CN=4.68), with one exception (FE: LD=0.83, CN=1.78, $p<.01$). As Facial Expression (FE) movement responses are thought to the reflect expressed emotion, the author hypothesized post-hoc that reduced FE scores in the experimental group may reflect the inability of children with learning disabilities to express their emotions, or a tendency to constrict or internalize these emotions leading to a deleterious impact on cognitive functioning. Additionally, as predicted, no significant differences were noted in Box 6 codes between experimental and control groups, suggesting that children in the learning disorder group demonstrated no more difficulties in rational thinking than their control group peers.

Along with secondary movement scores, the presence of Global Rejections was investigated with the clinical group producing a significantly higher percentage of this special score than their control group counterparts. Specific statistically significant differences were noted in Box 1 (LD: 0.19, CN: 0.04, $p<.05$), Box 2 (LD: 0.29, CN: 0.11, $p<.05$), Box 4 (LD: 0.49, CN: 0.2, $p<.01$), and Box 5 (LD: 0.44, CN: 0.17, $p<.01$). Given these findings, the author concluded that diminished cognitive performance resulting from psychoaffective disturbance is likely related to several factors, including self-concept (Box 1), general emotional sensitivity (Box 2), and reactions to authority (Box 4, Box 5). Overall, results suggested that children with learning disabilities demonstrate adequate cognitive abilities (M, Box 6 code), although may experience greater degrees of internal tension (FM, MI), may have a harder time expression emotions (FE), and likely experience greater psychoaffective disturbance (GR).

General Psychiatric Conditions

Benedetti, Bologna, Crosato, Favretti, Giusti, Lestingi, Sangiorgi, & Crisi (2008)

Benedetti and colleagues conducted exploratory analysis regarding the ability of the CWS to differentiate between normative and clinical groups diagnosed with various psychiatric conditions. 56 participants, 84% referred for clinical care and assessed while participating in inpatient psychiatric treatment, and 16% participating in medical-legal evaluation for forensic reasons, comprised the clinical group (age: $\bar{x}$=29.49, SD=13.38; range=10-55). No differentiation was made related to clinical diagnosis or mental health condition. The clinical group was 53.57% female (N=30) and 46.43% male (N=26). Educational level ranged from 5 years of academic instruction (17.8%), 6-8 years (44.6%), 9-13 years (33.9%), to over 13 years of academic instruction (3.57%). No significant differences were noted within the clinical group (in terms of age, gender, or education differences) as related to scores on CWS indices. The clinical group was compared with a group of control subjects (N=299) matched in age, gender, and educational level from the Italian standardization sample.

Differences between clinical group and control group participants were examined related to 12 major CWS indices (EC+%, AQ+%, FQ+%, Affective Stability Index, P%, P+%, O%, O+%, IIT-1, IM, AI, and IIT-2). Additionally, 14 content categories were investigated (Human, Object, Architecture, Symbol, Astronomical, Nature, Botany, Animal Simulacrum, Biology, Food, Fire, Human Simulacrum, Pathology, and Cloud). Lastly, 17 Special Score categories were evaluated (Personalized Answer, Self Criticism, Crossed Border, Reversed Box, Anxiety Stroke, Morbid, Incomplete Drawing, Inadequate Integration, Missing Union, Perseveration, Arbitrary Performance, Global Rejection, Partial Rejection, Interpreted Stimulus, Disproportion, and Transparency). Given the high number of comparisons undertaken, the Bonferroni post-hoc correction was applied (Bonferroni, .05: $p<.001$).

Presented in Table 2.15, significant differences between clinical and control groups were found in 4 out of 12 major scoring indices, 1 content category, and 7 out of 17 special score categories following Bonferroni correction.

Table 2.15
Significant Differences on Major CWS Indices, Content Categories, and Special Scores: Clinical (Psychiatric Diagnosis) Group vs. Control Group

Index, Content, Special Score	*Control* $\bar{x}$	*Control SD*	*Clinical* $\bar{x}$	*Clinical SD*	*t* (354 df)	*p*	*d*[1]
Indices							
Form Quality (FQ+%)	99.76	1.576	87.15	13.19	16.044	.000*	1.70[a]
Popular Responses (P%)	19.38	10.83	23.72	11.45	-2.743	.006	0.29[b]
Original Responses (O%)	0.24	1.618	7.32	9.34	-12.233	.000*	1.30[b]
O% with Good Form (O+%)	2.34	15.14	40.49	42.78	-12.015	.000*	1.27[b]
Anxiety Index (AI)	0.94	0.15	0.84	0.20	4.372	.000*	0.46[a]
Index of Inner Tension 2 (IIT-2)	5.22	1.34	4.79	1.58	2.152	.032	0.22[a]
Content Categories							
Human Content (H%)	12.35	10.98	17.89	16.54	-3.185	.002	0.33[b]
Object Content (OBJ%)	38.19	19.58	28.48	16.11	3.523	.000*	0.37[a]
Symbol Content (SIG%)	4.90	9.31	8.82	13.42	-2.692	.007	0.28[b]
Human Simulacrum (HS%)	12.50	0.00	14.58	4.86	-1.986	.05	0.21[b]
Special Scores							
Personalized Answer (PA)	0.02	0.16	0.35	1.02	-5.267	.000*	0.56[b]
Self Criticism (SC)	0.00	0.00	0.12	0.38	-5.603	.000*	0.56[b]
Anxiety Stroke (AS)	7.46	1.29	6.65	1.63	4.161	.000*	0.44[a]
Incomplete Drawing (ID)	0.02	0.14	0.18	0.42	-5.037	.000*	0.54[b]
Inadequate Integration (II)	0.03	0.19	0.18	0.42	-4.131	.000*	0.44[b]
Missing Union (MU)	0.16	0.40	0.00	0.00	2.966	.003	0.32[a]
Perseveration (PP)	0.00	0.00	0.25	0.60	-7.051	.000*	0.79[b]
Arbitrary Performance (AP)	0.02	0.12	0.23	0.59	-5.508	.000*	0.59[b]
Interpreted Stimulus (IS)	0.00	0.05	0.11	0.55	-3.096	.002	0.33[b]

Note. *Significance noted following Bonferroni ($p<.05$) post-hoc correction; $p<0.001$.
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]Control group values greater than clinical group.
[b]Clinical group values greater than control group.

As expected, CWS indices (FQ+%, O%, O+%, AI) demonstrated higher mean values in the control group as compared to the clinical group, reflecting more adequate adjustment to the environment and to reality, but also indicating higher awareness of experienced difficulties. Popular response

percentage (P%) approached significance, suggesting greater ability of control group participants to engage in shared conceptual thinking than clinical groups participants.

Regarding primary Contents, results found lower levels of Object content (OBJ%) in clinical subjects, suggesting some disconnection from relationships in the environment. Lower levels of Human (H%) and higher levels of Human Simulacrum (HS%) were further noted in the clinical group, indicating the tendency to develop poor interpersonal relationships and to defensively avoid emotional contact. It should be noted that Animal Simulacrum (AS) content, Pathological content (PAT), Food content (FD), and Cloud content (CLD) – all of which have been linked to lower functioning and pathological conditions in other research— also approached significance in this study, with the clinical group demonstrating higher means than control group participants.

Lastly, referring to Special Scores, results demonstrate that the mean values of 7 out of 17 categories were significantly different between the clinical and control groups. In particular, clinical group members produced a higher frequency of: (1) Personalized Answer [PA], indicating an egocentric view of the world and potentially weakened self/world and reality/fantasy boundaries; (2) Self Criticism [SC], indicating low self-esteem and insecurity; (3) Anxiety Stroke [AS], suggesting higher levels of internal tension and distress; (4) Incomplete Drawing [ID], suggesting marked insecurity and indecision, as well as possible detachment from reality; (5) Inadequate Integration [II], indicating an impaired sense of reality; (6) Perseveration [PP], indicating a stereotypical style of thinking; and (7) Arbitrary Performance [AP], indicating a reduced ability to perceive reality. It should be noted that two additional special scores approached significant, including Missing Union (MU), indicating psychoaffective inhibition of intelligence, and Interpreted Stimulus (IS), suggesting a tendency to avoid conflict with an associated high level of internal tension.

Individual box Codes were also analyzed by the investigators, to identify the ability of the Analysis of Sequence 1 to differentiate between clinical and control subjects. As described previously, in the Analysis of Sequence 1, each of the eight WDCT boxes are assigned a Code computed mathematically by summing the Evocative Character (EC) and Affective Quality (AQ) scores of each box. Six codes are possible in each box, two considered positive (C, PC) and four considered negative (NC, AC, AD, D). To analyze differences between clinical and control groups, the presence of positive codes (C and PC) in each group was investigated using *t*-test analysis. The control group demonstrated a statistically significant higher mean of positive codes per protocol ($\bar{x}$=5.22, *SD*=1.34) as compared to the clinical group ($\bar{x}$=4.78, *SD*=1.58; *t*=2.152, *df*=354, *p*=.032). From these results, the authors concluded that members of the clinical group produce significantly fewer positive box codes, suggesting decreased relatedness to the environment and less positive affective experiences, and a higher number of negative codes.

Overall, the authors concluded that the CWS demonstrates preliminary ability to discriminate between clinical (psychiatric) and normative groups on several levels: normatively scored calculations and indices, presence of Special Scores, Content of drawings (scored per normative scoring rules), and Code of individual WDCT boxes (calculated per CWS guidelines). Given these findings, it was suggested that the CWS could serve as a sensitive, brief, and non-intrusive screening tool in a variety of clinical settings.

Crisi (2009)

The convergent validity of the Wartegg Index of Psychopathology (WIP) was investigated via comparison between normal (control group) subjects (*N*=307), applicants deemed unfit for military service ("Unfit": *N*=265), victims of workplace harassment and/or abuse ("Work-abused": *N*=106), and individuals independently seeking outpatient psychotherapy ("Therapy-seeking: *N*=57). To determine differences between groups, percentages of each group's distribution within the four quadrants of the WIP were calculated, with differences between groups subsequently investigated via chi-square analysis.

The Wartegg Index of Psychopathology (WIP) is a graphic representation of personality structure, plotting state anxiety and distress on the vertical axis and trait personality integration and flexibility on the horizontal access. Given the intersection of these continuous axes, four distinct personality descriptions, or "quadrants," may be derived:

> Quadrant A: *Low experienced state distress; High degree of flexibility and integration.* Quadrant A generally represents autonomy and independence and lack of significant distress, tension, or dysphoria. Given the assumption of adaptation and health, it is expected that a significant proportion of the normal population will be categorized within Quadrant A.
>
> Quadrant B: *Low experienced state distress; Low degree of flexibility and integration*. Quadrant B generally represents detachment, isolation, and in extreme cases, psychosis. These symptoms are generally not accompanied by awareness of suffering or discomfort. Given these symptoms and descriptions, it is expected that a higher percentage of individuals unfit for military service (including detachment, inability to work collaboratively with others, limited self-knowledge, antisocial attitudes, depressive symptoms, and difficulties in reality testing) might be categorized in Quadrant B.
>
> Quadrant C: *High experienced state distress; High degree of flexibility and integration*. Quadrant C generally represents tendencies toward dependence. Moreover, given a well-functioning personality structure coupled with elevated state distress, trauma, adjustment, and adaptation disorders prevail in this Quadrant. Given this, it is expected that individuals experiencing workplace harassment or abuse may more readily belong to Quadrant C.
>
> Quadrant D: *High experienced state distress; Low degree of flexibility and integration.* Quadrant D generally represents long-standing and problematic patterns of ambivalence, discomfort, conscious distress, immaturity, and conflict. Anxious conditions and Cluster B personality disorders (Borderline, Histrionic, Narcissistic) are frequently found in this quadrant. Given the long-standing nature of symptoms, and the profoundly experienced distress experienced, it is expected that individuals seeking psychotherapy may frequently be categorized in Quadrant D.

In the normal population (Italian standardization sample, N=2,300), distribution of subjects by Quadrant yielded the following results: Quadrant A (42.7%), Quadrant B (10.7%), Quadrant C (21.1%), and Quadrant D (25.4%).

Results of chi-square analysis are presented in Table 2.16.

Table 2.16 Differences in Percentage Distribution of WIP Quadrants between Healthy, Unfit, Work-Abused, and Therapy-Seeking Groups							
	Healthy %	Unfit %	***p***	Work-Abused %	***p***	Therapy-seeking %	***p***
Quadrant A%	45	43.1	---	10.3	0.001	15.8	0.001
Quadrant B%	17.3	32.1	0.001	18.8	---	10.6	---
Quadrant C%	18.6	10.1	0.01	36.9	0.001	35.1	0.01
Quadrant D%	19.3	14.6	0.01	33.9	---	38.6	0.01

In considering the results, the distribution of healthy (control group) subjects demonstrated no significant differences from the overall Italian normative sample. Statistically significant differences were noted between the control group and experimental groups. Individuals unfit for military service demonstrated a higher percentage of Quadrant B classifications as hypothesized, although also demonstrated a high percentage of classification in Quadrant A. This was explained by the authors, given that lack of fitness for military duty does not equate with psychopathology; rather, the "unfit" individuals who were classified in Quadrant A did not meet additional external criteria for military service (i.e., fitness level, military orientation, shared vision, etc.) that were unrelated to psychopathology.

Victims of work abuse and harassment demonstrated a higher percentage of trauma and adjustment difficulties (Quadrant C) as hypothesized, with a concurrent reduction in healthy adaptation (Quadrant A). Lastly, individuals seeking psychotherapy demonstrated lower levels of healthy adaptation (Quadrant A), higher levels of dependency (Quadrant C), and significantly higher levels of felt distress (Quadrant D) as hypothesized. Given these findings, the authors concluded that the CWS WIP is able to effectively discriminate between groups.

Crisi & Dentale (2016)

Crisi and Dentale investigated the convergent validity of three CWS indices--Evocative Character (EC), Affective Quality (AQ), and Form Quality (FQ)—by examining the ability of these scores to differentiate between normal ("healthy"), clinically anxious, and psychotic conditions. The CWS protocols of 564 participants (290 males, 274 females) were analyzed, with participants classified into three distinct groups based upon previous psychiatric evaluation using the DSM-IV-TR: healthy, anxious, and psychotic. The healthy group (N=401) was comprised of 186 females and 215 males (age: $\bar{x}$=19.95, SD=3.19). The anxious group (N=56, 42 male, 14 female; age: $\bar{x}$=32.23, SD=9.42) was comprised of individuals diagnosed with obsessive-compulsive disorder (N=18), generalized anxiety disorder (N=24), and panic disorder (N=14). The psychotic group (N=107, all female; age: $\bar{x}$=37.80. SD=9.71) was comprised of individuals diagnosed with paranoid schizophrenia (N=41), schizoaffective disorder (N=11), undifferentiated schizophrenia (N=6), disorganized schizophrenia (N=13), schizophrenia (no subtype) (N=6), delusional disorder, erotomanic subtype (N=1), bipolar psychosis (N=5), and unspecified psychotic disorder (N=24). Significant age [$F(2, 563)=453.44$, $p<.001$] and gender [$\chi^2= 31.38$, $p<.001$] differences were noted between the healthy, anxious, and psychotic groups, so these variables were included as covariates for subsequent analyses.

Two specific hypotheses were tested to assess the convergent validity of CWS indices. First, since healthy subjects are expected to demonstrate greater social sensitivity, balanced emotionality, and greater reality testing ability than psychopathological counterparts, it was hypothesized that subjects from the normative group would demonstrate higher mean Evocative Character (measuring social

sensitivity), Affective Quality (measuring emotional balance), and Form Quality (measuring reality testing and connection to the environment) percentages than their psychopathological counterparts. Second, it was expected that subjects in the anxiety group would have higher averages on all three indices as compared to subjects in the psychotic group.

The power of EC and AQ in discriminating between groups was calculated using MANCOVA (controlling for age and gender). A non-parametric Quade test (controlling for age and gender) was conducted for FQ, as the distribution of scores demonstrated significant deviation from normal distribution. Significant age effects were noted for AQ (p<.05) and FQ (p<.001), but not for EC. Similarly, significant gender effects were noted for all categories (EC: p<.05; AQ: p<.001; FQ: p<.001). Sidak post-hoc comparisons were conducted for each category.

Results, summarized in Table 2.17, confirmed the first experimental hypotheses of the study, with mean scores of EC, AQ, and FQ significantly different between healthy, anxious, and psychotic groups in the expected directions. Specifically, healthy (normal) subjects demonstrated higher scores in all categories than pathological (anxious and psychotic groups), indicating better social adjustment (EC), more balanced emotionality (AQ), and adequate reality testing (FQ).

In regards to the second experimental hypothesis, anxious subjects were noted to exhibit higher scores on both EC and FQ categories than their psychotic counterparts, suggesting better social adaptability and reality testing. Although there was a trend in the predicted direction, no significant difference was discovered in terms of Affective Quality (AQ), suggesting less significant difference in emotional balance and regulation between the two groups.

Table 2.17
Evocative Character (EC), Affective Quality (AQ), and Form Quality (FQ) Comparisons between Healthy, Anxious, and Psychotic Groups

Category	Group	$\bar{x}$	*SD*	Skewness	Kurtosis	*F*	*p*	Partial eta squared	d^1*
EC	Psychotic	$58.72_{a,c}$	15.69	-.41	-.20	33.70	<.001	.11	$Psy/Hea=1.66_d$ $Anx/Hea=0.91_e$ $Psy/Anx=0.47_f$
	Anxious	$66.39_{b,c}$	16.71	-.41	-.29				
	Healthy	$77.17_{a,b}$	1.62	-.22	-.10				
AQ	Psychotic	57.92_a	14.12	.04	-.65	13.65	<.001	.05	$Psy/Hea=0.59_d$ $Anx/Hea=0.36_e$ $Psy/Anx=0.18_f$
	Anxious	60.73_b	15.77	-.39	-.16				
	Healthy	$65.77_{a,b}$	12.07	-.05	-.04				
FQ	Psychotic	$68.50_{a,c}$	16.60	-.23	-.53	12.42	<.001	.30	$Psy/Hea=2.67_d$ $Anx/Hea=1.33_e$ $Psy/Anx=1.11_f$
	Anxious	$85.93_{b,c}$	14.59	-1.39	2.50				
	Healthy	$99.79_{a,b}$	1.33	-6.98	52.10				

Note. Sidak adjustment for multiple comparisons:
$_a$mean difference significant at .01 level between Healthy and Psychotic groups;
$_b$mean difference significant at .01 level between Healthy and Anxious groups;
$_c$mean difference significant at .01 level between Anxious and Psychotic groups.
*Effect sizes calculated by current authors from previously published analyses via pair-wise comparisons between groups: Hea=healthy; Anx=anxiety; Psy=psychotic; ^1d effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988):
$_d$ Healthy group values greater than psychotic group.
$_e$ Healthy group values greater than anxiety group.
$_f$Anxiety group values greater than psychotic group.

The authors concluded that analyses support the convergent validity of EC, AQ, and FQ through demonstration of the indices' ability to differentiate between healthy and pathological groups. Further research regarding convergent and incremental validity was recommended.

Psychosis

Crisi, Testa, Carlesimo, & Maio (2010)

Given the work of Benedetti et al. (2008, see above), further investigation of the ability of CWS indices and computations to differentiate between clinical and normative groups was undertaken by Crisi and colleagues (2010). The study began with 350 clinical patients who were initially assessed and diagnosed by qualified psychiatrists according to the DSM-IV-TR (American Psychiatric Association, 2000). Following evaluation, each patient was individually administered a battery of psychodiagnostic tests including the WDCT according to CWS administration guidelines. Thirty-four cases were excluded from the research, 14 (4%) due to falling outside the age range of inclusion (<18 or >60), and 20 due to the inability to produce a valid Wartegg protocol (5.7%). The research included 316 patients, 184 (58%) of which were hospitalized, and 132 (42%) of which were participating in day treatment programming or outpatient psychotherapy.

The researchers focused on a subset of the overall clinical sample, specifically those clients diagnosed with psychotic disorders (N=151, 47.8% of the overall clinical sample); These 151 members of the Psychosis group (58 male, 93 female; age: $\bar{x}$=37.01, SD=9.60) were compared to a random sample of 299 subjects (150 male, 149 female; age: $\bar{x}$=20.94, SD=2.82) selected from the CWS standardization sample. As this constituted a preliminary data analysis, prior to investigating all clinical diagnostic groups, only differences between groups in terms of Order of Sequence (that is, the box order in which participants completed the test) and Content categories (that is, the presence or absence of specific contents, scored according to CWS guidelines) were investigated by the authors.

Two experimental hypotheses were developed, the first related to the Order of Sequence, and the second to specific Content categories. First, it was hypothesized that a higher number of psychosis group participants would follow the numerical Order of Sequence as compared to control group participants. This finding has been noted in previous research, suggesting that individuals without significant pathology are able to respond to the evocative character of the WDCT stimuli as they are generally in touch with and responsive to their internal psychoaffective dynamics. Individuals in the experimental psychosis group, by comparison, were expected to demonstrate reduced self-awareness, tendencies to insulate themselves from the evocative psychoaffective nature of the stimuli, and rigidity of thought process and behavior, as evidenced by inflexible adherence to the numerical Order of Sequence.

The second experimental hypothesis related to primary Content scoring categories. The authors hypothesized that contents related to human relationships, practical abilities and connection to the environment, sensitivity and relatedness to others (H, A, NAT, BOT, OBJ, ARC) would be higher in the control group, as compared to the experimental group. Conversely, contents connected to isolation, detachment, and withdrawal from the world and others (AS, ANA, HS, ICE, MIN, PAT, SIG) were hypothesized to occur more frequently in the experimental group.

In considering the first experimental hypothesis, related to the Order of Sequence, it is important to note that previous research has indicated that 92% of the normative sample follows an individualized Order of Sequence (Crisi, 1998, 2007); that is, when presented with the blank WDCT test form, the vast majority of test-takers follow a unique and personal system of completing the

test, not bound by numerical or methodical order. In the normative sample, only 7% of test-takers followed a *numerical* Order of Sequence (i.e., Box 1, Box 2, Box 3, etc.). As such, it is expected that in a healthy, well-balanced and flexibly integrated individual, a unique Order of Sequence will be followed.

In evaluating results related to their first hypothesis, the authors noted significant differences in approach to the test between the psychosis group and the control group, presented in Table 2.18.

Table 2.18
Differences in Individualized and Numerical Orders of Sequence: Psychosis Group vs. Control Group

Group	Numerical Order of Sequence		Individualized Order of Sequence		χ^2	*p*	Φ^1
	N (%)	χ^2	*N* (%)	χ^2			
Psychosis	98 (65%)	62.56	53 (35%)	26.53	134.07	.00001	.55
Control	36 (12%)	31.59	263 (88%)	13.40			

Note. Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Given these results, the authors' first experimental hypothesis was confirmed. In considering Order of Sequence, generally speaking, healthy individuals (who are better able to access inner psychoaffective dynamics, and are therefore more responsive and adaptable) let themselves be lead during test completion by the test stimuli, responding to the evocative character of each box (similar to the concept of "card pull" on the Rorschach Inkblot Test). As such, individualized Orders of Sequence would be expected, reflecting each participant's unique and personally-driven response to the test. Conversely, results of the current study suggested that individuals in the psychosis group were less able to access internal dynamics, and were less self-aware. Moreover, a strong tendency to follow the numerical Order of Sequence was suggestive of an extreme defense against anxiety, as well as a cognitive rigidity that may serve to protect and insulate participants from disturbing evocative character.

In regards to the authors' second experimental hypothesis, differences between groups in terms of Primary Content categories were investigated. It should be noted that Content categories are rigorously scored according to CWS guidelines, and this process should not be mistaken for less structured "content analysis." In investigating the hypothesis, the frequency of protocols not containing any contents related to human relationships, practical abilities, connection to the environment, and sensitivity and relatedness to others (H, A, NAT, BOT, OBJ, ARC) was calculated for each group. As expected, the control group (61.06%) differed significantly from the clinical group (48.20%, χ^2=6.41, p<.01), producing more of these adaptive contents. Similarly, the same analysis was completed for contents connected to isolation, detachment, and withdrawal from the world and others (AS, ANA, HS, ICE, MIN, PAT, SIG). Again, significant differences were noted between groups, with the experimental group (26.4%) producing these contents more frequently than the control group (21.1%, χ^2=3.05, p<0.01).

In further examining these content category differences between groups at the individual category level (as opposed to the aggregate protocol level described above), four statistically significant results were noted following application of the Bonferroni correction for multiple comparisons. These results are presented in 2.19.

Table 2.19
Significant Differences in CWS Content Categories: Psychosis Group vs. Control Group

Index, Content, Special Score	*Control* $\bar{x}$	*Control SD*	*Psychosis* $\bar{x}$	*Psychosis SD*	*t* *(354 df)*	*p*	*d*[1]
Human Content (H%)	12.76	11.98	8.24	10.88	-3.492	.001	0.37_a
Object Content (OBJ%)	37.89	18.49	27.37	19.88	-5.87	.00001*	0.62_a
Architecture Content (ARC%)	19.58	11.27	11.82	11.21	-6.261	.00001*	0.67_a
Symbol Content (SIG%)	4.59	8.66	26.04	23.53	14.813	.0001*	1.57_b
Pathological Content (PAT%)	1.40	3.95	3.64	8.02	4.020	.0001*	0.43_b

Note. *Significance noted following Bonferroni (p<.05) post-hoc correction; p<0.0039.
Effect sizes calculated by current authors from previously published analyses; [1]d effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]Control group values greater than psychosis group.
[b]Psychosis group values greater than control group.

Out of the 13 Content categories investigated, four successfully differentiated between groups in the direction predicted by the experimental hypotheses. Most significantly, Object (OBJ%) and Architectural (ARC%) contents were noted as higher in the control group as compared to the psychosis group. Increased OBJ content likely indicates that the individuals of the control group demonstrate a greater ability to relate themselves to the concrete and practical aspects of reality (as compared to the psychosis group members). Similarly, ARC content (in the normative level, as seen here) can indicate a personality that is adequately structured and internally supported. Control group subjects were also noted to demonstrate higher Human (H%) content scores (approaching significance) suggesting greater interpersonal connectivity, and to the contrary, higher levels of isolation and detachment in the psychosis group.

Two additional Content domains demonstrated statistical significance, suggesting higher pathology and difficulty on the part of the Psychosis group. As compared to their control group counterparts, the psychosis group demonstrated significantly higher Symbol (SIG%) content. SIG content is typically indicative of an individual's mental representations (as opposed to concrete reality), and this heightened degree of SIG content is likely related to a rigid analog style, including primary process thinking (which would be more expected in a thought-disordered individual). Lastly, the psychosis group demonstrated higher levels of Pathological (PAT%) content, suggesting a tendency to imbue their internal world with negative experiences and bad objects. Overall, the authors concluded that their experimental hypotheses were supported with further research related to Primary Content categories recommended.

Crisi, Testa, Carlesimo, Lops, & Maio (2011)

Further investigation of the ability of CWS scores to differentiate between clinical and normative groups was undertaken by Crisi and colleagues (2011). Beginning with the same initial sample and procedures noted above (Crisi, Testa, Carlesimo, & Maio, 2010), 106 cases were excluded from the overall sample, 24 (6.9%) due to falling outside the age range of inclusion (<18 or >60), 28 for invalid CWS protocols (8.1%), and 54 (18.3%) due to failure to understand the test instructions. As it has been previously established that only 0.5% of the normative population experiences difficulty

with test instructions (Crisi, 1998, 2007), the fact that 18.3% of clinical cases were unable to grasp the demands of the test indicate severe levels of affective distress and cognitive confusion. Over one quarter (26%) of individuals diagnosed as meeting criteria for a psychotic disorder, and 11% of individuals diagnosed was meeting criteria for a personality disorder were noted to experience difficulties comprehending test instructions, and were subsequently excluded from the study.

Following exclusions, 241 patients were included in the study, 136 (56%) of whom were hospitalized, and 105 (44%) of whom were participating in day treatment programming or outpatient psychotherapy. Clinical participants were grouped into five categories, based upon diagnostic features as determined by psychiatric evaluations: 1) Psychosis (*N*=107, 44.4%; 74 female, 33 male; age: $\bar{x}$=37.8, *SD*=9.706); 2) Personality Disorder (*N*=54, 22.4%; 17 female, 37 male; age: $\bar{x}$=32.98, *SD*=8.732); 3) Anxiety Disorder (*N*=56, 23.2%; 14 female, 42 male; age: $\bar{x}$=32.23, *SD*=9.416); 4) Mood Disorder (*N*=19, 7.9%; 4 female, 15 male; age: $\bar{x}$=38.42, *SD*=9.634), and 5) Other Disorder (*N=5*, 2.1%; 2 female, 3 male; age: $\bar{x}$=2.758, *SD*=6.943). It should be noted that a subset of the personality disorder group (*N*=37, 15.3% of total group) was diagnosed with borderline personality disorder. For the purposes of the current study, only the first three groups were considered (psychosis, personality disorder, anxiety disorder) due to limited number of subjects falling within the remaining two categories. These three groups were compared to a random sample of 401 subjects (215 male, 186 female; age: $\bar{x}$=20.94, *SD*=2.82) selected from the 2005 CWS standardization sample. Differences between groups were investigated in terms of Order of Sequence (that is, the box order in which participants completed the test), major indices of the CWS, Content categories, and Special Scores.

Six specific experimental hypotheses were investigated by the authors, each specifically in relationship to differences between the control and psychosis groups. While the anxiety disorder and personality disorder groups were included in analyses, given their relatively small sizes, inclusion was considered exploratory in nature, allowing only preliminary comparison between groups.

Hypothesis 1: Given previous research, the psychosis group would demonstrate a higher frequency of numerical Order of Sequence than the other groups. Moreover, all clinical groups would demonstrate a higher frequency of numerical Order of Sequence than the control group participants.

Regarding Order of Sequence, similar to the study described above, significant differences (χ^2=108.321, p<.00001) were noted in the frequency of *numerical* Orders of Sequence between the control group and each of the clinical groups (psychosis, personality disorder, anxiety disorder). As predicted, the psychosis group demonstrated the highest frequency of *numerical* Orders of Sequence, followed by the personality disorder and anxiety disorder experimental groups. The authors asserted that the higher percentage of *numerical* (non-individualized) Orders of Sequence in the more pathological groups confirm the rigidity, internal distress, need for structure, heightened anxiety, and defensive attempts related to interpersonal protection from unwanted stimulation suggested by the included diagnoses. Results from chi-square analysis between the four groups are provided in Table 2.20.

Table 2.20
Frequency of Distribution of Numerical Order of Sequence between Control, Psychosis, Anxiety Disorder, and Personality Disorder Groups

Group	Frequency [Numerical OoS] (χ^2)	% [Numerical OoS]	Frequency [Individual OoS] (χ^2)	% [Individual OoS]	χ^2	*p*	*V*[1]
Control	31 (5.65)	7.7	370 (24.96)	92.3	108.1321	.00001	.41
Psychosis	54 (13.45)	50.5	53 (59.47)	49.5			
Anxiety Disorder	15 (0.48)	25.9	41 (2.11)	74.1			
Personality Disorder	14 (0.37)	26.8	40 (1.64)	73.2			

Note. Effect sizes calculated by current authors from previously published analyses; [1]*V* effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Hypothesis 2: Experimental groups will demonstrate more negative values on the main indices of the CWS as compared to the control group. Of the clinical groups, the psychosis group will produce the most negative values. In terms of Content scoring, results from the previous study (Crisi, Testa, Carlesimo, & Maio, 2010) were expected to be replicated.

In terms of comparisons between groups related to major CWS indices, exploratory analysis was conducted between groups with authors anticipating significant effect sizes. Highest discriminatory power was expected between the control group and psychosis group. In terms of specific indices of the CWS, it was expected that the control group (as compared to the psychosis group) would demonstrate 1) higher Evocative Character (EC+%); 2) higher Affective Quality (AQ+%); 3) higher Form Quality (FQ+%); 4) higher Affective Stability Index (A/F Ratio); 5) lower Popular (P%) percentage; 6) higher Popular responses with good form (P+%); 7) lower Original responses (O%); 8) higher Original responses with good form (O+%); 9) higher Index of Inner Tension-1, Impulsivity Index (IM), and Anxiety Index (AI); 10) lower Index of Suicidal Tendencies; and 11) higher level of primary Movement (M) responses.

In terms of Content categories, similar to previous research, the authors hypothesized that control group participants would demonstrate greater frequency of Human (H%), Object (OBJ%), and Architecture (ARC%) content as compared to clinical groups, and lower frequency of Symbol (SIG%), Abstract (ABS%), and Pathology (PAT%) content.

Evaluation results demonstrated statistically significant differences in 30 out of 40 comparisons between the control group and experimental groups, following Bonferroni post-hoc correction. In terms of indices and computations, 15 statistically significant differences were noted, occurring in the expected direction, with effect size ranging from medium/large (EC+%, FQ+%, P%, P+%, O%, O+%, IIT-1, AI, IIT-2 and IST) to small (Affective Stability Index, M). Regarding Primary Contents, 7 comparisons were significant and in the expected direction, with effect sizes ranging from large (OBJ%, ARC% and SIG%) to medium/small (H%, ABS%, PAT%, and CLD%). Lastly, in terms of Special Scores, 9 comparisons demonstrated significant differences between the control and clinical groups in the expected direction with effect sizes ranging from large/medium (ID, II, PP, AP, PR, MU) to small (PA, GR, IS, DS). Full results are presented in Table 2.21.

Table 2.21
Effect Sizes of Major CWS Indices and Content Categories in Psychosis, Anxiety Disorder, and Personality Disorder Groups as compared to Control Group

Index, Content, Special Score [Control $\bar{x}$, *SD*]	Psychosis			Anxiety Disorders			Personality Disorders		
	$\bar{x}$ *(SD)*	$t_{(506)}$ *(p)*	*d Cohen*	$\bar{x}$ *(SD)*	*t (p)*	*d Cohen*	$\bar{x}$ *(SD)*	*t (p)*	*d Cohen*
Indices									
Evocative Character (EC+%) [77.17, 10.62]	58.72 (15.69)	-14.291 (.000*)	1.40[3]	66.39 (16.71)	-6.552 (.000*)	.079[2]	72.59 (13.96)	-2.855 (.005)	0.48[1]
Affective Quality (AQ+%) [65.77, 12.06]	57.92 (14.12)	-5.761 (.000*)	0.60[2]	60.73 (15.71)	-2.808 (.005)	0.36[1]	57.20 (15.92)	-4.697 (.000*)	0.76[2]
Form Quality (FQ+%) [99.79, 1.32]	68.50 (16.59)	-37.40 (.000*)	3.49[3]	85.93 (14.59)	-18.606 (.000*)	1.74[3]	86.65 (11.32)	-22.284 (.000*)	2.18[3]
Affective Stability Index: A [2.83, 0.67]	2.50 (0.79)	-4.26 (.000*)	0.44[1]	2.52 (0.97)	-2.957 (.003)	0.37[1]	2.44 (0.98)	-3.699 (.000*)	0.44[1]
Affective Stability Index: F [3.97, 0.19]	2.52 (0.76)	-34.03 (.000*)	3.01[3]	3.28 (0.71)	-15.700 (.000*)	1.51[3]	3.29 (0.61)	-16.874 (.000*)	1.83[3]
Popular Responses (P%) [19.62, 10.69]	22.60 (12.38)	2.471 (.014)	-0.26[1]	24.68 (13.31)	3.209 (.001*)	-0.42[1]	25.48 (12.94)	3.681 (.000*)	-0.57[2]
P% with Good Form (P+%) [99.82, 2.71]	83.98 (19.10)	-15.53 (.000*)	1.45[3]	91.53 (13.42)	-10.642 (.000*)	1.03[3]	91.44 (12.42)	-11.330 (.000*)	1.09[3]
Original Responses (O%) [0.24, 1.59]	7.92 (10.35)	14.26 (.000*)	-1.29[3]	3.59 (7.57)	7.760 (.000*)	-0.73[2]	7.91 (11.31)	12.745 (.000*)	-0.94[3]
O% with Good Form (O+%) [100.00, 0.00]	42.67 (30.72)	-5.86 (.000*)	3.73[3]	61.54 (30.13)	-4.014 (.0001*)	2.55[3]	71.63 (28.96)	-3.070 (.004)	2.74[3]
Index of Inner Tension 1 (IIT-1) [1.22, 0.31]	0.61 (0.40)	-16.38 (.000*)	1.66[3]	0.93 (0.42)	-6.075 (.000*)	0.78[2]	0.99 (0.45)	-4.586 (.000*)	0.72[2]
Impulsivity Index (IM) [0.28, 0.22]	0.22 (0.24)	-2.44 (.015)	0.26[1]	0.27 (0.21)	-0.342 (.0732)	0.05	0.29 (0.26)	0.131 (.896)	0.07
Anxiety Index (AI) [0.93, 0.17]	0.39 (0.28)	-24.66 (.000*)	2.37[3]	0.65 (0.29)	-10.087 (.000*)	1.19[3]	0.70 (0.30)	-8.046 (.000*)	1.03[3]
Index of Inner Tension 2 (IIT-2) *(first half of ratio)* [5.29, 1.32)	3.79 (1.40)	-10.29 (.000*)	1.10[3]	4.36 (1.73)	-4.772 (.000*)	0.61[2]	4.65 (1.61)	-3.282 (.001*)	0.59[2]
Index of Suicidal Tendencies (IST) [1.98, 1.63]	5.93 (1.88)	21.45 (.000*)	-2.25[3]	4.46 (2.05)	10.273 (.000*)	-1.47[3]	4.63 (1.97)	10.866 (.000*)	-1.34[3]
Primary Movement (M) [0.30, 0.68]	0.07 (0.34)	-3.44 (.001*)	0.46[1]	0.18 (0.47)	-1.300 (0.194)	0.21[1]	0.19 (0.64)	-1.179 (.239)	0.32[1]
Animal Movement (FM) [0.10, 0.32]	0.01 (0.09)	-2.73 (.006)	0.41[1]	0.00 (0.00)	-2.212 (.027)	0.59[2]	0.04 (0.19)	-1.316 (.189)	0.10
Inanimate Movement (MI) [0.56, 0.77)	0.37 (0.57)	-2.31 (.021)	0.27[1]	0.46 (0.63)	-0.877 (.381)	0.13	0.52 (0.74)	-0.361 (.719)	0.01
Facial Expression (FE) [0.17, 0.41)	0.03 (0.16)	-3.55 (.000*)	0.50[2]	0.02 (0.13)	-2.794 (.005)	0.57[2]	0.02 (0.13)	-2.732 (.007)	0.46[1]
Content Categories									
Human Content (H%) [12.76, 11.96]	7.54 (11.41)	-4.04 (.000*)	0.45[1]	8.96 (10.04)	-2.269 (.024)	0.35[1]	10.75 (12.97)	-1.144 (.253)	0.21[1]
Object Content (OBJ%) [37.89, 18.49]	26.95 (21.07)	-5.21 (.000*)	0.55[2]	21.93 (16.57)	-6.120 (.000*)	0.91[3]	26.05 (15.92)	-4.483 (.000*)	1.00[3]
Architecture Content (ARC%) [19.58, 11.27]	11.48 (11.90)	-6.53 (.000*)	0.70[2]	12.41 (9.99)	-4.516 (.000*)	0.67[2]	13.34 (11.24)	-3.823 (.000*)	0.59[2]
Symbol Content (SIG%) [4.59, 8.66]	27.72 (26.91)	14.63 (.000*)	-1.30[3]	26.88 (26.97)	12.608 (.000*)	-1.25[3]	20.36 (21.15)	9.985 (.000*)	-1.34[3]
Astronomical Content (AST%) [3.72, 6.25]	5.91 (9.55)	2.84 (.005)	-0.28[1]	4.45 (7.43)	0.989 (.370)	-0.12	3.94 (7.21)	0.235 (.815)	-0.15
Nature Content (NAT%) [5.49, 8.16]	3.19 (6.28)	-2.70 (.007)	0.32[1]	4.63 (7.15)	-0.742 (.458)	0.11	4.86 (8.20)	-0.528 (.598)	-0.01

Botany Content (BOT%) [6.00, 8.12]	4.64 (9.09)	-1.50 (.134)	0.16	8.40 (11.02)	1.976 (.049)	-0.25[1]	5.85 (8.81)	-0.127 (.899)	-0.02
Abstract Content (ABS%) [0.00, 0.00]	2.12 (7.40)	5.74 (.000*)	-0.57[2]	0.67 (3.70)	3.641 (.000*)	-0.36[1]	0.69 (3.77)	3.710 (.000*)	-0.51[2]
Cloud Content (CLD%) [0.47, 2.37]	1.00 (3.61)	1.84 (.066)	-0.18	0.22 (1.67)	-0.744 (.457)	0.12	1.73 (4.55)	3.200 (.001*)	-0.38[1]
Pathological Content (PAT%) [1.40, 3.95]	2.34 (6.67)	1.83 (.066)	-0.18	3.57 (7.80)	3.311 (.001*)	-0.37[1]	3.27 (6.09)	3.032 (.003)	-0.40[1]
Note. Effect sizes calculated by current authors from previously published analyses: [1]small effect size (.2-.49); [2]medium effect size (.5-.79); [3]large effect size (>.8; Cohen, 1988). *Significance noted following Bonferroni (p<.05) post-hoc correction; p<0.00125.									

Regarding experimental hypothesis #2, 15 investigated variables demonstrated significant effect sizes (ranging from medium to large), with the psychosis group typically showing the most difference as compared to the control group. In examining the indices, the control group demonstrated 1) higher Evocative Character (EC+%), generally suggesting greater relatedness and connection with the environment; 2) higher Affective Quality (AQ+%), indicating balanced affective regulation and integration; 3) significantly higher Form Quality (FQ+%), related to cognitive processes, control, and reality testing; 4) higher Affective Stability Index (A/F Ratio), indicating an appropriate level of affective openness and experience coupled with suitable levels of cognitive control and ego strength; 5) lower Popular (P%) responses suggesting the clinical groups demonstrated higher levels of stereotypical thinking; 6) higher Popular responses with good form (P+%), related to the ability to share in a common way of thinking with others while remaining connected to reality; 7) lower Original responses (O%), indicating the potential for the clinical groups to demonstrate idiosyncratic, unique, and divergent manners of thinking; 8) higher Original responses with good form (O+%), with similar interpretation as P+% above; 9) higher Index of Inner Tension-1, Impulsivity Index (IM),and Anxiety Index (AI), as these markers of internal distress are less apparent in psychotic clients as compared to the normative group; 10) lower Index of Suicidal Tendencies, suggesting increased likelihood of suicidal ideation or attempt in clinical groups; and 11) higher percentage of Primary Movement (M) as expected, reflecting greater levels of internal cognitive resources. The control group was further noted to demonstrate a significantly higher percentage of Facial Expression (secondary movement). While this effect was not predicted, previous studies have also demonstrated this tendency, suggesting facial expression responses may indicate greater awareness of internal emotional states on the part of the control/normative group (as compared to clinical groups).

Regarding Content categories, 7 investigated contents indicated significant (medium or large) effect sizes. As predicted, control group subjects demonstrated greater frequency of Human (H%), Object (OBJ%), and Architecture (ARC%) content as compared to clinical groups, suggesting greater adaptability to the environment, social connectedness, and practicality. Conversely, clinical groups demonstrated more frequent Symbol (SIG%), Abstract (ABS%), and Pathology (PAT%) content, indicating tendencies toward primary process thinking, disconnection from the environment, increased levels of internal pain, and poor mental representations. Additionally, while not specifically hypothesized, it was noted that subjects in the personality disorder group demonstrated higher frequency in Cloud content (CLD%), suggesting greater internalized distress that is likely to have reached conscious awareness. Overall, the authors concluded that experimental hypothesis #2 was confirmed.

Hypothesis 3: Experimental groups will demonstrate higher frequency of CWS Special Scores as compared to the control group. Of the experimental groups, the psychosis group will demonstrate the highest frequency of special scores. Moreover, a specific constellation of Special Scores

thought to be associated with thought disorder (Personalized Answer, Contamination, Arbitrary Performance, Disproportion, and Transparency) will demonstrate significantly higher frequency in the psychosis group as compared to the control group.

Exploratory analysis was completed, investigating the frequency of all CWS Special Scores in the three clinical groups (psychosis, personality disorder, anxiety disorder) as compared to the control group. Results are presented in 2.22.

Table 2.22
Effect Sizes of Major CWS Special Scores in Psychosis, Anxiety Disorder, and Personality Disorder Groups as compared to Control Group

Index, Content, Special Score [Control $\bar{x}$, *SD*]	Psychosis			Anxiety Disorders			Personality Disorders		
	$\bar{x}$ *(SD)*	$t_{(506)}$ *(p)*	*d* Cohen	$\bar{x}$ *(SD)*	*t (p)*	*d* Cohen	$\bar{x}$ *(SD)*	*t (p)*	*d* Cohen
Personalized Answer (PA) [0.03, 0.18]	0.23 (0.90)	4.195 (.000*)	-0.37[1]	0.00 (0.00)	-1.212 (.226)	0.32[1]	0.13 (0.61)	2.521 (.012)	-0.42[1]
Reversed Box (RB) [0.16, 0.58]	0.08 (0.31)	-1.25 (.211)	0.16	0.07 (0.32)	-1.080 (.281)	0.19	0.06 (0.23)	-1.270 (.205)	0.32[1]
Incomplete Drawing (ID) [0.02, 0.17]	0.31 (0.53)	8.98 (.000*)	-0.80[3]	0.11 (0.36)	2.812 (.005)	-0.31[1]	0.19 (0.51)	4.623 (.000*)	-0.43[1]
Inadequate Integration (II) [0.05, 0.22]	0.58 (0.91)	10.48 (.000*)	-0.93[3]	0.32 (0.60)	6.325 (.000*)	-0.65[2]	0.46 (0.81)	8.074 (.000*)	-0.89[3]
Missing Union (MU) [0.18, 0.41]	0.05 (0.25)	-3.10 (.002)	0.39[1]	0.00 (0.00)	-3.200 (.001*)	0.86[3]	0.00 (0.00)	-3.143 (.002)	0.86[3]
Perseveration (PP) [0.00, 0.05]	0.32 (0.76)	8.24 (.000*)	-0.78[2]	0.07 (0.26)	4.749 (.000*)	-0.45[1]	0.06 (0.30)	3.226 (.001*)	-0.44[1]
Arbitrary Performance (AP) [0.02, 0.14]	1.13 (1.35)	16.08 (.000*)	-1.48[3]	0.71 (0.92)	13.796 (.000*)	-1.28[3]	0.83 (1.20)	12.817 (.000*)	-1.39[3]
Global Rejection (GR) [0.04, 0.25]	0.21 (0.65)	4.16 (.000*)	-0.37[1]	0.16 (0.62)	2.533 (.000*)	-0.26[1]	0.13 (0.43)	2.095 (.037)	-0.35[1]
Partial Rejection (PR) [0.01, 0.25]	0.43 (0.90)	8.10 (.000*)	-0.72[2]	0.41 (0.91)	7.000 (.000*)	-0.68[2]	0.17 (0.63)	3.235 (.001*)	-0.44[1]
Interpreted Stimulus (IS) [0.00, 0.05]	0.04 (0.23	2.75 (.006)	-0.24[1]	0.05 (0.22)	3.899 (.000)	-0.37[1]	0.00 (0.00)	-0.367 (.714)	0.10
Disproportion (DS) [0.01, 0.12]	0.06 (0.23)	2.49 (.013)	-0.23[1]	0.04 (0.18)	1.108 (.268)	-0.13	0.09 (0.04)	3.526 (.000*)	-0.28[1]
Transparency (TR) [0.01, 0.08]	0.05 (0.21)	2.91 (.004)	-0.26[1]	0.02 (0.13)	0.780 (.436)	-0.09	0.06 (0.23)	2.928 (.004)	-0.48[1]

Note. Effect sizes calculated by current authors from previously published analyses:; [1]small effect size (.2-.49); [2]medium effect size (.5-.79); [3]large effect size (>.8; Cohen, 1988).
*Significance noted following Bonferroni ($p<.05$) post-hoc correction; $p<0.00125$.

9 Special Scores demonstrated significant effect sizes when comparing the control group to each clinical group, with significance noted following Bonferroni post-hoc correction. As expected, the psychosis group demonstrated both the largest effect sizes and the most frequent differences from the control group. Specifically, clinical group members produced a higher frequency of: (1) Personalized Answer (PA), indicating their egocentric view of the world and their weakened self/world and reality/fantasy boundaries; (2) Incomplete Drawing, suggesting marked insecurity and indecision, as well as possible detachment from reality; (3) Inadequate Integration (II), indicating an impaired sense of reality; (4) Missing Union (MU), indicating potential psychoaffective inhibition of intelligence, noted solely in the anxiety disorder group; (5) Perseveration (PP), indicating a stereotypical style of thinking; (6) Arbitrary Performance (AP), indicating a reduced ability to perceive reality; (7) Global Rejection (GR), indicating areas of personality too conflictual, problematic, or "split off" to access; (8) Partial Rejection (PR),

suggesting significant ambivalence and conflict in specific areas of functioning; and (9) Interpreted Stimulus (IS), suggesting a tendency to avoid conflict with an associated high level of internal tension, again noted only in the Anxiety Disorder group. As expected, in each case, Special Scores were significantly more frequent in clinical groups than the control group.

To further investigate experimental hypothesis #3, frequency of the presence of Special Scores indicating thought disorder (including Arbitrary Performance [AP], Personalized Answer [PA], Contamination [CO], Disproportion [DS], and Transparency [TR]) were summed across all WDCT boxes of the psychosis group and the control group. Chi-square analysis, presented in Table 2.23, indicates a statistically significant difference between groups in the frequency of these special scores related to thought disturbance. Protocols of 82% of psychosis group members contained problematic Special Scores, whereas on 7.5% of protocols of control group members contained these same Special Scores. Given these findings, experimental hypothesis #3 was considered fully supported.

Table 2.23
Frequency of Distribution of Special Scores Indicative of Thought Disturbance (AP, PA, CO, DS & TR) Between Psychosis and Control Groups

Group	Frequency [Present] (χ^2)	% [Present]	Frequency [Absent] (χ^2)	% [Absent]	χ^2	*p*	Φ^1
Psychosis	88 (160.43)	82.2	19 (48.54)	17.8	264.73	.00001	.72
Control	30 (42.81)	7.5	371 (12.95)	92.5			

Note. χ^2 computed for total WDCT boxes possible per group (Control: *N=401* multiplied by 8 boxes per protocol = 3,208 total boxes; Psychosis: *N=107* multiplied by 8 boxes per protocol = 856 total boxes).
Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Hypothesis 4: Control group will demonstrate more positive Global Assessment classification than experimental groups, with Psychosis group demonstrating more negative classifications than other experimental groups.

In addition to examination of Order of Sequence, Indices, Content categories, and Special Scores, the authors also examined overall Global Assessment of the protocols in each group to determine if differences exist. It was hypothesized that experimental groups would demonstrate higher percentages of negative classifications as compared to the control group, with the Psychosis group demonstrating the highest frequency of negative classifications.

As described earlier, in determining the Global Assessment, each protocol is assigned an overall classification based specific CWS scores and indices, ranging from "No Pathology" (NOP) to "Pathological" (PTL). This classification rubric is divided into five categories, two of which are considered positive codes (NOP + "Low Symptoms" [LSI]), one of which is considered neutral ("Moderate Symptoms" [MSI]) and two of which are considered negative codes (PTL + "Severe Symptoms" [SSI]). Chi-Square analysis was conducted, indicating significant differences between groups existed (χ^2=299.6455, p<0.00001) in terms of Global Assessment. Results are presented in Table 2.24.

Table 2.24
Frequency of Distribution of Positive (+), Neutral (+/-) and Negative (-) Global Assessment Classifications between Psychosis, Anxiety Disorder, Personality Disorder, and Control Groups

Group	Fr. + (χ^2)	% +	Fr. +/- (χ^2)	% +/-	Fr. – (χ^2)	% -	χ^2	*p*	*V*[1]
Control	365 (22.97)	91	32 (11.48)	8	4 (51.32)	1	299.6455	.00001	.49
Psychosis	25 (34.08)	23.4	21 (2.03)	19.6	61 (12.92)	57			
Personality Disorder	23 (7.02)	40.7	22 (24.06)	38.9	11 (0.92)	20.4			
Anxiety Disorder	25 (4.60)	46.4	14 (4.98)	25	15 (6.25)	28.6			

Note. +Positive Assessment; +/-Neutral Assessment; -Negative Assessment.
Effect sizes calculated by current authors from previously published analyses; [1]*V* effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Results of analyses demonstrate statistically significant differences between groups in terms of distribution of Global Assessment classifications. As predicted, the control group was noted to demonstrate the highest percentage of positive codes, followed by the anxiety disorder group. The psychosis group was noted to demonstrate the highest percentage of negative codes, followed by the anxiety disorder group. Each group differed significantly from the control group, demonstrating the effectiveness of the Global Assessment in identifying pathological conditions.

Hypothesis 5:In terms of placement on the Wartegg Index of Psychopathology (WIP), the psychosis group will demonstrate higher frequency of classification in Quadrant B than other groups, and overall less positive placement (lower frequency in Quadrant A and alpha Area of all Quadrants).
As described previously, the CWS co-locates individuals on a graphic depiction of personality functioning, based upon the degree of currently experienced distress (vertical axis) and level of personality flexibility and integration (horizontal axis), known as the WIP. Four quadrants result, labeled A, B, C, and D, respectively. Quadrants A and C (reflecting high flexibility and integration) are generally considered less pathological and therefore more "healthy." Quadrant B reflects low personality integration and flexibility paired with absence of significant internal distress, whereas Quadrant D reflects low personality integration paired with significant levels of tension and distress. Given the expectation of personality inflexibility coupled with lower distress levels, it was hypothesized that a higher frequency of psychosis group members would be classified in Quadrant B.As predicted, when compared to the control group, the psychosis group demonstrated a significantly higher frequency of classification in Quadrant B of the WIP. See Table 2.25.

Table 2.25
Frequency of Individuals Classified in WIP Quadrant B: Psychosis vs. Control Groups

Group	Frequency [Qr. B] (χ^2)	% [Qr. B]	Frequency [Not Qr. B] (χ^2)	% [Not Qr. B]	χ^2	*p*	Φ[1]
Psychosis	67 (82.92)	62.62	40 (22.92)	37.38	134.07	.00001	.51
Control	43 (22.13)	10.72	358 (6.11)	89.28			

Note. Effect sizes calculated by current authors from previously published analyses; [1]φ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988)

Overall frequency of classification in "positive" areas of the WIP was further compared between the psychosis and control groups. Positive classifications on the WIP are typically considered area *alpha* of all quadrants (the adaptive sphere of functioning), as well as all of Quadrant A. It was hypothesized that the psychosis group would demonstrate lower frequency of classification in these positive, adaptive areas of the WIP as compared to the control group. Results of chi-square analysis are presented in Table 2.26.

Table 2.26
Frequency of "Positive" WIP Classifications: Psychosis vs. Control Groups

Group	Frequency [Positive] (χ^2)	% [Positive]	Frequency [Negative] (χ^2)	% [Negative]	χ^2	*p*	Φ^1
Psychosis	41 (5.59)	38.32	66 (6.92)	61.68	15.84	.0001	.18
Control	240 (1.49)	59.85	161 (1.85)	40.15			

Note. "Positive" WIP classification is considered placement in Quadrant A, or the *alpha* area of Quadrant B, C, and D. These areas represent adaptive functioning. Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

As predicted, the psychosis group demonstrated significantly fewer positive classifications on the WIP as compared to the control group. This suggests overall less flexible, integrated, and adaptive functioning on the part of the psychosis group.

Hypothesis 6: The psychosis group will demonstrate fewer overall positive box Codes (C or PC) than the control group.

In determining areas of integration, freedom from conflict, strength, and adaptive functioning, each box of the WDCT is assigned a positive or negative Code. As described previously, Codes are calculated via summation of Affective Quality and Evocative Character scores for each WDCT box, with six possible Codes resulting. Two of these Codes are positive (C, PC) whereas four are considered negative (AC, AD, NC, D). A negative Code suggests less sensitivity, integration, and adaptive strength in the area of functioning in which it occurs.

To investigate experimental hypothesis #6, the presence of positive WDCT box Codes was investigated. That is, given the theory that individuals diagnosed with psychotic disorders demonstrate less flexibility, integration, and environmental sensitivity, it was hypothesized that psychosis group members would produce fewer positive box Codes. The number of positive box Codes (C and PC) was summed across protocols, with chi-square results presented in Table 2.27.

Table 2.27
Frequency of WDCT Boxes with Positive Codes: Psychosis vs. Control Groups

Group	Frequency [Positive] (χ^2)	% [Positive]	Frequency [Negative] (χ^2)	% [Negative]	χ^2	*p*	Φ^1
Psychosis	406 (30.13)	47.4	450 (49.64)	52.6	101.04	.00001	.16
Control	2123 (8.04)	66.2	1085 (13.24)	33.8			

Note. Positive codes are considered C and PC.
Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

As predicted, members of the psychosis group produced fewer positive box Codes (47.4%) than control group members (66/2%), suggesting reduced flexibility, sensitivity, and adaptive functioning.

Overall, the authors concluded that all six experimental hypotheses were confirmed, with the CWS clearly differentiating between psychosis group and control group members in terms of index scores, Content scoring, Special Scores, Order of Sequence, Global Assessment, and box Codes. While exploratory, further differences were noted between the control group and remaining clinical groups (anxiety disorder, personality disorder). While these differences were significant, they were less robust as compared to noted differences between the control and psychosis groups. The authors concluded that more research was needed to clarify these differences.

Depression and Suicidality

Daini, Manzo, Pisani, & Tancredi (2010)

Daini and colleagues used the CWS to investigate differences between depressed subjects with no history of suicidal attempt, psychiatric patients who had attempted suicide, and matched control group subjects. Exploratory in nature, the authors reported their intention to assess whether the CWS demonstrates validity in assessing the risk of suicidality. No specific experimental hypotheses were reported. Per the authors, while self-reports are typically used in hospital settings, a projective methodology for determining risk of suicide was warranted, given the intersection of projective methods, psychopathology, and therapeutic intervention. Given this, the CWS was chosen given that it is "simpler and quicker to administer in emergency situations compared to other projective tests" (p. 173).

The current study comparatively analyzed responses to each box of the WDCT, as well as differences between group means related to major calculated indices of the CWS (including EC+%, AQ+%, FQ+%, P%, IIT-1, IM, AI, and Global Rejection) to determine if the CWS is able to differentiate between psychiatric and control groups. While the CWS includes a computer-generated constellation of scores to indicate suicidal risk and intrapsychic indicators of suicidal tendencies (Index of Suicidal Tendencies, IST), review of this published study indicates that the authors did not include the IST as a variable to be studied in their research. Future studies related to suicidality will benefit from specific inclusion of the IST.

Three groups of subjects participated in the study, labeled as Attempted Suicide (AS), Depression (D), and Control (C). Subjects from the clinical groups (AS, D) were currently receiving hospital-based treatment, whereas control group subjects had never participated in psychiatric care. The attempted suicide (AS) group was comprised of 25 subjects (5 male, 20 female; age: $\bar{x}$=36.58, *SD*=14.6) previously diagnosed with various psychiatric disorders, including depression, eating disorders, personality disorders, obsessive-compulsive disorder, or delirium. The depression (D) group was comprised of 29 subjects (5 male, 24 female; age: $\bar{x}$=37.28, *SD*=10.48) presenting with clinically relevant states of depression, some comorbid with other mood and eating disorders. Lastly, the control (C) group was comprised of 29 subjects (9 male, 20 female; age: $\bar{x}$=39.95, *SD*=11.26). No clinically significant differences in age or gender were noted between groups.

To evaluate differences between groups, as related to major indices of the CWS, analysis of variance (ANOVA) was conducted, with Duncan's post-hoc test utilized to explore differences in terms of group means. Results are presented in Table 2.28.

Table 2.28
Mean Differences in CWS Indices:
ANOVA by Group (Attempted Suicide, Depression, Control)

Variable	AS (N=25) $\bar{x}$ (*SD*)	D (N=29) $\bar{x}$ (*SD*)	C (N=29) $\bar{x}$ (*SD*)	*F*	*Duncan Post-Hoc*	*d*[1]
Evocative Character (EC+%)	68.50 (17.07)	64.34 (16.18)	71.77 (16.33)	0.293	-	0.14
Affective Quality (AQ+%)	58.75 (17.68)	63.58 (13.16)	69.40 (12.08)	3.739	AS<C*	0.56
Form Quality (FQ+%)	72.84 (19.87)	75.19 (22.97)	79.31 (13.37)	0.800	-	0.27
Popular Responses (P%)	37.25 (23.67)	25.86 (13.54)	24.35 (11.85)	4.640	AS>C** AS>D**	0.40
Index of Inner Tension (IIT-1)	0.52 (0.47)	0.61 (0.57)	0.59 (0.43)	0.240	-	0.04
Impulsivity Index (IM)	0.17 (0.20)	0.18 (0.20)	0.18 (0.20)	0.70	-	0.07
Anxiety Index (AI)	0.38 (0.39)	0.42 (0.42)	0.43 (0.36)	0.93	-	0.10
Global Rejection (GR)	0.52 (1.0)	0.14 (0.74)	0.03 (0.18)	3.387	AS>C*	.11

Note. *p<0.05; **p<0.01.
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).

Table adapted from *Daini, Manzo, Pisani, & Tancredi (2010).*

In evaluating results, significant differences were noted in terms of Affective Quality (AQ+%), Popular Responses (P%), and the number of special score Global Rejection (GR). In terms of Affective Quality, those who had attempted suicide demonstrated significantly lower AQ than matched control group subjects. When investigated further, the difference between groups occurred primarily in the area of ego functioning and thought (Adaptive Area, F=3.574; p<.5), with no significant difference noted between groups in the more internal, psychoaffective domain (Affective Area). Similarly, the attempted suicide group demonstrated significantly higher levels of Popular responses (P%) than both the depression and control group. Again, these differences, when investigated further, depended solely on the area of ego functioning and rational thought (Adaptive Area, F=4.689, p<.01). Lastly, the attempted suicide group was noted to demonstrate a significantly higher number of Global Rejections (GR, in which nothing is drawn in a WDCT Box), reflecting a complete denial or avoidance of specific areas of personality, as compared to the control group. Analysis of the frequency of GR responses by box indicated no significant findings; that is, while AS group members were more likely than other groups to produce Global Rejections, there was no statistical trend related to which WDCT box the GR was produced within.

In considering these results, the authors posited that while depressed clients and those who have attempted suicide share many commonalities, the significant difference between them evident in this study appears related to thought process. That is, subjects who had previously attempted suicide appeared to gloss over internal thoughts about their lives, concerns, and feelings, instead subscribing to the conventional way of thinking (higher P%). To potentially counter negative affective states (lower AQ+%), this thinking style may reflect attempts at rationalization, divesting the self from both affective and interpersonal experiences. This hypothesis is further supported by all significant

differences occurring in the area of ego functioning (ADP area, related to rational thought, view of self, and other-relatedness), as well as the higher number of Global Rejections presented by the attempted suicide group. The authors suggested that this last finding, the increase in GR responses, may indicate a tendency of participants to avoid relationships and self-disclosure of internal feeling states—a finding support by research conducted with suicide survivors.

Trauma and Adjustment Disorders

Crisi (2010a, in Bianchi di Castelbianco & Di Renzo)

Crisi and colleagues conducted exploratory analysis to investigate the impact of trauma on a group of adolescents, hypothesizing that the CWS would sensitively identify affective and interpersonal changes resulting from exposure to traumatic experiences. 104 adolescents (48 male, 56 female) aged 14-18 ($\bar{x}$=16.23, *SD*=1.243) were included in the clinical sample. Each had been exposed to the significant earthquake that occurred in L'Aquila, Italy in April 2009, and had subsequently sought treatment at a mobile mental health clinic. Each participant was assessed within three months of the traumatic event, completing several self-report questionnaires along with the WDCT according to CWS administration guidelines.

Results of the clinical group CWS protocols were compared to a group of 400 subjects from the 2005 CWS standardization sample (186 male, 215 female; See Appendix A for description of CWS standardization sample). In terms of major CWS indices, four primary and central scores were investigated, each relevant to clinical interpretation: 1) Evocative Character (EC+%), which measures an individuals sensitivity and engagement with the environment; 2) Affective Quality (AQ+%), which indicates an individual's affective receptivity and expression; 3) Form Quality (FQ+%), which illustrates a person's cognitive and rational functioning, thought processes, and commonality in thinking with others; and 4) Index of Suicidal Tendencies (IST), a constellation of scores and markers suggesting depressive tendencies, internal tension, limited coping, and heightened impulsivity. It was expected that the control group would demonstrate higher values of EC, AQ, and FQ, whereas the trauma-exposed group would demonstrate higher values on the IST. Results of *t*-test analyses are presented in Table 2.29.

Table 2.29
Significant Differences on Major CWS Indices:
Clinical (Trauma-Exposed) vs. Control Groups

Index	***Control*** $\bar{x}$	***Control SD***	***Clinical*** $\bar{x}$	***Clinical SD***	***t*** *(503 df)*	***p***	***d***[1]
Evocative Character (EC+%)	77.17	10.62	64.58	18.51	-9.052	.0001	0.80[a]
Affective Quality (AQ+%)	65.77	12.07	56.40	15.39	-6.638	.0001	0.59[a]
Form Quality (FQ+%)	99.79	1.33	93.31	15.23	-8.427	.0001	0.75[a]
Index of Suicidal Tendencies (IST)	1.98	1.64	3.42	2.24	7.362	.0001	0.66[b]

Note. Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]Control group values greater than clinical group.
[b]Clinical group values greater than control group.

Given that major CWS indices have been noted to stabilize at approximately age 14, no significant differences between the adolescent population and the young adult control group were expected;

however, overall results suggested significant impact of trauma on the adolescents' functioning (as compared to expected normative values), as predicted. Specifically, lower Evocative Character (EC+%) is likely to reflect increased difficulty on the part of the clinical group subjects, in relating to the environment, engaging in reciprocal social interactions, and remaining open to affective situations. Similarly, lower Affective Quality (AQ+%) suggest adolescents in the clinical group were more likely to experience greater depressive symptoms, including limited spontaneity, reduced motivation, and higher levels of overall emotional constriction or avoidance.

A significant difference was noted in Form Quality (FQ+%) between groups, which typically relates to an individual's reality testing, judgment, cognitive self-control, and ego functions. As each of the assessed adolescents was in school at the time of assessment, with no notable learning difficulties or mental health diagnoses related to psychosis or thought disturbance, such a significant difference in Form Quality likely indicated a current condition related to high stress and difficulties in adjustment, potentially resulting in a relaxation of rational control mechanisms, and less cognitive ability to manage emotions and impulses. Lastly, the Index of Suicidal Tendencies (IST), which indicates a depressive state that may produce intrapunitive self-harm behaviors, was noted to be significantly higher in the clinical group. It should be noted that the average value for the IST in the normative population is 1.98, and the clinical group mean was equal to 3.42. This suggested a significantly higher level of depressive affect and internal discomfort in the clinical group, as compared to normative expectations.

In addition to examining major CWS indices, the authors further evaluated categorical differences between groups related to overall Global Assessment classifications assigned to each participant's CWS protocol. As discussed above, in completing the CWS scoring mechanics, each protocol is assigned an overall Global Assessment classification, ranging from "No Pathology" (NOP) to "Pathological" (PTL). This classification system is divided into five categories, two of which are considered positive (NOP + "Low Symptoms" [LSI]) and three of which are considered negative (PTL + "Moderate Symptoms" [MSI] and "Severe Symptoms" [SSI]). Clinical and control groups were compared based on their classification within this Global Assessment rubric. The authors hypothesized that the clinical (trauma-exposed) group would produce fewer positive Global Assessment classifications, and a higher frequency of negative classifications as compared to the control group. Results are presented in Table 2.30.

Table 2.30
Frequency of Distribution of Positive and Negative Global Assessment Classifications: Clinical (Trauma-Exposed) vs. Control Groups

Group	Frequency [Positive] (χ^2)	% [Positive]	Frequency [Negative] (χ^2)	% [Negative]	χ^2	p	Φ^1
Clinical	62 (4.52)	59.6	42 (16.3)	40.4	25.87	.00001	.23
Control	332 (1.17)	83.0	69 (4.16)	17.0			

Note. Positive Codes: NOP + LSI; Negative Codes: MSI+SSI+PP.
Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Results demonstrate a significant difference between the Global Assessment classifications of the clinical and control groups, with the clinical group demonstrating significantly more negative overall classifications. In considering interpretation of these categories, No Pathology (NOP)

typically indicates adequate adjustment with the absence of significant symptoms or impairments. Similarly, while some mild affective difficulties may be present in a Low (LSI) classification, no considerable impairment is noted (due to the role of containing defenses). As expected, 83% of the control group, free from pathology, fell within these healthy categories. Conversely, the Medium (MSI) and Severe (SSI) codes indicate problematic affective, behavioral, and interpersonal symptoms that result in increasing levels of functional impairment until achieving a significant degree of mental health pathology (PTL). While only 17% of the control group fell within this most negative classification category, 40.4% of the clinical group were classified in the negative range (MSI+SSI+PTL) suggesting a significant increase in symptoms and impairment for adolescents exposed to trauma.

Given these findings, box-by-box analysis was conducted by the authors, investigating differences between clinical and control groups in terms of positive and negative Codes produced in each WDCT box. As described previously, Codes for each box are calculated by summing the Affective Quality and Evocative Character scores for each WDCT box. Six Codes are possible following this mathematical computation, two of which are considered positive (C, PC) and four of which are considered negative (AC, AD, NC, D). It was hypothesized that trauma-exposed adolescents would produce more negative codes in those areas likely to be impacted by trauma, including self-concept and self-esteem (Box 1), relationships and socialization (Boxes 2, 4, and 8), and functional energies (Boxes 3 and 5). Differences were not anticipated in Box 6, given no expected impact on cognitive ability and intellectual functioning, and Box 7, given that box values in this area tend overall to be low during adolescence due to developmental factors. Results of box-by-box analysis are presented in Table 2.31.

Table 2.31
Frequency of Distribution of Positive and Negative Codes by WDCT Box: Clinical (Trauma-Exposed) vs. Control Groups

Box	Group	Frequency [Positive] (χ^2)	% [Positive]	Frequency [Negative] (χ^2)	% [Negative]	χ^2	*p*	Φ^1
Box 1	Clinical	58 (2.82)	56	46 (6.43)	44	11.66	.0006	.15
	Control	293 (0.73)	73	108 (1.67)	27			
Box 2	Clinical	50 (9.01)	48	54 (24.69)	52	42.43	.00001	.29
	Control	320 (2.34)	80	81 (6.40)	20			
Box 3	Clinical	44 (1.09)	42	60 (1.07)	58	2.71	.099	.07
	Control	206 (0.28)	51	195 (0.28)	49			
Box 4	Clinical	36 (9.29)	35	68 (12.43)	65	27.35	.00001	.23
	Control	253 (2.41)	63	148 (3.22)	37			
Box 5	Clinical	46 (7.66)	44	58 (15.10)	56	28.66	.00001	.24
	Control	289 (1.99)	72	112 (3.97)	28			

Box 6	Clinical	84 (0.06)	81	20 (0.30)	19	0.449	.5027	.03
	Control	335 (0.02)	84	66 (0.08)	16			
Box 7	Clinical	42 (1.17)	40	62 (1.07)	60	2.82	.092	.07
	Control	199 (0.30)	50	202 (0.28)	50			
Box 8	Clinical	58 (0.01)	56	46 (0.07)	44	0.040	.841	.01
	Control	228 (0.00)	57	173 (0.00)	43			

Note. Positive Codes: C+PC; Negative Codes: NC+AC+AD+D.
Effect sizes calculated by current authors from previously published analyses; $^{1}\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Results of the box-by-box comparison suggest overall lower functioning on the part of the clinical group, particularly in respect to psychoaffective issues, generally as predicted. It should be noted that Box 6, the box of rational thinking, intellect, and control (and therefore the most ego-driven of the CWS boxes), did not reach significance, suggesting the profound impact of trauma on affective experience (but not on intellect). Additionally, Box 7, which measures libidinal energies and ability to engage in emotionally intimate relationships, demonstrated no significant difference between groups; however, it should be noted that due to developmental tasks of adolescence (including questions of identity and intimacy), Box 7 Codes tend to be lower for all adolescents during this time.

In terms of the remaining boxes, a significant split between clinical and control groups was noted in Box 1, with the clinical group demonstrating significantly more negative Codes. Box 1 evokes an individual's feelings about the self and identity, with positive values linked to adequate ego functioning, harmony, and functioning in the environment. Lower scores, evidenced by the clinical group, in turn suggest lack of balance, disrupted sense of self, potential isolation and insecurity, and increased anxiety.

Box 3, which demonstrated a moderate (yet non statistically significant) deficit on the part of the clinical group, reflects an individual's ability to productively direct psychic energy in the service of adaptation to the environment, including goal-directed behaviors. In the trauma-exposed group, this ability appeared diminished and variable, consistent with trauma-related literature indicating an increase in "blocked energy" or depressive affect. Similarly, Box 5, which represents an individual's integration of aggressive energy into the personality structure, appears significantly decreased in the clinical group as compared to matched controls. As might be expected based upon literature, lower Box 5 Codes may suggest limited frustration tolerance, limited effectiveness in dealing with negative feedback, pessimism, a view of the environment as dangerous and aggressive, and internalized aggressive impulses (intrapunitive thoughts and behaviors).

Boxes 2 and 4 suggest decreased sensitivity and relatedness in relationships, as expected. However, Box 8 demonstrated a unique and significant similarity between trauma-exposed and control group adolescents. This box relates to the social dimension of functioning, including social skills, interpersonal orientation, and the ability to engage in mutual relationships. As noted above, the adolescents in the clinical group produced a similar distribution of positive codes in this box as compared to their control group counterparts. While not predicted in the current study, several authors have suggested the reactive function trauma serves within communities—that resilient individuals impacted by disaster or difficulty may tend to band together, support each other, and look to others who have shared their experience. Higher degrees of social support, reliance on

others, and sensitivity to the needs of others have been reported at the community-level post natural disaster (SAHMSA, 2014), which may, in part, suggest the reason behind the current findings. Given the exploratory nature of this study, further research in this area is recommended.

Crisi (2000)

The ability of the CWS to differentiate between groups was further evaluated in a study of children from economically disadvantaged neighborhoods versus matched children from the CWS normative sample. Given the higher rate of poverty, crime, and difficulties experienced by the experimental ("economically disadvantaged") group, it was hypothesized that these children's protocols would demonstrate higher levels of impairment and symptoms, and fewer indicators of overall positive health.

1,000 children (520 male, 480 female) aged 6-10 were included in the study, selected from an economically disadvantaged school within a poor urban neighborhood. Participants were divided into 5 equal bands, based upon age: 1) 6 years (*N*=232, 23.2% of experimental group; males: *N*=118, $\bar{x}$=6.42, *SD*=0.323; females: *N*=114, $\bar{x}$=6.45, *SD*=0.321); 2) 7 years (*N*=227, 22.7% of experimental group; males: *N*=129, $\bar{x}$=7.33, *SD*=0.613; females: *N*=98, $\bar{x}$=7.27, *SD*=0.819); 3) 8 years (*N*=248, 24.8% of experimental group; males: *N*=123, $\bar{x}$=8.32, *SD*=0.287; females: *N*=125, $\bar{x}$=8.36, *SD*=0.298); 4) 9 years (*N*=172, 17.2% of experimental group; males: *N*=85, $\bar{x}$=9.42, *SD*=0.303; females: *N*=87, $\bar{x}$=9.453 *SD*=0.287); and 5) 10 years (*N*=121, 12.1% of experimental group; males: *N*=65, $\bar{x}$=10.40, *SD*=0.315; females: *N*=56, $\bar{x}$=10.34, *SD*=0.311). No significant differences were discovered between groups regarding age and gender. The clinical group was compared to a sample of children selected from the CWS child normative sample (*N*=282; 113 female, age: $\bar{x}$=9.3602, *SD*=2.6043; 169 male, age: $\bar{x}$=9.0953, *SD*=2.595; range: 6-14; See Appendix A for further information).

Initial analyses examined differences in overall Global Assessment classifications between the economically disadvantaged and control group children. As described elsewhere in this chapter, in completing the CWS scoring mechanics, each protocol is assigned an overall Global Assessment classification, ranging from "No Pathology" (NOP) to "Pathological" (PTL). This classification system is divided into five categories, two of which are considered positive codes (NOP + "Low Symptoms" [LSI]) and three of which are considered negative codes (PTL + "Moderate Symptoms" [MSI] and "Severe Symptoms" [SSI]). Clinical and control groups were compared based on their classification within this Global Assessment rubric. See Table 2.32.

Table 2.32
Distribution of Global Assessment Classifications between Experimental (Economically Disadvantaged) and Control Groups

Global Assessment	Control		Experimental		$\chi^2_{(df=4)}$	*p*	*V*[1]
	N	%	N	%			
NOP (No Pathology)	95	32.8	188	18.8	51.95	.0001	.20
LSI (Low Impairment)	102	35.3	295	29.5			
MSI Medium Impairment)	60	20.8	234	23.4			
SSI (Severe Impairment)	26	8.9	191	19.1			
PTL (Pathological	6	2.0	92	9.2			

Note. Effect sizes calculated by current authors from previously published analyses; [1]*V* effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Chi-square analysis clearly illustrates significant differences in overall level of impairment and mental health symptoms experienced by the experimental and control groups. To better understand these differences, investigation of differences in positive Global Assessment classifications (NOP + LSI) and negative Global Assessments (PTL + SSI + MSI) were further studied. Overall, the experimental group demonstrated a significantly higher percentage of negative codes (51.7%) as compared to the control group (31.9%), with statistically significant results (χ^2=35.5017, p<.00001). This suggests, as hypothesized, that the children of the disadvantaged neighborhood demonstrated greater difficulties, symptoms, and impairments in their mental health functioning, as compared to their matched control group counterparts.

To further understand these differences, evaluation of differences occurring in each WDCT box was subsequently undertaken. It was noted that the experimental group demonstrated a higher than expected (as compared to normative data) number of Delay (D) and Negative Compensation (NC) box Codes. Negative Compensations have high diagnostic meaning, indicating strong internal conflict or difficulty, which have likely been "covered-up" or compensated for behaviorally. As such, difficulties may not be apparent at the behavioral level (given a demonstrated behavioral adaptation masking internal conflict). In the experimental group, the highest percentages of NC codes occurred in Boxes 2, 3, and 4, suggesting tendencies to overwork (without success), potential school phobia, possible conflict due to false assertiveness, and attachment challenges masked via social behaviors. Similarly, higher than expected levels of Delay (D) codes were discovered in the experimental group, specifically in Boxes 7 (24.4%), 5 (20.2%), 6 (18.7%), and 3 (18.7%). These findings suggest that children in dangerous and disadvantaged neighborhoods demonstrate strong repression of affectivity, higher levels of rigidity, difficulties planning and organizing or making sense of the world, and ego-centric self-protection from intimacy. High levels of concurrent Delay in Boxes 5 and 7 are indicative of strong internal conflict, including neurotic personality organization.

Lastly, specific differences between groups in terms of negative Codes by box were investigated. The presence of the two most negative codes, described above (D + NC) were, studied for each box, with significant differences noted in 5 out of 8 boxes between the experimental and control groups. Results are presented in Table 2.33.

Table 2.33
Distribution of Negative Codes by Box between Experimental (Economically Disadvantaged) and Control Groups

Box	Control		Experimental		$\chi^2_{(df=1)}$	p	Φ^1
	N	%	N	%			
1	47	16.2	151	15.1	.2332	.629	.01
2	40	13.8	188	18.8	4.676	.030	.06
3	27	9.4	480	38.0	140.411	.00001	.33
4	80	27.6	246	24.6	.5691	.4506	.02
5	38	13.2	120	12.0	.2751	.5999	.02
6	5	1.7	81	8.1	14.61	.0001	.10
7	54	18.6	317	31.7	20.988	.00001	.13
8	9	3.1	97	9.7	12.88	.0003	.10

Note. Negative valuation includes presence of Delay (D) and Negative Compensation (NC) in each box.
Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

As evident from Table 2.32, children in the experimental group demonstrated significantly more negative Codes in the boxes related to sensitivity, social skills, and the ability to enter into mutual relationships (Box 2, Box 7, Box 8), and rational control, planning and organization, and achievement of goals (Box 3, Box 6). Taken with other results, experimental hypotheses were considered confirmed, with the author concluding that disadvantaged neighborhoods (including crime, poverty, and general struggle) results in less effective adaptation and a higher degree of conflict or manifest mental health symptoms. The CWS was determined to effectively discriminate between groups, highlighting behavioral symptoms and impairments (Delay), as well as more internal, behaviorally masked symptoms (Negative Compensation).

Crisi and Pastore (2006, 2010)

Crisi and Pastore, using CWS indices, examined the characteristics of work-abused victims compared to a group of control subjects. Estimates by the International Labor Organization (Harnois, Gabriel, & World Health Organization, 2000) at the time the study was completed indicated that approximately 9% of European workers (approximately 13 million) had been the victim of work-related abuses. Work abuse is defined as a systematic and long standing (more than 6 months) pattern of psychological abuse inflicted on an individual by another individual or group in the workplace. Results of work abuse were documented to include lower self-esteem, high degree of personal distress, increased anxiety, social withdrawal, potential difficulties thinking and concentrating, and other affective impairments.

Given these previous findings, several experimental hypotheses were developed by the authors:

1) Participants in the work-abused group would demonstrate lower self-esteem as evident by less positive (and therefore more negative) Codes in Box 1, the box of self-concept.
2) Participants in the work-abused group will demonstrate tendencies toward social withdrawal, as evident by less positive Codes in Box 8
3) Participants in the work-abused group would demonstrate difficulties reacting to frustrations and obstacles, similar to previous research with individuals who have experienced traumatic events, evidenced through lower Codes in Box 5.
4) Participants in the work-abused group will demonstrate a lower level of functioning in the social-affective environment, as evidenced by lower EC+%, and potentially decreased levels of rational control and logical thinking ability (as evidenced by lower FQ+%, higher P%, lower P+ and O+ percentages, and higher number of special scores, specifically Arbitrary Performance, Global Rejection, Partial Rejection).
5) Participants in the work-abused group will demonstrate higher levels of internal distress, as evidenced by higher Anxiety Index, higher Index of Inner Tension-1, higher levels of Secondary Movement, and higher number of Special Scores related to tension (Self Criticism and Anxiety Stroke).

Participants in the study included 107 individuals who had experienced work abuse (55 female, 52 male), divided between two age bands (35-45, 45-60). Participants were spread across four types of employment: civil servants (*N*=25; 14 females, 11 males); civil management (*N*=27*;* 13 females, 14 males); private employees (*N*=28; 14 females, 14 males); and privately employed executives (*N*=27; 14 females, 13 males). Each group was equally divided between age ranges. This experimental group was compared to 299 subjects (149 female, 150 male; $\bar{x}$=20.94, *SD*=2.82) randomly selected from the 2005 CWS standardization sample. While few differences were noted within the experimental group, significant differences were noted between the experimental and control groups.

In regards to the first three experimental hypotheses, Codes of three specific boxes were analyzed,

thought to be related to trauma: 1) Box 1, which relates to an individual's feelings regarding the self, the identity, and personal attributes; 2) Box 5, which indicates how an individual reacts to frustrations and the degree to which aggressive impulses are integrated into the personality structure; and 3) Box 8, which relates to the psychoaffective dynamics of socialization and the manner in which individuals successfully navigate their environment and those they encounter. For each box, chi-square analysis was completed, comparing the frequency of positive, ambivalent, and negative codes between groups.

As discussed earlier, using the CWS methodology, each box of a WDCT protocol is assigned a code based upon a mathematical summation of Affective Quality and Evocative Character scores. Six codes are possible in each box: 2 positive (C, PC), 2 ambivalent (AC, AD), and 2 negative (NC, D). Results of chi-square analysis for Boxes 1, 5, and 8 are provided in Table 2.34.

Table 2.34
Distribution of Positive, Ambivalent, and Negative Codes in Boxes 1, 5, and 8: Work-Abused vs. Control Group

Box	*Valuation*	*Work-Abuse*		*Control*		χ^2
		N	*%*	*N*	*%*	*p*
Box 1	Positive	62	58.0	218	72.9	.01
	Ambivalent	36	33.6	72	24.0	.06
	Negative	9	8.4	9	3.1	.05
Box 5	Positive	50	46.7	212	70.9	.0001
	Ambivalent	34	31.8	76	25.4	---
	Negative	23	21.5	11	3.7	.001
Box 8	Positive	66	61.7	169	56.5	---
	Ambivalent	30	28.0	119	39.8	.05
	Negative	11	10.3	11	3.7	.05

In regards to these three boxes, significant differences were noted in each box, with control group participants demonstrating more positive Codes in Box 1 and Box 5, and work-abused individuals demonstrating more negative Codes in Box 1, Box 5, and Box 8. As predicted by the authors, this suggests lower self-esteem in individuals who have experienced work-abuse (Box 1), higher levels of internalize anger and tension (Box 5), and negatively impacted social relationships (Box 8). In general, differences in both statistical indices and box Codes suggested overall adaptation challenges to the situation or environment on the part of the work-abused participants, internal discomfort, challenges in thinking clearly and logically, and higher levels of anxiety. These finding are consistent with overall box Codes at the protocol level, as Table 2.35 details, suggesting general levels of dysfunction and higher levels of psychopathology.

Table 2.35
Distribution of Codes of All Boxes: Work-Abused vs. Control Group

Box Valuation	*Work-Abuse*		*Control*		χ^2
	N	*%*	*N*	*%*	*p*
Choice (C)	242	28.3	807	33.7	.005
Ambivalent Choice (AC)	134	15.6	328	13.7	---
Negative Compensation (NC)	51	5.9	61	2.6	.001
Positive Compensation (PC)	127	27.7	754	31.5	0.05
Ambivalent Delay (AD)	116	13.6	318	13.3	---
Delay (D)	76	8.9	124	5.2	.001

As noted above, work-abused participants demonstrated a significantly lower number of positive Codes and a significantly higher number of negative Codes than their control group counterparts at the protocol level. This suggests the negative impact of on-going trauma may have on personality integration, flexibility, and adaptability. While statistically significant differences were noted in 6 CWS Boxes (1, 3, 4, 5, 6, and 8), the most pronounced impact appears in Boxes 1, 5, and 8 (as previously described in Table 2.34).

To investigate experimental hypotheses 4 and 5, differences between groups experimental and control groups were investigated regarding indices related to environmental functioning, logical thinking and rational control, and internal tension and distress. Results of these analyses are presented in Table 2.36.

Table 2.36
Significant Differences on Major CWS Indices Between Work-Abused ("Work") and Control ("Control") Groups

Index, Content, Special Score	*Work* $\bar{x}$	*Work SD*	*Control* $\bar{x}$	*Control SD*	*F (2, 404)*	*p*	*d*[1]
Indices							
Evocative Character (EC+%)	70.56	15.81	77.14	11.95	19.95	.000**	0.50$_b$
Affective Quality (AQ+%)	60.07	12.01	64.48	11.88	10.80	.000**	0.37$_b$
Form Quality (FQ+%)	86.07	14.45	99.76	1.58	261.17	.000**	1.82$_b$
Popular Responses (P%)	25.72	13.21	19.38	10.84	23.89	.000**	0.55$_a$
P% with Good Form (P+%)	91.68	15.00	99.82	3.02	73.31*	.000**	0.96$_b$
Original Response (O%)	4.82	7.03	0.24	1.62	111.06	.000**	1.18$_a$
O% with Good FQ (O+%)	71.71	25.09	100.00	0.00	8.73*	.000**	0.33$_b$
Index of Inner Tension 1 (IIT-1)	1.14	0.39	1.22	0.29	4.44	0.04	0.23$_b$
Anxiety Index (AI)	0.82	0.28	0.94	0.15	30.91	.000**	0.63$_b$
Secondary Movement (m)	0.73	0.83	0.00	0.95	4.12	.04	0.22$_a$
Special Scores							
Self Criticism SS (SC)	1.10	1.61	0.00	0.00	141.37	.000**	1.33$_a$
Anxiety Stroke SS (AS)	6.46	2.23	7.46	1.30	31.37	.000**	0.63$_a$
Arbitrary Performance (AP)	0.22	0.57	0.00	0.02	34.70	.000**	0.66$_a$
Global Rejection (GR)	0.11	0.44	0.00	0.07	20.28	.000**	0.51$_a$
Partial Rejection (PR)	0.16	0.73	0.00	0.02	8.79	.000**	0.33$_a$

Note. *Cases excluded; **Significance noted following Bonferroni ($p<.05$) post-hoc correction: $p<0.003$.
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
$_a$Work group values greater than control group.
$_b$Control group values greater than work group.

Results suggest that work-abused subjects differ significantly from control subjects in several main indices in the expected direction. Primarily, work-abused subjects demonstrated a statistically lower EC+%, indicating a lower level of adjustment within the affective-relational environment, as well as decreased sensitivity to social context and situations. Additionally, work-abused subjects

demonstrated lower FQ+%, potentially indicating a lower level of cognitive functioning (including possible concentration, thinking, and rational problem solving), likely connected to a traumatic condition in the workplace. Work-abused subjects were noted to produce protocols with higher P%, suggesting rigidity and conformity (with lower P+%), as well as innovative responses without an adequate interpretation of reality (Higher level of O% with negative Form Quality, lower O+%) suggesting some disconnection from the environment and internal retreat into fantasy. Work-abused participants further demonstrated high percentages of Special Scores related to general psychoaffective difficulty (Global Rejection, Partial Rejection). Lastly, work-abused subjects were noted to demonstrate significantly higher tendency to ruminate and interpret situations in an inadequate and potentially illogical manner (AP), suggesting a retreat from the environment and a developing idiosyncratic view of the world as a result of trauma.

In regards to hypothesis 5, specifically related to internal distress, results were mixed. No significant differences were noted between the experimental and control groups as related to special score Anxiety Stroke, Anxiety Index (AI), or Index of Inner Tension-1 (IIT-1). However, the work-abused participants produced higher rates of Secondary Movement and other Special Scores, as predicted. Specifically, higher frequency of Self Criticism was noted in the work-abused sample, indicating pronounced feelings of insecurity, indecision, and fear. Moreover, while not predicted, Incomplete Drawing special scores were noted as higher in the work-abused subjects, further supporting the hypothesis related to insecurity and distress. This suggests that AI and IIT-1 are measuring trait anxiety, which may rise in the control group during test administration, and may not always effectively differentiate between characterologically anxious and transiently anxious individuals. Further exploration of this finding is recommended.

Overall, test results confirm experimental hypotheses 1-4, with mixed results noted for hypotheses 5. Given this, these results are generally consistent with previous literature related to work-abuse, demonstrating the Crisi Wartegg System's sensitivity in evaluating psychoaffective functioning.

Following investigation of the experimental hypotheses, further exploratory analysis was conducted on 78 additional CWS variables. Statistically significant differences were noted in 45 of 78 studied CWS variables following Bonferroni post-hoc correction, presented in Table 2.37.

Table 2.37
Exploratory Analysis of Significant Differences on Major CWS Indices Between Work-Abused ("Work") and Control (Control) Subjects

Index, Content, Special Score	***Work $\bar{x}$***	***Work SD***	***Control $\bar{x}$***	***Control SD***	***F (2, 404)***	***p***	***d[1]***
Indices							
Affective Stability Index (A/F): A	2.50	0.68	2.81	0.65	17.41	.000**	0.47_b
Affective Stability Index (A/F): F	3.23	0.73	3.94	0.35	171.04	.000**	1.47_b
Index of Inner Tension 2 (IIT-2) *(first half of ratio)*	4.5	1.64	5.22	1.35	20.28	.000**	0.51_b
Index of Inner Tension 2 (IIT-2) *(second half of ratio)*	3.50	1.64	2.78	1.35	20.28	.000**	0.51_a
Content Categories							
Object Content (OBJ%)	31.07	18.84	38.09	19.68	10.25	.000**	0.36_b

Symbol Content (SIG%)	13.93	17.15	4.89	9.31	45.60	.000**	0.76_a
Nature Content (NAT%)	2.80	5.77	5.89	8.21	12.88	.000**	0.40_b
Abstract Content (ABS%)	3.97	11.97	0.00	0.00	33.07	.000**	0.65_a
Special Scores							
Personalized Answer (PA)	0.26	0.77	0.02	0.16	26.34	.000**	0.58_a
Inadequate Integration SS (II)	0.13	0.34	0.03	0.19	14.11	.000**	0.42_a
Perseveration (PP)	0.11	0.42	0.00	0.00	21.46	.000**	0.52_a
Stimulus Repetition (SR)	0.31	0.99	0.00	0.15	5.36	.02	0.26_a
ADP and AFF Areas							
ADP AQ+% (AQ+%[ADP])	58.66	14.22	64.86	14.58	14.41	.000**	0.43_b
ADP FQ+% (FQ+%[ADP])	87.87	14.29	99.84	1.74	202.22	.000**	1.60_b
ADP: Affective Stability Index (A/F): A	1.19	0.44	1.37	0.47	11.26	.000**	0.38_b
ADP: Affective Stability Index (A/F): F	1.66	.044	1.98	0.16	116.45	.000**	1.22_b
ADP Popular (P%[ADP])	28.05	17.64	19.77	14.33	23.17	.000**	0.54_a
ADP Popular with Good FQ (P+%[ADP])	91.76	14.93	99.79	3.21	62.97*	.000**	0.90
ADP Original (O%[ADP])	3.91	9.21	0.00	0.00	53.90	.000**	0.83_a
ADP IIT-1 (IIT-1[ADP])	1.13	0.43	1.23	0.35	6.33	.01	0.28_b
ADP AI (AI[ADP])	0.79	0.31	0.93	0.18	32.37	.000**	0.64_b
ADP Number of Special Scores	5.67	1.90	5.15	1.50	8.15	.000**	0.32_a
ADP Choice Codes (C)	1.35	0.85	1.53	0.80	3.87	0.05	0.22_b
ADP Negative Compensation Codes (NC)	0.21	0.43	0.11	0.33	7.05	.01	0.30_a
AFF EC+% (EC+%[AFF])	64.99	19.75	79.12	17.39	48.31	.000**	0.78_b
AFF FQ+% (FQ+%[AFF])	84.36	17.89	99.71	2.36	210.94	.000**	1.63_b
AFF: Affective Stability Index (A/F): A	1.30	0.48	1.44	0.50	5.78	.02	0.27_b
AFF: Affective Stability Index (A/F): F	1.57	0.46	1.96	0.22	129.84	.000**	1.28_b
AFF Popular (P%[AFF])	23.65	20.14	19.10	17.04	5.09	.02	0.25_a
AFF Popular+ (P+%[AFF])	92.61	16.75	100.00	0.00	37.60	.000**	0.69_b

AFF Original (O%[AFF])	5.68	11.37	0.48	3.20	51.45	.000	0.80_a
AFF Original+ (O+%[AFF])	72.00	25.33	100.00	0.00	8.35*	.01	0.33_b
AFF AI (AI[AFF])	0.85	0.28	0.95	0.15	21.96	.000**	0.53_b
AFF IIT-2 (IIT-2[AFF]) *(first half of ratio)*	2.04	1.05	2.65	0.97	30.06	.000**	0.62_b
AFF IIT-2 (IIT-2[AFF]) *(second half of ratio)*	1.96	1.05	1.35	0.97	30.06	.000**	0.62_a
AFF Number of Special Scores	6.08	1.95	5.14	1.34	30.58	.000**	0.62_a
AFF Number of Rejections (GR/PR)	0.21	0.61	0.07	0.45	7.84	.01	0.32_a
AFF Primary Movement	0.28	0.61	0.15	0.45	5.70	.02	0.27_a
AFF Secondary Movement	0.39	0.63	0.55	0.75	3.89	.05	0.22_b
AFF Choice Codes (C)	0.93	0.68	1.17	0.71	9.93	.000**	0.36_b
AFF Negative Compensation Codes (NC)	0.26	0.48	0.10	0.31	16.34	.000**	0.46_a
AFF Positive Compensation Codes (PC)	1.11	0.87	1.47	0.85	14.19	.000**	0.42_b
AFF Delay Codes (D)	0.52	0.69	0.31	0.54	10.70	.000**	0.37_a

Note. *Cases excluded; **Significance noted following Bonferroni ($p<.05$) post-hoc correction: $p<0.009$.
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]Work group values greater than control group.
[b]Control group values greater than work group.

Results suggested that work-abused subjects differ significantly from control subjects, with several post-hoc hypotheses put forth by the authors. In terms of Primary Content category percentages, work-abused subjects demonstrated a less pronounced disposition to practical abilities than control subjects (lower OBJ% and higher ABS%), a more defensive attitude toward the testing situation (Higher SIG%), and a lower adjustment to social demands (lower NAT%). They further demonstrated high percentages of Special Scores related to self-referential introspection and egocentrism (PA). Overall, more negative functioning was noted for the work-abused group in both the Adaptive (ADP) area, indicating challenges in social-environmental functioning, and the Affective (AFF) area, indicating deficits in emotional regulation, control, and integration. Given these findings, the impact of work abuse on individuals appears pervasive and significant, with the need for more research indicated.

Eating Disorders

Daini, Lai, Festa, Maiorino, Pertosa, & De Risio (2006)

Using the CWS, the psychological and psychopathological characteristics of individuals with eating disorders were evaluated by Daini and colleagues, focusing specifically on impulsivity. The WDCT, according to the CWS method was administered given the quantitative scoring system, relatively brief administration time, and proven validity of the coding system (D'Amore, 2004; in Daini et al.,

2006). It was hypothesized that the CWS would differentiate between anorexic, bulimic, and control groups, with increased depression and impulsivity noted in the bulimic group, and affective overcontrol, perfectionism, and impulsivity noted in the anorexic group.

One hundred female pre-treatment outpatients selected to participate in hospital-based intervention for eating disorders were included in this study. Eating disorder diagnoses were established via psychiatric interview utilizing the DSM-IV (American Psychiatric Association, 1994) and the *Eating Disorder Evaluation Scale* (EDES, Vandereycken, 1993). The clinical (eating disorder) group was subdivided by diagnosis: anorexia, restricting type (ANrt, N=35; age: $\bar{x}$=21.4, SD=6.5); anorexia, purging type (ANpt, N=22; $\bar{x}$ and SD not reported), and bulimia (BL, N=43; age $\bar{x}$=24.3, SD=6.6). The Control (CN) group (age: $\bar{x}$=24.4, SD=6.5) was comprised of 81 females matched to the clinical groups on educational and socioeconomic status. It was noted that the mean age of the ANrt group was significantly lower than the bulimia and control groups (p<.01), with age subsequently considered as a covariate in statistical analyses.

To determine differences between groups related to impulsivity and anxiety, three CWS indices were compared by MANCOVA: the Index of Inner Tension (IIT-1), which measures overall distress and internal discomfort; the Anxiety Index (AI), which measures general levels of internal and manifest anxiety; and the Impulsivity Index (IM), which measures cognitive, affective, and behavioral impulsivity. Results of these analyses are presented in Table 2.38.

Table 2.38
Mean Differences in CWS Indices [Index of Inner Tension (IIT-1), Anxiety Index (AI), and Impulsivity Index (IM)]
MANCOVA by Group (Anorexic RT, Anorexic PT, Bulimic, Control)

Index	Anorexic RT (ANrt) $\bar{x}$ (*SD*)	Anorexic PT (ANpt) $\bar{x}$ (*SD*)	Bulimic (BL) $\bar{x}$ (*SD*)	Control (CN) $\bar{x}$ (*SD*)	*F* (3,176)	*LSD Post-Hoc*	*d*[1]
Index of Inner Tension (IIT-1)	1.20 (0.4)	0.98 (0.3)	1.02 (0.5)	0.91 (0.4)	4.0**	ANrt>ANpt** ANrt>CN**	0.45
Anxiety Index (AI)	0.80 (0.2)	0.66 (0.2)	0.69 (0.3)	0.62 (0.2)	4.1**	ANrt>ANpt* ANrt>BL* ANrt>CN**	0.39
Impulsivity Index (IM)	0.48 (0.3)	0.37 (0.3)	0.37 (0.3)	0.28 (0.2)	5.5**	ANrt>CN** BL>CN*	0.44

Note. *p<.05; **p<.01.
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).

Table adapted from *Daini, Lai, Festa, Maiorino, Pertosa, & De Risio (2006).*

Results demonstrated clear significant differences between clinical groups, as well as between clinical and control groups. In terms of overall inner tension and experienced distress and discomfort (IIT-1), the restricting anorexic (ANrt) group demonstrated significantly higher scores than both the purging anorexic (ANpt) and control (CN) group. Similarly, in terms of experienced anxiety, the restricting anorexic (ANrt) group demonstrated significantly higher scores than the purging anorexic (ANpt), bulimic (BL), and control (CN) groups. Lastly, both the restricting anorexic (ANrt) and bulimic (BL) groups demonstrated higher levels of impulsivity than the control

(CN) group. These findings generally supported the authors' experimental hypotheses in that the bulimic group demonstrated higher impulsivity than the control group (although did not demonstrate higher levels of internal distress, as predicted), and the anorexic group demonstrated significant over-control (restricting type) and impulsivity. From this data, the authors concluded that the CWS clearly differentiates between clinical and control groups in terms of anxiety, impulsivity and experienced distress.

To investigate hypotheses related to increased presence of rationalization, perfectionism, and overcontrol in the ANrt population, the distribution of five Primary Contents (Human [H%], Animal [A%], Object [OBJ%], Nature [NAT%], and Abstract [ABS%]) and one Special Score (Global Rejection [GR]) was evaluated using chi-square to determine differences between groups. Results demonstrate higher frequency of Abstract (ABS%) content in restricting anorexic patients (ANrt) as compared to bulimic and control group participants (χ^2=15.4, p<.01). Moreover, restricting anorexic participants further demonstrated fewer Global Rejections (χ^2=29.3, p<.01). Together, per the authors, these results suggested that restricting anorexic patients demonstrate strong rationalization defenses and perfectionism, as supported by prior research and hypothesized in the current study. No additional differences between experimental and control groups were noted in regards to the Primary Contents analyzed.

To further investigate the overall hypotheses related to depression, the investigators further evaluated differences between groups in Affective Quality (AQ), Evocative Character (EC), and Form Quality (FQ) scores in each WDCT box, hypothesizing that significant differences might identify specific intrapsychic challenges or internal conflict experienced by the members of each group. Results are presented in Table 2.39.

Table 2.39
Mean Differences in CWS Indices (EC+%, AQ+%, FQ+%) by Box
MANCOVA by Group (Anorexic RT, Anorexic PT, Bulimic, Control)

Box	Index	Anorexic RT (ANrt) $\bar{x}$ (SD)	Anorexic PT (ANpt) $\bar{x}$ (SD)	Bulimic (BL) $\bar{x}$ (SD)	Control (CN $\bar{x}$ (SD)	*F* (3,176)	*LSD Post-Hoc*	*d*[1]
.Box 1	**EC+%**	0.80 (0.4)	0.82 (0.3)	0.83 (0.3)	0.89 (0.3)	0.6		0.19_a
	AQ+%	0.76 (0.4)	0.66 (0.3)	0.69 (0.4)	0.81 (0.3)	2.1	CN>BL* CN>ANpt*	0.32_a
	FQ+%	0.86 (0.3)	0.77 (0.4)	0.88 (0.3)	0.93 (0.2)	2.0		0.29_a
Box 2	**EC+%**	0.77 (0.4)	0.77 (0.4)	0.80 (0.3)	0.74 (0.4)	0.2		0.10_b
	AQ+%	0.67 (0.7)	0.60 (0.6)	0.74 (0.7)	0.78 (0.8)	1.5	CN>ANpt*	0.24_a
	FQ+%	0.86 (0.3)	0.82 (0.3)	0.94 (0.2)	0.91 (0.2)	1.4		0.10
Box 3	**EC+%**	0.61 (0.3)	0.79 (0.3)	0.60 (0.3)	0.56 (0.3)	3.4*	ANpt>ANrt* ANpt>BL* ANpt>CN**	0.28_b
	AQ+%	0.64 (0.3)	0.57 (0.2)	0.58 (0.2)	0.68 (0.2)	1.9	CN>BL*	0.31_a
	FQ+%	0.90	0.73	0.76	0.89	4.2**	ANrt>ANpt*	0.33

		(0.2)	(0.4)	(0.3)	(0.2)		ANrt>BL* CN>ANpt* CN>BL**	
Box 4	**EC+%**	0.53 (0.5)	0.68 (0.4)	0.64 (0.4)	0.52 (0.5)	1.1		0.19[b]
	AQ+%	0.59 (0.3)	0.48 (0.3)	0.46 (0.3)	0.58 (0.3)	1.8		0.21
	FQ+%	0.81 (0.3)	0.73 (0.4)	0.84 (0.3)	0.88 (0.3)	1.4	CN>ANpt*	0.23[a]
Box 5	**EC+%**	0.82 (0.3)	0.74 (0.3)	0.80 (0.3)	0.85 (0.3)	1.4		0.24[a]
	AQ+%	0.54 (0.2)	0.54 (0.1)	0.55 (0.3)	0.59 (0.2)	0.5		0.18[a]
	FQ+%	0.77 (0.4)	0.73 (0.4)	0.79 (0.3)	0.92 (0.2)	4.8**	CN>BL* CN>ANrt* CN>ANpt**	0.55[a]
Box 6	**EC+%**	0.83 (0.3)	0.77 (0.4)	0.89 (0.3)	0.89 (0.3)	1.4		0.17[a]
	AQ+%	0.56 (0.2)	0.48 (0.2)	0.60 (0.2)	0.59 (0.2)	1.5	CN>ANpt*	0.12
	FQ+%	0.86 (0.3)	0.79 (0.4)	0.87 (0.3)	0.92 (0.2)	1.8	CN>ANpt*	0.28[a]
Box 7	**EC+%**	0.50 (0.4)	0.51 (0.5)	0.49 (0.5)	0.51 (0.5)	0.2		0.08[a]
	AQ+%	0.73 (0.3)	0.57 (0.3)	0.64 (0.4)	0.68 (0.3)- 0.23	1.2		0.07
	FQ+%	0.91 (0.2)	0.75 (0.4)	0.81 (0.3)	0.90 (0.3)	2.4	CN>ANpt* ANrt>ANpt*	0.23
Box 8	**EC+%**	0.73 (0.3)	0.80 (0.3)	0.79 (0.3)	0.75 (0.3)	0.4		0.07
	AQ+%	0.64 (0.4)	0.61 (0.3)	0.65 (0.4)	0.73 (0.8)	0.8		0.22[a]
	FQ+%	0.97 (0.1)	0.86 (0.3)	0.95 (0.1)	0.97 (0.1)	3.0*	ANrt>ANpt* BL>ANpt* CN>ANpt**	0.20[a]
Total	**EC+%**	0.70 (0.1)	0.74 (0.2)	0.73 (0.1)	0.71 (0.1)	0.5		0.08
	AQ+%	0.64 (0.2)	0.56 (0.1)	0.61 (0.1)	0.68 (0.2)	4.1**	CN>ANpt**	0.45[a]
	FQ+%	0.87 (0.2)	0.77 (0.3)	0.86 (0.1)	0.92 (0.1)	6.1**	ANrt>ANpt* BL>ANpt* CN>ANpt**	0.51[a]

Note. *p<.05; **p<.01.
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]Control group values greater than (or equal to) all clinical groups.
[b] All clinical group values greater than control group.

Table adapted from *Daini, Lai, Festa, Maiorino, Pertosa, & De Risio (2006).*

Results indicated that relatively few differences occurred in the scoring domain of Evocative Character (EC), which measures ego processes in relation to the environment. The authors concluded that little difference exists between individuals diagnosed with eating disorders as compared to the normal population in this perceptual and goal-directed domain. The notable exception to this occurred in Box 3, which measures the individual's ability to utilize psychic resources to reach goals and objectives. In this box, restricting anorexics (ANrt) demonstrated statistically significant higher scores than all other groups (ANpt, BL, CN). This suggests that in the domain of planning capacity, restricting anorexics may demonstrate greater ability.

Contrary to the Evocative Character, the Affective Quality scores significantly differentiated between groups. Overall, control (CN) group participants presented with higher AQ scores than clinical groups, confirming research indicating higher rates of depressive and negative affect in patients with eating disorders. Specifically in this area, bulimics (BL) demonstrated statistically lower AQ in Box 1 (related to self-evaluation and esteem), and Box 3 (related to planning capacity and goal achievement). Purging anorexics (ANpt) demonstrated statistically lower AQ in Box 2 (related to femininity, vitality, and sensitivity).

As related to Form Quality, or the ability to practically and conventionally represent the reality of one's thoughts and ideas, significant differences occurred between groups. In total, control (CN) group participants demonstrated the highest Form Quality, with purging anorexics (ANpt) demonstrating significantly lower FQ scores than all other groups. Specifically, restricting anorexics (ANrt), bulimics (BL), and purging anorexics (ANpt) demonstrated significantly lower FQ in Box 5, related to regulation adaptive and integration of aggressive impulses. These results, as well as other significant findings, are summarized in Table 2.40.

Table 2.40
Relative Strengths and Deficits within the Eating Disorder Clinical Groups

Box	Anorexic-Restricting (ANrt)	Anorexic- Purging (ANpt)	Bulimic (BL)
Box 1			Deficit (AQ)
Box 2		Deficit (AQ)	
Box 3		Strength (EC)	Deficit (AQ)
Box 4			
Box 5	Strength (FQ)	Deficit (FQ)	Deficit (FQ)
Box 6			
Box 7	Strength (FQ)		
Box 8	Strength (FQ)		

Note. Table prepared through analysis of relative strengths presented in Table 2.38 *within* the clinical group. Without exception, control group scores exceeded clinical group scores.

Considering the results from the current study, the authors reached several conclusions. First, given the lack of variability in Evocative Character (EC), the adaptive and associative processes necessary for participation in psychotherapy were presumed equal between individuals diagnosed with eating disorders and the normal population, confirming research supporting successful therapeutic intervention in these diagnostic areas. Significant indications of depression were noted (AQ) in the clinical groups, with depressive affect linked to personality identity, self-realization and goal achievement, and respecting social norms in the bulimic group. Additionally, both anorexic and bulimic participants demonstrated difficulty in adaptively utilizing aggressive and assertive

impulses to overcome obstacles (FQ, Box 5). Purging anorexic patients demonstrated deficits as related to femininity and sexuality (Box 2), and a higher level of rational functioning (Box 3).

Both anorexic and bulimic members of the clinical group demonstrated greater internal distress, anxiety, and impulsivity as compared to the control group, with restricting anorexic patients appearing more anxious and perfectionistic (AI, ABS content, GR), and bulimic patients demonstrating higher levels of depression (AQ) and impulsivity (IM), as initially hypothesized. Overall, the authors asserted that their experimental hypotheses were confirmed.

Psychiatric Symptoms Related to Medical Conditions

Crisi, Vari, Velotti, Carlesimo, Guzzi, & Zavattini (2014)

Convergent validity of the CWS was further assessed in a sample of dermatology patients affected by psoriasis (a chronic skin disease) in regards to their experience of depression and negative affective states. 42 adult patients diagnosed with psoriasis and receiving treatment at the Dermatology Unit of the *Policlinico Umberto I-Sapienza University, Rome* were compared with 42 matched control group participants. As presented in Table 2.41, no significant demographic differences were noted between groups, with the exception of educational level (with experimental group demonstrating less education than control).

Table 2.41
Descriptive Statistics (Age, Gender, Education, Relationship Status, Smoking Habits) Clinical Group vs. Control Group

	Clinical Group (*N*=42)		Control Group (*N*=42)		Analysis
	$\bar{x}$	*SD*	$\bar{x}$	*SD*	*p*
Age	42.05	11.21	37.90	14.21	0.142[1]
Gender					0.126[2]
Male	24	57.1	16	38.1	
Female	18	42.9	26	61.9	
Education					0.005[2]
Secondary School	10	23.8	1	2.4	
High School	19	45.2	17	40.5	
University	13	31.0	24	57.1	
Relationship Status					0.192[2]
Yes	30	71.4	35	83.3	
No	12	28.6	7	16.7	
Smoking Habits					0.251[2]
Yes	17	40.5	12	28.6	
No	25	59.5	30	71.4	

Note. [1]t-test; [2]Chi Square

In reviewing research related to psychopathology in dermatological patients, the authors noted that depression, anxiety, and suicidal ideation are common in patients affected by skin diseases (Kurd & Gelfand, 2009; Picardi, Lega, & Tarolla, 2013; Olivier, Robert, Daihung, Urbà, Catalin, Hywel, Kurd, Troxel, Crits-Christoph, & Gelfand, 2010). Additionally, authors have found that patients with psoriasis experienced increased thoughts of suicide (21.2%) as compared to health controls (6.8%; Zacharie, Zacharie, Lei, & Pederson, 2004).

Participants were administered the WDCT according to CWS guidelines, with tests independently scored by two trained coders (inter-rater reliability coefficient, r_w=0.826). In considering previous research findings, CWS indices (AQ+%, FQ+%, EC+%, Affective Stability Index) and the presence of CWS Special Scores were examined. These indices were selected given their relative ease in calculation (i.e., possible to do by hand quickly and easily), with the goal of developing an efficient screening tool for use in the hospital setting as an alternative to potentially minimizing self-report measures. With this aim in mind, only these specific indices, as well as the presence of Special Scores, were investigated.

Given research on individuals with dermatological conditions, it was hypothesized that clinical group participants would demonstrate lower values on all investigated indices as compared to control group participants, indicating depressive affect, increased sensitivity, and reduced cognitive control. Moreover, a higher percentage of Special Scores, specifically Global Rejections, was predicted for the clinical group, suggestive of increased levels of internal distress and psychoaffective conflict, as compared to the control group.

Results of analyses, presented in Table 2.42, demonstrated lower AQ+% and FQ+% in the clinical group, as well as a lower Affective Stability Index, as predicted. Contrary to expectations, while the control group demonstrated a higher mean value of EC+% than the clinical group, no significant difference was noted.

Table 2.42
CWS Indices (AQ+%, FQ+%, EC+%, Affective Stability Ratio): Comparison between Clinical and Control Group

	AQ+%		**FQ+%**		**EC+%**	
	Median (IR)	*p*	*Median (IR)*	*p*	*Median (IR)*	*p*
Clinical (Psoriasis)	56.00 (19)	.004[1]	92.50 (18)	.015[1]	63.00 (25)	.160[1]
Control	63.00 (13)		100.00 (6)		69.00 (18)	
	Affective Stability Index: A			**Affective Stability Index: F**		
	Mean	*SD*	*p*	*Mean*	*SD*	*p*
Clinical (Psoriasis)	2.40	0.64	<0.05[2]	3.29	0.78	<0.05[2]
Control	2.71	0.64		3.74	0.41	

Note: [1]*Mann-Witney U test;* [2]*t-test*

Additionally, of the 19 CWS Special Scores evaluated, two demonstrated significant differences between clinical and control groups. As predicted, Global Rejection (GR) was noted to be scored more frequently in the clinical group ($\bar{x}$ =47.00) as compared to the control group ($\bar{x}$ =38.00; U=1.071, p=.002). Arbitrary Performance (AP) was also noted to be scored more frequently in the clinical group ($\bar{x}$ =47.57) as compared to the control group ($\bar{x}$ =37.43; U=1.095, p=.025). Using the post-hoc Bonferroni correction for multiple comparisons (p<0.002), only Global Rejection remained statistically significant following correction, as predicted by the authors.

In considering the results, the authors suggested that the current research utilizing the CWS is consistent with previous research suggesting increased depression in patients with chronic skin diseases. This is evidenced by lower AQ+% in the clinical group as compared to the control group. Similarly, the clinical group evidenced more Global Rejections, which has been linked to depressive symptoms and suicidal ideation in previous research (Daini, 2010). Tendencies toward suicidal ideation are also confirmed via lower Affective Stability Ratios in the clinical group, suggesting attempts at emotional suppression and limited self-control. Lastly, a higher rate of Arbitrary

Performance in the clinical group, coupled with a lower FQ+%, may indicate difficulties in thinking clearly, potentially due to intrusive affective experiences; however, as AP differences did not reach significance, this finding requires further investigation. No significant difference between groups was noted related to Evocative Character (EC+%), suggesting similar levels of defensiveness and sensitivity to the environment in the clinical and control groups. Given these findings, the authors concluded that the specific CWS indices studied appear sensitive in the measurement of negative affective states and anxiety, which a patient may be unwilling or unable to self-report on standard questionnaires used within the hospital setting.

CWS and Attachment

Maio, Ricci, & Crisi (2008)

In another study of convergent validity, the relationship between attachment-related personality traits, as measured by the Wartegg Index of Psychopathology (WIP) Quadrant and styles of attachment as measured by the Separation Anxiety Test (SAT; Klagsbrun & Bowlby,1976; Italian revision: Attili, 2001) were explored. In consideration of the WIP, each subject may be classified as falling within one of four "Quadrants" (as described above; see Crisi, 2009) by graphically plotting state anxiety and distress on a vertical axis and trait personality flexibility and integration on a horizontal access. This overlay of state versus trait characteristics results in four quadrants (labeled A, B, C, and D, respectively), each of which corresponds to particular personality traits and attachment styles, as described below:

> Quadrant A: In Quadrant A, characteristics of autonomy and independence predominate, indicative of a Secure attachment classification.
>
> Quadrant B: In Quadrant B, characteristics of detachment and isolation predominate, indicative of an Insecure-Dismissing attachment classification.
>
> Quadrant C: In Quadrant C, characteristics of dependency predominate, indicative of an Insecure-Preoccupied attachment classification.
>
> Quadrant D: In Quadrant D, characteristics of conflict and confusion predominate, indicative of an Unresolved attachment style.

CWS and SAT results of 30 clinically referred female subjects (aged 6-14) were evaluated. 47% of subjects (*N*=14) were referred for mental health intervention subsequent to school difficulties, including school refusal or learning disorders. 53% of subjects (*N*=16) were referred for emotional disturbances, including anxiety, depression, or relational difficulties. No significant differences in age, socio-economic status, or intellectual ability were noted between groups. Experimental hypotheses asserted that subject's attachment classification (as determined by the SAT) would be correlated with the CWS WIP Quadrant reflective of that attachment style. Specifically, SAT Secure classifications were predicted to be correlated with WIP Quadrant A; SAT Avoidant classifications with QIP Quadrant B; SAT Ambivalent classifications with WIP Quadrant D; and SAT Disorganized classifications with WIP Quadrants B and C. Moreover, it was hypothesized that individuals with high coping abilities (as indicated by the SAT) would demonstrate greater interpersonal sensitivity and lower levels of anxiety and distress, as measured by CWS indices. That is, high coping individuals would demonstrate higher EC+%, lower Index of Inner Tension-1 scores (IIT-1) and lower Anxiety Index (AI scores), as compared to low coping ability counterparts. Results, summarized in Table 2.43, demonstrate that the greatest percentage of subjects classified as having a Secure attachment style on the SAT (50%) fell within CWS WIP Quadrant A, indicating autonomy, independence, and security of attachment. Similarly, the greatest number of subjects (72%) who were classified as having an Insecure-Disorganized attachment style on the SAT fell within the CWS WIP Quadrants B (36%) suggesting detachment/ dismissing styles and C (36%) suggesting dependent/ preoccupied styles. Given the small number of participants in this pilot study

(*N*=30), no statistically significant differences were noted (χ^2= 9.727, *p*=.373).

Table 2.43
Relationship between SAT Attachment Style and CWS WIP Quadrant: Percentage of Agreement

	SAT Attachment Classification			
	Secure	Insecure-Avoidant	Insecure-Ambivalent	Insecure-Disorganized
Quadrant A ("Autonomy")	50%	70%	50%	19%
Quadrant B ("Detachment")	38%	15%	25%	36%
Quadrant C ("Dependency")	---	---	25%	36%
Quadrant D ("Unresolved")	12%	15%	---	9%

Note. Shaded areas indicate findings supportive of experimental hypotheses.

While preliminary data, these percentages suggest that the greatest number of individuals classified with Secure attachment styles fall within Quadrant A, whereas the greatest number of individuals with Insecure attachment styles fall within Quadrants B and C. While more research is needed on the relationship between WIP Quadrant and attachment classification, initial results support the clinical interpretation of the CWS WIP classification schema.

Significant differences in CWS indices were discovered between individuals with high and low-coping abilities as determined by the SAT. Specifically, individuals classified as having high coping abilities demonstrated higher levels of EC+%, suggesting greater interpersonal and social sensitivity. While this difference did not reach statistical significance, likely due to small sample size, a preliminary trend is noted which would benefit from further study. Moreover, individuals rated as having higher coping skills demonstrated lower levels of experienced distress on the CWS Index of Inner Tension-1 (IIT-1, *p* approaching significance) and lower levels of anxiety as measured by the CWS Anxiety Index (AI, *p* approaching significance). Statistical analyses are presented in Table 2.44.

Table 2.44
Relationship between Coping Skills and CWS Indices: Evocative Character (EC+%), Index of Inner Tension-1 (IIT-1), Anxiety Index (AI)

Index	Coping Style	*N*	$\bar{x}$	*SD*	*F*	*t*	*df*	*p*	*d*
EC+%	high	20	73.8	16.39	.171	.898	28	.377	0.35[a]
	low	10	68.3	14.50					
IIT-1	high	20	0.82	0.36	1.324	-1.838	28	.077*	0.67[b]
	low	10	1.12	0.50					
AI	high	20	0.54	0.28	.109	-1.743	28	.092*	0.68[b]
	low	10	0.73	0.25					

*Note. *approaching significance.*
Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]High coping group values greater than low coping group.
[b]Low coping group values greater than high coping group.

Initial results of this pilot study supported the proposed experimental hypotheses, specifically those regarding the relationship between CWS and SAT indices. The authors suggested that these findings support preliminary construct and convergent validity of the CWS in relation to attachment classification and description of coping skills.

Di Riso, Crisi, Bianchi di Castelbianco, di Renzo, Vichi, Racinaro, & Morellini (2013)

In this study, the authors investigated CWS-derived personality factors of parents classified by the Adult Attachment Projective Picture System (AAP; George & West, 2012) as falling within the Pathological Mourning framework (Bowlby, 1980; George and West, 2012), as compared to a normative sample. As George and West describe:

> When mourning is incomplete, representational models of the self and attachment figures remained unchanged and failures of the past continue to 'haunt' the present. The failure to accomplish reintegration, in which the individual's internal works is experienced as incoherent and chaotic, begets helplessness, futility, and isolation. Bowlby referred to this state as *pathological mourning* and described two main forms that he believed were powerful explanatory tools of understanding debilitating psychiatric and physical symptoms. (p. 199)

Fifty parents (35 mothers, 15 fathers) of children diagnosed with learning disabilities were initially included in a larger study investigating the emotional availability of parents as predictive of learning challenges in children. Parents were administered the Adult Separation Anxiety Checklist (ASA-CL; Manicavasaga, Silove, & Curtis, 1997), Child Behavior Checklist (CBCL; Achenbach, 1991), and the AAP.

Results of initial analysis suggested the following attachment classifications: 56% (*N*=28) of parents were classified as Unresolved/Disorganized, 21% (*N*=10) Dismissing, 11.5% (*N*=6) Preoccupied, and 11.5% (*N*=5) Secure. Twenty-six of 50 parents (50.9%) of learning disabled children fell within the framework for Pathological Mourning with 58% demonstrating exaggerated distortions of healthy mourning (unresolved pathological mourning and/or pathological mourning with personal suffering) and 42% utilizing defensive processes to block the course of healthy mourning (failed mourning). In describing these 26 parents, traumatic experiences, feelings of helplessness, emotional dysregulation, and impaired capacity to make and maintain social relationships were noted.

Following classification based upon the AAP, specific CWS indices were next examined to determine differences between the Pathological Mourning parent group and the normal population. Studied indices included the CWS Global Assessment classification, Index of Inner Tension-2 (IIT-2), Wartegg Index of Psychopathology (WIP), WIP Quadrant, Form Quality Percentage (FQ+%), Popular Response Percentage (P%), practical Content percentages (OBJ%+ARC%), and the Index of Suicidal Tendencies (IST). Each index is discussed in turn below.

CWS Global Assessment Code: As discussed earlier, using a constellation of CWS indices and calculations, each individual can be classified in one of five incremental Global Assessment categories, ranging from no pathology to significant psychopathology. In order from highest functioning to highest level of symptoms/impairment, the five categories are: Not Pathological (NOP), Low level of symptoms (L-VA), Moderate symptoms and impairment (M-VA), Severe symptoms and impairment (S-VA) and Pathological (PTL).

Index of Inner Tension-2 (IIT-2): The IIT-2 measures the degree of integration and flexibility evident in an individual's personality structure. The more positive the IIT-2, the better able the subject is to utilize various aspects and strengths of personality, solve problems, regulate emotions, consistently navigate relationships, and meet the obstacles of daily life.

Wartegg Index of Psychopathology (WIP): The WIP (described earlier in this chapter, see Crisi 2009), graphically plots state anxiety and distress on a vertical axis and trait personality

flexibility and integration on a horizontal access. An individual's WIP may be considered "positive" if falling within adaptive areas of functioning (appropriate flexibility, average distress) or "negative" if significantly inflexible, dis-integrated, and/or distressed.

WIP Quadrant: Based upon the calculation of the WIP, each individual is classified as falling within one of four diagnostic Quadrants: A (low distress, high integration, characterized by autonomy and independence); B (low distress, low integration; characterized by detachment); C (high distress, high integration; characterized by dependency); and D (high distress, low integration; characterized by conflict and ambivalence). Each quadrant further relates to a specific attachment style: A (Secure); B (Dismissing); C (Preoccupied); D (Unresolved/ Disorganized).

Form Quality (FQ+%): FQ+% provides a global indication of an individual's cognitive efficiency, rational control, reality testing, and planning and organizational abilities.

Popular Response (P%): P% reflects a client's ability to share a common worldview and conventional way of thinking with others.

Practical Contents (OBJ%+ARC%): OBJ% and ARC% suggest an individual's ability to think practically and functionally.

Index of Suicidal Tendencies (IST): The IST is a constellation of symptoms that together arrive at a numerical value indicating risk of suicidal ideation or behavior. The IST can range from 0-23, with a clinical cutoff of 8. In the normative population, the mean IST score equals 1.98.

Results of chi-square analysis examining the differences between groups in terms of Global Assessment classifications are presented in Tables 2.45.

Table 2.45
Difference in Global Assessment Classifications between Pathological Mourning Parent Groups and Control Group

Index	Group	*N* Present	% Present	*N* Absent	% Absent	$\chi^2_{(df=4)}$	*p*	*V*[1]
Global Assessment: NOP	Pathological Mourning	3	11.5	23	88.5	11.025	.026	.30
	Control	37	39.8	82	60.2			
Global Assessment: L-VA	Pathological Mourning	9	34.6	17	65.4			
	Control	34	36.6	85	63.4			
Global Assessment: M-VA	Pathological Mourning	8	30.8	18	69.2			
	Control	12	12.9	107	87.1			
Global Assessment: S-VA	Pathological Mourning	4	15.4	22	84.6			
	Control	7	7.5	112	92.5			
Global Assessment: PTL	Pathological Mourning	2	7.7	24	92.3			
	Control	3	3.2	116	96.8			

Note. Effect sizes calculated by current authors from previously published analyses; [1]*V* effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Results of further analyses examining the differences in categorical and numerical CWS Indices

between normal and pathological mourning parent groups are presented in Tables 2.46 and 2.47, respectively.

Table 2.46
Difference In Categorical Classifications on CWS Indices between Pathological Mourning Parent and Control Groups

Index	Group	*n* Present	% [Present]	*n* Absent	% [Absent]	$\chi^2_{(df=1)}$	*p*	Φ^1
Positive Global Assessment (NOP+LSI)	Pathological Mourning	12	46.1	14	53.9	8.77	0.003	.27
	Control	71	76.4	22	23.6			
IIT-2 (Positive)	Pathological Mourning	7	26.9	19	73.1	13.17	.000	.33
	Control	62	66.7	31	33.3			
WIP (Positive)	Pathological Mourning	7	26.9	12	73.1	8.45	.004	.26
	Control	55	59.1	38	40.9			
WIP: Quadrant A (Secure)	Pathological Mourning	8	30.7	18	69.3	3.18	.07	.16
	Control	46	49.4	47	50.6			
WIP: Quadrant B (Dismissing)	Pathological Mourning	6	23.1	20	76.9	4.10	.042	.18
	Control	8	8.9	85	91.1			
WIP: Quadrant C (Preoccupied)	Pathological Mourning	5	19.2	21	80.8	0.01	.9120	.009
	Control	17	18.9	76	81.1			
WIP Quadrant D (Unresolved)	Pathological Mourning	7	26.9	19	73.1	0.02	.731	.009
	Control	22	23.7	71	76.3			

Note. Effect sizes calculated by current authors from previously published analyses; $^1\varphi$ effect sizes: small: .1-.29; medium: .3-.49; large: >.5 (Cohen, 1988).

Results are suggestive of significant differences between the pathological mourning group, as compared to the normative population. Specifically, both the AAP and CWS suggested less secure (Dismissing, Unresolved) patterns of attachment. That is, the pathological mourning group demonstrated less Secure attachment classifications (Quadrant A) and a greater percentage of Dismissing classifications (Quadrant B). Moreover, parents who met pathological mourning criteria demonstrated higher levels of emotional distress and impairment as evidenced by the Global Assessment classification (lower NOP and L-VA, higher M-VA, S-VA, PTL), IIT-2 (significantly less positive), and Index of Suicidal Tendencies (significantly more negative, see Table 2.47).

Table 2.47 Differences in CWS Numerical Indices (Means) between Pathological Mourning Parent and Control Groups						
Index	**Normal Group**	**Pathological Mourning**	***t***	***df***	***p***	***d*[1]**
Form Quality (FQ+%)	99.9	93.54	-8.832	117	0.00001	1.63_a
Popular Percentage (P%)	19.71	29.27	4.024	117	0.00001	0.74_b
Practical Content (OBJ%+ARC%)	62.62	48.07	-3.683	117	0.00001	0.07_a
Index of Suicidal Tendencies (IST)	2.04	3.96	4.819	117	0.00001	0.73_b
Note. Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988). [a]Normal group values greater than pathological mourning group. [b]Pathological mourning group values greater than normal group.						

In general, the pathological mourning group demonstrated lower levels of coping resources as evidenced by less cognitive efficiency (lower FQ+%), less conventional thinking and relationships (lower P%), and limited practical thinking and problem solving (OBJ%+ARC%). Overall, while concordance between the CWS and the AAP had been noted in numerous individual clinical case studies, these results statistically validate the convergence of the AAP and the CWS in the assessment of attachment attributes.

Special Topics of Study

Daini, Bernardini, & Panetta (2007)

Daini and colleagues investigated psychological aspects of infertility utilizing the CWS, citing that most previous studies had relied on self-report questionnaires to gain access to the inner world of women experiencing challenges in becoming pregnant. Sixty-one women, divided into three groups based upon cause of infertility, were studied: women experiencing consecutive miscarriages (CM, N=21; age: $\bar{x}$=30.81, SD=1.18); women diagnosed with organic sterility (OS, N=20; age: $\bar{x}$=28.50, SD=1.48); and women with unexplained sterility (US, N=20; age: $\bar{x}$=28.40, SD=1.16). These three groups of women were compared to a control group (CO, N=30; age: $\bar{x}$=30.03, SD=0.78). Clinical groups were selected from an urban hospital, where they were assessed with the CWS pre-treatment. Control group participants were selected at random from women presenting at the same hospital for an outpatient gynecological visit. No significant differences related to age or socioeconomic status between groups were noted.

Statistical analyses investigated differences between groups using MANCOVA procedures, specifically as related to scores on major CWS indices, including Evocative Character (EC+%), Affective Quality (AQ+%), Form Quality (FQ+%), Index of Inner Tension (IIT-1), Anxiety Index (AI), and Impulsivity Index (IM). Additionally, responses to each WDCT Box were analyzed, as well as the Order of Sequence followed by each participant. It was noted that while 76% of the Control group completed Box 1 first in the Order of Sequence, only 50% of the two sterility groups and 24 percent of the Consecutive Miscarriage (CM) group completed Box 1 first (χ^2=13.36, p=.004). Per the authors, this may indicate the negative impact of infertility on a woman's self-esteem, as well as a potential defensive tendency to avoid thinking about the self.

Multivariate analysis demonstrated significant differences (*F:* 3.218, $p<.001$) between groups as related to Form Quality (FQ+%, $p=.0001$) and Evocative Character (EC+%, $p=.006$). No significant differences were noted in terms of Affective Quality, with observed trends indicating lower mean scores in the consecutive miscarriages (CM) group. This lack of significant difference in AQ+% was hypothesized to relate to research linking infertility to alexithymia. The authors note that the differences in FQ+% and EC+% mean scores (lower in the clinical groups as compared to control groups) confirm the emotional impact of infertility on the imaginative life, engagement with the environment, and rational ability to plan.

Additional differences between clinical groups were noted. The unexplained sterility (US) group appeared generally commensurate with the control group in most indices, although demonstrated higher levels of anxiety (AI: CM: $\bar{x}=0.01\pm.04$; OS: $\bar{x}=0.04\pm0.1$, US: $\bar{x}=0.14\pm0.2$; CO: $\bar{x}=0.07\pm0.1$; $F=3.27$, $p=.05$). This may suggest generally adequate integration and functioning, with underlying tension related to currently unclear etiological issues of sterility. The organic sterility (OS) group demonstrated less adaptation, integration of self, and ability to enter into social relationships (lower EC+% in Boxes 4, 5, 6, and 7, $p<.05$). The consecutive miscarriage (CM) group appeared most impacted by challenges in maintaining a positive self-identify (EC+% Box 1: CM: $\bar{x}=0.78\pm.04$; OS: $\bar{x}=0.95\pm0.1$, US: $\bar{x}=0.87\pm.03$; CO: $\bar{x}=0.94\pm0.1$; $F=2.15$, $p=.05$);. Lastly, women impacted by organic sterility and consecutive miscarriages demonstrated lower affective stability (A/F Ratio: CM: $\bar{x}=0.81\pm1.0$; OS: $\bar{x}=0.75\pm1$, US: $\bar{x}=0.02\pm1.2$; CO: $\bar{x}=0.19\pm1$; $F=2.96$, $p=.05$). This likely indicates greater affective reactivity coupled with less apparent mental control. The utility of the CWS to access internal functioning that might not otherwise be apparent via self-report was noted.

Daini, Petrongolo, Manzo, & Bernardini (2012)

In this study, the authors investigated similarities and differences between student and professional nurses, hypothesizing that the length of time spent in professional practice increases empathy in professional nurses; that is, professional nurses are more empathic, whereas student nurses are more practical. Moreover, it was hypothesized that professional nurses may demonstrate higher rates of stress and or anxiety as related to student nurses. Given this higher stress level, professional nurses were predicted demonstrate higher rates of depression as evidenced by previous research on the topic. The CWS was selected for use in this study, given its ease of administration, and the likelihood it would "offer a wide range of personality features" (p. 19).

Two groups of individuals participated in this study: Professional nurses ($N=111$; 49 male, 62 female; age: $\bar{x}=33$, $SD=5.19$) working in an Italian hospital and Student nurses ($N=210$; 102 male, 108 female; age: $\bar{x}=24.46$, $SD=4.37$) selected from two graduate programs in nursing. The majority of participants in both groups were of Italian descent, with less than 3% of each group originating outside the country.

In comparing groups, the authors investigating differences between groups on major CWS indices (Evocative Character [EC+%], Affective Quality [AQ+%], Form Quality [FQ+%], Index of Inner Tension-1 [IIT-1], Impulsivity Index [IM], Anxiety Index [AI], and Popular percentages [P%]) and Content categories (Architecture [ARC%], Nature [NAT%], Object [OBJ%], and Symbol [SIG%]). Professional nurses' scores exceeded Student nurses' scores on five statistically significant indices. Evocative Character (EC+%; Pro: $\bar{x}=74.10\pm12$; Stu: $\bar{x}=66.79\pm15$, $F=5.78$, $p=.01$), Popular Responses (P%; Pro: $\bar{x}=21.33\pm12.49$; Stu: $\bar{x}=10.39\pm10.61$, $F=25.52$, $p=.000$), Architecture content (ARC%; Pro: $\bar{x}=8.05\pm10.23$; Stu: $\bar{x}=5.74\pm7.66$, $F=7.02$, $p=.008$), Nature content (NAT%; Pro: $\bar{x}=4.09\pm7.68$; Stu: $\bar{x}=2.72\pm5.98$, $F=3.98$, $p=.048$), and Symbol content (SIB%; Pro: $\bar{x}=12.31\pm12.91$; Stu: $\bar{x}=9.66\pm11.50$, $F=3.67$, $p=.05$). Conversely, Student nurses' scores exceeded Professional nurses' scores on Object content (OBJ%; Stu: $\bar{x}=49.21\pm21.01$, Pro: $\bar{x}=39.63\pm20.05$;

F=12.74, p=.000).

Further analysis demonstrated that Boxes 1, 2, and 4 accounted for the most differentiation between groups in terms of Evocative Character scores, with professional nurses scoring higher than student nurses. Professional nurses were further found to produce higher Affective Quality scores in boxes 1, 6, and 7 than their student counterparts. Lastly, in investigating Order of Sequence, it was noted that professional nurses tended to select Box 1 as the first box drawn as compared to student nurses (χ^2=16.75, p=.02).

The authors concluded that the test results suggest that professional nurses tend to possess more defenses than student nurses. While both groups' EC+% fell within the normal range, suggesting connection and responsiveness to the environment and intact associative processes, significantly higher scores in the professional group demonstrate increased sensitivity and receptivity. This heightened sensitivity appears particular to the professional nurses sense of self (Box 1), vitality (Box 2), and relationship to authority or external expectations (Box 4). To counteract this increased sensitivity (also evidenced by higher NAT%, higher ARC%), an important component of their caregiving capacity, professional nurses demonstrated tendencies toward rational detachment and self-protection (increased P%, increased SIG%, higher AQ in Box 6). Lastly, higher percentages of Object (OBJ%) content (associated with lower Human content percentage) observed in student nurses may indicate a practical, task-based approach to their work and those they encounter, possibly due to their limited professional experience.

Summarizing, the authors reported that the data of the current study converges with that of previous studies among health care workers, primarily that emotional detachment serves as a protective factor. In terms of the second hypothesis, no significant depressive tendencies were noted in the professional nursing group.

CWS in Selection Settings

Ravecca (2003)

Ravecca and colleagues in the Italian Navy investigated the convergent validity of the CWS in an applied setting, specifically focused on the selection of appropriate candidates for the Italian armed forces. Drawing from a sample of military applicants, 331 male subjects (aged 19-29) were selected for evaluation. All candidates were Caucasian, and had successfully completed a high school education. Convergent validity related to both clinical/ personality factors relevant to military selection and aptitude for military service was investigated.

CWS Agreement with Other Test Measures

To better understand the utility of the CWS in military selection, comparisons at the global level were studied to determine agreement between measures regarding clinical/ personality factors of candidates. Overall CWS Global Assessment classifications (derived from the clinical scoring software) were compared to global ratings of both the MMPI-2 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 1989) and the Guildford-Zimmerman Temperament Survey (GZTS; Guildford & Zimmerman, 1949). To facilitate such a comparison, independent psychiatrists rated each subject's performance on the MMPI-2 and GZTS as falling in one of five categories. Categories were derived using pre-established criteria related to MMPI-2 and GZTS test scores indicative of low, moderate, or severe symptoms. This five-point categorization schema corresponded to that of the computer-generated five-category CWS Global Assessment, derived from scored CWS indices and calculations, allowing easy comparison between test results and psychiatric ratings.

For the purposes of this study, the five ordinal categories included:

> 1) NOP (Non-Pathological), characterized by appropriate individual and social adaptive processes. This category corresponds to a judgment of general adequacy for military service.
>
> 2) L-VA (Low Symptoms/Impairments), characterized by psychoaffective difficulties that are ameliorated by the presence of appropriate defenses and result in little negative impact on individual global functioning. Without other negative results, this category corresponds to a judgment of general adequacy for military service, similar to NOP.
>
> 3) M-VA (Medium Symptoms/Impairments), indicating problematic and conflicting aspects of personality that likely impair the psychoaffective balance of the individual. This category suggests the need for further psychodiagnostic evaluation prior to making a selection decision
>
> 4) S-VA (Severe Symptoms/Impairments), indicating severely problematic and conflicting aspects of personality that likely negatively influences the individual's behavior. This category corresponds to a judgment of inadequacy for military service.
>
> 5) PTL (Pathological), characterized by the presence of overt psychopathology and significant impairment in functioning, especially if other indices suggest negative outcomes. This category corresponds to a judgment of inadequacy for military service.

In comparing the classifications between tests, concordance was determined between the CWS and the MMPI-2 in 76.1% of cases. Similarly, agreement was noted between the CWS and the GZTS in 74.6% of cases. Given the overall goal of selection, the five categories were subsequently condensed into three categories (Select, Further Investigation Needed, and Don't Select), with significant improvement in concordance between measures noted. In considering three classification categories, agreement between the CWS and MMPI-2 rose to 98.5%, whereas agreement between the CWS and the GZTS rose to 98.2%. The author concluded that results of this study support the convergent validity of the CWS with both the MMPI-2 and Guilford-Zimmerman test.

CWS Agreement with Expert Ratings

In terms of aptitude for military service, additional analysis was undertaken, comparing candidates' rating in five qualitative domains derived from the CWS clinical software: 1) Overall maturity level; 2) Emotional regulation; 3) Cognitive abilities; 4) Interpersonal skills; and 5) Group affiliation. Independent psychiatrists concurrently rated candidates on these categories, blind to the evaluations derived from the CWS. Each candidate was rated on each domain qualitatively as positive (+), neutral (+/-), or negative (-). Given that 331 candidates were included in the research, and each was rated on five domains, 1,655 points of comparison were analyzed. Of these, 84.2% demonstrated agreement between the CWS and independent psychiatric ratings.

In evaluating results, positive and negative predictive values were calculated to determine the extent of agreement between the CWS Global Assessment and expert psychiatric ratings. Results strongly

validated the utility of the test in screening and selection based upon the Global Assessment ratings. Positive and neutral ratings were combined, resulting in a dichotomous decision rating (positive/neutral vs. negative). Utilizing this approach, both Positive and Negative Predictive Values were excellent for the CWS (PPV=0.969; NPV=1.0; sensitivity=1.00; specificity=0.33).

Overall, the study suggested increased flexibility and efficiency of selection via inclusion of the CWS with self-report measures. Further, results were noted to suggest that the utility of the CWS is high when each derived component is analyzed separately, but demonstrates higher incremental validity when clinical/personality, aptitude, and descriptive components are fully integrated to inform the global process of selection.

CWS Content Categories

Crisi & Carlesimo (2008)

Crisi and Carlesimo (2008) examined differences in personality profiles between subjects whose WDCT protocols were positive for Human (H), Food (FD) and Cloud (CLD) contents as compared to subjects that did not depict these content categories. Similar to scoring systems developed for the Rorschach Inkblot Method, the presence or absence of specific content categories on the CWS are hypothesized to be indicative of important elements of personality structure and interpersonal relatedness. For each content studied, specific experimental hypotheses were developed, described below. In examining the experimental hypotheses, 2293 male subjects (aged 18-25; a subsection of the Italian CWS normative sample) were included in analyses.

In conducting analyses, specific indices and classifications contained within the CWS were utilized. Care was taken to study scores and indices demonstrating limited overlap with investigated content areas, although it must be considered that some degree of relationship is inherent in all test variables. Specifically studied variables included, but were not limited to:

> Global Assessment: As previous described, each CWS protocol is assigned an overall Global Assessment classification, ranging from "No Pathology" (NOP) to "Pathological" (PTL). This classification system is divided into five categories, two of which are considered positive (NOP + "Low Symptoms" [LSI]), one of which is considered to represent moderate difficulties ("Moderate Symptoms" [MSI]), and two of which are considered negative (PTL + "Severe Symptoms" [SSI]). During evaluation, the presence of positive and negative classifications was considered for each content area.
>
> WIP Quadrant: As described previously (see Crisi, 2009), the CWS co-locates individuals on a graphic depiction of personality functioning, based upon the degree of currently experienced distress (vertical axis) and level of personality flexibility and integration (horizontal axis). Four quadrants result, labeled A, B, C, and D, respectively. Quadrants A and C (reflecting high flexibility and integration) are generally considered less pathological and therefore more "healthy."
>
> Box Codes: As noted above, in each Box on the CWS, a Code is assigned based upon the mathematical summation of Evocative Character and Affective Quality scores. Two Codes (C and PC) are considered positive, whereas four Codes (NC, AC, AD, and D) are considered negative.
>
> Affective Stability Index (A/F Ratio): The A/F Ratio depicts affective levels ("A",

left side of ratio) as compared to cognitive resources and control ("F", right side of the ratio). Based on normative data, it is expected that F will be greater than A by 1-2 points, indicating sufficient cognitive control to modulate affective reactions to the environment. Higher levels of affect ("A") are typically related to higher expressed emotion, immaturity, overwhelming emotions, and reactivity to the environment.

Index of Suicidal Tendencies (IST): The IST is a constellation of CWS scores and calculations that are indicative of internal tension, depressive affect, low frustration tolerance, self-punitive thinking style, and poor affect regulation. Normative values range from 0-3, with higher values suggesting increased distress and pathology.

Index of Inner Tension-2 (IIT-2): The IIT-2 numerically measures the degree of flexibility and integration inherent in an individual's personality structure. To calculate the IIT-2, each box in a protocol is assigned a Code based upon a numerical formula combining Affective Quality and Evocative Character scores. Six Codes are possible per box, with 2 considered positive (C, PC) and four considered negative (NC, AC, AD, D). The IIT-2 creates a ratio of positive Codes on the left to negative Codes on the right. The higher the value on the left, the more general levels of health and integration are anticipated. Normatively, the left side of the ratio exceeds the right.

In general, as summarized below, overall results confirmed experimental hypotheses, supporting the usefulness and the construct validity of CWS Content categories.

Human Content

In evaluating the hypotheses related to the presence of Human content, 1518 males with Human content present in their protocols were compared to 775 males without Human content. Overall, the authors hypothesized that the presence of Human (H%) content is an important factor in the CWS and is able to effectively discriminate between subjects. Specific hypotheses included:

1) The presence of Human (H%) content will be associated with greater social skills as assessed by analyzing other CWS indices, including higher Popular (P%), lower Object content (OBJ%) and higher frequency of positive Code of Box 8 (the box of social skills and relatedness).
2) Individuals with Human (H%) content are more likely to demonstrate healthy functioning overall, as evident by a higher frequency of positive Global Assessment classification.
3) Individuals with Human (H%) content are more likely to demonstrate healthier attachment style, as evident by higher frequency of classification in WIP Quadrants A and C.

Results of analyses, related to the three Human (H%) content experimental hypotheses, are presented in Table 2.48.

Table 2.48
Differences between Groups on Select CWS Indices:
Clients with Human ("H%") Content vs. Those without Human Content ("No H%")

Scoring Phenomena	*"No H" %*	*"H" %*	χ^2	*p*	*Φ*[1]
Global Assessment (Positive Classification: NOP +LSI)	55%	82%	189.88	.00001	.29
WIP Quadrant (Positive Placement: A + C)	52%	61%	17.06	.0001	.08
Box 8 Valuation (Positive: C + PC)	50%	63%	35.26	.00001	.14

Index	*"No H" $\bar{x}$*	*"No H" SD*	*"H" $\bar{x}$*	*"H" SD*	*t* *(2,291 df)*	*p*	*d*[2]
Popular Responses (P%)	13.29	9.09	19.60	11.34	-13.417	.000	0.56[a]
Object Content (OBJ%)	47.56	20.00	38.45	17.88	11.079	.000	0.46[b]

Note. Effect sizes calculated by current authors from previously published analyses; [1]*φ* effect sizes: small: .1-.29; medium: .3-.49; large: >.5; [2]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]"No H" group values greater than "H" group.
[b]"H" group values greater than "No H" group.

In each case, the experimental hypothesis was confirmed as predicted by the authors, suggesting that the absence of Human content in a CWS protocol is likely indicative of affectively driven interpersonal difficulties, and may contribute to overall reduced functioning in individuals.

Additionally, for each protocol, 103 calculations and indices inherent to the CWS interpretation strategy were calculated for exploratory analytical purposes, with independent samples *t*-test used to analyze differences between "H" and "No H" groups. Results, presented in Appendix C, suggested that the presence of Human content differentiates to a statistically significant level between the "H" group and the "No H" group in 59 out of 103 evaluated indices. In evaluating results, it is important to note at all indices demonstrating significantly significant differences were related to affective and interpersonal functioning. That is, no indices measuring cognitive difficulties were found to be significantly different between groups. The authors noted that this constellation of differences suggests that the presence or absence of Human content differentiates well between clients experiencing interpersonal and affective difficulties as compared to those who are not.

Food Content

Experimental investigation of the ability of Food (FD) content alone to discriminate between individuals was investigated with a subset of the overall sample. 195 male participants who depicted Food content in the CWS protocols were compared to 209 matched participants without Food content. Overall, the authors hypothesized that the presence of Food (FD%) content is an important factor in the CWS and is able to effectively discriminate between subjects in terms of overall health and attachment style; however, when the positive Global Assessment classification of participants with Food content (79%) was compared to those without Food content (79%), no statistical significance was noted. Similarly, when positive WIP Quadrant placement of those with Food content (64%) was compared to those without Food content (62%), no statistical significance was noted. These results suggest that Food content alone does not discriminate between subjects. Given these initial results, further hypotheses were elaborated:

1) The presence of Food (FD%) content will be associated with passivity, dependence and immaturity as evidenced by increased Anxiety Index (AI), higher frequency of Architectural content (ARC%), and higher frequency of Human contents (H%).
2) The presence of Food (FD%) content will be associated with reduced autonomy and challenges in managing relationships with authority, as indicated by the increased frequency of negative Codes in Box 4 (the box of autonomy and relation to authority).

Results of the analyses related to the Food content experimental hypotheses are presented in Table 2.49.

Table 2.49
Differences between Groups on Select CWS Indices:
Clients with Food ("FD") Content vs. Those Without Food Content ("No FD")

Index	*"No FD" $\bar{x}$*	*"No FD" SD*	*"FD" $\bar{x}$*	*"FD" SD*	*t (402 df)*	*p*	*d*[1]
Anxiety Index (AI)	0.91	0.16	0.95	0.14	2.666	.01	0.26[b]
Architectural Content (ARC%)	19.69	11.86	16.55	10.53	4.050	.000	0.50[a]
Human Content (H%)	12.35	11.87	9.89	9.71	3.197	.001	0.32[a]

Scoring Phenomena	*"No FD" %*	*"FD" %*	*χ^2*	*p*	*Φ*[2]
Box 4 Valuation (Positive: C + PC)	33%	24%	5.22	.01	.22

Note. Effect sizes calculated by current authors from previously published analyses; [1]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8; [2]φ effect sizes: small: 0.1-0.29; medium: 0.3-0.49; large: >0.5 (Cohen, 1988).\
[a]"No FD" group values greater than "FD" group.
[b]"FD" group values greater than "No FD" group.

Given these findings, individuals who produce Food content responses on the CWS appear to exhibit greater tendencies towards dependence, insecurity, and anxiety than those who produce no Food responses; however these findings are less robust and less evident at the overall protocol level (given the lack of significance in differences in Global Assessment and WIP classification between groups) than other studied Content categories.

Additionally, for each protocol, 103 calculations and indices inherent to the CWS interpretation strategy were calculated for exploratory analytical purposes, with independent samples *t*-test used to analyze differences between "FD" and "No FD" groups. Results, presented in Appendix C, suggest that Food content is less able to discriminate between participants in isolation with only 33 of 103 analyzed indices demonstrating statistically significant differences.

Cloud Content
Experimental investigation of the hypotheses related to the presence of Cloud content was analyzed utilizing a subset of the overall Italian normative sample. 117 male participants' protocols containing Cloud content were compared to 2176 matched participants without Cloud content. Overall, the authors hypothesized that the presence of Cloud (CLD%) content is an important factor in the CWS and is able to effectively discriminate between subjects in terms of overall health, anxiety states, internal tension, and distress. Specific hypotheses included:

1) The presence of Cloud (CLD%) content will be associated with less positive overall health and functioning, as evidenced by fewer positive Global Assessment classifications, fewer classifications in WIP Quadrants A and C, and fewer overall positive box Codes as compared to the control group.
2) The presence of Cloud (CLD%) content will be associated with indicators of affective distress and tension, including a more negative IIT-2, higher Index of Suicidal Tendencies (IST), and lower Human (H%) content.
3) The presence of Cloud (CLD%) content will be associated with higher anxiety and lower ability to enter into mutual relationships, as evidenced by higher frequency of negative Codes in Box 2 (related to sensitivity, relatedness, and vulnerability). As the experimental sample was comprised only of males, it was noted by the authors that Box 2 (generally thought of as specifically representing relationship with women) would likely be most sensitive to discerning anxiety and distress.

Results of analyses related to the Cloud content experimental hypotheses are presented in Table 2.50.

Table 2.50
Differences between Groups on Select CWS Indices:
Clients with Cloud ("CLD") Content vs. Those Without Cloud Content ("No CLD")

Scoring Phenomena	***"No CLD" %***	***"CLD" %***	χ^2	***p***	***Φ[1]***
Global Assessment (Positive Classification: NOP +LSI)	82%	67%	16.0047	.0001	.08
WIP Quadrant (Positive Placement: A + C)	47%	39%	2.8989	.08	.03
Overall Positive Box Codes (C+PC)	61%	54%	18.7084	.00001	.09
Box 2 Valuation (Positive: C + PC)	76%	35%	96.666	.00001	.20

Index	***"No CLD" $\bar{x}$***	***"No CLD" SD***	***"CLD" $\bar{x}$***	***"CLD" SD***	***t*** *(2,291 df)*	***p***	***d[2]***
Affective Stability Index: Affective Quality	2.67	0.72	2.06	0.72	8.809	.000	1.14[a]
Human Content (H%)	12.27	11.67	8.41	11.14	3.487	.000	0.50[a]
Popular Responses (P%)	17.63	11.09	14.44	9.74	3.047	.002	0.43[a]
Index of Suicidal Tendencies (IST)	1.8	.12	4.8	.42	68.97	.0001	9.80[b]
Index of Inner Tension-2 *(left side: positive codes)*	4.91	1.42	4.34	1.41	4.220	.000	0.60[a]
Index of Inner Tension -2 *(right side: negative codes)*	3.08	1.42	3.65	1.41	-4.220	.000	0.60[b]

Note. Effect sizes calculated by current authors from previously published analyses; [1]***φ*** effect sizes: small: .1-.29; medium: .3-.49; large: >.5; [2]*d* effect sizes: small: 0.2-0.49; medium: 0.5-0.79; large: >0.8 (Cohen, 1988).
[a]"No CLD" group values greater than "CLD" group.
[b]"CLD" group values greater than "No CLD" group.

In general, experimental hypotheses were confirmed, with the presence of Cloud content discriminating between normal and pathological protocols. Findings suggest that individuals who

produce Cloud content are more likely to present with traits of pathological personality organization, accompanied by high level of anxiety, difficulties in navigating affective relationships, and potential deficits in their ability to control feelings and emotions.

Additionally, for each protocol, 103 calculations and indices inherent to the CWS interpretation strategy were calculated for exploratory analytical purposes, with independent samples *t*-test used to analyze differences between "CLD" and "No CLD" groups. Results, presented in Appendix C, suggest that Cloud content is less able than Human content to discriminate between participants at the individual index level, with only 34 of 103 analyzed indices demonstrating statistically significant differences. However, Cloud content was noted to differentiate between pathological and non-pathological at the overall protocol level (i.e., WIP, Global Assessment, overall box codes), indicating its usefulness in identifying pathology and symptomatology.

Overall, Human Content and Cloud Content appear to discriminate between subjects, whereas Food Content does not in isolation. Absence of Human content is related to greater difficulties in social relationships; presence of Food content suggests passivity, insecurity, and dependence; and presence of Cloud content indicates higher levels of pathology including anxiety and difficulties regulating emotions. Future research on the relationship of these Content areas to external measures (rather than other CWS indices) is recommended.

Clinical Utility of the CWS: Published and Presented Case Studies

In addition to the research presented above, several case studies utilizing the CWS have been presented and/or published, suggesting the clinical utility of the CWS. These case studies have investigated convergences between the CWS and the Rorschach Comprehensive System (Exner, 1997, 2003) the CWS and the Adult Attachment Projective (AAP), the CWS and the MMPI-2, and the CWS and the Social Adaptation of Self Evaluation Scale (SASS; Bosc, Dubini, & Polin, 1997) and "Scale for the Rapid Dimensional Evaluation (SSIRAD; Pancheri, Biondi, Gaetano, & Picardi, 1999). Overall, these case studies have suggested incremental validity of the CWS in conjunction with other measures, highlighting the utility of the measure to assess clients with affective disturbance, suicidal ideation, pre-psychotic thinking, and varied levels of psychopathology.

Crisi & Shorey (2009)

The convergence between the CWS and the Rorschach Comprehensive System (CS, Exner, 1997, 2003) results was presented by Crisi and Shorey. In this case, the primary author assessed a 19-year-old male who had recently attempted suicide, administering both the Wartegg according to CWS guidelines and the Rorschach according to Comprehensive System guidelines. The primary author independently scored the CWS whereas the second author independently scored the Rorschach. Only after written interpretations were prepared were the interpretive findings integrated.

Given the markedly different methods of information-gathering inherent to the two tests (i.e., the Rorschach relies primarily on verbalization, whereas the WDCT relies primarily on drawing), the authors noted a potential lack of clarity in the extent of "convergence and divergence" between the two measures. Results of the independent interpretations are summarized in Table 2.51.

Table 2.51
Comparisons between Rorschach Comprehensive System and CWS Interpretation in a 19-year-old Male Client

Comparative Domain	*Rorschach CS Results*	*CWS Interpretive Results*
Ideation	This client has a pessimistic view of the world that appears to be intruding into his thought processes and he prefers to deal with his negative affect through intellectualization. He also has difficulty thinking logically or coherently, and as a result, may arrive at faulty conclusions, which he then clings to in an inflexible manner.	This client is inclined to face situations with a strong tendency to rationalization and to rigidity. Such tendency should decline in highly emotional situations and his performances might suffer. At present it is important to highlight the presence of pessimistic polarizations of thought.
Cognitive Mediation	This client is capable of recognizing conventional modes of responding but he has difficulty in reality testing, may misperceive events, and may form mistaken impressions about the motives of other people.	His thought is appropriate and adequate to the common way of thinking with emphasis of conformism because of strong traits of dependency. It is possible that the client actively avoids his anxiety with characteristics of repetition and compulsion.
Information Processing	This client takes in less information than is needed to avoid arriving hasty ill conceived conclusions. As a result he may be careless in his approach to problems, and not be flexible enough to make needed corrections. By extension, he may perform well below his ability.	
Capacity for Control and Stress Tolerance	This client lacks a consistent and well-defined coping style and currently has fewer psychological resources than are needed to cope with the demands of everyday life in other than a very restricted repertoire of behaviors and situations.	The Index of Suicidal Tendencies is positive (>8): This client has a very low and weak tolerance for stress. There is a depressive condition that, together with a marked tendency to intropunitive reactions, make real the risk that this client should have strong self-aggressive or self-punitive behaviors.
Affect	This client is prone to depression and mood disturbances. He then copes with his emotional distress through intellectualization. This may further interfere with the clarity of his thinking in affect-laden situations.	The client has deep traits of immaturity arising from a structure of personality characterized by conflict and neuroticism with high and intense states of anxiety. The behavior of the client is mainly characterized by reactivity.

Self-Perception	This client appears to have a defensive tendency to over-value himself to the detriment of attending to the needs of others. Having some awareness that his inflated self-esteem is not reality based, and having underlying feelings of inadequacy, he avoids close self-examination.	Under an apparent tendency to overestimate and overvalue the self, the subject is very insecure and indecisive. This client likely doesn't share the interests of his peers.
Interpersonal Perception	This client shows interest in other people and generally expects positive interactions with others, but may have limited capacity to form close and lasting attachments.	The client shows a deep social dependency. He does not share sincere and genuine interests of his peer group. His relationships with women are the origin of strong anxiety and fears.

Comparisons of the Rorschach and Wartegg data, and resulting independent interpretations, revealed a high level of concordance in relation to affective experience, negative intrusive ideation, coping strategies, and ability to recognize conventional reality. The authors noted that this convergence between the tests is impressive given the different modes of response inherent to the two tests (spoken language vs. drawing).

Reports from the two measures diverged, however, in relation to information about thought processes and ideation. The Rorschach protocol identified problems with thinking and reality testing that were not identified with the Wartegg. The Wartegg, in contrast, was more sensitive in detecting a risk for suicide with this client (Rorschach S-CON = 5; Wartegg Index of Suicidal Tendencies > 8). The authors concluded that the data from this single case study is compelling enough to warrant further larger scale and better-controlled research.

Toivakka & Crisi (2014)

Similarities between the CWS and the Rorschach Comprehensive System (CS; Exner, 1997, 2003) were further presented by Toivakka and Crisi. In this case, the CWS was evaluated with minimal information about the client (14 year old female), with results of the CWS compared to that of the Rorschach. Independently generated interpretive results from each test were evaluated in terms of five comparisons: 1) Diagnosis; 2) Affect; 3) Interpersonal Relationships; 4) Conventionality of Thinking; and 5) Constellations of Scores. In each comparison, a strong degree of clinical concordance was discovered. Conclusions are summarized in Table 2.52, with Rorschach Comprehensive System data presented in the left-hand column, and CWS interpretation presented on the right.

Table 2.52
Comparisons between Rorschach Comprehensive System and CWS Interpretation in a 14-year-old Female Client

Comparative Domain	*Rorschach CS Results*	*CWS Interpretive Results*
Diagnosis	"In the RCS the main clinical impression is of an acute depression with posttraumatic features and developmental risk for a borderline personality disorder."	"In the CWS, clear signs of a reactive depression appeared and many elements indicate that the depressive condition is likely connected with a traumatic event (including sexual or physical violence)."
Affect	FC:CF+C= 8:2 Pure C= 0 SumC':WSumC= 2:6.0 Afr = 0.33 S = 2 Blends:R = 11:28 CP = 0	This client presents as "in control" emotionally; however, underneath this presentation, she is likely very immature (AQ+%=75 and Ratio A/F=4/4) and shows a high level of impulsivity (Impulsivity Index=0.75). Traits of egocentrism, tendency to be "explosive," and gross immaturity result in a tendency towards attempted overcontrol. Such a condition could be the origin of psychosomatic troubles. We should say that this girl "fights against her own dependency from others" but she's not aware of her dependency.
Interpersonal Relationships	COP=1 AG=0 GHR:PHR= 4:9 a:p= 11:5 Food=0 SumT=0 Human Content=10 Pure H=8 PER=1 Isolation Index= 0.36	Regarding social skills, we can say that we have found a strong need to lean on others, including marked dependency traits. Moreover, she seeks confirmation and support from others (resulting in reduced autonomy). This girl is very sensitive (she shows an elevated value in the Evocative Character, EC+% =94). Such a high score is generally found in severely disturbed subjects, who show excessive sensitivity and receptivity to external stimuli. Therefore, these individuals may lack filters, or, in other words, lack intact defense mechanisms. So she is likely to have her feelings easily hurt by the words or actions of others. Traits of hypervigilance are possible.
Conventional Thinking	XA%= 0.75 WDA%= 0.67 X-%= 0.25 S-=1 P=4 X+%= 0.36 Xu%= 0.39	Generally, high scores in Popular and Human Responses are found in highly conformistic individuals who demonstrate an excessive adherence to conventional thinking. In the present case, the high score obtained (P%=63% and H%=50%) appears more correlated with depression and strong dependency on others and a strong need for approval. Another element is consistent with this observation: the high percentage of Architectural contents (=38%).
Constellations of Scores*	PTI = 3 DEPI = 5	PTI = NO DEPI = YES

	CDI = 3 S-CON = N/A HVI= YES OBS = NO	CDI = YES S-CON = N/A HVI= YES OBS= NO
Note. *While the CWS does not incorporate the Comprehensive System constellations, the primary author extrapolated from CWS indices, estimating whether the client would meet criteria for each cluster based upon performance on the test.		

The authors concluded that the Rorschach and CWS demonstrate a high level of convergence in relation to overall clinical description of clients across multiple domains. Given the differences in information gathering inherent in the tests (i.e., verbalization versus drawing), this concordance reinforces the validity of both measures in the complex assessment of personality.

Crisi (2011a)

As part of a larger presentation, a pilot study evaluating the concordance between the CWS and the Adult Attachment Projective (George & West, 2012) was presented at the Society for Personality Assessment annual convention (2011). Two main questions were proposed by the researchers: 1) Is there concordance between the AAP and the CWS in terms of attachment classification; and 2) Is there incremental validity inherent in using both the AAP and CWS in the same assessment battery?

In considering the AAP, attachment classification is generally divided into four categories: Secure, Dismissing, Preoccupied, and Unresolved. This classification is arrived at via standardized scoring at the linguistic level. High convergence between the AAP and the Adult Attachment Interview (AAI) has been reported by George and West (2011). More specifically, 97% convergence between AAP-AAI for secure versus insecure classifications (*kappa* =.80, *p*=.000) and 92% convergence for the four major attachment groups (*kappa*=.89, *p*=.000) has been reported.

In regards to the CWS, attachment style is evident via examination of the Wartegg Index of Psychopathology (WIP). As described elsewhere, the WIP is a graphic plotting of a client's experienced distress and internal tension (vertical axis) as related to the degree of personality integration, flexibility, and adaptability (horizontal axis). Upon plotting this data, client's are classified into one of four quadrants, each with accompanying attachment style: A) Adaptation (Secure), B) Detachment (Dismissing), C) Dependency (Preoccupied), and D) Ambivalence (Unresolved).

To examine the concordance between AAP classification and CWS quadrant, a case study was evaluated. In addition, examination of several additional factors was considered to determine incremental validity, identifying coding dimensions of the AAP and indices of the CWS that measure similar constructs. First, the presence or absence of verbalized personal experience was noted. Second, content was analyzed for the presence of Agency of Self, Connectedness, and Synchrony, concepts specific to the evaluation of attachment. Lastly, attachment defenses (Deactivation, Cognitive Disconnection, and Segregated Systems) were analyzed. Table 2.53 presents results of this analysis.

Table 2.53 Comparison of AAP Coding Dimension and CWS Clinical Indices			
AAP Coding	**Presence/Absence**	**Corresponding CWS Scoring**	**Presence/Absence**
Discourse			
Personal Experience	NO	Personalized Answer (PA)	YES *(Box 4)*
Content			
Agency of Self	ABSENT	Valuation Box 3 & 5	ABSENT *(poor value, Box 3 & 5)*
Connectedness	LOW	H%, Positive AQ+%	LOW *(low H%, neg. AQ+%)*
Synchrony	LOW	Valuation of Box 8	LOW *(Box 8 ambivalent)*
Defenses			
Deactivation	YES	Quadrant B	YES *(Quadrant B)*
Cognitive Disconnection	YES	Quadrant C, high FQ+%, Incomplete Drawing (ID)	NO *(No Quadrant C, average FQ+%, no ID)*
Segregated Systems	YES	Valuation Box 2 & 4	YES *(Ambivalence in 2 & 4)*

Given these preliminary results, concordance between AAP attachment classification (Dismissing) and CWS quadrant (Quadrant B, Detachment) was discovered. Moreover, incremental validity of both measures is suggested, given unique information provided by both. That is, while the AAP better captured the client's thinking related to behaviors and attitudes, as well as tendencies toward cognitive disconnection, the CWS provided clinically relevant affective information related to the client's emotional experiences with parental figures (via analysis of Boxes 2 and 4).

Crisi (2010b)

Convergence between the CWS and the MMPI-2 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 1989) was presented in a case study focusing on a 38-year-old female client. During the course of an evaluation, the client was administered both the CWS and the MMPI-2, with resulting MMPI-2 code-types compared to CWS scores and indices.

Based upon MMPI-2 results (considered valid), the client produced a "2-7-8" code-type, oftentimes related to depression with psychotic features or thought disturbance. The 2-7-8 code-type has been described by Nichols (2001) as:

> Moderate to severe depression with suicidal ideation, anxiety, fearfulness and phobias, anhedonia, obsessional worry and rumination, compulsions, self-depreciation, pessimism… Feelings of worthlessness, helplessness, hopelessness, inadequacy… Markedly intropunitive; may be compulsively self-critical and self-accusatory… Look for manifest psychotic ideation, thought disorder, hallucinations and delusional ideations; a history of social isolation; and identity and sexual concerns. Maybe severe suicide risk. (pp. 258-259)

Similar to the description of the 2-7-8 code-type, based upon CWS scoring, the client was noted to meet criteria for 9 out of 13 indicators of psychotic depression, presented in Table 2.54.

Table 2.54 CWS Criteria Met for Psychotic Depression		
Criterion	***Client Score***	***Classification***
EC+% ≤ 50	50	MET
AQ+% ≤ 44	44	MET
Box 3 valuation: NC, AD, or D	NC	MET
IIT-1 negative	positive	NOT MET
IIT-2 negative	1/7 (neg)	MET
Quadrant B, area B	B, beta	MET
# of Primary Contents <4	1	MET
H% low	0%	MET
P% high	0%	NOT MET
OBJ% ≥ 63	100	MET
Presence of GR or PR	Not present	NOT MET
IST ≥ 8	9	MET
Opposite Direction (OD)	Not present	NOT MET

In this case, significant concordance was noted between MMPI-2 and CWS interpretation, suggesting the convergence of the measures. Given the strong agreement evident, further studies are recommended.

Rizzo, Della Villa, & Crisi (2015)

Published in *Computers in Human Behavior*, this case study investigated the utility of the CWS with a 17-year-old boy suffering from Problematic Internet Use (PIU) as described by Young (1996). Convergence of the CWS with both the Social Adaptation Self-Evaluation Scale (SASS; Bosc, Dubini, & Polin, 1997) and the Scale for Rapid Dimensional Assessment (SSIRAD; Pancheri, Biondi, Gaetano, & Picardi, 1999) was investigated. The analysis of the CWS and self-report (SASS) and expert report (SSIRAD) questionnaires included in the study, along with collected life history data, suggested that the client's withdrawal and cyber-dependence likely represented initial symptoms of a pre-psychotic state.

In this study, the WDCT was individually administered, per CWS guidelines, with resulting descriptive, categorical, and attitudinal calculations generated. The Social Adaptation Self-Evaluation Score (SASS) was completed as a self-report instrument by the client, measuring behavior in four broad areas of social functioning: 1) work; 2) spare time; 3) family; and 4) ability to cope with resources/finances. The SASS further assesses motivation, self-perception, and satisfaction. The Scale for the Rapid Dimensional Evaluation (SSIRAD) was compiled by a psychologist, based upon review of client's life history data. The SSIRAD evaluates clients on ten dimensions: 1) apprehension-fear; 2) sadness-demoralization; 3) rage-aggressiveness; 4) obsessionality; 5) apathy; 6) impulsiveness; 7) reality distortion; 8) thought disorganization; 9) somatic preoccupation-somatization; and 10) activation.

Interpreted results of the three instruments are summarized in Table 2.55.

Table 2.55 Comparisons between CWS, SASS, and SSIRAD Results		
SASS	***SSIRAD***	***CWS***
Social Relationships significantly lower than normal. Relationships likely restricted to family members.	Inability to be reassured by others. Decreased interest and initiative.	Marked difficulties in social interaction; Inability to form close relationships with peers. Behavior based on isolation and detachment. Potentially disorganized attachment classification.
Self Esteem significantly lower than normal; Feelings of inadequacy are likely.	Distrust of himself and his abilities. Feeling of imminent danger.	Feelings of insecurity, inferiority, inadequacy likely to present in interactions with authority.
Evidence of Frustration	Sense of constraint.	Intrapunitive reactions toward frustration; Inability to adequate (adaptively) express frustration.
	Severe psychic anxiety and tension.	Internal tension and distress.
	Reduction of creativity and energy; pessimism; decreased pleasure. Affective constriction	Basic depressive state, resulting in discomfort, and limiting performance and behavior
	Stereotypical, rigid, and de-realized thinking. Idiosyncratic and unusual beliefs, thoughts, decisions	Formal, rigid, closed ideational process. Synthesis and judgment are decidedly inadequate.

The authors noted the similarities between the client-report (SASS), clinician-report (SSIRAD), and projective methodology (CWS), suggesting both the convergent validity of the measures, as well as the incremental validity of the CWS. Additional strengths of the CWS were noted, in particular, that given this client's affective interference in communicating, the CWS provided a clear and comprehensive manner in which to better understand his personality structure, symptoms, strengths, and current functioning.

CONCLUSIONS

In reviewing presented data on the reliability of the CWS, interrater agreement has been clearly established in multiple studies, both in Italian and English. Preliminary test-retest reliability data is promising, and similar to that of other performance-based personality measures, although more conclusive research is needed.

Summarized studies further suggest the convergent validity of the CWS with other established measures, including the MMPI-2, the Rorschach Inkblot Test, the AAP, the Guilford-Zimmerman Temperament Survey, and others. Moreover, the CWS has demonstrated the ability to differentiate between normal and clinical groups, with studies focusing on a variety of ages, conditions, and mental health diagnoses. High levels of concordance with expert ratings have further been noted. As with any measure, further validity research is recommended.

Lastly, clinical experience with the CWS suggests practical advantages to the measure, incremental validity, and therapeutic applications. These advantages have been elaborated in presented and published case studies, as well as through formal case consultations with clinicians using the measure. With continued use, and the development of a vibrant CWS community of clinicians, further research, development, and growth of the measure is anticipated.

References

Achenbach, T. M. (1991). *Child behavior checklist/4-18*. University of Vermont, Department of Psychiatry.

Alves, I. C. B., Dias, A. R., Sardinha, L. S., & Conti, F. D. (2010). Precisão entre juízes na avaliação dos aspectos formais do teste de Wartegg [Accuracy among judges in evaluating the formal aspects of the Wartegg test]. *Aletheia*, (31), 54-65.

American Psychiatric Association. (1994). *Diagnostic and statistical manual, 4th ed. (DSM-IV).* American Psychiatric Association, Washington.

American Psychiatric Association. (2000). *Diagnostic and statistical manual, 4th ed., Text Revision (DSM-IV-TR).* American Psychiatric Association, Washington.

Attili, G. (2001). *Ansia da separazione e misura dell'attaccamento normale e patologico: versione modificata e adattamento italiano del Separation anxiety test (SAT) di Klagsbrun e Bowlby (12 tavole per la somministrazione del test)*. [Anxiety from separation and measurement of normal and pathological attachment: Modified version and Italian adaptation of the Klagsbrun and Bowlby Separation Anxiety Test (SAT) (12 tables for the administration of the test)]. Milan: Unicopli.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of psychosomatic research, 38*(1), 23-32.

Bailly, D., Dechoulydelenclave, M. B., & Lauwerier, L. (2002). Hearing impairment and psychopathological disorders in children and adolescents. Review of the recent literature. *L'Encephale*, *29* (4, pt.1), 329-337.

Balboni, G., & Cubelli, R. (2004). La specificità delle tecniche proiettive ei limiti del dibattito sulle loro proprietà psicometriche [The specificity of projection techniques and the limits of the debate on their psychometric properties]. *Psicologia Clinica dello Sviluppo*, *8*(2), 195-216.

Barker, D. H., Quittner, A. L., Fink, N. E., Eisenberg, L. S., Tobey, E. A., & Niparko, J. K. (2009). Predicting behavior problems in deaf and hearing children: The influences of language, attention, and parent–child communication. *Development and Psychopathology, 21*(02), 373-392.

Benedetti, A., Bologna, M., Crosato, S., Favretti, L., Giusti, E., Lestingi, L., Sangiorgi, A.M., Crisi, A. (2008). Il test di Wartegg nei servizi di consulenza [The Wartegg test in the Consultant services], *Psichiatria e Psicoterapia*, *27* (2), 128-132.

Bianchi, F., Crisi, A., Palermo, M.T. (2002). Ansia, sordità e serotonina [Anxiety, deafness and serotonin]. *Babele, anno VI* (21), 32-35.

Bianchi, F., Crisi, A., Palermo, M.T., Palladino, G., Sommaruga, L., Sgueglia, F., Vichi, P. (2002). Ilrischio psicopatologico nella sordità [The psychopathological risk in deafness]. *Babele, anno VI* (21), 36-40.

Bianchi, F., Crisi, A., & Di Renzo, M. (1996). *Il test di Wartegg nell'età evolutiva. Un contributo psicodiagnostico allo studio clinico della balbuzie, della sordità e dei disturbi dell'apprendimento* [The Wartegg Test in childhood. A clinical contribution to the study of stammering, deafness, and learning disorders]. Rome, Italy: Edizioni Magi.

Bianchi di Castelbianco, F., Crisi, A., & Di Renzo, M. (1993). *Il Test di Wartegg nell'età evolutiva* [The Wartegg Test in Childhood]. Rome: Armando.

Bornstein, R. F. (2009). Heisenberg, Kandinsky, and the heteromethod convergence problem: Lessons from within and beyond psychology. *Journal of Personality Assessment*, *91*(1), 1-8.

Bosc, M., Dubini, A., & Polin, V. (1997). Development and validation of a social functioning scale, the Social Adaptation Self-evaluation Scale. *European Neuropsychopharmacology 7*(1), S57-S70.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss: Sadness and depression.* New York: Basic Books.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, *41*(3), 687-699.

Butcher, J., Graham, J., Ben-Porath, Y., Tellegen, A., Dahlstrom, W.G., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory, Second Edition (MMPI-2).* University of Minnesota Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81.

Caspi, A., Hariri, A. R., Holmes, A., Uher, R., & Moffitt, T. E. (2010). Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American Journal of Psychiatry*, *167*(5), 509-527.

Checinski, K. (1991). Preliminary findings of the study of the prevalence of psychiatric disorder in prelingually deaf adults living in the community. *In Proceedings of the Mental Health and Deafness conference*, St. George's Hospital Medical School, London.

Chiorri, C. (2011). *Teoria e tecnica psicometrica: costruire un test psicologico* [Psychometric theory and technique: Build a psychological test]. New York: McGraw-Hill.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin, 70*(4), 213.

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

Crisi, A. (1998). *Manuale del Test di Wartegg* [Manual for the Wartegg Test]. Rome, Italy: Edizioni MaGi.

Crisi, A. (1999, July). *Similitudes between the Evocative Character of the Wartegg panels and those of Rorschach plates*. Paper presented at the meeting of the International Rorschach Society, Amsterdam, The Netherlands.

Crisi, A. (2000). Test di Wartegg [The Wartegg test]. In F. Bianchi & M. Di Renzo (Eds.), *Vivere bene la scuola: Esplorazione del disagio infantile nel mondo scolastico* [Living well the school: Exploring childhood discomfort in the scholastic world] (pp. 41-50). Rome, Italy: ES MaGi.

Crisi, A. (2002). "La valutazione psicodiagnostica Wartegg nell'ambito della sordità", [The clinical assessment of deafness with the Wartegg Test]. *Babele, anno VI* (21), 26-31.

Crisi, A. (2006, October). *Il test di Wartegg: un contributo allo studio dei Disturbi dell'Apprendimento* [The Wartegg Test: Contributions to the Study of Learning Disorders]. Paper presented at the meeting of the *Istituto di Ortofonologia*, Rome, Italy.

Crisi, A. (2007). *Manuale del test di Wartegg, 2nd ed.* [Manual for the Wartegg Test, 2nd Edition]. Rome, Italy: Edizioni MaGi.

Crisi, A. (2008b, July). *Rorschach and Wartegg: Two Twin Tests.* Paper presented at the meeting of the International Rorschach Society Convention, Leuven, Belgium.

Crisi, A. (2009, March). *The Wartegg Index of Psychopathology.* Poster presented at the meeting of the Society for Personality Assessment, Chicago, Il.

Crisi A., (2010a). "Il Test di Wartegg applicato ad un gruppo di adolescenti de L'Aquila" [The Wartegg test applied to a group of adolescents of L'Aquila]. In F. Bianchi di Castelbianco & M. Di Renzo (Eds.), *Le 398 meravigliose maestre de L'Aquila* [The 398 wonderful teachers of L'Aquila] (pp. 257-264). Rome, Italy: Edizioni MaGi.

Crisi, A. (2010b, July). *The forgotten test: The Wartegg Drawing Completion Test.* Workshop presented at the meeting of the Rorschach Summer Seminars, Montreux, Switzerland.

Crisi, A. (2011a, March). AAP and Wartegg: A pilot study on attachment. In A. Lis (Chair), *Adult Attachment Projective Picture System (AAP): New Perspectives in the Assessment of Adult Attachment Pattern*. Symposium conducted at the meeting of the Society for Personality Assessment, Boston, Ma.

Crisi, A., (2011b). *Reliability and Validity of the Crisi Wartegg System.* Unpublished manuscript.

Crisi, A., & Carlesimo, S. (2008, July). *The Human, Food, and Cloud contents: Discriminative ability in clinical personality assessment.* Paper presented at the meeting of the International Rorschach Society, Leuven, Belgium.

Crisi, A., & Dentale, F. (2016). The Wartegg Drawing Completion Test: Inter-rater Agreement and Criterion Validity of Three New Scoring Categories. *International Journal of Psychology and Psychological Therapy*, *16*(1), 85-92.

Crisi, A., Janson, H., & Palm, J. (2016). *Test-Retest reliability of the Crisi Wartegg System: Initial analysis.* Unpublished research.

Crisi, A., Palm, J., & Lops, I. (2016). *Interrater reliability of the Crisi Wartegg System: An initial United States sample.* Unpublished research.

Crisi, A., & Pastore, L. (2006). "La valutazione psicodiagnostica: Uno studio pilota condotto con il test di Wartegg" [The clinical assessment: A pilot study with the Wartegg test] (pp. 131-158). In L. Pastore (Ed.) *Il fenomeno del Mobbing. Aspetti psicologici e strumenti di ricerca-intervento*. Milan: Franco Angeli.

Crisi, A., & Pastore, L. (2010, March). *Main characteristics of work-abused clients examined through the Wartegg Test*. Paper presented at the meeting of the Society for Personality Assessment, San Jose, Ca.

Crisi, A., & Shorey, H. (2009, March). *Comparing projective measures: A case study using the Wartegg and the Rorschach.* Poster presented at the meeting of the Society for Personality Assessment, Chicago, Il.

Crisi, A., Testa, F., Carlesimo, S., Lops, I., & Maio, S. (2011, March). *The Wartegg Drawing Completion Test in a Sample of 350 Psychiatric Clients: The indices.* Paper presented at the meeting of the Society for Personality Assessment, Boston, Ma.

Crisi, A., Testa, F., Carlesimo, S., & Maio, S. (2010, March) *Psychopathology and projective techniques: Content analysis through the Wartegg Drawing Completion Test in a sample of 350 psychiatric clients.* Paper presented at the meeting of the Society for Personality Assessment, San Jose, Ca.

Crisi, A., Vari, C., Velotti, P., Carlesimo, S., Guzzi, C., & Zavattini, G.C. (2014, July). *Depression and negative affect in dermatology patients: The use of the WDCT according to the Crisi Wartegg System*. Paper presented at the meeting of the International Rorschach Society, Istanbul, Turkey.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

Daini, S., Bernardini, L., & Panetta, C. (2007). A personality perspective on female infertility: An analysis through Wartegg test. *SIS Journal of Projective Psychology & Mental Health*, *14*(2), 135-144.

Daini, S., Lai, C., Festa, G. M., Maiorino, F., Pertosa, M., & De Risio, S. (2006). Impulsivity in eating disorders: Analyzed through Wartegg test. *Journal of Projective Psychology & Mental Health*, *13*(2), 107-117.

Daini, S., Manzo, A., Pisani, F., & Tancredi, A. (2010). Attempted Suicide: Psychopathology and Wartegg Test indicators. *SIS Journal of Projective Psychology & Mental Health*, *17*(2), 171-177.

Daini, S., Petrongolo, L., Manzo, A., & Bernardini, L. (2012). Personality and Nursing Work: A Comparison between Professional and Student Nurses. *SIS Journal of Projective Psychology & Mental Health*, *19*(1), 18-24.

de Graaf, R., & Bijl, R. V. (2002). Determinants of mental distress in adults with a severe auditory impairment: Differences between prelingual and postlingual deafness. *Psychosomatic Medicine, 64*(1), 61-70.

Denmark, J. C. (1966). Mental illness and early profound deafness. *British Journal of Medical Psychology*, *39*(2), 117-124.

Di Riso, D., Crisi, A., Bianchi di Castelbianco, F., Di Renzo, M., Vichi, P., Racinaro, L., & Morellini D. (2013, March). *Attachment, personality and learning disabilities: A study of learning disabled children and their parents*. Symposium presented at the meeting of the Society for Personality Assessment, San Diego, Ca.

D'Amore, G. (2004). *Comparison between some Rorschach Inkblot Test indexes and Wartegg Test.* Proceedings of the 35th annual meeting of the Society for Psychotherapy Research.

Exner, J. (2003) *The Rorschach: A Comprehensive System. Volume 1: The Rorschach, basic foundations and principles of interpretation.* New York: Wiley and Sons. (1st edition, 1997).

Fellinger, J., Holzinger, D., Dobner, U., Gerich, J., Lehner, R., Lenz, G., & Goldberg, D. (2005). Mental distress and quality of life in the hard of hearing. *Social Psychiatry and Psychiatric Epidemiology, 40*, 737-742.

Fontana U. (1984): *L'Uso del test di Wartegg nella pratica clinica* [Using Wartegg's test in clinical work]. Verona: CISERP.

Fontana, U. (2005): *Una tecnica da non dimenticare. Il WZT validato e allargato per il clinico di oggi* [A method not to be forgotten. Validation of the WZT for today's clinicians]. Venice: Edizioni Quaderni ISRE.

Fontana, U. (2013). *La psicoanalisi imbavagliata* [Psychoanalysis gagged]. Libreriauniversitaria. It Edizioni.

Gallagher, T. M. (1999). Interrelationships among children's language, behavior, and emotional problems. *Topics in Language Disorders, 19*(2), 1-15.

Gardziella, M. (1985). *Wartegg-piirustustesti. K.sikirja* [The Wartegg Drawing Test. A Handbook.] Jyväskylä: Psykologien Kustannus Oy.

George, C., & West, M. L. (2011). The adult attachment projective picture system: Integrating attachment into clinical assessment. *Journal of Personality Assessment, 93*(5), 407-416.

George, C., & West, M. L. (2012). *The Adult Attachment Projective Picture System: Attachment theory and assessment in adults*. New York: Guilford Press.

Guilford, J. P., & Zimmerman, W. S. (1949). The Guilford–Zimmerman Temperament Survey: Manual. Beverly Hills, CA: Sheridan Supply Co.

Harnois, G., Gabriel, P., & World Health Organization. (2000). *Mental health and work: impact, issues and good practices.* International Labor Organization. WHO: Geneva, Switzerland.

Retrieved from: http://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcms_108152.pdf

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*(3), 278.

Hindley, P. A., Hill, P. D., McGuigan, S., & Kitson, N. (1994). Psychiatric disorder in deaf and hearing impaired children and young people: A prevalence study. *Journal of Child Psychology and Psychiatry*, *35*(5), 917-934.

Hindley, P., Kitson, N., & Leach, V. (2000). Forensic psychiatry and deaf people. *Mental Health and Deafness*, 206-231.

Iqbal, Z., & Hall, R. (1991). Mental health services for deaf people: A need identified. *Public Health*, *105*(6), 467-473.

Jorm, A. F., Prior, M., Sanson, A., Smart, D., Zhang, Y., & Easteal, S. (2000). Association of a functional polymorphism of the serotonin transporter gene with anxiety-related temperament and behavior problems in children: A longitudinal study from infancy to the mid-teens. *Molecular Psychiatry*, *5*(5), 542.

Jung C.G. (1960). A review of the complex theory. *Collected Works, Vol. VIII*. Bollingen Series XX. Princeton, New Jersey: Princeton University Press.

Kitson, N., & Fry, R. (1990). Prelingual deafness and psychiatry. *British Journal of Hospital Medicine*, *44*(5), 353.

Klagsbrun, M., & Bowlby, J. (1976). Responses to separation from parents: A clinical test for young children. *British Journal of Projective Psychology & Personality Study*.

Kurd, S.K., & Gelfand, J.M. (2009). The prevalence of previously diagnosed and undiagnosed psoriasis in US adults: Results from NHANES 2003-2004. *Journal of the American Academy of Dermatology*, *60* (2), 218- 224.

Kvam, M. H., Loeb, M., & Tambs, K. (2007). Mental health in deaf adults: Symptoms of anxiety and depression among hearing and deaf individuals. *Journal of Deaf Studies and Deaf Education, 12*(1), 1-7.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Laucht, M., Treutlein, J., Blomeyer, D., Buchmann, A. F., Schmid, B., Becker, K., & Banaschewski, T. (2009). Interaction between the 5-HTTLPR serotonin transporter polymorphism and environmental adversity for mood and anxiety psychopathology: Evidence from a high-risk community sample of young adults. *International Journal of Neuropsychopharmacology*, *12*(6), 737-747.

Maio, S., Ricci, C., & Crisi, A. (2008, July). *Attachment styles in projective techniques: A pilot study.* Paper presented at the meeting of the International Rorschach Society, Leuven, Belgium.

Manicavasaga, V., Silove, D., & Curtis, J. (1997). Separation anxiety in adulthood: A phenomenological investigation. *Comprehensive Psychiatry*, *38*(5), 274-282.

Margoob, M. A., & Mushtaq, D. (2011). Serotonin transporter gene polymorphism and psychiatric disorders: Is there a link? *Indian Journal of Psychiatry*, *53*(4), 289.

Marschark, M., Mouradian, V., & Halas, M. (1994). Discourse rules in the language productions of deaf and hearing children. *Journal of Experimental Child Psychology*, *57*(1), 89-107.

Mattlar, C.-E. (2005, July). *An overview of the Wartegg Zeichentest (WZT): The structure and use thereof, and of research supporting its utility.* Paper presented at the meeting of the International Rorschach Society, Barcelona, Spain.

Melke, J., Landén, M., Baghei, F., Rosmond, R., Holm, G., Björntorp, P., Westberg, L., Hellstrand, M., & Eriksson, E. (2001). Serotonin transporter gene polymorphisms are associated with anxiety□related personality traits in women. *American Journal of Medical Genetics*, *105*(5), 458-463.

Nichols, D. (2001). *Essentials of MMPI-2 assessment*. New York: Wiley.

Olivier, C., Robert, P. D., Daihung, D., Urbà, G., Catalin, M. P., Hywel, W., Kurd, S., Troxel, A. B., Crits-Christoph, P, & Gelfand, J. M. (2010). The risk of depression, anxiety, and suicidality in patients with psoriasis: A population-based cohort study. Archives of dermatology, 146(8), 891-895.

Padden, C. A., & Humphries, T. (1990). *Deaf in America*: *Voices from a Culture*. Harvard University Press.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*(1), 75-92.

Pancheri, P., Biondi, M., Gaetano, P., & Picardi, A. (1999). Costruzione della 'SSIRAD', una scala per la valutazione rapida dimensionale [Construction of the 'SSIRAD', a scale for rapid dimensional evaluation]. *Riv. Psichiatria, 34*, 72-83.

Picardi, A., Lega, I., & Tarolla, E. (2013). Suicide risk in skin disorders. *Clinics in dermatology, 31*(1), 47-56.

Rainer, J. D. (1963). *Family and Mental Health Problems in a Deaf Population. Final Report of the Mental Health Project for the Deaf.* New York: New York State Psychiatric Institute, Department of Medical Genetics.

Randolph, J. J. (2005). Free-Marginal Multirater Kappa (multirater K [free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. *Online Submission*. Retrieved from http://files.eric.ed.gov/fulltext/ED490661.pdf

Ravecca, F. (2003, February). *Il processo di selezione del personale in una Forza Armata che cambia: sperimentazione del test Wartegg all'Accademia Navale di Livorno.* [The Selection proceeding in an Armed Force which is changing: Experimentation with the Wartegg test at the Navy Academy of Livorno]. Paper presented at the Annual Convention of Italian Society of Psychopathology (SOPSI).

Ridgeway, S. M. (1997). Deaf People and Psychology: Some Preliminary Findings. *Deaf Worlds, 13*, 9-18.

Rizzo, A., Della Villa, L., & Crisi, A. (2015). Can the Problematic Internet Use evolve in a pre-psychotic state? A single case study with the Wartegg. *Computers in Human Behavior*, *51*, 532-538.

Roivainen, E., & Ruuska, P. (2005). The use of projective drawings to assess alexithymia. *European Journal of Psychological Assessment*, *21*(3), 199-201.

Rorschach, H. (1942). *Psychodiagnostics: A diagnostic test based on perception*. Bern, Switzerland: *Verlag Hans Huber*. (Original work published 1921)

Scarpellini, C., (1962), *Diagnosi di personalità col reattivo di realizzazione grafica.* [Diagnosis of personality with the drawing completion test]. Contributi dell'Istituto di Psicologia dell'Università Cattolica del Sacro Cuore [Contributions from the Institute of Psychology of the Catholic University of the Sacred Heart], pp. 1-82.

Scarpellini, C. (1964). Diagnostic differentiel a travers deux tests de personalite: le Rorschach et le Wartegg. [Differential diagnostic through two personality tests: The Rorschach and the Wartegg.]. *Bulletin de Psychologie Scolaire et d'Orientation, 13*, 79-92.

Siegel, S., & Castellan, N. J. (1988). Nonparametric systems for the behavioural sciences. *McGraw Hill International Editions*.

Soilevuo Grønnerød, J., & Grønnerød, C. (2012). The Wartegg Zeichen Test: A literature overview and a meta-analysis of reliability and validity. *Psychological Assessment, 24 (2),* 476-89.

Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Substance Abuse Treatment, US (2014). *Trauma-Informed Care in Behavioral Health Services*. Treatment Improvement Protocol (TIP) Series, No. 57. Rockville, MD. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK207191/.

Toivakka, H., & Crisi, A. (2014, March). *Rorschach and Wartegg as tools in a case of adolescent assessment.* Paper presented at the meeting of the Society for Personality Assessment, Arlington, Va.

Turner, O., Windfuhr, K., & Kapur, N. (2007). Suicide in deaf populations: a literature review. *Annals of General Psychiatry*, *6*(1), 6-26.

Vaccari, C., & Marschark, M. (1997). Communication between parents and deaf children: Implications for social-emotional development. *Journal of Child Psychology and Psychiatry, 38*(7), 793-801.

Vandereycken, W. (1993). The Eating Disorder Evaluation Scale (EDES). *Eating Disorders, 1*(2), 115-122.

Warrens, M. J. (2010). Inequalities between multi-rater kappas. *Advances in Data Analysis and Classification*, *4*(4), 271-286.

Werngren-Elgström, M., Dehlin, O., & Iwarsson, S. (2003). Aspects of quality of life in persons with pre-lingual deafness using sign language: subjective wellbeing, ill-health symptoms, depression and insomnia. *Archives of Gerontology and Geriatrics*, *37*(1), 13-24.

World Health Organization. (2017). *Fact sheet: Deafness and hearing loss.* Retrieved from: http://www.who.int/mediacentre/factsheets/fs300/en/

Young, K. S. (1996). Psychology of computer use: XL. Addictive use of the Internet: a case that breaks the stereotype. *Psychological Reports, 79* (3), 899-902.

Zacharie, R., Zacharie, C.O., Lei, U., & Pedersen, A,F. (2004). Medication and health care service utilization related to depressive symptoms in older adults with psoriasis. *Journal of Drugs in Dermatology, 3* (6), 661–666.

Zhang, K., Xu, Q., Xu, Y., Yang, H., Luo, J., Sun, Y., Sun, N., Wang, S., & Shen, Y. (2009). The combined effects of the 5-HTTLPR and 5-HTR1A genes modulates the relationship between negative life events and major depressive disorder in a Chinese population. *Journal of Affective Disorders*, *114*(1), 224-231.

NOTES

[i] We are thankful to Isabella Lops who contributed with data analysis and suggestions for the text in this section.

[ii] We are thankful to Harald Janson who contributed with data analysis and suggestions for the text in this section.