

## Precisão entre juízes na avaliação dos aspectos formais do teste de Wartegg

**Irai Cristina Boccato Alves,  
Augusto Rodrigues Dias,  
Luís Sérgio Sardinha,  
Fábio Donini Conti**

**Resumo:** A questão da clareza e objetividade dos critérios utilizados na avaliação e interpretação dos testes psicológicos é uma das preocupações dos profissionais da Psicologia que trabalham com a avaliação psicológica. O objetivo deste trabalho consistiu em verificar em que medida os critérios de avaliação delineados para alguns aspectos formais do Teste de Completamento de Desenhos de Wartegg (WZT) estão definidos adequadamente. Participaram desta pesquisa dois juízes devidamente treinados nos critérios desenvolvidos pela autora, que avaliaram 191 protocolos do teste. Para determinar a precisão entre juízes foi calculado o coeficiente Kappa, para cada campo em cada uma das variáveis estudadas, tendo variado 0,66 a 1,00. Estes resultados permitiram concluir que os critérios propostos, de um modo geral, se mostraram claros e objetivos para os aspectos formais estudados, possibilitando o seu emprego de forma relativamente segura na avaliação de um protocolo do WZT.

**Palavras-chave:** técnicas projetivas; teste de Wartegg; precisão de avaliadores; avaliação de desenhos.

### **Raters reliability in the assessment of formal aspects of the Wartegg test**

**Abstract:** The question of clearness and objectivity of the criteria in the assessment and interpretation of psychological tests has been a concern between Psychology professionals who work with psychological assessment. The purpose of this work was to verify if the assessment criteria for some formal aspects of The Wartegg Test (WZT) are adequately defined. Two trained judges evaluated 191 test protocols. The Kappa coefficient was calculated to determinate the rater's reliability to each studied variable and for each WZT field. The coefficients between raters ranged from 0.66 to 1.00. It can be concluded that the proposed criteria for evaluation of the formal aspects were clear and objective and they will permit their use in a secure form in the WZT assessment.

**Keywords:** projective techniques; Wartegg test; raters reliability; drawing assessment.

---

### **Introdução**

Para que um teste psicológico seja considerado em condições de uso, ele necessita apresentar estudos atualizados relativos aos seus parâmetros psicométricos, em especial os que evidenciem sua validade e precisão. Tal condição passou a ser determinante a partir da Resolução 002/2003 do Conselho Federal de Psicologia (C.F.P.), que define e regulamenta o uso, a elaboração e a comercialização dos testes psicológicos. Como consequência desta Resolução, na atualidade, uma série de instrumentos se encontra com a utilização suspensa devido à ausência de tais estudos, em particular aqueles que são classificados como técnicas ou métodos projetivos (Alves, 2004; Sardinha, 2008).

Dentro deste universo inclui-se o Teste de Completamento de Desenhos de Wartegg (WZT). Este instrumento caracteriza-se como uma técnica projetiva gráfica, de completamento de desenhos, que pretende avaliar a personalidade por meio das produções realizadas livremente pelos examinandos a partir de oito sinais gráficos dispostos em oito campos (Berlinck, 2000). O WZT encontra-se na lista de testes com parecer “não favorável” devido à escassez de estudos de validade e precisão com a população brasileira.

O Teste de Wartegg foi muito usado no Brasil, conforme pode ser constatado na pesquisa de Noronha, Beraldo e Oliveira (2003), na qual foi indicado como o 5º colocado entre os testes mais usados pelo psicólogo (N = 52) na sua prática profissional. Por outro lado, a literatura internacional sobre o WZT é muito ampla, mas é de difícil acesso, conforme foi apontado pelo levantamento apresentado por Berlinck (2000, 2006) e Ramon (2006), principalmente porque alguns dos estudos mais completos foram realizados por Takala (citados por Heidberg, 1981) na Finlândia.

Também podem ser destacados os trabalhos de Kinget (1952) nos Estados Unidos e o de Biedma e D'Alfonso (1955/1973) na Suíça. Kinget foi responsável pela sistematização dos critérios de avaliação do teste e realizou estudos de validade, comparando os resultados do WZT com os dados de um questionário baseado no Inventário de Personalidade de Benreuter.

No Brasil, a literatura sobre o WZT é pequena e existem poucos dados de pesquisa disponíveis sobre a nossa realidade, que foram resumidos por Alves (2008). Algumas dessas pesquisas serão apresentadas a seguir.

Os únicos estudos normativos do Teste de Wartegg desenvolvidos para adultos no Brasil foram os de Berlinck (2000; 2006), cujo objetivo foi estabelecer critérios para a aplicação, avaliação e interpretação para pessoas com diversos níveis de escolaridade. A autora propôs critérios objetivos para avaliação, baseados principalmente no trabalho de Kinget (1952).

Em 1999, Gullo, Reis e Siqueira compararam as características de originalidade avaliadas pelo Teste de Wartegg e pelo Teste Pensando Criativamente com Figuras de Torrance, em universitários, tendo obtido uma correlação positiva e fraca (0,21) entre as características de originalidade avaliadas pelos dois testes.

Salazar, Tróccoli e Vasconcelos (2001) compararam os resultados do fator Desempenho do IFP-R e os desenhos do campo 5 do Wartegg, entre o campo 8 e o fator Afiliação e entre o tipo de Sequência do WZT e o fator Ordem, de uma amostra de 723 participantes com nível de escolaridade superior. Os protocolos do WZT foram avaliados por dois juízes e foram correlacionados os resultados de cada avaliador e da média entre eles com os fatores do IFP-R. As correlações obtidas foram próximas de zero, indicando não haver relação entre os aspectos avaliados nos dois testes. As correlações entre os juízes variaram entre 0,18 e 0,78, as quais apontam para a necessidade de critérios objetivos. Concluem que os dois testes parecem abordar aspectos diferentes da personalidade, bem como apresentam pressupostos teóricos diferentes, o que pode explicar os resultados encontrados. É importante ressaltar que não se sabe exatamente como foram avaliadas as características do WZT para chegar às conclusões apresentadas.

Investigando a validade de critério do WZT com o Teste de Zulliger, em relação à variável movimento, Berlinck (2002) obteve uma correlação significativa de 0,349 entre os dois testes. Embora a correlação não seja muito alta, esta indica a validade da interpretação dessa variável nos dois testes, que se relaciona à criatividade, empatia, espontaneidade e poder de adaptação ao meio externo.

Em outro estudo de validade, Cruz, Ruschel, Meazzi, Monteiro e Fagundes (2003) avaliaram a persistência e o desempenho no IFP e no campo 3 do Wartegg, que avalia ambição, desejo de crescimento e perseverança. Encontraram que um resultado favorável no campo 3 estava relacionado a maior persistência no IFP, o que sugere tendência a terminar um trabalho, mesmo quando este é difícil. Foram encontradas diferenças entre homens e mulheres, com resultados mais altos para os homens.

Muitos questionamentos surgem quando se aborda a validade e a precisão das técnicas projetivas, em especial a adequação ou não destes parâmetros psicométricos para este tipo de instrumento. A esse respeito, Vane e Guarnaccia (1989) assinalam que os procedimentos (métodos) utilizados para o estabelecimento de tais parâmetros foram criados para testes cujos resultados são expressos de forma quantitativa e em dimensões únicas, diferentemente do que ocorre com as técnicas projetivas, em que os resultados dependem muito da subjetividade do avaliador. Contudo, ressaltam a necessidade de estabelecer critérios padronizados para a avaliação. Cabe salientar que, a principal distinção entre os testes objetivos e os projetivos reside no fato de que os primeiros têm por objetivo informar o quanto um indivíduo tem de um determinado traço, estado ou fator; enquanto que os projetivos seriam considerados meios de se obter informações sobre a pessoa avaliada (Weiner, 2000). Entretanto, independente da problemática exposta, é impossível não tratar as questões de validade e precisão dessas técnicas, em virtude de ser necessário verificar se elas fazem aquilo que se propõem a fazer e com que consistência o fazem (Alves, 2006).

No que se refere à precisão das técnicas projetivas deve ser dada atenção especial à escolha do método a utilizar. Os métodos mais comuns para o estabelecimento da precisão de um teste são (*teste-reteste*, *formas paralelas* e *divisão em metades* ou *consistência interna*), entretanto, apresentam dificuldades quando aplicados às técnicas projetivas. No caso do *teste-reteste* a dificuldade estaria relacionada a mudanças em aspectos da personalidade que podem ocorrer com o passar do tempo ou como consequência de pressões externas e internas, que vão depender do intervalo de tempo entre as aplicações. Se o intervalo for curto, o reteste vai mostrar se as flutuações em um breve período de tempo alteram o resultado do teste. Se a correlação for alta, ela vai indicar que estão sendo avaliados aspectos mais estáveis da personalidade. Em relação às *formas paralelas*, a dificuldade estaria ligada à criação de formas equivalentes, pois os estímulos nestas duas formas deveriam avaliar as mesmas características. No caso da *divisão em metades* a dificuldade estaria centrada na divisão de um teste projetivo em duas metades equivalentes (Alves, 2006).

Depreende-se, portanto, que avaliar a precisão de uma técnica projetiva pelos métodos mais convencionais pode se tornar uma tarefa difícil e até mesmo inócua, afinal toda avaliação, que em seu processo envolva aspectos da subjetividade do indivíduo que a realiza, está mais propensa a erros e, conseqüentemente, pode ser considerada mais

vulnerável e imprecisa. Desse modo, torna-se necessária a busca de um método mais adequado às características peculiares destes instrumentos (Sardinha, 2000). Este método estaria baseado na concordância entre as avaliações efetuadas por dois ou mais juízes independentes (Alves, 2006).

Comumente denominada de precisão ou fidedignidade do avaliador (juízes independentes), este método prevê que uma amostra dos protocolos do teste seja pontuada independentemente, de acordo com critérios previamente definidos, por dois ou mais examinadores. Os escores atribuídos por cada avaliador a cada examinando devem ser correlacionados, sendo os coeficientes resultantes, medidas da fidedignidade do avaliador (Anastasi & Urbina, 2000; Dias, 2005). No Brasil, alguns estudos foram desenvolvidos para aferir a precisão do WZT fazendo uso deste método, entre os quais se destacam os trabalhos de Silva (2004) e Ramon (2006).

Silva (2004) realizou dois estudos de concordância entre avaliadores. O primeiro envolveu a avaliação de 93 protocolos quanto às variáveis “Atmosfera e Envolvimento”, por dois profissionais treinados. A variável “Atmosfera” foi avaliada no campo “2” do WZT, seguindo os critérios propostos por Kinget (1952), e classificada em três níveis previamente definidos: atmosfera negativa, neutra ou positiva. Para a variável “Envolvimento” foram considerados os desenhos executados nos oito campos, no que se refere à atenção na produção demonstrada por meio da realização cuidadosa e presença de detalhes adicionais e enriquecedores. Esta variável foi classificada em três níveis alto, médio e baixo.

No segundo estudo foi solicitado a dois psicólogos clínicos com experiência no WZT que avaliassem as seguintes características complexas da personalidade: recursos intelectuais; habilidades sociais, flexibilidade, organização, nível de aspiração, nível de energia, controle emocional, segurança e assertividade de 30 protocolos. Para cada característica os profissionais seguiram critérios previamente definidos e as classificavam nos níveis 1 (baixo), 2 (médio) e 3 (alto).

Quanto aos resultados, no primeiro estudo Silva (2004) observou que as avaliações relativas à variável “atmosfera” apresentaram correlação de 0,91 e de 0,89, para a variável “envolvimento”. No segundo estudo, a análise de características complexas da personalidade atingiu correlação média de 0,88, sendo que as correlações máximas (1,00) foram para as características “energia” e “assertividade” e a correlação mais baixa (0,81) para a característica “controle emocional”. Com base nestes dados, a autora concluiu que profissionais treinados atingem altos níveis de concordância na avaliação do WZT, tanto na avaliação de características de execução quanto de características de personalidade mais complexas.

Ramon (2006) efetuou um estudo, relativo à precisão do WZT, dividido em duas etapas. A primeira voltada a verificar a precisão da classificação de 27 variáveis e a segunda direcionada para a precisão da interpretação de três características: relacionamento interpessoal, afetividade e controle emocional e ambição. Participaram do estudo 18 psicólogos com experiência entre três e 30 anos na avaliação do WZT, que analisaram cinco protocolos.

Na primeira etapa deste estudo (precisão da classificação), os resultados mostraram que das 27 variáveis classificadas, sete apresentaram correlações altas (acima de 0,70),

quatro obtiveram correlações satisfatórias (acima de 0,60), cinco correlações medianas (entre 0,50 e 0,60) e 11 variáveis correlações abaixo de 0,50. Em relação à segunda etapa (precisão das interpretações), o autor obteve precisão satisfatória para as três características consideradas, entretanto, este resultado foi obtido com 10 dos 18 juízes. Em suas conclusões, Ramon observou que foi possível estabelecer a precisão no sistema de classificação do WZT para algumas variáveis, sendo necessário o desenvolvimento de novos estudos, principalmente em relação à precisão das interpretações, para permitir que avaliadores diferentes façam interpretações semelhantes de um mesmo protocolo.

Como é possível observar, os resultados encontrados nos estudos de Silva (2004) e Ramon (2006) são animadores e indicam que, em relação à precisão do WZT, seria primordial o domínio ou conhecimento que o profissional possua dos critérios utilizados, principalmente no que se refere à classificação e interpretação dos dados. Na verdade, a questão da experiência é uma preocupação antiga dos profissionais da Psicologia. Van Kolck (1984) e Alves (2006) apontam que para uma avaliação são necessários profissionais com bastante experiência na área, que possam traduzir melhor os dados fornecidos por um teste.

Concorda-se que a experiência seja um fator importante em um processo avaliativo, mas não pode ser o único (Sardinha, 2000). Esta experiência encobre uma falta de parâmetros claros de avaliação dos testes, pois *“muitas técnicas pecam por não apresentar um sistema de análise e interpretação que seja suficientemente preciso de modo a permitir alto grau de concordância quando avaliados por profissionais diferentes”* (Villemor-Amaral, 2006, p. 168-169). Villemor-Amaral (2006) acrescenta ainda, que, para ampliar a gama de indicadores de validade das técnicas projetivas, é preciso criar sistemas de análise que garantam a precisão entre avaliadores. Assim, os psicólogos devem se esforçar para criar critérios objetivos e que possam ser utilizados de maneira mais uniforme e ampla por toda a categoria profissional.

Foi neste sentido que Berlinck (2000) realizou um estudo sobre o WZT. Seu trabalho teve como meta estabelecer critérios objetivos para a aplicação, avaliação e interpretação do teste, a partir de uma ampla revisão da literatura sobre o mesmo. Em termos da classificação dos aspectos formais do WZT, propõe que sejam avaliadas cinco categorias e suas respectivas derivações, a saber: a) traçado ou linha, em que se avaliam a pressão (forte, média, fraca e mista), o tipo de linha (reta, curva e reta/curva), a continuidade do traçado (contínuo ou descontínuo) e a qualidade da linha (reforço, retoque e tremor); b) o tamanho do desenho (grande, médio, pequeno e constrição); c) sombreado (presença ou ausência); d) composição do desenho em que são avaliados a organização (bidimensional e tridimensional), a repetição, duplicação e a recorrência, o movimento, a orientação, a clareza, a originalidade e a popularidade e; e) outros em que se enquadram a transparência, o desenho sobre moldura, a visão de pássaro e a mudança de posição. Entretanto, apesar de valioso, este sistema de classificação ainda não foi submetido a um estudo que verifique a sua clareza e objetividade. Entende-se que estudos desta natureza trariam uma maior confiabilidade ao sistema proposto.

Foi a partir desta constatação que o presente trabalho se originou. Seu objetivo consistiu em verificar em que medida alguns critérios de classificação dos aspectos formais do WZT, delineados por Berlinck (2000), estão adequadamente definidos. Em outros

termos, se são claros e objetivos o suficiente para serem compreendidos e utilizados pelos profissionais que irão realizar uma avaliação (Primi, Miguel, Couto & Muniz, 2007).

## **Método**

### *Participantes*

Participaram do presente estudo 191 sujeitos, dos quais 100 eram do sexo feminino, com as idades variando entre 18 e 54 anos (média de 28,6 anos, DP de 8,71), e 91 do sexo masculino com idades entre 18 e 50 anos (média de 28,0, DP de 9,36). Em termos de escolaridade 2,1% (n=4) do total não a informaram, 11,5% (n=22) indicaram possuir o ensino básico completo, 12,6% (n=24) o ensino médio incompleto e, 73,8% (n=141) o ensino médio completo. Participaram também, dois juízes treinados no método de pontuação desenvolvido por Berlinck (2000) para a avaliação dos aspectos formais do WZT. Estes juízes eram profissionais devidamente inscritos no CRP 06 – São Paulo.

### *Material*

Foram utilizadas as folhas de aplicação padronizadas do Teste de Completamento de Desenhos de Wartegg publicadas pela editora (CETEPP), lápis preto N° 2 e borracha para cada examinando. A folha de aplicação é constituída de oito quadrados ou campos, havendo em cada quadrado um sinal ou estímulo gráfico, para ser completado pelo examinando da forma e sequência que desejar.

### *Procedimento*

A coleta de dados foi realizada em um único dia e de forma coletiva, sem limite de tempo. Os sujeitos eram parte de um grupo de 700 candidatos a um concurso público para o cargo de Agente Comunitário em uma cidade da Grande São Paulo no ano de 2003, realizado antes da publicação da lista definitiva de testes pelo CFP. Estes foram subdivididos em grupos de aproximadamente 20 indivíduos por sala e todos assinaram o termo de consentimento livre-esclarecido. As instruções de aplicação do WZT utilizadas foram as padronizadas por Wartegg (1987).

Para o estudo de precisão, inicialmente os dois juízes se submeteram a um treinamento na codificação dos aspectos formais delineados por Berlinck (2000), selecionados para esta pesquisa. Para tanto, 10 protocolos do WZT (metade de cada sexo), que não fizeram parte de amostra, foram utilizados. Cada juiz indicou a presença ou ausência de cada um dos aspectos formais avaliados para cada um dos oito campos dos protocolos considerados, com base nas definições de Berlinck. Em seguida, os juízes discutiram as concordâncias e discordâncias das classificações de modo a dirimir as dúvidas.

Os 191 protocolos do WZT foram entregues aos dois juízes obedecendo a seguinte distribuição: juiz 1 recebeu os 100 protocolos de sujeitos do sexo feminino para a avaliação e o juiz 2 recebeu os 91 protocolos dos sujeitos masculinos. Posteriormente, os juízes trocaram os protocolos do WZT, para avaliar os restantes. Os juízes indicaram a presença ou ausência de cada uma das variáveis para cada campo em todos os protocolos.

Os juízes consideraram os seguintes aspectos formais e respectivas definições:

1. **Pressão do lápis:** a) forte: “*caracterizada por marcas profundas, escuras e visíveis principalmente nas costas do papel*” (Berlinck, 2000, p. 94); b) média: linhas intermediárias, não há traços fortes, nem leves; c) fraca: realizada com linhas extremamente delicadas e leves e; d) mista: presença de traços fortes, médios e fracos em diversas combinações em um mesmo desenho;

2. **Tipo de linha:** a) reta: aparenta ter sido traçada com o auxílio de uma régua; b) curva: traçado com características de flexibilidade em sua construção, bem como graciosidade e fluência; c) mista: quando os dois tipos de linhas estão presentes num mesmo desenho;

3. **Continuidade da linha:** a) contínua: traçado constituído por uma linha sem interrupções, feito de tal forma que dá a ideia de que o lápis não foi levantado do papel; b) descontínua: realizada com interrupções, paradas ou quebras nos traçados, subdividindo-se em: b1) traçado interrompido: com uma ou mais interrupções ao longo das linhas; b2) tracejado: linha substituída por pequenos traços que mostram aparente continuidade; b3) pontilhado: quando a linha contínua é substituída por pontos. Para o presente estudo, estes três tipos foram agrupados em uma única variável, a saber: traçado interrompido.

4. **Qualidade da linha:** a) reforço ou retoque: caracterizado pela presença de linhas repassadas, dando a impressão de maior largura ou cobrindo linhas leves e finas; b) Linha tremida: avaliada pelas constantes mudanças não intencionais do traçado, ou seja, ondulações características de tremores.

5. **Tamanho do desenho:** a) grande: quando o desenho ocupa mais do que  $\frac{3}{4}$  do campo (de 13 a 16 células); b) médio: entre  $\frac{1}{4}$  a  $\frac{3}{4}$  do campo (de 5 a 12 células); c) pequeno: desenho em área menor que  $\frac{1}{4}$  do campo (de 1 a 4 células). Para fazer a distinção entre os três tamanhos foi utilizado o crivo de avaliação, que subdivide a área de um campo do WZT em 16 células de  $1\text{cm}^2$ , proposto por Biedma e D’Alfonso (1973) e adotado por Berlinck (2000).

6. **Sombreado nos desenhos:** a) presença; b) ausência, considerando-se as superfícies de colorido mais escuro, contrastando com zonas de luz.

7. **Movimento nos desenhos:** a) presença; b) ausência. O desenho transmite uma sensação de que as figuras estão em movimento.

8. **Transparência nos desenhos:** a) presença; b) ausência. Constitui a inclusão de partes no desenho, que não seriam visíveis na realidade.

## Resultados e discussão

A análise estatística foi feita como auxílio do SPSS versão 17 para Windows. Foram calculados os coeficientes Kappa entre os dois juízes de cada variável por campo do WZT e depois as médias desses coeficientes para cada variável. O coeficiente Kappa de Cohen (1960) é uma medida estatística de concordância entre juízes para escalas nominais (qualitativas, isto é, em que existe a avaliação em categorias). Ele é considerado uma maneira mais adequada de avaliação do que o simples cálculo de porcentagens de acordo, porque leva em conta a possibilidade de concordância por acaso (Cohen 1960; Wikipedia, 2009). A interpretação do coeficiente Kappa pode ser classificada como

*concordância quase perfeita*, quando os valores estão entre 0,81 e 1,00; *concordância substancial*, quando estão entre 0,61 e 0,80; *concordância moderada* entre 0,41 e 0,60; *concordância fraca ou pequena* entre 0,21 e 0,40; *concordância leve* entre 0,0 e 0,20 e, *nenhuma correlação* quando forem menores do que zero (0), indicando *ausência de acordo* (Landis & Koch, 1977).

Além deste sistema de classificação, adotou-se como critério mínimo para indicar clareza e objetividade das definições dos aspectos formais avaliados a correlação de 0,60, ou seja, *moderada* (Conselho Federal de Psicologia – CFP, 2003; Dancey & Reidy, 2006). Desse modo, as correlações inferiores a 0,60 indicariam falta de clareza e objetividade e, conseqüentemente, a necessidade de redefinição dos critérios de avaliação. Salieta-se que este valor de correlação é o mínimo aceito para coeficientes de precisão pelo Conselho Federal de Psicologia (CFP), de acordo com a Resolução 02/2003.

Nas Tabelas de 1 e 2 são apresentados os coeficientes de concordância Kappa entre os juizes e as médias para os aspectos formais avaliados.

Tabela 1 – Coeficientes Kappa para cada campo da pressão do traçado, tipo, continuidade, descontinuidade e qualidade da linha

| Aspectos formais   | Campos |      |      |      |      |      |      |      | M    |
|--------------------|--------|------|------|------|------|------|------|------|------|
|                    | 1      | 2    | 3    | 4    | 5    | 6    | 7    | 8    |      |
| Pressão            | 0,84   | 0,80 | 0,85 | 0,83 | 0,86 | 0,79 | 0,86 | 0,88 | 0,84 |
| Tipo de linha      | 0,93   | 0,85 | 0,82 | 0,87 | 0,91 | 0,93 | 0,91 | 0,94 | 0,89 |
| Linha contínua     | 0,90   | 0,84 | 0,90 | 0,86 | 0,84 | 0,83 | 0,97 | 0,88 | 0,88 |
| Linha descontínua  | 0,90   | 0,83 | 0,88 | 0,88 | 0,84 | 0,95 | 0,93 | 0,99 | 0,90 |
| Qualidade da linha | 0,92   | 0,89 | 0,84 | 0,86 | 0,91 | 0,91 | 0,90 | 0,85 | 0,88 |

Na Tabela 1 verifica-se que todos os coeficientes foram positivos e variaram de 0,79 a 0,99, sendo todas acima do critério mínimo adotado pela Resolução 02/2003 do CFP para a precisão. Os maiores coeficientes encontrados foram para linha descontínua, no campo 8 com  $kappa = 0,99$ , linha contínua, no campo 7 com  $kappa = 0,97$  e o menor foi  $kappa = 0,79$  no campo 6 (pressão do lápis). No caso do maior coeficiente observa-se a presença da concordância positiva quase perfeita (0,99), relativa a linha descontínua no campo 8 do WZT.

Em relação ao menor coeficiente ( $kappa = 0,79$ , no campo 6 da pressão do lápis), ocorreu a maior discrepância entre os avaliadores, embora possa ser classificado como uma concordância significativa. Foram calculadas as médias dos coeficientes dos oito campos para cada um dos aspectos formais. Estas variaram entre 0,84 (pressão do lápis) e 0,90 (linha descontínua). Estes coeficientes podem ser considerados pelo critério adotado como concordâncias quase perfeitas.

Só é possível fazer comparações em relação à literatura com o estudo de Ramon (2006), no qual a precisão foi obtida por meio das correlações de Pearson entre quatro

pares de juizes, que para a pressão variaram de 0,13 a 0,62, sendo três delas superiores a 0,56. Contudo em seu trabalho o autor não especificou como foi avaliada a pressão, nem ao que as correlações se referem, isto é, se para cada campo e para cada intensidade de pressão, o que torna difícil uma comparação dos resultados.

Ramon (2006) também investigou a continuidade da linha, tendo encontrado correlações muito baixas (entre 0,14 e 0,31), todas não significantes. Deve-se lembrar que o autor estudou as variáveis linhas contínuas e descontínuas, mas não ofereceu uma definição, nem ilustrações das mesmas para os juizes, o que fez com que cada juiz utilizasse seu próprio parâmetro, o que pode ter levado à ausência de correlações significantes. Para a linha reforçada ou trêmula Ramon (2006) obteve correlações entre 0,21 e 0,60, que foram muito inferiores às da presente pesquisa.

Tabela 2 – Coeficientes kappa para cada campo do tamanho, sombreado, movimento e transparência nos desenhos

| Aspectos formais | Campos |      |      |      |      |      |      |      |      |
|------------------|--------|------|------|------|------|------|------|------|------|
|                  | 1      | 2    | 3    | 4    | 5    | 6    | 7    | 8    | M    |
| Tamanho          | 0,90   | 0,88 | 0,86 | 0,78 | 0,83 | 0,90 | 0,91 | 0,90 | 0,87 |
| Sombreado        | 0,87   | 0,85 | 0,77 | 0,81 | 0,85 | 0,75 | 0,92 | 0,87 | 0,84 |
| Movimento        | 0,96   | 0,89 | 0,85 | 0,82 | 0,95 | 0,91 | 1,00 | 0,93 | 0,92 |
| Transparência    | 1,00   | 0,93 | 0,74 | 0,88 | 0,85 | 0,86 | 0,70 | 0,66 | 0,83 |

Na Tabela 2 todos os coeficientes Kappa foram positivos, sendo que os maiores foram para movimento (Kappa = 1,00) no campo 7 e transparência no campo 1 (Kappa = 1,00) e, os menores, foram para transparência (Kappa = 0,66) no campo 8 e no campo 7 (Kappa = 0,70).

Estes dados são importantes, pois se pode considerar no geral que, na avaliação de um determinado aspecto formal dos desenhos do WZT, os dois juizes usaram o mesmo critério, o que leva a uma consistência nas conclusões que podem ser tiradas a partir dos mesmos. Os coeficientes podem ser considerados como indicando concordância substancial ou quase perfeita, uma vez que variaram entre 0,66 e 1,00, com médias que variaram entre 0,83 para o sombreado até 0,92 para o movimento.

Em seu estudo, Ramon (2006) obteve para o tamanho dos desenhos correlações entre 0,67 e 0,80, indicando que essa é uma das variáveis mais precisas nas duas pesquisas, mesmo que o autor não tenha estabelecido parâmetros para ela. As variáveis movimento e sombreado também foram investigadas por Ramon (2006), que estabeleceu como movimento os vários tipos propostos por Kinget (1952), também considerados na presente pesquisa, incluindo movimento humano, animal, inanimado, cósmico e mecânico. As correlações obtidas variaram de 0,62 a 0,84, um pouco menores do que as desta pesquisa. Quanto ao sombreado ele estabeleceu as classificações leve, moderado e escuro, tendo encontrado correlações entre 0,37 e 0,73, também mais baixas do que as desta pesquisa.

O outro estudo brasileiro sobre a fidedignidade de avaliadores do WZT foi realizado por Silva (2004), que investigou variáveis diferentes das abordadas na presente pesquisa, o que torna impossível a comparação com seu estudo. A única conclusão que pode ser considerada em comum com essa autora é a da importância do domínio ou experiência do profissional sobre os critérios utilizados.

### **Conclusão**

Este trabalho é parte de um projeto mais amplo que pretende verificar a adequação de todos dos critérios propostos por Berlinck (2000) para a avaliação do WZT em relação à precisão entre avaliadores. O objetivo deste trabalho consistiu em verificar em que medida os critérios delineados por Berlinck para alguns aspectos formais do Teste de Completamento de Desenhos de Wartegg (WZT) estão definidos adequadamente, isto é, se eles são claros e objetivos o suficiente para serem compreendidos e utilizados pelos profissionais que irão realizar a avaliação do teste.

Os resultados foram promissores, pois os coeficientes obtidos foram classificados entre concordância substancial ou quase perfeita, de acordo com o critério de Landis e Koch (1977), bem como ficaram acima do critério mínimo adotado (0,60) pela Resolução 02/2003 do CFP para os estudos de precisão. Tais dados permitem afirmar que os critérios propostos por Berlinck (2000) para os aspectos formais do WZT podem ser utilizados com considerável precisão, mesmo os que apresentaram as correlações mais baixas.

Também é importante lembrar que os resultados mais altos obtidos nesta pesquisa em relação aos relatados por Ramon (2006) devem-se principalmente a dois aspectos. Nesta pesquisa as variáveis e suas classificações foram claramente definidas fazendo com que os juízes tivessem parâmetros para fazer suas avaliações. O segundo aspecto, que muito contribuiu para os resultados obtidos, foi a realização de um treinamento e da discussão dos critérios entre os juízes para que os dois estabelecessem parâmetros comuns para as variáveis. Assim, pode-se concluir que para se tirar conclusões confiáveis do WZT é necessário que os avaliadores conheçam bem as definições das variáveis, bem como sejam submetidos a uma treinamento na avaliação para evitar classificações altamente subjetivas que somente poderão prejudicar as interpretações de qualquer técnica projetiva.

### **Referências**

- Alves, I. C. B. (2004). Técnicas projetivas: questões atuais na Psicologia. Em: C. E. Vaz & R. L. Graeff (Orgs.), *Técnicas projetivas: produtividade em pesquisa* (pp. 361-366). Porto Alegre: SBRo.
- Alves, I. C. B. (2006). Considerações sobre a validade e precisão nas técnicas projetivas. Em: A. P. P. Noronha, A. A. A. Santos & F. F. Sisto (Orgs.), *Facetas do fazer em avaliação psicológica* (pp.173-190). São Paulo: Vetor.
- Alves, I. C. B. (2008). Pesquisas com o teste de Wartegg no Brasil. Em: A. E. Villemor-Amaral & B. S. G. Werlang (Orgs.), *Atualizações em métodos projetivos para avaliação psicológica* (pp. 345-363). São Paulo: Casa do Psicólogo.

- Anastasi, A., & Urbina, S. (2000). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Berlinck, V. R. (2000). *O Teste de Completamento de Desenhos de Wartegg em universitários de São Paulo*. Dissertação de Mestrado. Instituto de Psicologia. Universidade de São Paulo, São Paulo.
- Berlinck, V. R. (2002). Estudo comparativo de indicadores de movimentos no Wartegg Zeichen Test e no Teste de Zulliger?. Resumo dos *Anais do 1º Congresso Brasileiro de Psicologia: Ciência e Profissão*. Conselho Regional de Psicologia, São Paulo, Identificador 286. Disponível: <[www.bvs-psi.org.br/congre/busca.asp](http://www.bvs-psi.org.br/congre/busca.asp)> Acesso: 15.10.09.
- Berlinck, V. R. (2006). *O Teste de Completamento de Desenhos de Wartegg em profissionais adultos com nível de escolaridade fundamental e médio*. Tese de Doutorado. Instituto de Psicologia da Universidade de São Paulo, São Paulo.
- Biedma, C. J., & D'Alfonso, P. G. (1973). *A linguagem do desenho. Versão modificada do teste de Wartegg*. São Paulo: Ed. Mestre Jou. (Original publicado em 1955)
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Conselho Federal de Psicologia – CFP. (2003). *Resolução 002: Define e regulamenta o uso, a elaboração e a comercialização de testes psicológicos e revoga a Resolução CFP nº 025/2001*. Brasília-DF.
- Cruz, D. V. A., Ruschel, M., Meazzi, R., Monteiro, J. K., & Fagundes, P. M. (2003). Investigando possíveis relações entre o IFP e o Wartegg no processo seletivo. *Resumos do I Congresso Nacional de Avaliação Psicológica e IX Conferência Internacional de Avaliação Psicológica: Formas e Contextos*. Instituto Brasileiro de Avaliação Psicológica, Campinas, p.203.
- Dancey, C. P., & Reidy, J. (2006). *Estatística sem matemática para psicologia: Usando SPSS para Windows*. Porto Alegre: Artmed.
- Dias, A. R. (2005). *Avaliação da criatividade por metáforas*. Dissertação de Mestrado. Universidade São Francisco, Itatiba, SP.
- Gullo, A. S. M., Reis, C. L., & Siqueira, L. G. G. (1999). Originalidade figural: comparação entre os Testes de Wartegg e pensando criativamente com figuras. [Resumo]. Em: Sociedade Brasileira de Psicologia (Org.), *Resumos de comunicações científicas. XXIX Reunião Anual de Psicologia*. (p.234). Campinas: SBP.
- Heidgerd, E. (1981). Pesquisas nas técnicas gráficas. Em: E. F. Hammer (Org.), *Aplicações clínicas dos desenhos projetivos* (pp. 368-387). Rio de Janeiro: Ed. Interamericana.
- Kinget, M. (1952). *The Drawing-Completion Test: a projective technique for the investigation of personality*. Chicago, New-York: Grune & Stratton.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Noronha, A. P. P., Beraldo, F. N., & Oliveira, K. L. (2003). Instrumentos psicológicos mais conhecidos e utilizados por estudantes e profissionais de Psicologia. *Psicologia Escolar e Educacional*, 7(1), 47-56.
- Primi, R., Miguel, F. K., Couto, G., & Muniz, M. (2007). Precisão de avaliadores na avaliação da criatividade por meio da produção de metáforas. *Psico-USF*, 12(2), 197-210.

- Ramon, R. R. (2006). *Wartegg: Precisão entre avaliadores e evidência de validade com o Método de Rorschach*. Dissertação de Mestrado. Universidade São Francisco, Itatiba.
- Salazar, A., Tróccoli, B. T., & Vasconcelos, T. (2001). Investigação das correlações entre medidas de personalidade projetiva e objetiva. [Resumo]. Em: Sociedade Brasileira de Psicologia (Org.), *Resumos de comunicações científicas*. XXXI Reunião Anual de Psicologia (p.244). Rio de Janeiro: SBP.
- Sardinha, L. S. (2000). Estudo sobre aspectos de personalidade de usuários e não usuários de drogas através do Rorschach. Em: Sociedade Brasileira de Rorschach (Org.), *Livro dos Anais do II Congresso Nacional da Sociedade Brasileira de Rorschach e outros métodos projetivos* (p.119-123). Porto Alegre/RS: AGE Editora
- Sardinha, L. S. (2008). As emoções de dependentes de drogas: estudo através do Rorschach. Em: Sociedade Brasileira de Rorschach (Org.), *Livro de Programas e Resumos do V Encontro da Associação Brasileira de Rorschach e Métodos Projetivos* (p.549-558). Ribeirão Preto: ABR.
- Silva, M. C. V. M. (2004). Estudos de validação e precisão do WZT de Wartegg. (manuscrito não publicado).
- Vane, J. B., & Guarnaccia, V. J. (1989). Personality theory and personality assessment measures: How helpful to the clinician? *Journal of Clinical Psychology*, 45(1), 5-19.
- Van Kolck, O. L. (1984). *Testes projetivos gráficos no diagnóstico psicológico*. São Paulo: EPU.
- Villemor-Amaral, A. E. (2006). Desafios para a cientificidade das técnicas projetivas. Em: A. P. P. Noronha, A. A. A. Santos & F. F. Sisto (Orgs.), *Facetas do fazer em avaliação psicológica* (pp.163-171). São Paulo: Vetor.
- Wartegg, E. (1987). *Teste de Wartegg – WZT – Diagnóstico de camadas. Livros I e II*. São Paulo: Casa do Psicólogo.
- Weiner, I. B. (2000). *Princípios da interpretação do Rorschach*. São Paulo: Casa do Psicólogo.
- Wikipedia – The Free Encyclopedia (2009). Cohen's Kappa. Disponível: <[http://en.wikipedia.org/wiki/cohen%27s\\_kappa](http://en.wikipedia.org/wiki/cohen%27s_kappa)> Acesso: 25.09.2009.

---

**Recebido em maio de 2009**

**Aceito em setembro de 2009**

---

**Irai Cristina Boccato Alves:** Psicóloga; Doutora em Psicologia Escolar e do Desenvolvimento Humano (Universidade de São Paulo/ USP). Professora do curso de Psicologia (Universidade de São Paulo/ USP).

**Augusto Rodrigues Dias:** Psicólogo; Mestre em Psicologia (Universidade São Francisco – USF). Especialista em Educação a Distância (Faculdades SENAC/SC). Professor dos cursos de Psicologia e Gestão de Recursos Humanos (Centro Universitário Paulistano – UniPaulistana) e do curso de Psicologia da Universidade do Grande ABC – UniABC/SP).

**Luís Sérgio Sardinha:** Psicólogo; Mestre em Educação Arte e História da Cultura (Universidade Mackenzie). Doutorando em Psicologia Escolar e do Desenvolvimento Humano (Universidade de São Paulo/ USP).

**Fábio Donini Conti:** Psicólogo; Mestrando em Psicologia Escolar e do Desenvolvimento Humano (Universidade de São Paulo/ USP). Professor do curso de Psicologia (Universidade Guarulhos/UnG).

**Endereço eletrônico para correspondência:** iraicba@usp.br