UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE PSICOLOGIA

ANTONIO CARLOS PACHECO E SILVA NETO

**Fidedignidade do Sistema Compreensivo do Rorschach:
revisão e estudo da estabilidade temporal em
adultos da cidade de São Paulo**

São Paulo

2008

ANTONIO CARLOS PACHECO E SILVA NETO

**Fidedignidade do Sistema Compreensivo do Rorschach:
revisão e estudo da estabilidade temporal em
adultos da cidade de São Paulo**

Tese apresentada ao Instituto de Psicologia da
Universidade de São Paulo para obtenção do título de
Doutor em Psicologia

Área de Concentração: Psicologia Escolar e do
Desenvolvimento Humano

Linha de Pesquisa e orientação: Avaliação Psicológica

Orientadora: Profa. Dra. Eda Marconi Custódio

São Paulo

2008

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTES
TRABALHOS, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA
FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catálogo na publicação
Serviço de Biblioteca e Documentação
Instituto de Psicologia da Universidade de São Paulo

Silva Neto, Antonio Carlos Pacheco e.

Fidedignidade do sistema compreensivo do Rorschach: revisão e estudo
da estabilidade temporal em adultos da cidade de São Paulo / Antonio
Carlos Pacheco e Silva Neto; orientadora Eda Marconi Custódio. -- São
Paulo, 2008.

261 p.

Tese (Doutorado – Programa de Pós-Graduação em Psicologia. Área de
Concentração: Psicologia Escolar e do Desenvolvimento Humano) –
Instituto de Psicologia da Universidade de São Paulo.

1. Teste de Rorschach 2. Precisão do teste 3. Medidas da personalidade
4. Técnicas projetivas 5. Psicometria 6. Avaliação psicológica I.
Título.

BF698.8.R5

FOLHA DE APROVAÇÃO

Antonio Carlos Pacheco e Silva Neto

Fidedignidade do Sistema Compreensivo do Rorschach: revisão e estudo da estabilidade temporal em adultos da cidade de São Paulo

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo para a obtenção do título de doutor em Psicologia

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

DEDICATÓRIA

Aos meus pais, Maria Lúcia e Antonio Carlos.

À Isabela e Beatriz, sobrinhas queridas.

AGRADECIMENTOS

À Profa. Dra. Eda Marconi Custódio, pela orientação segura e atenciosa.

À Profa. Dra. Regina Gattas Fernandes do Nascimento, pelo incentivo constante ao meu envolvimento com o Rorschach. Agradeço sua contribuição ao participar da banca de qualificação, o empréstimo do programa RIAP, utilizado para levantamento dos resultados dos protocolos e, acima de tudo, sua amizade.

À Profa. Dra. Iraí Boccato Alves, pela contribuição oferecida ao participar da banca de qualificação.

À minha família e aos meus amigos, por estarem presentes ao longo do processo de realização do doutorado, fornecendo o apoio tão necessário.

À Lígia Caran, pelas portas abertas. À Leila Caran, pela gentil acolhida.

Aos professores Ricardo Franklin Ferreira, Ines Bianca Loureiro e Rosa Maria Rizzo, coordenadores do curso de Psicologia, pela autorização para a realização deste trabalho.

À Isabel Venâncio, pela divulgação da pesquisa aos funcionários.

À Maria Filomena Porfírio e Vera Lúcia da Silva, pelo eficiente agendamento das entrevistas para aplicação do Rorschach.

À Maria do Carmo Sobral, pelos cafés e chás que tornaram os momentos da coleta de dados menos desgastantes.

Às secretárias da Clínica de Psicologia, pela atenção no agendamento das salas de atendimento.

Aos participantes, sem os quais não haveria pesquisa.

Aos professores do Instituto de Psicologia da USP, pelos cursos de alto nível oferecidos no programa de pós-graduação.

À Maria Olívia Martins Rosa, por me ajudar com os aspectos burocráticos e resolver os entraves.

À Lígia Furusawa, pelas dicas sobre a USP.

Aos bibliotecários do IP/USP pela ajuda com a pesquisa da literatura.

Ao Prof. Dr. Gregory Meyer (Universidade de Toledo, EUA) pelo modelo de pesquisador criativo e rigoroso e de pessoa atenciosa e generosa. Seus artigos e o contato pessoal muito contribuíram para esta pesquisa. O resumo em inglês foi significativamente melhorado com sua ajuda.

Ao Prof. Cato Grønnerød (Universidade de Joensuu, Finlândia), pelo envio de artigos e informações adicionais relativas à sua metanálise da estabilidade temporal do Rorschach.

Ao Prof. Harald Janson (Universidade de Oslo, Noruega), pelo envio do programa Rorschach Research Utilities, utilizado nesta pesquisa.

À Cláudia Goto, pelas sugestões para melhora do texto e paciência com o estresse no final da preparação desta tese.

Ao colega rorschachista Paulo Francisco de Castro, pelas informações sobre legislação dos testes psicológicos.

À Érika Leonardo de Souza, pelo incentivo à realização deste trabalho.

Ao analista do comportamento Roberto Banaco, por presenciar e fazer-acontecer.

RESUMO

SILVA NETO, A. C. P. **Fidedignidade do Sistema Compreensivo do Rorschach: revisão e estudo da estabilidade temporal em adultos da cidade de São Paulo.** 2008. Tese (Doutorado). Instituto de Psicologia. Universidade de São Paulo, 2008.

Neste estudo, avaliamos a estabilidade temporal de 59 variáveis principais do Sistema Compreensivo do Rorschach (SCR). Nossa amostra foi composta por 32 adultos não-pacientes da cidade de São Paulo, participantes voluntários. Eles poderiam receber os resultados da avaliação após a coleta dos dados. Predominaram mulheres (75%), indivíduos solteiros (50%), das classes A (41%) e B (41%), com idades de 19 a 58 anos e média de 13 anos de instrução. Vinte e cinco participantes (78%) eram funcionários da universidade onde se realizou a pesquisa, dois (6%) eram alunos e cinco (16%) eram conhecidos dos funcionários. Utilizamos um delineamento de teste-reteste com intervalo de 3 a 4 meses entre as testagens. Os protocolos foram coletados e codificados pelo autor da pesquisa. Um segundo avaliador codificou, de modo independente, 10 protocolos do teste e 10 protocolos do reteste, sorteados. A fidedignidade inter-codificadores foi substancial ($\kappa > 0,60$) para a maioria das variáveis. A média das correlações de teste-reteste para as 59 variáveis centrais do SCR foi $r = 0,61$, o que corresponde a um nível moderado de estabilidade temporal. Investigamos também a direcionalidade das proporções e a consistência das categorias interpretativas. Entre 44% e 70% dos participantes permaneceu na mesma categoria interpretativa no teste e no reteste. Os resultados de estabilidade temporal encontrados foram menores do que os das pesquisas originais do SCR, mas semelhantes aos de um estudo francês recente. A estabilidade das constelações foi alta, com pelo menos 88% dos participantes mantendo o mesmo status negativo ou positivo no teste e no reteste. Fatores como a restrição de faixa de valores e assimetria não parecem explicar a menor estabilidade temporal encontrada na nossa amostra, mas a distribuição dos valores das variáveis, incluindo a maior frequência de participantes com escores iguais a zero e a presença de escores extremos, pode ter contribuído. Na nossa amostra, o nível de engajamento na tarefa aparentemente foi semelhante ao da amostra normativa da cidade de São Paulo, mas menor do que nas pesquisas originais do SCR. O menor engajamento na tarefa pode ter contribuído para a menor estabilidade temporal encontrada. Fatores relacionados à situação de testagem também parecem ter desempenhado importante papel nos resultados. Os usuários brasileiros devem ter cuidado ao interpretar os resultados obtidos com o SCR, principalmente frente a protocolos com baixo nível de engajamento na tarefa (identificados por R baixo e Λ alto). Mais pesquisas brasileiras com o SCR são necessárias, inclusive para investigar procedimentos que promovam maior engajamento na tarefa, e que assim possam contribuir para uma maior fidedignidade e validade dos resultados obtidos.

Descritores: Teste de Rorschach; Precisão do teste; Medidas da personalidade; Técnicas projetivas; Psicometria; Avaliação Psicológica.

ABSTRACT

SILVA NETO, A. C. P. **Reliability of the Rorschach Comprehensive System: review and study of temporal stability in adults from the city of São Paulo.** 2008. Thesis (Doctoral). Instituto de Psicologia. Universidade de São Paulo, 2008.

We have evaluated the temporal stability of 59 Rorschach Comprehensive System (RCS) core variables. Our sample was composed of 32 nonpatients adults from the city of São Paulo, who volunteered to participate. They could have access to assessment results after the data collection. Participants were predominantly women (75%), single individuals (50%), from socioeconomic classes A (41%) and B (41%), with ages from 19 to 58 and 13 years of education on average. Twenty-five participants (78%) were employees from the university where the research was accomplished, two (6%) were students at the university and five (16%) were acquaintances of the employees. We used a test-retest design with a 3 to 4 months retest interval. All records were collected and codified by the author. A second rater independently coded 10 records from test and 10 records from retest, randomly selected. Interrater reliability was substantial ($\kappa > .60$) for most RCS variables. The mean test-retest correlation for the 59 core variables was $r = .61$, which indicates a moderate level of temporal stability. We also investigated directionality of proportions and categorical consistency. Between 44% and 70% of the participants were in the same interpretive category on test and retest. These results were lower than the original RCS research, but similar to the findings of a recent French study. Temporal stability for constellations was high, with at least 88% of the participants keeping the same negative or positive status on test and retest. Range restriction and skewness did not seem to explain the lower stability found in our sample, but the distributions of values for some variables, with a high frequency of participants with scores of zero and outliers, may have contributed. Task engagement in our sample apparently was similar to that found for the normative sample of adults from São Paulo, but lower than in the original RCS research. Lower task engagement may have contributed to the lower temporal stability. Factors of the testing situation seem to have played an important role in the results also. Brazilian users should take care in interpreting RCS results, mainly when task engagement is low (indicated by low R and high Lambda). More research with RCS in Brazil is needed, also to investigate procedures for obtaining higher task engagement, which may contribute to higher reliability and validity of test results.

Keywords: Rorschach Test; Test Reliability; Personality Measures; Projective Techniques; Psychometrics; Psychological Assessment.

LISTA DE ABREVIATURAS E SIGLAS

ABEP	Associação Brasileira de Empresas de Pesquisa
CFP	Conselho Federal de Psicologia
CRPs	Conselhos Regionais de Psicologia
ICC	Intraclass Correlation Coefficient ou Coeficiente de Correlação Intraclass
GHQ	General Health Questionnaire ou Questionário de Saúde Geral de Goldberg
MMPI	Minnesota Multiphasic Personality Inventory
RIAP	Rorschach Interpretive Assistance Program (programa de computador)
RRU	Rorschach Research Utilities (programa de computador)
SATEPSI	Sistema de Avaliação dos Testes Psicológicos
SCR	Sistema Compreensivo do Rorschach
SPSS	Statistical Package for Social Sciences (programa de computador)

INFORMAÇÃO IMPORTANTE

A Lei 4119, de 27 de agosto de 1962 regulamenta a profissão de Psicólogo e no artigo 13º, 1º parágrafo dispõe que: **“Constitui função privativa do psicólogo o uso de métodos e técnicas para diagnóstico psicológico”** e, portanto, não pode ser exercida por outras pessoas sem essa formação e que não estejam inscritas no Conselho Regional de Psicologia.

O Código de Ética Profissional, publicado em agosto de 2005, estabelece as responsabilidades do Psicólogo:

- Artigo 1º (deveres fundamentais dos Psicólogos), Alínea i, estabelece que cabe a todo psicólogo **“zelar para que a comercialização, aquisição, doação, empréstimo, guarda e forma de divulgação do material privativo do psicólogo, sejam feitas conforme os princípios deste código”**;
- Artigo 2º, estabelecendo o que é vedado ao Psicólogo, conforme alínea h, impede o profissional de **“interferir na validade e fidedignidade de instrumentos e técnicas psicológicas, adulterar seus resultados ou fazer declarações falsas”**;
- Artigo 15º, determina que o Psicólogo deverá **“zelar pelo destino de seus arquivos pessoais, inclusive todo material psicológico (testes)”** e
- Artigo 18º estabelece que: **“o Psicólogo não divulgará, ensinará, cederá, emprestará ou venderá a leigos instrumentos e técnicas psicológicas, que permitam e facilitem o exercício ilegal da profissão”**.

Diante do exposto, decidiu-se que a presente versão fosse reduzida para ser disponibilizada na internet. Embora pequena, tal redução foi necessária, pois houve a supressão de informações que possibilitassem a leigos compreenderem aspectos da aplicação ou interpretação do Método de Rorschach.

Informamos que a versão completa da Tese, disponível para Psicólogos ou estudantes de Psicologia, encontra-se na biblioteca do Instituto de Psicologia da Universidade de São Paulo ou diretamente com o autor.

Antonio Carlos Pacheco e Silva Neto

SUMÁRIO

1. INTRODUÇÃO: A AVALIAÇÃO PSICOLÓGICA E A QUALIDADE DOS TESTES PSICOLÓGICOS	11
2. TESTES PSICOLÓGICOS	13
3. FIDEDIGNIDADE DOS TESTES PSICOLÓGICOS	15
3.1. Fidedignidade de teste-reteste / estabilidade temporal	17
3.2. Fidedignidade de formas paralelas	18
3.3. Consistência interna.....	19
3.4. Fidedignidade inter-avaliadores	19
3.5 Fidedignidade intra-avaliadores	20
4. MÉTODO DAS MANCHAS DE TINTA DE RORSCHACH.....	21
4.1. Natureza do Rorschach.....	21
4.2. Os Sistemas norte-americanos do Rorschach.....	22
4.3. O Sistema Compreensivo do Rorschach	24
5. FIDEDIGNIDADE DO RORSCHACH.....	33
5.1. Consistência interna do Rorschach.....	35
5.2. Concordância das interpretações do SCR.....	36
5.3. Fidedignidade inter-codificadores do SCR	37
5.4. Acurácia das codificações do SCR.....	47
5.5. Estabilidade temporal do SCR.....	51
6. PESQUISAS BRASILEIRAS COM O SCR	104

7. JUSTIFICATIVA	107
8. OBJETIVO	107
9. MÉTODO	108
9.1. Amostra	108
9.2. Instrumentos	109
9.3. Procedimento	109
10. RESULTADOS	114
10.1. Características da amostra obtida	114
10.2. Porcentagem de respostas iguais	116
10.3. Fidedignidade inter-codificadores	118
10.4. Estatísticas descritivas das variáveis do SCR.....	130
10.5. Comparações entre os resultados de teste e reteste	139
10.6. Estabilidade temporal do SCR.....	140
10.7. Problemas estatísticos interferindo nas estimativas da estabilidade temporal	162
10.8. A estabilidade temporal e o engajamento na tarefa.....	168
10.9. Protocolos de participantes que conheciam o examinador.....	170
10.10. Protocolos com poucas respostas	171
11. DISCUSSÃO	174
12. CONCLUSÕES	189
13. REFERÊNCIAS	190
APÊNDICE A – CORRELAÇÃO	205

APÊNDICE B – CÓDIGOS E VARIÁVEIS DO SCR.....	217
APÊNDICE C – PORCENTAGEM DE CONCORDÂNCIA E KAPPA COMO MEDIDAS DA FIDEDIGNIDADE INTER-CODIFICADORES.....	227
APÊNDICE D – DIFERENTES MODOS DE AVALIAR A FIDEDIGNIDADE INTER-CODIFICADORES DO SCR.....	232
APÊNDICE E – COEFICIENTE IOTA.....	236
APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE E INFORMADO	238
APÊNDICE G – ANÁLISE DAS DISCORDÂNCIAS NAS CODIFICAÇÕES DAS RESPOSTAS POPULARES	239
APÊNDICE H – COMPARAÇÃO DOS RESULTADOS NO TESTE E NO RETESTE DE 6 PROTOCOLOS COM MENOS DE 14 RESPOSTAS	250

1. INTRODUÇÃO: A AVALIAÇÃO PSICOLÓGICA E A QUALIDADE DOS TESTES PSICOLÓGICOS

Em nosso dia-a-dia, convivemos com pessoas e avaliamos seus comportamentos e suas características, para tomar decisões sobre como vamos nos relacionar com elas. Esta avaliação cotidiana ou informal baseia-se principalmente nas impressões de quem avalia, muitas vezes então levando a conclusões equivocadas. Por outro lado, algumas situações importantes de tomada de decisão sobre os indivíduos – por exemplo, decisões sobre a sanidade mental e decisões de seleção de pessoal – demandam que a avaliação gere resultados confiáveis e corretos. Assim sendo, existe a necessidade de procedimentos que garantam uma avaliação de qualidade, denominada avaliação psicológica científica (PASQUALI, 2001). Como parte da avaliação psicológica científica, foram desenvolvidos instrumentos de avaliação: os testes psicológicos.

A questão da qualidade dos testes psicológicos em uso no Brasil vem ganhando importância recentemente. Em 1997, o Conselho Federal de Psicologia (CFP) divulgou as **Notas para Orientar a Intervenção do CPF na Questão dos Instrumentos de Avaliação Psicológica** e formou a Câmara Setorial Interinstitucional dos Instrumentos de Avaliação Psicológica, composta por laboratórios de pesquisa, editoras, grandes usuários de testes, entidades da Psicologia e docentes.

A proposta era a construção de uma política nacional para os instrumentos de avaliação psicológica que fosse capaz de combinar os interesses presentes nesse campo, negociando as parcerias que contemplassem o benefício coletivo e que pudesse enfrentar os vários problemas existentes em relação à pesquisa, criação, validação, produção comercial, comercialização, ensino, uso e defesa social da utilização. (CONSELHO FEDERAL DE PSICOLOGIA, 2007)

Em dezembro de 2000, ocorreu o I Fórum Nacional de Avaliação Psicológica, organizado pelo Conselho Federal de Psicologia (CFP) e Conselhos Regionais de Psicologia (CRPs). Em seguida, foi publicada a **resolução 025/2001**, que definiu teste psicológico como método de avaliação privativo do psicólogo e regulamentou sua elaboração, comercialização e uso. Esta resolução foi posteriormente substituída pela **resolução 02/2003**, que versa sobre o mesmo conteúdo. Desde a promulgação destas resoluções, exige-se que os testes psicológicos

apresentem critérios de qualidade para terem sua utilização aprovada no Brasil. Gislene Maia de Macedo, conselheira do CFP, comenta:

Até então não havia uma regulamentação específica sobre o assunto, que pudesse nortear os critérios mínimos que os instrumentos precisam possuir para serem considerados testes psicológicos. Agora, estamos reunindo um quadro de referências dos conceitos, princípios e procedimentos dos testes psicológicos, com base em documentos de critérios internacionais, reconhecidos pela comunidade científica e capazes de atestar a qualidade de um teste. Isso implica em maior credibilidade dos serviços que os psicólogos prestam à sociedade, ao mesmo tempo em que reafirma o compromisso social da psicologia quando busca o aperfeiçoamento de suas técnicas e instrumentos (CONSELHO FEDERAL DE PSICOLOGIA, 2007).

Com a finalidade de avaliar a qualidade dos testes, também foi criada a Comissão Consultiva em Avaliação Psicológica formada por seis psicólogos com ampla experiência e competência na área.

Todo o processo de controle da qualidade dos testes psicológicos culminou no estabelecimento de um Sistema de Avaliação dos Testes Psicológicos (SATEPSI) pelo CFP.

A questão da qualidade dos instrumentos de avaliação psicológica encontra-se em pauta e crescem as pesquisas sobre os testes psicológicos. O presente estudo situa-se neste movimento de pesquisa sobre os instrumentos de avaliação psicológica. Escolhemos avaliar a fidedignidade do método das manchas de tinta de Rorschach, utilizado segundo o Sistema Compreensivo (Exner, 2003).

Apresentamos, a seguir: 1) a definição de teste psicológico e seus parâmetros de qualidade (especialmente a fidedignidade); 2) o método das manchas de tinta de Rorschach e o Sistema Compreensivo; 3) as pesquisas sobre fidedignidade do Sistema Compreensivo do Rorschach (SCR).

2. TESTES PSICOLÓGICOS

Anastasi e Urbina (2000, p. 18) definem: “Um teste psicológico é essencialmente uma medida objetiva e padronizada de uma amostra de comportamento”. Já a American Psychological Association (1974, p. 2) afirma que

Um teste é um caso especial de um procedimento de avaliação. Pode ser concebido como um conjunto de tarefas ou perguntas visando eliciar determinados tipos de comportamentos quando apresentadas sob condições padronizadas, e gerar escores que terão propriedades psicométricas desejáveis, como alta fidedignidade e alta validade.

Notamos nestas definições a menção à objetividade, padronização, fidedignidade e validade, conceitos que explicamos a seguir.

Os testes psicológicos são instrumentos de avaliação que buscam fornecer conclusões seguras sobre os aspectos que avaliam. Neste sentido, a partir dos resultados obtidos, todo e qualquer avaliador deve chegar a conclusões semelhantes. Este aspecto de independência da opinião pessoal do avaliador corresponde à objetividade (CRONBACH, 1996).

Padronização implica um teste ser aplicado sempre do mesmo modo em todos os indivíduos, incluindo utilizar o mesmo material-estímulo e as mesmas instruções para a realização da tarefa, aplicar o mesmo procedimento de coleta dos dados e interpretar os resultados segundo regras pré-determinadas. Nas palavras de Anastasi e Urbina (2000, p. 20):

Padronização implica uniformidade de procedimento na aplicação e na pontuação do teste. Para que os escores obtidos por pessoas diferentes sejam comparáveis, as condições de testagem devem obviamente ser as mesmas para todas. Essa exigência é apenas uma aplicação especial da necessidade de condições controladas em todas as observações científicas. Em uma situação de teste, a única variável independente é frequentemente o indivíduo que está sendo testado. (grifo nosso)

A fidedignidade de um teste refere-se à consistência dos resultados obtidos. Trata-se de comparar o teste consigo mesmo, ou seja, comparar os resultados obtidos em diferentes situações, a fim de verificar o quanto se repetem (ANASTASI; URBINA, 2000).

A validade de um teste refere-se ao grau em que um teste mede o que se propõe medir (ANASTASI; URBINA, 2000). Weiner (1996, p. 206) complementa:

Clínicos e pesquisadores geralmente concordam que conhecer o que um teste mede consiste em identificar como os escores que ele gera correlacionam-se com outras variáveis ou fenômenos que têm implicações para determinados propósitos de avaliação psicológica.

Como o foco do nosso estudo é a fidedignidade, fornecemos a seguir uma apresentação mais detalhada deste aspecto dos testes psicológicos.

3. FIDEDIGNIDADE DOS TESTES PSICOLÓGICOS

Um teste psicológico busca avaliar determinada(s) característica(s) psicológica(s) dos indivíduos. Contudo, condições ambientais, condições físicas e psicológicas do testando, características da interação entre o aplicador e o testando etc. – ou seja, outras características que aquela(s) que o teste quer avaliar – interferem nos resultados obtidos. Assim, os resultados de qualquer teste contêm certo *erro de medida*, devido às influências variadas presentes no momento da testagem¹. Cabe sublinhar que estes erros de medida não decorrem de falhas ao aplicar o teste, mas são devidos a fatores casuais (aleatórios); evidentemente, erros decorrentes de falhas ao aplicar o teste também podem ocorrer, mas não são objeto desta discussão.

O escore obtido num teste é denominado *escore observado*. O escore que seria obtido caso não houvesse erro de medida é denominado *escore verdadeiro*. Segundo Kubiszyn e Borich (2003, p. 322), a teoria psicométrica clássica estabelece que:

$$\text{Escore observado} = \text{Escore verdadeiro} \pm \text{Erro de medida}$$

Evidentemente, o que temos são os escores observados; os escores verdadeiros e os erros de medida são conceitos teóricos, que levam em conta o fato de que os resultados variam em diferentes momentos. De qualquer modo, é importante que o usuário de um teste conheça o quanto os resultados obtidos flutuam de um momento para outro, ou o quanto de erro de medida os escores observados costumam conter, ou o quanto os escores observados encontram-se mais próximos ou menos próximos dos escores verdadeiros. Todas estas questões são correlatas: a pergunta sobre o quanto os resultados flutuam seria uma pergunta sobre a consistência ou *fidedignidade* dos escores; a pergunta sobre a magnitude do erro de medida ou sobre o quanto os escores observados encontram-se próximos aos escores verdadeiros refere-se à *precisão* dos resultados. Na teoria psicométrica, encontramos o termo

¹ Nesta exposição, a referência a medidas indica que estamos tratando de testes que utilizam resultados numéricos (escores), ou seja, testes quantitativos ou psicométricos. De fato, tradicionalmente, esta discussão sobre as características de qualidade dos testes psicológicos costuma utilizar como referência os testes psicométricos.

fidedignidade para se referir a esta característica da consistência dos resultados de um teste. Por vezes, encontramos também os termos *precisão* e *confiabilidade*.

Fidedignidade é o atributo de algo que é digno de fé ou merecedor de crédito (FERREIRA, 1988). O termo *fidedignidade* é utilizado, no contexto dos testes psicológicos, para indicar que os escores, por serem consistentes, merecem crédito – o usuário do teste deve acreditar que os resultados são corretos em vez de equivocados.

Como o usuário do teste pode confiar em seus resultados, podemos também dizer que os resultados apresentam *confiabilidade*, característica daquilo em que se pode ter confiança, definida como: segurança íntima de procedimento; crédito, fé; boa fama; segurança e bom conceito (FERREIRA, 1988). Contudo, segundo Fachel e Camey (2000, p. 160), devemos evitar utilizar *confiabilidade* no contexto dos testes psicológicos, já que

[...] a palavra ‘confiabilidade’ como tradução da palavra inglesa *reliability* não deveria ser usada neste contexto, pois por *confiabilidade* entende-se a área da engenharia que trata da ‘confiança’ que podemos ter em sistemas em geral, incluindo sistemas de segurança.

Precisão significa exatidão de medida, funcionamento sem falhas, regularidade na execução (FERREIRA, 1988). Refere-se, no contexto dos testes psicológicos, aos escores observados serem próximos aos escores verdadeiros. Contudo, Tavares (2003, p. 126) comenta sobre a inadequação do termo *precisão*, associado a uma “visão positivista ingênua de que há *precisão* [exatidão, identidade entre o escore e a característica considerada] na avaliação”, sugerindo a utilização de *fidedignidade* ou *confiabilidade*.

O termo em inglês é *reliability*, derivado de *reliable*, que significa seguro, confiável (HOUAISS; CARDIM, 1996). Neste trabalho utilizamos o termo *fidedignidade*. O importante é sublinhar o significado de consistência dos escores, que justamente traz segurança e permite que o usuário do teste confie nos resultados obtidos. Na definição de Anastasi e Urbina (2000, p. 84):

‘Fidedignidade’ ou ‘confiabilidade’ refere-se à consistência dos escores obtidos pelas mesmas pessoas quando elas são reexaminadas com o mesmo teste em diferentes ocasiões, ou com diferentes conjuntos de itens equivalentes, ou sob outras condições variáveis de exame.

De modo ainda mais direto, Nunnally (1967) afirma que a fidedignidade refere-se ao grau em que os resultados obtidos se repetem.

Em algumas ciências, como a Física, medidas repetidas são utilizadas para se estimar o escore verdadeiro e o erro de medida: um mesmo objeto é medido diversas vezes, a média dos escores é tomada como sendo o escore verdadeiro e o desvio-padrão dos escores é tomado como sendo o erro de medida (neste caso, chamado erro-padrão de medida). Neste contexto, o coeficiente de fidedignidade corresponde à variância dos escores verdadeiros dividida pela variância dos escores observados (CRONBACH, 1996; TRAUB; ROWLEY, 1991).

Em Psicologia não se costuma utilizar diversas medidas repetidas; o procedimento mais utilizado consiste em tomar duas medidas de um mesmo grupo de indivíduos e calcular a correlação² entre os resultados nas duas testagens (ANASTASI; URBINA, 2000, CRONBACH, 1996). Esta correlação é equivalente ao coeficiente de fidedignidade obtido por meio de medidas repetidas (TRAUB; ROWLEY, 1991).

Existem diferentes tipos de fidedignidade, apresentados a seguir.

3.1. Fidedignidade de teste-reteste / estabilidade temporal

A fidedignidade de teste-reteste refere-se ao grau em que os resultados permanecem semelhantes em duas ou mais aplicações do teste. Para se estimar esta fidedignidade, o mesmo teste é aplicado a um mesmo grupo de indivíduos em dois momentos diferentes.

A fidedignidade de teste-reteste mostra a extensão em que os escores em um teste podem ser generalizados para ocasiões diferentes; quanto maior a fidedignidade, menos susceptíveis serão os escores às mudanças aleatórias diárias nas condições dos testandos ou do ambiente de testagem. (ANASTASI; URBINA, 2000, p. 90).

² Uma explicação sobre correlação encontra-se no Apêndice A.

Quando o intervalo entre as testagens é breve, é mais provável que as pessoas não tenham mudado nas características avaliadas, então a correlação de teste-reteste informa sobre o quanto os escores do teste são consistentes em diferentes momentos, o que corresponde à *fidedignidade de teste-reteste* propriamente dita. As mudanças nos resultados de uma situação de testagem para a outra decorrem de erros de medida associados à situação de testagem. Quando o intervalo entre as testagens é longo, pode ter ocorrido mudança na característica que está sendo avaliada. Nesta situação, o termo *estabilidade temporal* vem sendo usado para indicar o quanto a característica avaliada é estável ao longo do tempo. As mudanças nos resultados de uma situação de testagem para a outra decorrem das mudanças na característica avaliada e também dos erros de medida associados a fatores situacionais (GRØNNERØD, 2003). Na literatura, encontramos os termos *fidedignidade de teste-reteste* e *estabilidade temporal* muitas vezes utilizados de modo intercambiável. Neste trabalho, utilizaremos *estabilidade temporal* para nos referirmos aos resultados de estudos com delineamento do tipo teste-reteste, tanto para intervalos breves quanto para intervalos longos.

3.2. Fidedignidade de formas paralelas

A forma paralela de um teste corresponde a um teste equivalente, que gera resultados semelhantes aos obtidos com o teste original. O principal interesse em se desenvolver formas paralelas seria para utilizá-las quando há necessidade de reaplicar o teste, pois a utilização da forma paralela reduz o problema de a memória das respostas dadas na primeira testagem interferir com as respostas fornecidas no reteste.

A avaliação da fidedignidade de formas paralelas consiste, então, em aplicar duas formas supostamente paralelas de um mesmo teste e comparar os resultados obtidos – se as correlações forem suficientemente altas, conclui-se que as formas são paralelas. Quando as duas formas são aplicadas em sucessão imediata, a correlação vai mostrar a fidedignidade entre as formas, não entre as ocasiões. Os erros de medida, neste caso, representam flutuações no desempenho ao se utilizar uma forma ou outra, mas não flutuações ao longo do tempo.

As formas paralelas também podem ser aplicadas com um intervalo de tempo entre a aplicação de cada forma. Este procedimento, que é uma combinação do procedimento

de teste-reteste com o procedimento de formas paralelas, leva em consideração os dois tipos de interferência – o momento da aplicação e a forma do teste – e, portanto, estima a o quanto os resultados são consistentes apesar destes dois fatores (ANASTASI; URBINA, 2000).

3.3. Consistência interna

Este procedimento busca superar um problema que os procedimentos de teste-reteste e de formas paralelas apresentam: a interferência nos resultados devidas à memória das respostas e as diferenças entre as formas, respectivamente. Assim, em vez de aplicar o teste em dois momentos ou utilizar duas formas, a consistência interna é estimada a partir de apenas uma aplicação do teste.

Para se estimar a consistência interna, divide-se o teste em partes e calcula-se a correlação entre os resultados obtidos em cada uma dessas partes. No método das duas metades (split-half), o teste é dividido em duas metades equivalentes e é calculada a correlação entre os escores em cada uma das metades. No método de Kuder-Richardson avalia-se o quanto cada item do teste correlaciona-se com os outros itens. Este tipo de fidedignidade permite avaliar o grau de homogeneidade dos itens do teste (ANASTASI; URBINA, 2000).

3.4. Fidedignidade inter-avaliadores

Este tipo de fidedignidade refere-se ao grau em que diferentes avaliadores pontuam ou classificam as respostas de um teste de modo semelhante (ANASTASI; URBINA, 2000). Uma mesma amostra de protocolos de um teste é pontuada independentemente por dois examinadores e avalia-se o grau de concordância dessas pontuações (esta é a fidedignidade do avaliador em relação às pontuações, mas também existe a fidedignidade do avaliador em relação à interpretação dos resultados: comparam-se as interpretações de diferentes avaliadores para um mesmo conjunto de protocolos).

Um teste psicológico bem construído deve apresentar regras claras para a pontuação das respostas, garantindo, com isto, que diferentes avaliadores as pontuem de modo

semelhante. São justamente os estudos de fidedignidade inter-avaliadores que indicam se as regras de pontuação são ou não são claras.

3.5 Fidedignidade intra-avaliadores

Este tipo de fidedignidade refere-se ao grau em que um mesmo avaliador pontua as respostas de um teste de modo semelhante em diferentes momentos.

4. MÉTODO DAS MANCHAS DE TINTA DE RORSCHACH

O método das manchas de tinta de Rorschach³ é um instrumento de avaliação da personalidade. Foi desenvolvido pelo psiquiatra suíço Hermann Rorschach e consiste em um conjunto padronizado de manchas de tinta, que são apresentadas aos testandos para que digam com o que elas se parecem.

Respostas a manchas de tinta vinham sendo estudadas por diversos pesquisadores, mas os trabalhos detinham-se nos conteúdos das respostas, utilizados para a avaliação da imaginação. A originalidade de Hermann Rorschach foi estudar as respostas considerando outras características além de seu conteúdo. Ele desenvolveu um sistema de classificação da estrutura das respostas às manchas de tinta: um conjunto de códigos era usado para representar a área da mancha utilizada [*localização*], outro conjunto de códigos para as principais características da mancha responsáveis pela imagem percebida [*determinante*] e outros códigos para o tipo de objeto percebido [*conteúdo*]. Rorschach descobriu que esses elementos estruturais das respostas relacionavam-se com características de personalidade. Em 1921, publicou o **Psicodiagnóstico**, reportando os resultados de suas pesquisas (EXNER, 2003).

4.1. Natureza do Rorschach

[Trecho suprimido na versão online]

³ Neste trabalho, nos referimos ao método das manchas de tinta de Rorschach como método de Rorschach, técnica de Rorschach ou simplesmente Rorschach.

[Trecho suprimido na versão online]

4

4.2. Os Sistemas norte-americanos do Rorschach

Um ano após a publicação do **Psicodiagnóstico**, Hermann Rorschach veio a falecer. Após sua morte, alguns de seus colegas continuaram a utilizar o método das manchas de tinta. A técnica foi divulgada na Europa e, posteriormente, nos Estados Unidos.

⁴ [Nota de rodapé suprimida na versão online]

Exner (2003) conta a história do Rorschach nos Estados Unidos:

Ainda no início da década de 1920, David Levy, psicólogo norte-americano, estudou na Suíça com Oberholzer, colega de Hermann Rorschach, quando então aprendeu o método das manchas de tinta. Por meio de Levy, Samuel Beck conheceu a técnica e começou a realizar as primeiras pesquisas nos Estados Unidos. Posteriormente, sua colega Marguerite Hertz também se interessou pelo método e realizou suas pesquisas. Beck e Hertz publicaram suas dissertações sobre o Rorschach em 1932.

Com a Primeira Guerra Mundial, os Estados Unidos receberam muitos refugiados europeus, dentre eles Bruno Klopfer. Em 1934, seus alunos da Universidade de Colúmbia solicitaram seminários sobre o Rorschach. A partir desses seminários, a carreira de Klopfer dirigiu-se cada vez mais para o método das manchas de tinta e ele passou a desenvolver sua própria abordagem da técnica.

Dentre os participantes do primeiro seminário de Klopfer estava Zygmunt Piotrowski, estudante polonês que realizava seu pós-doutorado nos Estados Unidos e que, com a invasão de seu país, decidiu permanecer na América. Em 1950, Piotrowski publicou **Perceptanalysis**, sua obra sobre o Rorschach.

David Rapaport também fugiu da Europa para os Estados Unidos, chegando em 1938, após completar seu doutorado na Hungria. Sua orientação era psicanalítica e seu maior interesse era o estudo dos processos de pensamento. O Rorschach e outros testes psicológicos passaram a ser utilizados neste estudo e Rapaport desenvolveu sua própria abordagem do método das manchas de tinta. Posteriormente, esta abordagem foi enriquecida com a contribuição de Roy Schafer, que em 1954 publicou **Psychoanalytic Interpretation in Rorschach Testing**.

Wood et al. (2003) situam o desenvolvimento do Rorschach no contexto da Psicologia norte-americana:

Nas décadas de 1920 e 1930, começaram a ser construídos os primeiros testes de personalidade, basicamente inventários, instrumentos que avaliavam traços de personalidade a partir de respostas do indivíduo a perguntas sobre suas características pessoais ou modos

típicos de agir. Os inventários apresentam limitações: sua validade depende de o indivíduo possuir um bom auto-conhecimento e estar disposto a responder honestamente às perguntas. Além disso, os primeiros inventários avaliavam poucas características, não fornecendo uma avaliação mais ampla da personalidade, e não ajudavam a identificar problemas psicológicos.

Neste cenário, desenvolveram-se outras técnicas de avaliação da personalidade, caracterizadas por instruções que permitem grande liberdade de resposta e por frequentemente utilizarem estímulos ambíguos. [Trecho suprimido na versão online]

O crescimento do Rorschach nos Estados Unidos nas décadas de 1940 e 1950 deveu-se à publicação de numerosos livros e artigos sobre o teste, ao entusiasmo de seus promotores, à influência da psicanálise – [trecho suprimido na versão online]

– e ao aumento da demanda por serviços psicológicos e necessidade de instrumentos para se detectar transtornos mentais.

4.3. O Sistema Compreensivo do Rorschach

Hermann Rorschach reconhecia que seu método das manchas de tinta ainda não estava completo. Assim sendo, outros autores modificaram e ampliaram a técnica para integrar novas conceituações e resultados de pesquisas. Estas revisões e acréscimos produziram diferentes modos de usar o instrumento. Nos Estados Unidos, desenvolveram-se cinco principais sistemas de aplicação, codificação e interpretação do Rorschach – os sistemas de Beck, Klopfer, Hertz, Piotrowski e Rapaport/Schaffer (EXNER, 2003).

A partir da década de 1960, John Ernest Exner Jr., psicólogo norte-americano, decidiu estudar estes cinco principais sistemas do Rorschach utilizados nos Estados Unidos. Buscava inicialmente compará-los e ver qual deles apresentaria uma melhor fundamentação.

Esta pesquisa resultou em seu livro **The Rorschach Systems**, publicado em 1969. Exner concluiu que, de modo geral, o Rorschach mostrava-se útil para gerar resultados relevantes para o trabalho clínico. Todos os sistemas apresentavam alguns aspectos bem fundamentados e outros aspectos que demandavam melhor fundamentação, pois não haviam sido pesquisados ou as pesquisas mostravam resultados contraditórios. Provavelmente, a conclusão mais importante foi que as diferenças entre os sistemas eram tantas que não se poderia falar em um único teste de Rorschach, mas sim seriam cinco diferentes testes, cada uma deles utilizando determinado enquadre (setting) e instruções fornecidas, códigos (símbolos) próprios para a classificação das respostas – com diferentes critérios de codificação nos diferentes sistemas – e postulados para a interpretação dos resultados (EXNER, 2003).

Exner também realizou pesquisa com usuários do Rorschach, encontrando que predominava a utilização dos sistemas de Beck e Klopfer. Vinte e dois por cento dos psicólogos que responderam à pesquisa disseram que haviam abandonado a codificação das respostas, realizando a avaliação apenas interpretando o conteúdo. Setenta e cinco por cento dos psicólogos que responderam à pesquisa personalizavam a codificação, combinando elementos de diferentes sistemas.

Não tendo encontrado um sistema que fosse melhor que os demais e tendo notado que a maioria dos usuários não seguia um único sistema, Exner decidiu integrar os diferentes sistemas. Em 1970, realizou uma revisão sistemática da literatura do Rorschach, encontrando que muitos estudos apresentavam problemas em seu delineamento e nas análises estatísticas. Seiscentos estudos adequados foram localizados. Com esta revisão da literatura, Exner observou que alguns estudos apresentavam resultados contraditórios em relação a outros, sendo um dos problemas a utilização de amostras pequenas – havia então a necessidade de replicações com amostras maiores. Exner e seus colaboradores realizaram diversas pesquisas, buscando assim suprir a fundamentação que faltava ao método das manchas de tinta. Em 1974, Exner publicou **The Rorschach: A Comprehensive System**. O Sistema Compreensivo

do Rorschach (SCR) tem este nome por abarcar e integrar os aspectos mais bem fundamentados dos cinco sistemas norte-americanos do Rorschach⁵.

Discutindo a importância do Sistema Compreensivo do Rorschach, Weiner (1998b, p. 218) coloca a seguinte pergunta:

Se os avaliadores sentiam-se satisfeitos com seus modos individuais de utilizar o Rorschach e eram capazes de formar impressões clínicas relevantes das pessoas examinadas, o que importava se poucos utilizavam o teste do mesmo modo?

Justifica uma utilização uniforme do Rorschach argumentando que a avaliação subjetiva das respostas por parte do avaliador seria insuficiente para uma psicologia clínica científica e respeitável:

Embora impressões subjetivas da singularidade individual enriqueçam a prática clínica e ampliem a capacidade do psicólogo de ser útil [aos clientes e à sociedade], a subjetividade sozinha é insuficiente para sustentar a psicologia clínica e o psicodiagnóstico como ciência e profissão (Weiner, 1983, 1991)⁶. Para evitar afundar-se na subjetividade e ser equivocadamente tomados por artistas ou filósofos, os psicólogos que trabalham com psicodiagnóstico necessitam introduzir substancial objetividade em suas testagens projetivas (WEINER, 1998b, p. 218).

Exner (1999a, p. 159) comenta:

Desde sua publicação em 1921, um amplo segmento da história do Rorschach, principalmente do final da década de 1940 até a metade da década de 1970, foi frequentemente marcado por considerável crítica baseada em argumentos de que evidências convincentes de fidedignidade e/ou validade do teste não haviam sido estabelecidas. [...]

⁵ Os símbolos usados na codificação das respostas, os critérios de codificação e as variáveis do Sumário Estrutural do SCR encontram-se no Apêndice B. [Suprimido na versão online]

⁶ WEINER, I. B. The future of psychodiagnosis revisited. **Journal of Personality Assessment**, 47: 451-461, 1983.

WEINER, I. B. Theoretical foundations of clinical psychology. In: **The clinical psychology handbook**. HERSEN, M.; KASDIN, E. A.; BELLACK, A. S. (editores). 2nd ed. New York: Pergamon, 1991. p. 26-44.

A amplitude das críticas fez com que alguns autores desautorizassem o uso da palavra *teste* ao discutir o Rorschach e adotassem nomenclaturas como técnica ou método (Klopfer & Kelly, 1942)⁷, ou mesmo o descrevessem como uma forma de entrevista não-estruturada (Zubin, Eron & Schumer, 1965)⁸. Muitos defensores do teste contra-argumentavam que princípios específicos de medida poderiam não ser aplicáveis ao estudo do Rorschach porque seus dados somente podem ser compreendidos num enquadre interpretativo global clínico (Holt, 1958, 1970)⁹. [...] Também é importante notar que muita pesquisa inicial do Rorschach, especialmente durante as décadas de 1940 e 1950, foi marcada por delineamentos que eram excessivamente simplistas ou confusos e por abordagens de análise dos dados que seriam julgadas como muito primitivas para os padrões contemporâneos. (grifo do autor)

O Rorschach encontrava-se, e ainda encontra-se, tensionado entre duas tradições da Psicologia. De um lado, uma tradição baseada em pesquisas empíricas (objetividade) e busca de leis gerais (abordagem nomotética); de outro, uma tradição baseada na construção de compreensão (subjetividade) da singularidade individual (abordagem idiográfica).

Autores como Beck e Hertz enfatizavam a necessidade da fundamentação psicométrica e empírica do Rorschach. Já Klopfer propunha que o Rorschach fosse validado clinicamente, ou seja, os aspectos psicométricos não seriam tão importantes, já que a interpretação teria sua validade comprovada quando convergisse com dados obtidos numa avaliação clínica (estudo de caso); além disso, a interpretação do Rorschach seria um processo mais artístico do que propriamente científico, requerendo talento e experiência do avaliador mais do que o seguimento de um procedimento uniforme e fundamentado em pesquisas (WOOD et al., 2003).

⁷ KLOPFER, B.; KELLEY, D. **The Rorschach Technique**. Yonker, NY: World Book, 1942.

⁸ ZUBIN, J.; ERON, L. D.; SCHUMER, F. **An experimental approach to projective techniques**. New York: Wiley, 1965.

⁹ HOLT, R. Clinical and statistical prediction: A reformulation and some new data. **Journal of Abnormal and Social Psychology**, v. 56, p. 1-12, 1958.

HOLT, R. Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? **American Psychologist**, v. 25, p. 337-349, 1970.

Em nossa opinião, esta estratégia de validação do Rorschach proposta por Klopfer apresenta problemas: o procedimento para se chegar à interpretação global é subjetivo, ou seja, depende do avaliador mais do que de regras objetivas para a interpretação dos resultados; existe a dificuldade de se determinar o que seria convergência ou divergência entre as conclusões pelo Rorschach e as conclusões do estudo de caso, principalmente porque esses dois procedimentos fornecem diversas informações sobre características do indivíduo, muitas vezes cada um deles informando sobre aspectos diferentes; frequentemente um mesmo avaliador realizava a interpretação do Rorschach e o estudo de caso, não havendo independência entre o instrumento a ser validado e o critério de comparação (problema de contaminação do critério). Conforme discutido por Arkes (1981) e Garb (1998), devemos considerar também que o raciocínio clínico está sujeito a diversas falhas, o que contribui para desaconselhar esta abordagem à validação.

No final da década de 1940, Cronbach publicou seu importante artigo sobre os métodos estatísticos aplicados ao Rorschach. Já naquela época, afirmou que

[...] pesquisas estatísticas com o teste de Rorschach não são apenas justificáveis, mas indispensáveis. A flexibilidade do pensamento clínico cria hipóteses excelentes, mas estas hipóteses só podem ser tomadas como verdadeiras a partir de estudos controlados. [...] Mesmo que o clínico estudando uma pessoa não utilize estatísticas, ele emprega generalizações sobre o teste que devem estar embasadas em evidências geradas cientificamente. (CRONBACH, 1949, p. 393)

Outra situação que trouxe confusão para a compreensão do Rorschach foi sua inclusão dentre os (assim chamados) métodos projetivos.

[Trecho suprimido na versão online]

[Página suprimida na versão online]

Com esta denominação *métodos projetivos* surgiu uma tendência equivocada de separar os testes psicológicos em dois grupos distintos: os testes objetivos e os testes (ou métodos) projetivos. Meyer (1997a, p. 298) comenta:

O precedente histórico nos deixou com os termos objetivo para descrever o MMPI¹⁰ e projetivo para descrever o Rorschach. Esses termos são errôneos e obscurecem em vez de esclarecer. Em particular, implicam que o MMPI é factual e preciso, enquanto sugerem que o Rorschach é subjetivo e não-factual. Na realidade, ambos devem ser considerados medidas factuais, mas imprecisas, da personalidade, cada um dos quais apto para quantificar certos tipos de construtos da personalidade (grifos nossos).

Em 2006, Meyer foi ainda mais enfático, afirmando a necessidade de se suspender a utilização dos termos *objetivo* e *projetivo* para descrever diferentes tipos de testes psicológicos. Propôs como substitutos para o termo *testes projetivos*: tarefas de execução (performance tasks), tarefas comportamentais (behavior tasks), métodos construtivos (constructive methods), medidas de resposta livre (free response measures), testes expressivos de personalidade (expressive personality tests).

O Sistema Compreensivo buscou superar uma série de equívocos e dicotomias (ainda presentes) em relação ao Rorschach. Trata-se de um instrumento cuja avaliação inclui aspectos objetivos e subjetivos, estruturais e temáticos, nomotéticos e idiográficos, e que pode ser considerado um teste psicológico e também um método para se obter diversas informações sobre o funcionamento psicológico do indivíduo (WEINER, 1998a).

¹⁰ MMPI é a abreviação de Minnesota Multiphasic Personality Inventory.

Em relação ao método das manchas de tinta de Rorschach (em inglês, Rorschach Inkblot Method - RIM), Weiner (1998a, p. 55) comenta que

O Sistema Compreensivo do Rorschach, desenvolvido por Exner (J. E. Exner, 1991, 1993; J. E. Exner e Weiner, 1995)¹¹ está fundamentado em três pilares: aplicação padronizada, codificação objetiva e fidedigna, e uma base de dados normativos representativa. A objetividade sistemática introduzida no teste de Rorschach por J. E. Exner (1974)¹², após mais de 50 anos de utilização altamente variável e freqüentemente caótica do instrumento, tornou possível [...] que os dados de pesquisa acumulados demonstrassem o rigor psicométrico do método das manchas de tinta de Rorschach (RIM). Além de ampliar a utilização proveitosa do RIM em estudos empíricos, o advento do Sistema compreensivo permitiu que os profissionais que utilizam o Rorschach formulassem conclusões clínicas com maior segurança do que antes era possível e que as comunicassem com maior eficiência (Weiner, 1997a, 1998c)¹³.

[...] Os dados do Rorschach obtidos, codificados e apresentados de acordo com as diretrizes do Sistema Compreensivo fornecem aos clínicos um conjunto de códigos e índices fidedignos e fáceis de observar, que dispõem de correlatos – empiricamente demonstrados e conceitualmente significativos – com aspectos do funcionamento da personalidade.

O SCR representa “o esforço mais ambicioso e frutífero já realizado de colocar o Rorschach em bases psicometricamente sólidas” e “uma das maiores contribuições do trabalho de Exner é o oferecimento de um sistema Rorschach uniforme que permite a

¹¹ EXNER, J. E. Jr. **The Rorschach: A Comprehensive System**. Volume 2: Interpretation. 2nd ed. New York: Wiley, 1991.

EXNER, J. E. Jr. **The Rorschach: A Comprehensive System**. Volume 1: Basic Foundations. 3rd ed. New York: Wiley, 1993.

EXNER, J. E. Jr.; WEINER, I. B. **The Rorschach: A Comprehensive System**. Volume 3: Assessment of children and adolescents. 2nd ed. New York: Wiley, 1995.

¹² EXNER, J. E. Jr. **The Rorschach: A Comprehensive System**. New York: Wiley, 1974.

¹³ WEINER, I. B. Current status of the Rorschach Inkblot Method. **Journal of Personality Assessment**, v. 68, p. 5-19, 1997.

WEINER, I. B. Rorschach Inkblot Method. In: **The use of psychological testing in treatment planning and outcome evaluation**. MARUIISH, M. (editor). 2nd. ed. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

comparabilidade entre os achados de pesquisas de diferentes investigadores”. (ANASTASI; URBINA, 2000, p. 341-342).

O Sistema Compreensivo foi apresentado em 1974 e desde então se tornou a abordagem mais utilizada nos Estados Unidos: na metade da década de 1980, aproximadamente 35% dos psicólogos que usavam o Rorschach o adotaram; no final daquela década, o número era de 40%, aproximadamente o dobro do número de usuários dos sistemas Beck e Klopfer; no final de 1990, o SCR era ensinado em aproximadamente 75% dos cursos de graduação (WOOD et al, 2003). Piotrowski (1996) encontrou que o Sistema Compreensivo é o mais utilizado em pesquisas com o Rorschach nos Estados Unidos. O SCR já é dominante na maioria dos outros países ou está em processo de adoção pelos psicólogos que realizam psicodiagnóstico com o Rorschach, conforme pode ser observado pelas publicações no periódico internacional **Rorschachiana** e nos anais dos congressos da Sociedade Internacional de Rorschach (WEINER, 1998b).

Como o foco de nosso trabalho é a fidedignidade do SCR – mais especificamente a estabilidade temporal –, a seguir apresentamos os estudos de fidedignidade do Sistema Compreensivo.

5. FIDEDIGNIDADE DO RORSCHACH

Propostas para a avaliação da fidedignidade do Rorschach

Holzberg (1977) menciona quatro atitudes frente à fidedignidade do Rorschach:

- 1) A fidedignidade não seria uma questão adequada ao Rorschach, em função do modo como este é usado na prática clínica – trata-se de um método qualitativo e informal, semelhante a uma entrevista aberta.
- 2) O importante seria estabelecer a validade do Rorschach, já que esta, uma vez comprovada, implicaria em fidedignidade adequada. Segundo esta proposta, havendo validade, as oscilações nos resultados não demonstrariam falta de consistência do instrumento, mas sim oscilações na personalidade.
- 3) A fidedignidade deve ser investigada, mas de um modo que respeite a especificidade do Rorschach, técnica que não avalia apenas um aspecto ou alguns aspectos isolados da personalidade, mas que considera diversos aspectos em conjunto e gera informações complexas sobre diversas características do funcionamento psicológico.
- 4) A fidedignidade deveria ser estudada como em qualquer teste.

A primeira atitude propõe uma utilização mais qualitativa e subjetiva do Rorschach. Cabe dizer que, mesmo ao se utilizar o Rorschach deste modo mais informal e qualitativo, é viável e indicado realizar estudos sobre a fidedignidade. Podemos, por exemplo, comparar as interpretações produzidas por diferentes avaliadores para ver se concordam – este tipo de estudo será comentado no item concordância das interpretações, a seguir.

A segunda atitude apresenta uma proposta facilmente questionável, pois desconsidera o fato de que todo instrumento de avaliação apresenta erros de medida: os resultados flutuam devido a fatores aleatórios, e a magnitude destes erros deve ser conhecida para uma utilização mais bem fundamentada dos resultados.

Em relação à terceira atitude para lidar com a singularidade do Rorschach, considerando seu caráter holístico e seus aspectos quantitativos e qualitativos, foram propostas diversas abordagens: avaliar a fidedignidade do Sumário Estrutural; usar escalas de

pontuação (rating scales) para avaliar aspectos da interpretação e então comparar as interpretações de diferentes avaliadores e estabelecer a fidedignidade; usar técnicas de pareamento (matching techniques), ou seja, fornecer protocolos de teste e reteste de diversos indivíduos e solicitar que juízes estabeleçam os pares de protocolos que seriam do mesmo indivíduo, ou solicitar que os juizes pareiem as interpretações formuladas por diferentes avaliadores.

O Sistema Compreensivo desenvolve-se buscando demonstrar a fidedignidade psicométrica do método das manchas de tinta – seguindo, então, a quarta atitude mencionada por Holzberg (1977).

O Sistema Compreensivo e a fidedignidade do Rorschach

Nesta parte, revisamos os estudos da fidedignidade do Sistema Compreensivo do Rorschach. Existem estudos com outros sistemas, a maioria deles mais antigos, sendo mais difícil a obtenção dos periódicos e livros em que se encontram. Em função desta maior dificuldade de acesso aos estudos e também por uma questão de economia, optamos por revisar apenas os estudos com o SCR.

Antes de apresentarmos os estudos de fidedignidade do SCR, cabe mencionar um estudo de Parker, Hanson e Hunsley (1988). Trata-se de uma metanálise de artigos publicados no **Journal of Personality Assessment** e no **Journal of Clinical Psychology**, entre 1970 e 1981, sobre os testes MMPI, Rorschach e WAIS. A metanálise incluiu pesquisas com diversos sistemas do Rorschach e os autores concluíram que, de modo geral, a fidedignidade dos três testes é aceitável e aproximadamente equivalente. A fidedignidade (consistência interna e concordância inter-avaliadores) do Rorschach foi estimada em 0,86 e a estabilidade temporal em 0,85.

Também cabe mencionar o único estudo com informações sobre fidedignidade intra-avaliadores que encontramos. Voigt e Dana (1964) utilizaram 10 estudantes de psicologia clínica, que codificaram 150 respostas do Rorschach em duas ocasiões. A fidedignidade intra-avaliadores, ou consistência entre as duas codificações realizadas pelos avaliadores, foi estimada em 73%.

5.1. Consistência interna do Rorschach

Existem pesquisas sobre a consistência interna com outros sistemas do Rorschach. Contudo, Exner (1978, p. 63) comenta que

[...] a demonstração da fidedignidade das duas metades para o teste [de Rorschach] coloca diversos problemas [...]. O problema com a fidedignidade das duas metades deve-se ao fato de que os estímulos do teste são diferentes entre si. Eles são diferentes em termos dos níveis de dificuldade e complexidade, e são provavelmente diferentes em termos dos diferentes tipos de respostas que induzem ('stimulus pull'). Entretanto, alguns estudos de fidedignidade das duas metades produziram resultados que estão longe de serem pouco respeitáveis. [...]

Embora a técnica das duas metades elimine a desvantagem da exposição prévia às manchas e, de fato, controle o efeito da prática, provavelmente não é uma boa medida da 'solidez' do Rorschach.

Como este tipo de fidedignidade não parece adequado ao Rorschach, o Sistema Compreensivo não apresenta estudos de consistência interna.

Localizamos apenas um estudo com o título de consistência interna, mas que não avalia este aspecto segundo a definição tradicional: Williamson (1988) estudou a consistência interna de grupos de variáveis utilizadas na interpretação do SCR. Os grupos de variáveis escolhidos avaliam os seguintes aspectos: iniciativa e complexidade cognitiva; rigidez e disfunção cognitiva; convencionalidade e acurácia perceptiva; características da ideação; características do afeto; características da auto-imagem; percepção interpessoal; Constelação de Suicídio; Índice de Esquizofrenia; Índice de Depressão. Os resultados indicaram que alguns dos grupos de variáveis apresentam boa consistência interna, enquanto outros apresentam baixa consistência interna. Em outras palavras, alguns grupos de variáveis apresentam índices que se referem a um mesmo tipo de característica do funcionamento psicológico (por exemplo, *Zf*, *W* e *DQ+* apresentam alta consistência interna, pois avaliam o esforço de organização perceptiva), enquanto outros grupos de variáveis apresentam índices que avaliam características independentes uma das outras (por exemplo, *Afr* e *FC : CF+C* referem-se a dois aspectos distintos e independentes da afetividade, a responsividade às situações afetivas e o grau de modulação afetiva, respectivamente).

5.2. Concordância das interpretações do SCR

Meyer, Mihura e Smith (2005) publicaram um estudo sobre a concordância das interpretações. Utilizaram 55 protocolos, a partir dos quais avaliadores deveriam pontuar 29 características de personalidade pré-selecionadas, ou organizar estas 29 características numa ordem para cada paciente (técnica Q). A pesquisa primeiramente utilizou os 3 autores como avaliadores; depois um conjunto de avaliadores, com cada um interpretando uma parte dos protocolos; por fim, 8 avaliadores interpretando os mesmos 55 protocolos. Os resultados encontrados indicam que o nível de concordância entre diferentes avaliadores que utilizam o SCR é semelhante ao nível de concordância entre avaliadores utilizando outros testes em Psicologia e Medicina. Concluíram que o Rorschach pode ser interpretado de modo fidedigno.

Pesquisas anteriores ao Sistema Compreensivo utilizaram o procedimento de pareamento para avaliar a concordância das interpretações. Esta metodologia apresenta alguns problemas: certas interpretações podem ser pareadas baseando-se em aspectos superficiais, quando na verdade diferem em significado; afirmações genéricas podem ser pareadas mais facilmente, dando impressão de maior fidedignidade do que há de fato. A heterogeneidade ou homogeneidade dos protocolos a serem pareados, o grau de perícia do avaliador e o tempo disponível para realizar a tarefa são fatores que influenciam o pareamento e, consequentemente, os resultados de fidedignidade obtidos a partir de estudos deste tipo, que avaliam não apenas a fidedignidade das interpretações, mas também a fidedignidade do avaliador que faz o pareamento (ZUBIN; ERON; SCHUMER, 1965). Holzberg (1977) comenta ainda que cada intérprete pode enfatizar determinados aspectos em vez de outros, e o modo como cada avaliador procede para construir a interpretação (por exemplo, utilizando sinais ou levantando hipóteses e validando-as) também produz maior variedade de interpretações, o que muitas vezes dificulta o pareamento. O estudo de Meyer, Mihura e Smith (2005) buscou contornar estes problemas, utilizando 29 características de personalidade pré-selecionadas, que deveriam ser pontuadas por todos os avaliadores.

5.3. Fidedignidade inter-codificadores do SCR

Conforme exposto anteriormente, a fidedignidade inter-avaliadores refere-se ao grau em que as pontuações realizadas por diferentes avaliadores são semelhantes. No Rorschach, não se trata propriamente de pontuar as respostas, mas sim de codificá-las, ou seja, atribuir a elas um conjunto de símbolos que resumem suas características estruturais. Usamos, então, as denominações fidedignidade inter-codificadores ou fidedignidade das codificações.

Uma das preocupações ao desenvolver o Sistema Compreensivo foi criar critérios de codificação claros, que então apresentassem adequada fidedignidade inter-codificadores:

[...] nenhuma categoria de codificação foi incluída no 'novo' sistema [Sistema Compreensivo] a menos que um nível mínimo de 0,85 pudesse ser facilmente alcançado para grupos de 10 a 15 codificadores ao longo de pelo menos 10 a 20 protocolos nos quais a categoria-alvo ocorresse com uma frequência razoável. Isto causou uma rejeição inicial de muitas categorias de classificação aparentemente úteis, que foram posteriormente incluídas no sistema, com seus critérios de codificação revisados. (EXNER, 2003, p.25).

Em 1991, a questão da fidedignidade das codificações recebeu um destaque importante no **Journal of Personality Assessment**, principal periódico de estudos sobre o Rorschach. Irving Weiner, na época o editor deste periódico, publicou uma nota determinando que, para um artigo ser aceito para publicação, deveria apresentar comprovação mínima da fidedignidade das codificações: pelo menos 20 protocolos deveriam ser codificados por mais de um avaliador e a porcentagem de concordância das codificações por segmentos (localização, determinantes, qualidade formal, conteúdos etc.) deveria ser de pelo menos 80% (WEINER, 1991).

Hellreich (1995, apud Nakata, 2000)¹⁴ utilizou estudantes de Psicologia com diferentes níveis de experiência para investigar a fidedignidade inter-codificadores. Estes avaliadores codificaram respostas com diversos graus de dificuldade. A autora encontrou que o nível de treinamento dos avaliadores afetou a fidedignidade das codificações e que as menores concordâncias das codificações ocorreram para os determinantes, qualidade formal e códigos especiais. Além disso, o nível de complexidade das respostas interferiu, com respostas mais complexas apresentando menor fidedignidade inter-codificadores.

Um dos problemas desta pesquisa foi a utilização de respostas do Workbook¹⁵. Os avaliadores provavelmente conheciam muitas destas respostas, por utilizarem o Workbook para aprender a codificar. Também não se sabe se o conjunto de respostas selecionadas para estudo da fidedignidade das codificações é semelhante a protocolos reais considerados em sua totalidade. Além disso, os codificadores eram alunos de uma mesma faculdade, o que deve ter contribuído para maior concordância das codificações (NAKATA, 2000).

Wood, Nezworski e Stejskal (1996a) publicaram **The Comprehensive System for the Rorschach: a critical examination**, artigo no qual revisaram o Sistema Compreensivo e apontaram alguns problemas, sendo um deles a fidedignidade inter-codificadores. Segundo estes autores, a porcentagem de concordância não seria um índice adequado para se estimar a fidedignidade. Haveria a necessidade de se controlar o problema da concordância obtida por mero acaso e, para tanto, o coeficiente kappa seria a medida indicada¹⁶. Além disso, o Sistema Compreensivo apresentaria apenas informações sobre a fidedignidade de códigos isolados, faltando informações sobre a fidedignidade das variáveis do Sumário Estrutural. Outro ponto destacado é que os resultados provêm de pesquisas com

¹⁴ HELLREICH, C. L. **Rorschach intercoder reliability**. Tese de doutorado. California School of Professional Psychology, 2000.

¹⁵ Nos Estados Unidos, existem os livros do Sistema Compreensivo (**The Rorschach: A Comprehensive System**), atualmente em sua quarta edição (Exner, 2003), e o **Rorschach Workbook for the Comprehensive System**, apostila que resume as orientações para a aplicação do Rorschach e para a codificação das respostas, bem como apresenta exercícios de codificação com gabarito de correção, atualmente em sua quarta edição. No Brasil, o **Manual de classificação do Rorschach para o Sistema Compreensivo** (Exner, 1999b) corresponde à tradução do Workbook (terceira edição).

¹⁶ Uma explicação sobre o problema da concordância obtida ao acaso e o coeficiente kappa encontra-se no Apêndice C.

avaliadores extremamente bem treinados e com maior controle experimental do que na prática clínica cotidiana; faltaria conhecer a concordância entre psicólogos clínicos comuns em seu trabalho cotidiano – o que os autores denominaram fidedignidade inter-avaliadores de campo (field interrater reliability).

Em resposta às críticas, Exner (1996b) explicou que o que vinha sendo denominado porcentagem de concordância (percentage of agreement), na verdade corresponde à porcentagem de codificações corretas, devendo então ser denominado porcentagem correta (percentage correct). Neste procedimento, mais bem explicado numa comunicação pessoal de Exner reportada em McGrath, Pogge e Strokes (2005), três ou mais avaliadores da diretoria de Rorschach Workshops (apelido da Rorschach Research Foundation, organização fundada por Exner) primeiramente estabelecem um consenso sobre as codificações corretas de um conjunto de respostas. Depois, outros avaliadores codificam estas respostas e é realizada a comparação de suas codificações com as codificações corretas. O conjunto de respostas apresentaria diversas ocorrências dos códigos em estudo, evitando-se assim os problemas de códigos com baixa taxa de ocorrência e alta concordância obtida por mero acaso. Na visão de Exner, este procedimento seria suficiente para se construir critérios de codificação claros, que apresentassem boa concordância entre diferentes codificadores. A ausência de dados de fidedignidade para as variáveis do Sumário Estrutural não seria importante, pois a fidedignidade da codificação das respostas é adequada e o Sumário Estrutural é gerado a partir dos códigos das respostas. A questão da fidedignidade das codificações obtidas por psicólogos em seu trabalho cotidiano seria uma proposta interessante, mas que raramente é investigada nos diversos testes, e que, portanto, não deveria ser exigida do SCR.

Wood, Nezworski e Stejskal (1996b), em sua resposta às explicações de Exner, afirmam que o procedimento utilizado por ele não é uma medida da fidedignidade inter-codificadores (interscorers reliability), mas sim uma medida da acurácia das codificações (scoring accuracy). Assim sendo, a fidedignidade inter-codificadores seria desconhecida, havendo necessidade urgente de pesquisas. Novamente, reafirmaram a necessidade de se utilizar kappa em vez de porcentagem de concordância ou porcentagem correta. A fidedignidade deveria ser reportada para todas as categorias de codificação e para todos as

variáveis do Sistema Compreensivo e pesquisas de fidedignidade inter-avaliadores de campo seriam necessárias.

Esta série de artigos, publicada no periódico **Psychological Science**, deu início ao que veio a ser conhecido como a Controvérsia do Rorschach. Vários artigos de críticos e de defensores do Rorschach foram publicados em diversos periódicos científicos e também na mídia popular, incluindo o jornal **The New York Times** e a revista **Scientific American**. A discussão atingiu seu ápice entre 2001 – com a publicação da série especial no periódico **Psychological Assessment** – e 2003 – com a publicação do livro **What's wrong with the Rorschach**, de Wood, Nezworski, Lillienfeld e Garb. Apesar das tentativas de se chegar a um acordo sobre os méritos e limitações do Rorschach, a controvérsia ainda permanece, agora com publicações mais esporádicas. De qualquer modo, um aspecto importante desta discussão foi promover a realização de novas pesquisas, que preencheram lacunas no conhecimento sobre o SCR.

A fidedignidade inter-codificadores pode ser avaliada de diferentes modos: ao nível das respostas, pode-se investigar a fidedignidade de decisões dicotômicas de codificação, de decisões policotômicas de codificação e a fidedignidade dos segmentos da codificação (localização, determinantes, conteúdos etc.); ao nível do protocolo, pode-se investigar a fidedignidade dos escores do Sumário Estrutural e das categorias interpretativas. Uma explicação sobre os diferentes modos de avaliar a fidedignidade inter-codificadores do SCR encontra-se no Apêndice D.

McDowell e Acklin (1996) decidiram investigar empiricamente a melhor medida para se avaliar a concordância inter-avaliadores: porcentagem de concordância ou kappa. Utilizaram 20 protocolos de estudantes universitários (college), com um total de 412 respostas, e estudaram a fidedignidade dos segmentos de codificação. Concluíram que kappa é a medida de fidedignidade inter-avaliadores mais adequada, por fornecer estimativas mais conservadoras e não afetadas pela concordância obtida ao acaso.

Meyer (1997b), em resposta às críticas de Wood, Nezworski e Stejskal (1996a, 1996b), afirmou que a porcentagem correta é essencialmente idêntica à porcentagem de concordância, por se tratar de uma comparação entre duas codificações independentes. Apresentou uma metanálise de dados publicados sobre fidedignidade inter-codificadores, que

indica que o Sistema Compreensivo possui excelente concordância das codificações, avaliada por segmentos e utilizando-se kappa: valores de 0,72 a 0,96, com média = 0,86. Destaca ainda que o objetivo ao investigar a concordância inter-avaliadores é saber se os critérios de codificação das respostas são claros. Assim, utilizar avaliadores com menor treinamento e num contexto da prática clínica cotidiana (field interrater reliability) não seria adequado para se investigar este aspecto da clareza dos critérios.

Discordamos de Meyer quando ele afirma que o objetivo do estudo de concordância inter-avaliadores é verificar a clareza dos critérios de codificação – este pode ser um dos objetivos deste tipo de estudo, que deve, então, ser realizado com avaliadores bem treinados. Em nossa opinião, ambos os tipos de situação – contexto de pesquisa e contexto da prática clínica cotidiana – trazem informações interessantes sobre o funcionamento do Rorschach e a consistência das codificações.

Em resposta a este artigo de Meyer (1997b), Wood, Nezworski e Stejskal (1997) insistem na necessidade de se conhecer a fidedignidade dos códigos individuais, em vez de se utilizar os segmentos de codificação.

Em nossa opinião, muitas vezes a discussão enveredou por argumentos teóricos ou metodológicos difíceis de serem resolvidas sem pesquisas empíricas. Seria mais sensato reconhecer que cada uma das estimativas de concordância (porcentagem de concordância e kappa) apresenta vantagens e desvantagens; o ideal seria reportar todas elas nos estudos. Naquele momento, os críticos tinham certa razão em solicitar que os proponentes do Sistema Compreensivo fornecessem informações mais completas sobre a fidedignidade das codificações – informações estas que foram então pesquisadas e reportadas.

Grønnerød (1999) investigou a fidedignidade inter-codificadores do Rorschach, utilizando diferentes estatísticas: porcentagem de concordância, kappa, correlações. Contudo, utilizou uma adaptação norueguesa do Sistema Klopfer – motivo pelo qual não incluímos este trabalho nesta parte de revisão dos estudos de fidedignidade inter-codificadores do SCR. De qualquer modo, sua discussão sobre as diferentes estatísticas de fidedignidade traz informações importantes:

Todas as estimativas de porcentagem de concordância ao nível das respostas alcançaram o critério de 0,80, embora existisse considerável discordância entre os avaliadores. Portanto, a porcentagem de concordância não seria uma boa medida da fidedignidade, sendo recomendada a utilização de kappa. De fato, kappa parece ser uma medida melhor, por descontar a concordância obtida por mero acaso. No entanto, kappa é muito sensível para códigos infreqüentes. Obter um kappa baixo para categorias de codificação com baixa taxa de ocorrência não indica que as codificações são pouco fidedignas, mas sim que a estimativa da fidedignidade é incerta. O autor encontrou também que a avaliação da concordância ao nível dos escores do Sumário Estrutural não acrescentou informações à avaliação ao nível das respostas.

Acklin et al. (2000) re-examinaram os protocolos de 20 estudantes universitários utilizados em trabalho anterior (McDowell e Acklin, 1996), agora considerando a fidedignidade das decisões de codificação, códigos individuais e escores do Sumário Estrutural. Também investigaram a fidedignidade inter-codificadores em uma amostra de 20 pacientes psiquiátricos. Concluíram que a porcentagem de concordância é uma medida inadequada de fidedignidade, porque produz estimativas não sensíveis a variações na taxa de ocorrência e grau de concordância devida ao acaso; kappa é muito mais sensível a variações na taxa de ocorrência e ao grau de concordância obtida ao acaso. Ao nível das respostas, deve-se utilizar kappa e, ao nível do protocolo, deve-se utilizar o Coeficiente de Correlação Intraclass¹⁷ – no caso, ICC (2,1), ou seja, o modelo de dois fatores, efeito aleatório para a variância e concordância absoluta como definição de correlação (two-way random effects model of variance and an absolute agreement definition of correlation). Com poucas exceções, a maioria dos códigos do Sistema Compreensivo demonstrou fidedignidade pelo menos substancial (valores de kappa ou do Coeficiente de Correlação Intraclass maiores do que 0,60).

¹⁷ O Coeficiente de Correlação Intraclass (Intraclass Correlation Coefficient - ICC) baseia-se nos componentes da análise de variância (BARTKO, 1966). A interpretação do ICC é semelhante à do kappa (ACKLIN et al., 2000).

Nakata (2000) investigou a fidedignidade inter-codificadores utilizando protocolos de indivíduos pacientes e não-pacientes. Supunha que os protocolos de pacientes teriam respostas mais complexas, mais códigos especiais e mais determinantes infrequentes, o que contribuiria para investigar a questão do nível de complexidade das respostas interferir com a fidedignidade inter-codificadores. Utilizou como codificadores estudantes de psicologia com treinamento semelhante, mas de diferentes faculdades. Encontrou fidedignidade substancial ou excelente para a maioria das variáveis do SCR. A fidedignidade ao nível do Sumário Estrutural, avaliada pelo ICC, foi maior do que ao nível das respostas, avaliada pelo kappa. A maioria das variáveis com fidedignidade baixa foram aquelas com baixa taxa de ocorrência, mas muitas variáveis com baixa taxa de ocorrência apresentaram fidedignidade adequada.

Meyer et al. (2002) utilizaram 8 amostras para estimar a concordância inter-avaliadores: 1) codificadores estudantes; 2) codificadores experientes; 3) codificadores clínicos frente a protocolos selecionados; 4) codificadores clínicos frente a protocolos da prática clínica cotidiana; 5) reunião das amostras de 1 a 4, que representa uma amostra de pacientes encontrados na prática clínica; 6), 7) e 8) amostras nas quais 10%, 20% e 30% dos códigos foram substituídos por códigos sorteados dentre todos os possíveis, deste modo introduzindo elementos que diminuiriam a concordância. Utilizaram kappa para o nível das respostas e ICC (1,1) para o nível do protocolo. Encontraram como medianas dos ICCs para as amostras mencionadas: 0,85, 0,96, 0,97, 0,95, 0,93, 0,95, 0,89 e 0,82. As amostras em que, deliberadamente, foram introduzidos “erros” nas codificações (códigos atribuídos por sorteio), ainda assim mostraram níveis de fidedignidade adequados.

As conclusões do estudo foram: 1) A fidedignidade ao nível das respostas fornece uma estimativa mais rigorosa do que ao nível do Sumário Estrutural; 2) Clínicos alcançam fidedignidade semelhante à de pesquisadores; 3) O Sumário Estrutural, por agregar dados, e assim anular parte das discordâncias, apresenta fidedignidade maior do que a do nível das respostas.

O nível mais adequado de análise depende do objetivo do estudo. Se o objetivo for determinar o grau em que dois avaliadores compreendem e concordam sobre as regras de codificação, o nível das respostas parece mais adequado (mas com cuidado ao se utilizar

amostras pequenas); se for avaliar a fidedignidade para pesquisa ou prática, o nível do Sumário Estrutural parece ser o mais indicado; se for documentar a fidedignidade como precursora da validade, o mais indicado é usar o nível dos segmentos.

Estimativas de fidedignidade como Kappa e ICC são afetadas pela taxa de ocorrência das variáveis investigadas, sendo que variáveis com baixa taxa de ocorrência (menor do que 5%) costumam gerar estimativas equivocadas, frequentemente mais baixas do que a fidedignidade verdadeira. Assim sendo, amostras pequenas, que costumam apresentar baixa taxa de ocorrência para muitas variáveis, frequentemente geram estimativas de fidedignidade pouco confiáveis (MEYER et al., 2002).

Meyer (2003) realizou um levantamento de metanálises sobre a fidedignidade e a validade de testes médicos e psicológicos. Esta comparação o levou a concluir que o Rorschach apresenta fidedignidade das codificações semelhante à de outros procedimentos em Psicologia e Medicina.

Viglione e Taylor (2003) apresentam dados de fidedignidade inter-avaliadores para 70 variáveis do SCR. A fidedignidade foi estimada utilizando 84 protocolos de diversos locais, representando uma amostra de campo ou da prática cotidiana (field applications). A maioria dos valores de fidedignidade encontrados foi excelente ($ICC > 0,80$). Os valores do ICC foram mais altos para variáveis com maior taxa de ocorrência; para variáveis com baixa taxa de ocorrência, a fidedignidade mostrou-se mais instável, apresentando alguns valores igualmente altos, mas outros valores baixos; haveria necessidade de amostras grandes para se obter estimativas estáveis de variáveis com baixa taxa de ocorrência.

A fidedignidade foi mais baixa para algumas variáveis ou decisões de codificação: qualidade formal, códigos especiais críticos, *FC* versus *CF*, subtipos de sombreado, *DQv* e *DQv/+* e algumas categorias de conteúdo. Portanto, alguns critérios de codificação ainda podem ser aprimorados, com isso melhorando ainda mais a fidedignidade inter-avaliadores. No sentido de aprimorar os critérios de codificação, Donald Viglione investigou divergências entre diferentes codificadores, discutiu problemas de codificação e elaborou diretrizes mais precisas para a codificação das respostas, trabalho que resultou em seu **Rorschach coding solutions – A reference guide for the Comprehensive System** (VIGLIONE, 2002).

McGrath, Pogge e Strokes (2005) comentam que diversos estudos demonstraram boa concordância das codificações realizadas por diferentes avaliadores em contextos de pesquisa, mas faltavam estudos das codificações em contextos da prática clínica cotidiana. Apresentam, então, dados da concordância das codificações realizadas por psicólogos clínicos ao avaliarem 84 protocolos de pacientes psiquiátricos adolescentes. Os protocolos foram coletados por psicólogos e estudantes, e a amostra apresentou menor *R* e valores maiores de *Lambda* do que a amostra de adolescentes do Sistema Compreensivo, ou seja, a amostra foi composta por protocolos mais defendidos ou mais simples do que os da amostra normativa. Os protocolos foram codificados pelos aplicadores e supervisionados; posteriormente, doutorandos em psicologia realizaram uma segunda codificação dos protocolos, independente da primeira. Oitenta e sete por cento dos códigos apresentaram fidedignidade aceitável ($\kappa > 0,60$), com a maioria dos coeficientes tendo nível excelente ($\kappa > 0,80$). Para a investigação da fidedignidade inter-codificadores ao nível do Sumário Estrutural, utilizaram o ICC (1,1) e encontraram resultados semelhantes aos do nível das respostas. A conclusão do estudo foi que codificações fidedignas são possíveis mesmo num contexto clínico.

De modo semelhante a outras pesquisas, identificaram o problema de variáveis com baixa taxa de ocorrência apresentarem menor estabilidade das estimativas de fidedignidade. A fidedignidade só se torna estável quando a taxa de ocorrência encontra-se próxima de 10%.

Cabe notar que Acklin et al. (2000) utilizaram o ICC (2,1) – modelo de dois fatores e efeito aleatório (two way random effects model) –, enquanto os demais trabalhos utilizaram o ICC (1,1) – modelo de um fator e efeito aleatório (one way random effects model). McGrath, Pogge e Strokes (2005, p. 206) justificam a utilização do ICC (1,1), citando Meyer et al. (2002) e comunicação pessoal com Shrout, um dos autores do ICC:

Embora esta estatística [ICC (1,1)] seja mais consistente quando nenhum avaliador codificou mais do que um protocolo, é geralmente considerada a mais adequada para circunstâncias em que diferentes alvos são avaliados por diferentes avaliadores.

Shrout e Flies (1979) observam que, para um mesmo conjunto de dados, ICC (1,1) costuma gerar valores menores do que ICC (2,1).

Mais recentemente, Janson e Olsson (2001, 2004) apresentaram uma nova medida de fidedignidade: o coeficiente iota¹⁸. Trata-se de uma medida multivariada de concordância para dados intervalares ou nominais, que é uma extensão do coeficiente kappa.

Os autores destacam que

Uma área em que uma medida multivariada de concordância para dados nominais poderia ser bastante útil é a da codificação objetiva dos testes de personalidade projetivos. Sistemas de codificação para materiais projetivos tipicamente envolvem certo número de decisões categoriais para cada resposta. Em tais casos, uma medida multivariada permitiria comparações do nível total de concordância entre estudos publicados, ou entre conjunto de juízes, e permitiria a avaliação contínua da qualidade das codificações em treinamentos ou em pesquisas em andamento. (JANSON; OLSSON, 2001, p. 279).

Ou seja, uma vantagem desta medida de concordância é possibilitar avaliar a concordância geral (overall agreement) mesmo para amostras que são muito pequenas, nas quais seria problemático estimar a concordância das categorias de codificação, devido ao problema das baixas taxas de ocorrência.

Especificamente para o SCR, Harald Janson desenvolveu o software **Rorschach Research Utilities (RRU)**¹⁹. A partir das codificações das respostas realizadas pelos avaliadores, este programa de computador calcula a concordância ao nível das respostas (concordância para decisões dicotômicas de codificação; concordância para decisões policotômicas de codificação; concordância para os segmentos da codificação; concordância geral) e ao nível do protocolo (variáveis do Sumário Estrutural), utilizando o coeficiente iota.

Sahly (2006) estudou a fidedignidade inter-codificadores de 487 protocolos de indivíduos não-pacientes. Os protocolos foram codificados por pelo menos dois avaliadores e o programa **Rorschach Research Utilities** foi utilizado para o cálculo da fidedignidade inter-

¹⁸ O coeficiente iota é representado pela letra grega minúscula ι . Iota é apresentado no Apêndice E.

¹⁹ O programa **Rorschach Research Utilities** funciona na plataforma Microsoft Windows. O usuário brasileiro deve ter o cuidado de modificar as configurações de idioma para English (United States) de modo a que as casas decimais sejam separadas por pontos (em vez de vírgulas), para que o programa funcione corretamente.

codificadores. Vinte e sete por cento das variáveis apresentaram concordância substancial ($\text{iota} > 0,60$) e 66% das variáveis apresentaram concordância excelente ($\text{iota} > 0,80$). Parte das variáveis com fidedignidade fraca ou regular ($\text{iota} > 0,20$ e $\text{iota} > 0,40$, respectivamente) apresentaram baixa taxa de ocorrência. A autora concluiu que o SCR pode ser codificado de modo fidedigno, embora alguns critérios de codificação ainda possam ser aprimorados.

Nesta pesquisa, boa parte dos protocolos foram codificados por dois avaliadores de um mesmo local. Apesar de terem codificado as respostas de modo independente, o fato de serem do mesmo local e anteriormente terem discutido codificações duvidosas, pode ter contribuído para adotarem certas regras de consenso sobre aspectos da codificação para os quais os critérios são imprecisos ou insuficientes. Provavelmente, isto contribuiu para uma maior concordância do que a que seria obtida se os avaliadores fossem de locais diferentes (SAHLY, 2006).

5.4. Acurácia das codificações do SCR

Uma outra questão correlata à fidedignidade inter-codificadores refere-se ao grau em que diferentes avaliadores codificam corretamente as respostas. Dois codificadores podem até mesmo concordar em suas codificações e ainda assim ambos estarem codificando incorretamente. Isto pode ocorrer, por exemplo, entre codificadores de um mesmo grupo de pesquisa, com uma mesma compreensão equivocada dos critérios de codificação. Meyer, Erdberg e Shaffer (2007) comentam que as codificações podem diferir em função de convenções locais de codificação. Aparentemente, codificações realizadas por avaliadores de um mesmo local apresentam maior concordância do que codificações realizadas por avaliadores de locais diferentes.

Para o estudo da acurácia das codificações, as codificações corretas precisam ser previamente estabelecidas. Mais comumente um conjunto de especialistas chega a um consenso sobre as codificações corretas para um conjunto de respostas; depois, as codificações de outros avaliadores são comparadas com as classificações corretas e o grau de acurácia das codificações é calculado.

Exner (1999a) comenta que a acurácia das codificações é importante em pesquisas, porque a validade das interpretações ou conclusões depende de as variáveis em estudo terem seus valores calculados a partir de códigos corretos. Cita dados de suas pesquisas com o SCR:

Quando a porcentagem de concordância correta, resposta a resposta, para cada uma das nove categorias de codificação possíveis [cada um dos segmentos da codificação] foi calculada para 15 codificadores e 400 respostas, e para 20 codificadores e 500 respostas, os resultados variaram de aproximadamente 89% a 98% de concordância correta para os 18 códigos especiais e entre 87% e 97% de concordância correta para os 23 determinantes. (EXNER, 1999a, p. 163).

Contudo, Guarnaccia et al. (2001) mencionam uma pesquisa, apresentada na Alumni Newsletter²⁰ (comunicação de Rorschach Workshops para seus alunos) de 1988, na qual se encontrou uma importante quantidade de erros nas codificações realizadas por alunos do curso de treinamento no SCR.

Preocupado com este problema e com a ausência de pesquisas publicadas, Guarnaccia et al. (2001) decidiram investigar a acurácia das codificações, ou seja, o grau em que diferentes avaliadores codificam corretamente as respostas. Utilizaram como codificadores 21 estudantes de psicologia e 12 psicólogos, que codificaram 20 respostas provenientes de protocolos de indivíduos não-pacientes e pacientes clínicos. Estas respostas e suas codificações corretas foram retiradas dos livros de Exner (1986) e Weiner (1998a), criadores do SCR. O nível de acerto dos codificadores mostrou-se próximo do aceitável para as respostas dos não-pacientes, mas claramente abaixo do aceitável para as respostas dos pacientes clínicos. Os codificadores obtiveram diferentes taxas de acerto para os diferentes segmentos da codificação. Alguns aspectos da codificação, tais como a localização e a qualidade evolutiva, são mais simples de serem codificados, o que levou a uma maior taxa de acerto; outros aspectos da codificação, como os determinantes e os códigos especiais, são mais difíceis de serem codificados, o que levou a uma menor taxa de acerto. A baixa taxa de acerto entre os psicólogos profissionais pode ter ocorrido por negligência ao participar da

²⁰ EXNER, J. E. Jr. Rorschach Workshops: Scoring issues. **Alumni Newsletter**, p. 4-8, 1988.

pesquisa; talvez, em situações em que apliquem os protocolos e os codifiquem, em vez de codificar respostas que foram selecionadas para uma pesquisa, apresentem maior taxa de acerto. De qualquer modo, a baixa taxa de acertos dos estudantes é preocupante. O problema do treinamento adequado para a codificação do Rorschach mostra-se então bastante sério e os autores chegaram a sugerir um processo de certificação dos usuários.

Os resultados do estudo de Guarnaccia et al. (2001) devem ser vistos com cautela, pois os autores utilizaram uma estatística idiossincrática para investigar a acurácia das codificações, subtraindo pontos para os erros (MCGRATH; POGGE; STROKES, 2005). Além deste problema, Hilsenroth, Charnas e Zoldan (2007) comentam outras limitações do estudo: 1) Foram utilizadas apenas 20 respostas, o que faz com que as estimativas de fidedignidade sejam pouco precisas devido à baixa taxa de ocorrência de muitos códigos; 2) O grau de dificuldade das respostas codificadas não foi avaliado; 3) Faltam dados sobre a experiência dos psicólogos com o Rorschach e sobre sua proveniência (eventualmente são do mesmo local que os estudantes codificadores, o que explicaria os mesmos níveis de fidedignidade obtidos por psicólogos e estudantes).

Em 2006, Hilsenroth e Charnas prepararam seu **Manual de Treinamento para Fidedignidade Inter-avaliadores**, composto por protocolos previamente codificados por usuários experientes do Rorschach, estabelecendo-se as codificações corretas por consenso. Utilizando este material, Hilsenroth, Charnas e Zoldan (2007) desenvolveram um procedimento de treinamento da codificação das respostas do Rorschach: primeiramente, os alunos estudam os critérios de codificação de determinado segmento da classificação; os professores então apresentam exemplos de respostas em que os códigos estudados ocorrem e não ocorrem, bem como exemplos de respostas ambíguas, esclarecendo a codificação correta; após cada aula, os alunos codificam respostas de protocolos selecionados, sendo codificado apenas o segmento que estudaram na aula; as codificações dos alunos são avaliadas a cada aula, esperando-se um mínimo de 80% de codificações corretas; caso um aluno não atinja este nível de acerto, deve praticar mais, codificando outros protocolos e checando seu nível de acerto, até atingir 80%, somente então passando à próxima etapa.

A eficiência desta proposta de treinamento das codificações foi avaliada. Vinte e nove alunos que completaram as 27 horas de treinamento codificaram dois protocolos compostos por respostas selecionadas de modo a representar os diversos códigos do SCR. O primeiro protocolo continha 19 respostas de indivíduos não-pacientes e apresentava nível

médio de dificuldade; o segundo protocolo continha 20 respostas de pacientes psiquiátricos e apresentava nível altíssimo de dificuldade (o nível de dificuldade corresponde à média dos valores atribuídos por 20 usuários experientes do Rorschach). Foi calculada a porcentagem de concordância exata para cada segmento das respostas e o valor de kappa para a fidedignidade de cada segmento de codificação foi estimado, seguindo os procedimentos detalhados em Meyer (1999)²¹.

Para o primeiro protocolo de avaliação, todos os alunos obtiveram 80% ou mais de acerto total, bem como mais de 80% de acerto em cada segmento da codificação; o kappa estimado situou-se na faixa excelente ($\text{kappa} > 0,74$) para todos os segmentos exceto para notas Z ($\text{kappa} = 0,71$) e códigos especiais ($\text{kappa} = 0,73$), cujos valores situaram-se na faixa de boa a excelente. Para o segundo protocolo de avaliação, 23 alunos (79%) obtiveram 80% ou mais de acerto total. Dos 6 alunos com menos de 80% de acerto, quatro alcançaram os 80% após treinarem com mais um protocolo e dois necessitaram de mais dois protocolos de treino para então alcançar o nível esperado. Para os segmentos, o nível de acerto foi maior do que 80%, exceto para os determinantes (78%) e para os códigos especiais (65%); os valores de kappa foram excelentes ($\text{kappa} > 0,74$) para localizações, qualidade formal, pares, conteúdos e populares, de bom a excelente ($0,60 < \text{kappa} < 0,75$) para determinantes, qualidade formal e notas Z, e de regular a bom para os códigos especiais ($\text{kappa} = 0,56$). Os resultados são bastante superiores aos do estudo de Guarnaccia et al. (2001). Na literatura, os autores encontraram apenas o estudo de Meyer et al. (2002), que contém dados de uma amostra de quatro codificadores estudantes; os resultados são semelhantes, cabendo notar que foram codificados 66 protocolos, com um total de 1 407 respostas, número bastante adequado para gerar estimativas seguras de fidedignidade.

Hilsenroth, Charnas e Zoldan (2007) concluem que o Rorschach pode ser codificado de modo correto, desde que haja treinamento adequado. A estratégia de ensino da codificação por segmentos, estabelecendo-se 80% de acerto como mínimo aceitável para se passar ao próximo segmento parece contribuir para formar codificadores proficientes.

²¹ MEYER, G. J. Simple procedures to estimate chance agreement and kappa of the inter-rater reliability of response segments using the Rorschach Comprehensive System. **Journal of Personality Assessment**, v. 72, p. 230-255, 1999.

5.5. Estabilidade temporal do SCR

Estudos com delineamento do tipo teste-reteste avaliam o grau em que os resultados mantêm-se semelhantes em duas ou mais aplicações do teste. Nos estudos de teste-reteste do SCR, tanto com intervalos breves, quanto com intervalos longos, encontramos a denominação *estabilidade temporal*.

Exner (1978, p. 63-64) destaca a importância dos estudos de estabilidade temporal para o Rorschach:

[...] se é verdadeira a hipótese de que *a maioria ou todas as pessoas têm estilos preferenciais de resposta*²² que irão se manifestar na maioria de suas respostas; e se o Rorschach é um instrumento a partir do qual vislumbres destes estilos de resposta podem ser apreendidos, *então* estes mesmos estilos de respostas devem se evidenciar em testagens repetidas do Rorschach. (grifos do autor)

Os estudos de teste-reteste servem, então, para investigar em que grau os resultados do Rorschach se repetem em diferentes momentos. Espera-se que os resultados sejam estáveis, mesmo em intervalos relativamente longos, o que indicaria que os erros de medida são pequenos e que o teste avalia características estáveis da personalidade. O próprio conceito de personalidade implica em padrões de funcionamento psicológico que se repetem em diversas situações e em diferentes momentos – e o Rorschach é considerado um teste de avaliação da personalidade.²³

No entanto, cabe uma ressalva de Exner (1978, p. 65-66):

²² Estilo de resposta é a denominação utilizada por Exner para se referir ao que também é chamado traço, hábito ou disposição. Um estilo de resposta é um modo característico de responder às situações.

²³ Cabe observar que, além da estabilidade dos resultados – condição necessária para se afirmar que o teste avalia características estáveis do funcionamento psicológico –, existe a necessidade de pesquisas que indiquem que os estilos de resposta no Rorschach (estrutura das respostas às manchas de tinta) de fato expressam estilos de resposta em outras situações da vida diária (estrutura da personalidade). Existem diversos estudos que confirmam esta correspondência entre os estilos de resposta no Rorschach e os estilos de resposta em outras situações da vida – estes são os estudos de validade do Rorschach.

Uma das falácias relativas ao Rorschach é a noção de que a consistência temporal deve ser demonstrada para *todas* as variáveis. Em outras palavras, se o teste é ‘fidedigno’ dentro dos parâmetros deste tipo de fidedignidade, todas as suas características devem ser consistentes ao longo do tempo. Isto não é verdade. A falácia baseia-se na premissa de que cada um dos índices do Rorschach reflete um aspecto da pessoa que não muda. É verdade que muitos, ou até mesmo a maioria dos dados estruturais devem permanecer relativamente constantes ao longo do tempo; mas qualquer revisão completa das pesquisas sobre muitas das variáveis do Rorschach indica que algumas delas relacionam-se especificamente a fenômenos situacionais. [...] Spielberger (1966, 1971)²⁴ foi bastante convincente ao demonstrar que parte da ansiedade é ‘estado’ – ou seja, relaciona-se a situações –, enquanto outra parte é ‘traço’ – ou seja, uma característica mais durável da própria personalidade. Os mesmos conceitos de *estado* e *traço* aplicam-se aos dados estruturais do Rorschach. (grifos do autor)

[Trecho suprimido na versão online]

²⁴ SPIELBERGER, C. D. **Anxiety and behavior**. New York: Academic Press, 1966.

SPIELBERGER, C. D.; LUSHENE, R. E.; MCADOO, W. G. Theory and measurement of anxiety states. In: **Handbook of Modern Personality Theory**. CATTELL, R. B. (editor). Chicago: Aldine Press, 1971.

A questão da estabilidade temporal do Rorschach vinha sendo criticada há tempo. Holzberg (1977) mencionou alguns problemas na avaliação da estabilidade temporal do Rorschach: 1) A personalidade, objeto de avaliação pelo Rorschach, não seria assim tão estável, então a estabilidade dos resultados poderia ser baixa, sem necessariamente comprometer a validade do teste; 2) A disposição ou estado situacional (setting) do indivíduo pode variar de momento a momento, e a própria realização do teste e do reteste já traz mudanças no estado da pessoa; 3) a memória das respostas fornecidas pode aumentar artificialmente as correlações obtidas.

As pesquisas empíricas do Sistema Compreensivo responderam a estas críticas: 1) os resultados dos estudos de teste-reteste, mesmo em intervalos longos como três anos, mostraram correlações altas para a maioria das variáveis, indicando que o Rorschach avalia características estáveis da personalidade; 2) as correlações altas para a maioria das variáveis também indicam que fatores situacionais têm pouco impacto nos resultados; algumas mudanças no estado do indivíduo refletem-se em algumas variáveis associadas a fenômenos situacionais (EXNER; ARMBRUSTER; VIGLIONE, 1978).

O problema da interferência da memória nos estudos de teste-reteste também havia sido comentado por outros autores. A crítica mais freqüente é que, nos estudos com intervalos breves entre as testagens, os participantes lembravam e repetiam as respostas, o que levaria a correlações altas, não necessariamente expressando a verdadeira fidedignidade do teste. No entanto, Zubin, Eron e Schumer (1965) comentam que nem sempre a memória das respostas leva a estas serem repetidas no reteste; em alguns casos, a memória poderia fazer com que a pessoa buscasse modificar ou até não fornecer as mesmas respostas da testagem anterior. De qualquer modo, a interferência da memória prejudicaria a estimativa da estabilidade temporal do Rorschach.

Alguns estudos anteriores ao Sistema Compreensivo buscaram contornar este problema da memória das respostas. Kelley, Margulies e Barreva (1941, apud Grønnerød, 2003a)²⁵ compararam 12 protocolos obtidos imediatamente antes e depois de

²⁵ KELLEY, D. M.; MARGULIES, H.; BARREVA, S. E. The stability of the Rorschach method as demonstrated in electric convulsive therapy cases. **Rorschach Research Exchange**, v. 5, p.35-43, 1941.

eletroconvulsoterapia (ECT), procedimento que provoca amnésia para os acontecimentos imediatamente anteriores – ou seja, os testandos, ao passarem por ECT, esqueciam as respostas que haviam fornecido ao Rorschach. Encontraram semelhanças no padrão de resultados de cada indivíduo; contudo, houve algumas mudanças em alguns índices. Griffith (1951, apud Grønnerød, 2003a)²⁶ realizou teste e reteste em 4 pacientes com síndrome de Korsakoff, portanto com problemas na memória de retenção. Houve dificuldade para realizar o inquérito, mas o conteúdo das respostas no teste e no reteste mostrou-se semelhante, indicando que as respostas são estáveis. Wagner e Daubney (1976) selecionaram pacientes com problemas neurológicos supondo que isso tornava as mudanças de personalidade menos prováveis. Porém, não encontraram alta estabilidade entre os resultados do teste e do reteste realizado após três anos.

O estudo do Sistema Compreensivo com intervalo de reteste de 3 anos (Exner, Armbruster e Viglione, 1978) provavelmente reduziu significativamente o problema da memória das respostas, e as correlações de teste-reteste encontradas foram altas. Ainda assim, Aronow, Reznikoff e Moreland (1995, p. 221) comentam:

Tal como assinalado nos **Standards for Educational and Psychological Testing**²⁷, publicados pela American Psychological Association, ‘Estimativas da estabilidade baseadas em um reteste com a mesma forma [...] podem ser espuriamente elevadas devido aos efeitos da memória’ (1985, p. 21). Anastasi²⁸, de modo semelhante, nota que ‘a técnica de teste-reteste apresenta dificuldades quando aplicada aos testes psicológicos [...]. Para a grande maioria dos testes psicológicos, [...] o reteste com uma forma idêntica do teste não é uma técnica adequada para se encontrar um coeficiente de fidedignidade’ (1988, p. 18). Tal como afirmamos em uma publicação anterior (Reznikoff, Aronow e Rauchway, 1982)²⁹, acreditamos que devido à alta

²⁶ GRIFFITH, R. M. Test-retest similarities of the Rorschach of patients without retention. **Journal of Projective Techniques**, v. 15, p. 516-525, 1951.

²⁷ AMERICAN PSYCHOLOGICAL ASSOCIATION. **Standards for Educational and Psychological Testing**. Washington, DC: American Psychological Association, 1985.

²⁸ ANASTASI, A. **Psychological testing**. 6th ed. New York: Macmillan, 1988.

²⁹ REZNIKOFF, M.; ARONOW, E.; RAUCHWAY, A. The reliability of inkblot content scales. In: **Advances in personality assessment**. SPIELBERGER, C. D.; BUTCHER, J. D. (editores). Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

saliência dos perceptos, o Rorschach pode apresentar efeitos ainda mais sérios de memória do que outros testes psicológicos. Resumidamente, consideramos a fidedignidade das categorias do Sistema Compreensivo como não estabelecidas de modo algum.

É interessante a menção dos autores à alta saliência dos perceptos. Provavelmente estão se referindo ao fato de as características dos estímulos induzirem determinadas respostas. De fato, algumas importantes características das manchas – denominadas *critical bits* ou *distal features* – contribuem muito para que determinadas respostas sejam fornecidas (EXNER, 1996a). A saliência dos perceptos não é um problema, pois, apesar de as manchas possuírem características indutoras das respostas, apresentam suficiente ambigüidade para que diversas respostas potenciais possam ser percebidas e, algumas delas, selecionadas (EXNER; ARMBRUSTER; MITTMAN, 1978). A seleção das respostas que são fornecidas depende das características de personalidade do testando (EXNER, 2003). O próprio fato de, em um reteste, as respostas serem novamente percebidas e fornecidas, parece indicar, mais do que uma simples recordação, a repetição do funcionamento característico da personalidade.

Se o sujeito tivesse ‘visto’ apenas um ou dois objetos por mancha, a influência da memória poderia ser substancial; mas, como foi demonstrado, os sujeitos geralmente reportam apenas um ou dois objetos, selecionados a partir de um número muito maior de objetos que eles de fato perceberam ou identificaram como respostas potenciais (Exner, Armbruster e Mittman, 1978)³⁰. Assim, o fator memória não se baseia tanto no que foi visto na primeira testagem, mas sim no que foi reportado; e isto requer alguns engramas razoavelmente consistentes, especialmente se o intervalo de tempo entre a primeira e a segunda testagem é longo. [...] embora a memória possa desempenhar algum papel, a natureza deste papel é variável de um indivíduo para outro, e parece improvável que ela vá mudar as características básicas da personalidade ou dos estilos de resposta (EXNER; ARMBRUSTER; VIGLIONE, 1978, p. 475, grifo nosso).

³⁰ EXNER, J. E.; ARMBRUSTER, G. I.; MITTMAN, B. The Rorschach response process. **Journal of Personality Assessment**, v. 42, p. 27-38, 1978.

Além do estudo com reteste após 3 anos (Exner, Armbruster e Viglione, 1978), que supostamente reduziu os efeitos da memória, existem outros dois estudos em que foram solicitadas respostas diferentes no reteste, visando deste modo eliminar o problema da memória: Exner (1980) e Haller e Exner (1985). As correlações encontradas nos grupos que receberam instruções para fornecer respostas diferentes no reteste foram semelhantes às correlações nos grupos controle, que receberam as mesmas instruções padronizadas no teste e no reteste. Concluiu-se que o conteúdo das respostas pode mudar, mas sua estrutura mostra-se bastante estável.

Cabe ainda mencionar os parâmetros para se considerar uma correlação de teste-reteste como indicando estabilidade. Exner, Armbruster e Viglione (1978, p. 480) comentam:

É muito difícil, se não for impossível, definir com precisão que valor da correlação, ou faixa de valores, separa as variáveis 'estado' das 'traço'. Uma abordagem conservadora de medida provavelmente incluiria valores mínimos não menores do que 0,85 ou 0,80, mas é sempre arriscado tentar definir um valor específico como sendo um parâmetro definitivo.

Exner e Weiner (1995, p. 22) acrescentam:

O critério para consistência temporal adequada vem sendo definido de variados modos. Comumente, uma correlação de reteste com valor mínimo de 0,75 (responsável por 50% da variância) é satisfatória, se o intervalo de reteste é considerável, embora se requeira uma correlação mínima maior (0,80 ou 0,85), se o reteste ocorrer após um período breve de tempo.

Exner e Weiner (1995) não especificam o que seria um intervalo breve de tempo. Supomos, observando as diferentes pesquisas de estabilidade temporal de Exner e seus colaboradores, que intervalos breves sejam de dias e intervalos longos sejam de meses ou anos.

Apresentamos a seguir as pesquisas de estabilidade temporal do Sistema Compreensivo.

5.5.1. Pesquisas sobre a estabilidade temporal do SCR com crianças e adolescentes

Exner (1980) apresenta estudos de teste-reteste em crianças.

No primeiro estudo³¹, dois grupos de crianças foram comparados: um primeiro grupo de 30 crianças de 6 anos retestadas após 24 meses, e um segundo grupo de 25 crianças de 9 anos retestadas após 30 meses. Os resultados são apresentados na Tabela 1. As correlações foram maiores nas crianças de 9 anos, mas menores do que as encontradas em adultos retestados após 3 anos. Nas crianças de 6 anos, as 19 correlações reportadas apresentaram média = 0,46 e desvio-padrão = 0,25; nas crianças de 9 anos, média = 0,55 e desvio-padrão = 0,21.

Tabela 1 – Estabilidade temporal de 19 variáveis do SCR em duas amostras de crianças norte-americanas

Idade	6 anos	9 anos
Reteste	24 meses	30 meses
N	30	25
<i>R</i>	0,67	0,61
<i>Lambda</i>	0,18	0,39
<i>F</i>	0,51	0,69
<i>M</i>	0,48	0,62
<i>WSumC</i>	0,41	0,58
<i>EA</i>	0,19	0,45
<i>FM</i>	0,49	0,60
<i>m</i>	0,13	0,09
<i>SumSH</i>	0,08	0,29
<i>es</i>	0,20	0,57
<i>FC</i>	0,38	0,34
<i>CF+C</i>	0,27	0,35
<i>Afr</i>	0,51	0,79
<i>a</i>	0,86	0,81
<i>p</i>	0,42	0,29
<i>Popular</i>	0,77	0,74
<i>X+%</i>	0,84	0,86
<i>Zf</i>	0,55	0,68
<i>3r+(2)/R</i>	0,78	0,74

Fonte: Exner (1980) [primeiro estudo]

³¹ THOMAS, F.; EXNER, J. **Temporal consistency among six- and nine-year-olds over a 24 to 30 month interval.** Workshops Study No. 346. Estudo não publicado. Rorschach Workshops, 1980.

No segundo estudo³², 25 crianças de 8 anos de idade foram retestadas após 7 dias. Os resultados são apresentados na Tabela 2. As 26 correlações apresentaram média = 0,83 e desvio-padrão = 0,11, indicando, portanto, boa estabilidade.

Tabela 2 – Estabilidade temporal de 26 variáveis do SCR numa amostra de 25 crianças norte-americanas de 8 anos de idade retestadas após 7 dias

Variável	<i>r</i>	Variável	<i>r</i>
<i>R</i>	0,88	<i>es</i>	0,74
<i>Lambda</i>	0,82	<i>D-score</i>	0,93
<i>F</i>	0,79	<i>FC</i>	0,90
<i>M</i>	0,90	<i>CF</i>	0,76
<i>WSumC</i>	0,88	<i>C</i>	0,72
<i>EA</i>	0,85	<i>CF+C</i>	0,89
<i>FM</i>	0,75	<i>Afr</i>	0,91
<i>m</i>	0,49	<i>a</i>	0,91
<i>SumC'</i>	0,77	<i>p</i>	0,86
<i>SumT</i>	0,86	<i>Popular</i>	0,86
<i>SumV</i>	0,96	<i>X+%</i>	0,95
<i>SumY</i>	0,42	<i>Zf</i>	0,91
<i>SumSH</i>	0,70	<i>3r + (2)/R</i>	0,94

Fonte: Exner (1980) [segundo estudo]

No terceiro estudo³³, dois grupos foram retestados após aproximadamente 9 meses: o primeiro grupo com 20 crianças de 7 anos e o segundo grupo com 20 adolescentes de 15 anos de idade. Os resultados são apresentados na Tabela 3. Para as crianças, as correlações apresentaram média = 0,51 e desvio-padrão = 0,26, e somente 6 das 24 variáveis apresentaram correlações iguais ou maiores do que 0,75. Assim, concluiu-se que, mesmo em intervalos relativamente breves como este de 9 meses, ocorrem mudanças importantes no funcionamento psicológico das crianças. Já para os adolescentes, houve maior estabilidade: as correlações apresentaram média = 0,72 e desvio-padrão = 0,20, sendo que 13 das 24 variáveis investigadas apresentaram correlações iguais ou maiores do que 0,75, a maioria delas com

³² BRENT, P.; EXNER, J. **Temporal consistency among eight-year-olds retested after seven days.** Workshops Study No. 361. Estudo não publicado. Rorschach Workshops, 1980. Mais informações sobre este estudo são fornecidas em Exner (1986).

³³ ALINSKY, D.; EXNER, J. **Temporal consistency among seven- and fifteen-year-olds retested after nine months.** Workshops Study No. 352. Estudo não publicado. Rorschach Workshops, 1980. Mais informações sobre este terceiro estudo são fornecidas em Weiner e Exner (1995). Neste último trabalho, a correlação reportada para CF+C, para as crianças de 7 anos de idade, é 0,41 em vez de 0,46 (provável erro de digitação).

valores iguais ou maiores do que 0,80, semelhantes às correlações obtidas em adultos retestados após períodos mais longos.

Tabela 3 – Estabilidade temporal de 24 variáveis do SCR numa amostra de 20 crianças e numa amostra de 20 adolescentes, com reteste após 9 meses

Idade	7 anos	15 anos
<i>R</i>	0,71	0,80
<i>Lambda</i>	0,24	0,76
<i>F</i>	0,82	0,74
<i>M</i>	0,64	0,82
<i>WSumC</i>	0,37	0,70
<i>EA</i>	0,23	0,73
<i>FM</i>	0,53	0,70
<i>m</i>	0,06	0,17
<i>SumC'</i>	0,49	0,73
<i>SumT</i>	0,75	0,87
<i>SumV</i>	–	0,92
<i>SumY</i>	0,21	0,16
<i>SumSH</i>	0,11	0,48
<i>es</i>	0,18	0,64
<i>FC</i>	0,28	0,81
<i>CF+C</i>	0,46	0,71
<i>Afr</i>	0,61	0,89
<i>a</i>	0,77	0,85
<i>p</i>	0,39	0,64
<i>Sum5*</i>	0,79	0,88
<i>Popular</i>	0,88	0,83
<i>X+%</i>	0,86	0,83
<i>Zf</i>	0,62	0,79
<i>3r + (2)/R</i>	0,70	0,86

* Na época do estudo havia apenas 5 códigos especiais críticos.

Fonte: Exner (1980) [terceiro estudo]

A conclusão destes três estudos é que os resultados de crianças variam bastante ao longo do tempo. O desempenho em um momento não permite prever com segurança os resultados em momentos subsequentes, porque ocorrem mudanças na personalidade devidas ao processo de desenvolvimento. Somente na metade da adolescência é que a personalidade se estabiliza e, paralelamente, os resultados do Rorschach mostram-se mais estáveis. Apenas a variável *X+%* apresentou boa estabilidade em todos os grupos de crianças. [Trecho suprimido na versão online]

No quarto estudo³⁴, 55 crianças e adolescentes de 11 a 15 anos, pacientes psiquiátricos internados e ambulatoriais, sem deficiência intelectual nem problemas neurológicos, foram divididas em dois grupos. Para os participantes do grupo experimental pediu-se que se esforçassem para fornecer respostas melhores no reteste, enquanto que os participantes do grupo controle receberam as instruções originais. Não foram encontradas diferenças estatisticamente significantes entre os resultados dos dois grupos, mesmo com o grupo experimental tendo fornecido respostas diferentes no reteste, realizado após 2 dias.

No quinto estudo³⁵, 60 crianças não-pacientes de 8 anos foram divididas em dois grupos. O reteste foi realizado após 3 ou 4 dias e as crianças do grupo experimental foram instruídas a fornecer respostas diferentes, enquanto as crianças do grupo controle receberam as instruções originais. Os resultados são apresentados na Tabela 4. O grupo experimental forneceu apenas 14% de respostas iguais às da primeira testagem, enquanto o grupo controle forneceu 86%. As correlações de teste-reteste no grupo controle e no grupo experimental foram semelhantes.

Este último estudo também apresenta resultados sobre a direcionalidade das proporções. As proporções são índices do Sumário Estrutural que comparam os escores de duas variáveis. Por exemplo, a proporção $a : p$ compara o número de movimentos ativos com o número de movimentos passivos. Considerou-se que uma proporção apresentava direcionalidade quando um dos lados fosse maior do que o outro, exceto para *EB*, em que direcionalidade foi considerada quando a diferença entre os lados desta proporção fosse de pelo menos dois pontos. Considerou-se que havia estabilidade quando o participante apresentasse a mesma direcionalidade em ambas testagens ou ausência de direcionalidade em ambas testagens. Os resultados são resumidos na Tabela 5. Esta investigação da direcionalidade das proporções encontrou evidências de boa estabilidade.

³⁴ HULSTART, B. **The effects of a “second chance” instructional set on the Rorschach records of emotionally disturbed and culturally deprived children.** Tese de doutorado. Long Island University, 1979.

³⁵ THOMAS, E.; EXNER, J. **The effects of instructions to give different responses during retest on the consistency of the scores for eight-year-olds.** Workshops Study No. 372. Estudo não publicado. Rorschach Workshops, 1980

Tabela 4 – Estabilidade temporal em dois grupos de 30 crianças norte-americanas de 8 anos de idade retestadas após 3 ou 4 dias

Variável	Grupo Experimental	Grupo Controle
<i>R</i>	0,81	0,89
<i>Lambda</i>	0,67	0,82
<i>F</i>	0,59	0,81
<i>M</i>	0,91	0,89
<i>WSumC</i>	0,88	0,85
<i>EA</i>	0,84	0,88
<i>FM</i>	0,78	0,75
<i>m</i>	0,27	0,48
<i>SumSH</i>	0,44	0,71
<i>es</i>	0,38	0,70
<i>FC</i>	0,91	0,89
<i>CF+C</i>	0,83	0,87
<i>Afr</i>	0,92	0,91
<i>a</i>	0,94	0,90
<i>p</i>	0,82	0,87
<i>Popular</i>	0,76	0,84
<i>X+%</i>	0,84	0,94
<i>Zf</i>	0,89	0,87
<i>3r+(2)/R</i>	0,90	0,93

Fonte: Exner (1980) [quinto estudo]

Tabela 5 – Frequência e porcentagem de participantes com estabilidade em 5 proporções do SCR em dois grupos de 30 crianças norte-americanas de 8 anos de idade retestadas após 3 ou 4 dias

Proporção	Grupo Experimental	Grupo controle
<i>EB</i>	25 (83%)	27 (90%)
<i>eb</i>	15 (50%)	24 (80%)
<i>EA : es</i>	27 (90%)	24 (80%)
<i>a : p</i>	24 (80%)	25 (83%)
<i>FC : CF+C</i>	23 (77%)	28 (93%)

Fonte: Exner (1980) [quinto estudo]

Exner (1983) apresenta dados de um estudo³⁶ com 35 crianças de 9 anos retestadas após de 18 a 21 dias. Dados complementares deste estudo são apresentados em Exner (1986). Os resultados são apresentados na Tabela 6. As 26 correlações apresentam média = 0,80 e desvio-padrão = 0,20.

³⁶ THOMAS, E. E.; ALINSKY, D.; EXNER, J. E. **The stability of some Rorschach variables in 9-years-olds as compared with nonpatient adults**. Workshops Study No. 441. Estudo não publicado. Rorschach Workshops, 1982.

Tabela 6 – Estabilidade temporal em 35 crianças de 9 anos de idade retestadas após 3 semanas

Variável	<i>r</i>	Variável	<i>r</i>
<i>R</i>	0,87	<i>D-score</i>	0,91
<i>Lambda</i>	0,84	<i>FC</i>	0,84
<i>F</i>	0,80	<i>CF</i>	0,74
<i>M</i>	0,87	<i>C</i>	0,64
<i>WSumC</i>	0,87	<i>CF+C</i>	0,92
<i>EA</i>	0,87	<i>Afr</i>	0,91
<i>FM</i>	0,78	<i>a</i>	0,91
<i>m</i>	0,20	<i>p</i>	0,88
<i>SumC'</i>	0,74	<i>Popular</i>	0,89
<i>SumT</i>	0,92	<i>X+%</i>	0,92
<i>SumV</i>	0,93	<i>Zf</i>	0,92
<i>SumY</i>	0,17	<i>3r+(2)/R</i>	0,86
<i>es</i>	0,70	<i>FD</i>	0,81

Fonte: Exner (1983)

Exner, Thomas e Mason (1985) reportam dois estudos com crianças e adolescentes pacientes psiquiátricos. No primeiro estudo, com 29 adolescentes esquizofrênicos, com idades entre 12 e 15 anos e reteste realizado após entre 11 e 14 meses, encontraram boa estabilidade (correlações de 0,75 ou maiores) para 4 dos 5 dos indicadores de esquizofrenia. No segundo estudo, com 23 adolescentes deprimidos, com idades entre 11 e 14 anos, retestados após entre 9 e 13 meses, 3 das 5 variáveis associadas à depressão mostraram boa estabilidade. Assim, os autores concluíram que condições graves, tais como esquizofrenia e depressão, podem barrar o desenvolvimento, contribuindo para que padrões de funcionamento psicológicos associados à patologia se mantenham.

5.5.2. Pesquisas sobre a estabilidade temporal do SCR com adultos

Exner, Armbruster e Viglione (1978) realizaram um estudo de reteste em 100 indivíduos não-pacientes após 3 anos da realização do teste. As variáveis apresentaram boa estabilidade temporal, com a maioria das 19 correlações reportadas apresentando valores maiores que 0,75 (média = 0,79, desvio-padrão = 0,07, mediana = 0,80).

Avaliaram também a direcionalidade de algumas proporções: *EB*, *EA* : *es*, *eb*, *a* : *p*, *FC* : *CF+C*. Consideraram direcionalidade quando um dos lados da proporção fosse ao menos dois pontos maior do que o outro (por exemplo, uma proporção *a* : *p* = 5 : 4 seria

considerada como não apresentando direcionalidade, mas uma proporção $a : p = 6 : 4$ seria considerada como apresentando direcionalidade). Os resultados encontram-se resumidos na Tabela 7. Todas as proporções, exceto *eb*, mostraram boa estabilidade, com mais de 75% dos participantes apresentando a mesma direcionalidade no teste e no reteste, ou não-direcionalidade em ambas testagens. A menor estabilidade de *eb* é compreensível pelo fato de este índice incluir os determinantes movimento inanimado e sombreado difuso, associados a fenômenos situacionais.

Tabela 7 – Frequência e porcentagem de participantes com estabilidade em 5 proporções, numa amostra norte-americana de 100 adultos não-pacientes retestados após 3 anos

Proporção	<i>f</i> (%)
<i>EB</i>	86 (86%)
<i>EA : es</i>	82 (82%)
<i>eb</i>	58 (58%)
<i>a : p</i>	77 (77%)
<i>FC : CF + C</i>	78 (78%)

Fonte: Exner, Armbruster e Viglione (1978)

Exner (1978) apresenta diversos estudos de estabilidade temporal.

O primeiro estudo reportado é o reteste de 100 adultos não-pacientes após 3 anos, ou seja o mesmo estudo apresentado em Exner, Armbruster e Viglione (1978).

No segundo estudo³⁷, 25 adultos não-pacientes foram retestados após 7 dias. As 19 variáveis investigadas apresentaram correlações de teste-reteste com média = 0,79, desvio-padrão = 0,13 e mediana = 0,83, sendo que 13 variáveis apresentaram correlações maiores do que 0,75.

As informações sobre o estudo da direcionalidade sugerem boa estabilidade para todas as proporções, exceto *eb*: 18 participantes apresentaram direcionalidade no *EB* nas duas testagens e nenhum deles mudou de direção no reteste; somente 6 dos 25 participantes apresentaram direcionalidade no *eb* nas duas testagens; 21 apresentaram direcionalidade em

³⁷ EXNER, J. E.; BRIANT, E. A. **A Study of temporal consistency over a seven day period.** Workshops Study No. 205. Estudo não publicado. Rorschach Workshops, 1974.

EA : es em ambas testagens e nenhum deles mudou de direção no reteste; 20 apresentaram mais *FC* do que *CF+C* em ambas testagens; 6 com mais movimentos passivos do que movimentos ativos no teste continuaram a apresentar esta configuração no reteste.

No terceiro estudo³⁸, 25 adultos não-pacientes foram retestados após 60 dias. As correlações de teste-reteste foram semelhantes às obtidas para o reteste após 3 anos, com média = 0,79, desvio-padrão = 0,08 e mediana = 0,82, sendo que 13 das 19 variáveis investigadas apresentaram correlações maiores do que 0,75.

Novamente, encontraram boa estabilidade para todas as proporções exceto *eb*: 16 participantes apresentaram direcionalidade do *EB* nas duas testagens, com um deles mudando de direção no reteste; 18 apresentaram direcionalidade de *EA : es* nas duas testagens, mas um deles mudou de direção no reteste; 4 apresentaram mais movimentos passivos do que movimentos ativos no teste e permaneceram assim no reteste; 17 apresentaram mais *FC* do que *CF+C* em ambas testagens; apenas 7 apresentaram direcionalidade do *eb* em ambas testagens, sendo que um deles mudou de direção no reteste, e 15 dos 18 restantes apresentaram direcionalidade em uma das testagens, mas não na outra.

No quarto estudo³⁹, 25 pacientes ambulatoriais em lista de espera para tratamento foram retestados após 30 dias. Os resultados mostram mais variabilidade neste grupo do que nos não-pacientes. Não se pôde determinar se a maior variabilidade deveu-se à perturbação, ao erro amostral ou a um esforço deliberado para dramatizar maior desorganização de modo a obter tratamento mais rapidamente. As 19 variáveis investigadas apresentaram correlações de teste-reteste com média = 0,73, desvio-padrão = 0,12 e mediana = 0,74, sendo que 9 variáveis apresentaram correlações maiores do que 0,75. *M*, *X+%*, *Afr*, $3r+(2)/R$, *EA*, *Zf* e *R* apresentaram correlações de teste-reteste altas, semelhantes às encontrados em não-pacientes.

³⁸ EXNER, J. E.; ARMBRUSTER, G. L.; LEURA, A. V. **Temporal consistency among nonpatients over a 60 day interval.** Workshops Study No. 218. Estudo não publicado. Rorschach Workshops, 1975.

³⁹ LEURA, A. V.; WYLIE, J. R.; EXNER, J. E. **Reexamination of prospective patients who have been wait-listed during a 30 day period.** Workshops Study No. 231. Estudo não publicado. Rorschach Workshops, 1976.

Em relação à estabilidade das proporções, 15 participantes mantiveram a mesma direcionalidade no *EB* e 9 não apresentaram direcionalidade em ambas testagens; 14 apresentaram mais *CF+C* do que *FC* em ambas testagens; 7 que apresentaram mais movimentos passivos do que movimentos ativos no teste, continuaram assim no reteste; 19 apresentaram direcionalidade em *EA : es* em ambas testagens e apenas 2 deles mudaram de direção no reteste; somente 4 apresentaram direcionalidade no *eb* em ambas testagens, e 19 dos 21 restantes não apresentaram direcionalidade no teste, mas apresentaram direcionalidade no reteste.

No quinto estudo⁴⁰, 20 esquizofrênicos internados foram retestados após 10 dias. Esses pacientes estavam em observação antes de iniciarem o tratamento, mas 7 deles já estavam medicados antes da internação e a medicação foi mantida. As 19 correlações de teste-reteste pesquisadas apresentaram média = 0,80, desvio-padrão = 0,06 e mediana = 0,79, sendo que 15 variáveis apresentaram correlações iguais ou maiores do que 0,75.

A análise da direcionalidade das proporções sugere boa estabilidade, exceto para *eb*: 17 participantes apresentaram a mesma direcionalidade do *EB* no teste e no reteste; 16 apresentaram mais *CF+C* do que *FC* em ambas as testagens; 4 apresentaram mais movimentos passivos do que movimentos ativos no teste e no reteste; 11 apresentaram direcionalidade em *EA : es* em ambas testagens e os outros 9 não apresentaram direcionalidade em ambas testagens; 15 participantes apresentaram lado direito do *eb* maior no teste, mas somente 4 permaneceram com direcionalidade no reteste, embora 18 dos 20 participantes continuassem a manifestar mais *SumSH* do que *FM+m* no reteste.

No sexto estudo, compararam dois grupos de pacientes em tratamento: no primeiro grupo⁴¹, os participantes foram 35 sujeitos com depressão reativa testados até uma

⁴⁰ EXNER, J. E.; ZALIS, T.; SCHUYLER, W.; SCHUMACHER, J.; KUHN, B. **Reevaluation of newly admitted schizophrenic patients after a ten day period**. Workshops Study No. 228. Estudo não publicado. Rorschach Workshops, 1976.

⁴¹ EXNER, J. E.; WYLIE, J. R.; ARMBRUSTER, G. L. **Effects of brief treatment with anxious or depressed patients**. Workshops Study No. 216. Estudo não publicado. Rorschach Workshops, 1975.

semana antes do início de psicoterapia breve e retestados após 90 dias; no segundo grupo⁴², os participantes foram 30 sujeitos neuróticos testados até uma semana antes do início de psicoterapia dinâmica de longa duração e retestados após 180 dias. A psicoterapia breve deveria se completar em até 90 dias, mas 12 dos pacientes continuaram em terapia após este período. A psicoterapia longa não estipulava um prazo para o término do tratamento, que durou pelo menos 180 dias.

No grupo de psicoterapia breve, as 19 variáveis investigadas apresentaram correlações de teste-reteste com média = 0,73, desvio-padrão = 0,17 e mediana = 0,79, sendo que 12 variáveis apresentaram correlações maiores do que 0,75. A baixa estabilidade de algumas variáveis provavelmente indica que ocorreram mudanças em algumas características, devidas ao tratamento.

Ocorreram mudanças na direcionalidade em todas as proporções, exceto *EB*: 18 participantes que apresentaram *EB* com direcionalidade no teste mantiveram a direção no reteste; 21 apresentaram direcionalidade em *EA : es* no teste, mas 8 deles apresentaram mudança na direção no reteste; 24 apresentaram direcionalidade em *FC : CF+C* no teste, mas 12 deles mudaram de direção no reteste; 22 apresentaram direcionalidade do *eb* no teste e 15 deles continuaram a apresentar direcionalidade no reteste, mas com 10 mudando de direção; 9 apresentaram mais movimentos passivos do que movimentos ativos no teste e no reteste.

No grupo de psicoterapia longa as correlações de teste-reteste foram mais baixas, com as 19 variáveis estudadas apresentando média = 0,61, desvio-padrão = 0,16 e mediana = 0,64, sendo que apenas 4 variáveis apresentaram correlações maiores que 0,75. A correlação para *R* foi mais baixa do que nos demais estudos, o que está associado ao aumento das respostas no reteste, fator que também afetou as demais correlações.

Em relação à estabilidade das proporções, 6 participantes apresentaram mesma direcionalidade no teste e no reteste para *FC : CF+C*; 13 apresentaram direcionalidade para o *eb* no teste e no reteste, mas 8 mudaram de direção; 20 apresentaram direcionalidade para

⁴² EXNER, J. E.; WYLIE, J. R.; ARMBRUSTER, G. L. **A follow-up of patients in long term treatment after the first six months.** Workshops Study No. 223. Estudo não publicado. Rorschach Workshops, 1976.

EA : es no teste, sendo 14 com o lado esquerdo maior, e 18 continuaram a apresentar direcionalidade no reteste, mas 10 dos 14 com lado esquerdo maior mudaram de direção no reteste, e 3 dos 6 com lado direito maior também mudaram de direção; 17 apresentaram a mesma direcionalidade do *EB* no teste e no reteste, e 5 dos 11 sem direcionalidade no teste apresentaram direcionalidade no reteste; 24 apresentaram direcionalidade para *a : p* em ambas testagens, incluindo 7 com mais movimentos passivos.

Assim, concluiu-se, a partir destes dois estudos com pacientes, que algumas variáveis, especialmente as relacionadas às características afetivas (determinantes de cor e de sombreado), estiveram sujeitas a alterações decorrentes do tratamento; outras variáveis permaneceram estáveis (*Zf*, *X+%*, *Afr*, $3r+(2)/R$, *EA*, *Popular*). A direcionalidade de *EB* e de *a : p* permaneceu estável para a maioria dos indivíduos, mas a intervenção alterou *eb*, *FC : CF+C* e *EA : es*.

Haller e Exner (1985) estudaram 50 pacientes internados, com sintomas de depressão segundo os critérios do DSM-III e 15 ou mais pontos no Inventário de Depressão de Beck. O reteste foi realizado três ou quatro dias após o teste. Metade dos participantes foram retestados com o procedimento padrão e a outra metade recebeu instruções para fornecer respostas diferentes das do primeiro teste. Este segundo grupo repetiu aproximadamente 1/3 das respostas, enquanto no primeiro grupo 2/3 das respostas do reteste foram semelhantes às fornecidas na primeira testagem. Os resultados deste estudo são apresentados na Tabela 8.

Os resultados de fidedignidade de teste-reteste dos dois grupos foram semelhantes, com apenas 4 variáveis mostrando diferenças estatisticamente significantes entre os grupos: *Popular*, *Lambda*, *FM* e *m*. Os autores afirmam que as variáveis associadas à depressão apresentaram boa estabilidade, mesmo nesta amostra de indivíduos com quadros mais transitórios do que crônicos, mas notamos que as correlações encontradas nesta pesquisa, principalmente no grupo experimental, são mais baixas do que as em outros estudos: para o grupo controle, temos: média = 0,71, desvio-padrão = 0,13 e mediana = 0,74, com 13 das 28 correlações reportadas sendo iguais ou maiores do que 0,75; para o grupo experimental, temos: média = 0,66, desvio-padrão = 0,17 e mediana = 0,71, com 10 das 28 correlações reportadas sendo iguais ou maiores do que 0,75.

Tabela 8 – Estabilidade temporal em pacientes norte-americanos internados com sintomas de depressão, com o reteste realizado após 3 ou 4 dias

N Variável	Grupo controle 25	Grupo experimental 25
<i>R</i>	0,77	0,72
<i>Popular</i>	0,76	0,37
<i>M</i>	0,75	0,78
<i>FM</i>	0,28	0,72
<i>m</i>	0,84	0,34
<i>a</i>	0,71	0,78
<i>p</i>	0,51	0,68
<i>FC</i>	0,70	0,79
<i>CF</i>	0,51	0,33
<i>C</i>	0,50	0,43
<i>CF+C</i>	0,74	0,62
<i>WSumC</i>	0,68	0,69
<i>SumT</i>	0,86	0,84
<i>SumC'</i>	0,77	0,69
<i>SumV</i>	0,82	0,86
<i>SumY</i>	0,69	0,57
<i>SumSH</i>	0,74	0,73
<i>FD</i>	0,76	0,81
<i>F</i>	0,78	0,69
<i>Lambda</i>	0,82	0,40
<i>X+%</i>	0,82	0,72
<i>Afr</i>	0,57	0,67
$3r + (2) / R$	0,74	0,70
<i>Color-shading blends</i>	0,70	0,41
<i>Sum5*</i>	0,84	0,81
<i>MOR</i>	0,87	0,88
<i>EA</i>	0,71	0,76
<i>es</i>	0,69	0,78

Correlações iguais a 0,75 ou maiores
Correlações entre 0,50 e 0,74
Correlações menores do que 0,50

* Na época do estudo havia apenas 5 códigos especiais críticos.

Fonte: Haller e Exner (1985)

Exner (1986) apresenta um estudo⁴³ no qual a estabilidade temporal numa amostra de 50 adultos retestados após 1 ano foi comparada com a do estudo de 100 adultos retestados após 3 anos. Dados complementares deste estudo são apresentados em Exner (1999a). Os resultados foram semelhantes, com o grupo retestado após 1 ano apresentando correlações um pouco mais altas do que o grupo retestado após 3 anos: média das 41 correlações reportadas = 0,78, desvio-padrão = 0,14 e mediana = 0,82.

⁴³ EXNER, J. E.; THOMAS, E. E.; COHEN, J. B. **Temporal consistency of test variables for patient and nonpatient populations**. Workshops Study No. 276. Estudo não publicado. Rorschach Workshops, 1983.

Exner (1986) apresenta também dados de outro estudo⁴⁴, com 35 adultos não-pacientes retestados após 3 semanas. Para as 41 correlações de teste-reteste reportadas, temos: média = 0,81, desvio-padrão = 0,13 e mediana = 0,85.

Exner (1999a) retoma estudos anteriores, complementando os resultados. Na interpretação do Rorschach, os valores de uma variável são divididos em faixas (por exemplo, faixa de valores normativos, faixa de valores abaixo do esperado e faixa de valores acima do esperado), sendo que para cada faixa de valores corresponde uma hipótese interpretativa. Neste trabalho, Exner analisou quantos participantes se mantinham na mesma faixa de valores e quantos mudavam de faixa, denominando esta análise como estudo de consistência das categorias [interpretativas]. Os resultados são apresentados na Tabela 9.

Tabela 9 – Consistência das categorias interpretativas de 14 variáveis do Rorschach para dois grupos de adultos não-pacientes norte-americanos

			35 adultos não-pacientes				50 adultos não-pacientes			
			Teste		Reteste		Teste		Reteste	
					após 3 semanas				após 1 ano	
	A	B	A	B	A	B	A	B	A	B
<i>EB^a</i>	Introversivo	Extratensivo	16	13	16	13	22	18	21 ¹	18
<i>Lambda</i>	< 1,0	≥ 1,00	28	7	30 ²	5 ²	42	8	43 ¹	7 ¹
<i>EA</i>	6,5 a 10,5	≥ 11,0	21	8	19 ²	10 ²	35	11	33 ²	13 ²
<i>EA</i>	< 6,5	≥ 6,5	6	29	5 ¹	30 ¹	4	46	4	46
<i>es^b</i>	4 a 9	≥ 10	25	6	23 ²	8 ²	43	5	42 ¹	6 ¹
<i>Adjes^b</i>	4 a 9	≥ 10	32	2	32	2	46	3	43 ³	6 ³
<i>FC: CF+C</i>	equilibrado	desequilibrado	19	16	20 ¹	15 ¹	27	23	26 ¹	24 ¹
<i>Afr</i>	< 0,55	≥ 0,55	6	29	7 ¹	28 ¹	7	43	9 ²	41 ²
<i>X+%</i>	≥ 0,60	< 0,60	31	4	32 ¹	3 ¹	47	3	47	3
<i>Xu%</i>	< 0,21	≥ 0,21	19	16	18 ¹	17 ¹	38	12	36 ²	14 ²
<i>X-%</i>	< 0,20	≥ 0,20	32	3	29 ³	6 ³	44	6	43 ¹	7 ¹
<i>a : p</i>	p ≤ a+1	p > a+1	33	2	33	2	47	3	46 ¹	4 ¹
<i>WSum6</i>	< 7,0	≥ 7,0	29	6	30 ¹	5 ¹	43	7	43	7
<i>COP</i>	0 ou 1	≥ 2	12	23	11 ¹	24 ¹	17	33	18 ¹	32 ¹
<i>MOR</i>	< 3	≥ 3	31	4	32 ¹	3 ¹	44	6	46 ²	4 ²
<i>3r+(2)/R^c</i>	0,33 a 0,43	≥ 0,44	22	9	22	9	31	10	32 ¹	9 ¹
<i>3r+(2)/R</i>	< 0,33	≥ 0,33	4	31	4	31	6	44	4 ²	46 ²

Sobrescrito o número de participantes que mudaram de categoria do teste para o reteste.

^a ambíguais foram excluídos; ^b valores menores do que 4 foram excluídos; ^c valores menores que 0,33 foram excluídos.

Fonte: Exner (1999a)

⁴⁴ THOMAS, E. E.; ALINSKY, D.; EXNER, J. E. **The stability of some Rorschach variables in 9 year-olds as compared with nonpatient adults**. Workshops Study No. 441. Estudo não publicado. Rorschach Workshops, 1982.

A Tabela 10 reúne dados de diferentes estudos de teste-reteste com adultos não-pacientes e a Tabela 11 dados de estudos com adultos pacientes, com as variáveis agrupadas por aspectos do funcionamento psicológico: complexidade, estilos, controles; afeto; ideiação; mediação; processamento; interpessoal; auto-percepção. (Algumas pesquisas são apresentadas em diversos artigos ou livros do Sistema Compreensivo, e algumas das correlações de uma mesma pesquisa apresentam valores diferentes nestas várias apresentações. Tratam-se, provavelmente de erros ao editar o texto. Nestes casos, apresentamos os resultados que se repetem em maior número de trabalhos ou, quando apresentados em apenas dois trabalhos, os valores menores.)

Os diversos estudos com adultos não-pacientes permitem identificar um padrão das correlações de teste-reteste: a maioria das variáveis apresenta correlações em torno de 0,80, sendo então variáveis que avaliam traços de personalidade; movimento inanimado e sombreado difuso apresentam correlações na faixa de 0,20 a 0,40, refletindo estados de personalidade. As variáveis das quais estes últimos determinantes participam (*FM+m*, *SumSH* e *es*) apresentam correlações na faixa de 0,50 a 0,70 e esta menor estabilidade se deve justamente à presença do sombreado difuso e do movimento inanimado. Algumas variáveis (*FM*, *SumC'*, *CF*, *Blends*), com correlações nas faixas de 0,60 e 0,70, são predominantemente traço, embora também apresentem um componente estado.

Em comparação com os não-pacientes, as amostras dos pacientes mostram coeficientes de teste-reteste com valores menos semelhantes entre si para as variáveis *Lambda*, *WSumC*, *EA*, *FM*, *SumSH*, *FC*, *CF*, *C*, *Afr*, *movimento passivo* e *Índice de Egocentricidade*. Esta maior variabilidade das estimativas de estabilidade temporal mais provavelmente decorre de erro amostral, visto que as amostras são de tamanho pequeno. Os pacientes também apresentam coeficientes com valores menores do que os não-pacientes, indicando menor estabilidade em seu funcionamento psicológico. Chama particularmente a atenção a menor estabilidade de *Lambda* e das *respostas de cor* (*FC*, *CF*, *CF+C* e *WSumC*), variáveis associadas ao grau de simplificação perceptiva e à expressão do afeto, respectivamente.

Tabela 10 – Estabilidade temporal em 5 amostras de adultos não-pacientes norte-americanos

		Grupo				
		1	2	3	4	5
N		100	50	25	35	25
Intervalo		> 3 anos	> 1 ano	60 dias	3 semanas	7 dias
Variáveis						
Complexidade, estilo, controles						
<i>R</i>		0,79	0,86	0,84	0,84	0,86
<i>Lambda</i>		0,82	0,78	0,86	0,76	0,73
<i>EB</i>						
	<i>M</i>	0,87	0,84	0,85	0,83	0,81
	<i>WSumC</i>	0,86	0,82	0,78	0,83	0,85
<i>EA</i>		0,85	0,83	0,83	0,84	0,83
<i>FM</i>		0,72	0,77	0,74	0,72	0,63
<i>m</i>		0,39	0,26		0,34	0,28
<i>SumC'</i>		0,67	0,73		0,67	
<i>SumV</i>		0,81	0,87		0,89	
<i>SumT</i>		0,87	0,91		0,96	
<i>SumY</i>		0,23	0,31		0,41	
<i>eb</i>						
	<i>FM+m</i>	0,69		0,67		0,56
	<i>SumSH</i>	0,66		0,59	0,71	0,51
<i>es</i>		0,72	0,64	0,69	0,59	0,62
<i>Adjes</i>			0,82		0,79	
<i>D-score</i>		0,83	0,91		0,88	
Afeto						
<i>FC</i>		0,86	0,86	0,83	0,92	0,93
<i>CF</i>		0,66	0,58		0,68	
<i>CF+C</i>		0,79	0,81	0,73	0,83	0,82
<i>C</i>		0,51	0,56		0,59	
<i>Afr</i>		0,90	0,82	0,89	0,85	0,93
<i>Blends</i>		0,67	0,62		0,71	
Ideação						
<i>a</i>		0,86	0,83	0,82	0,87	0,91
<i>p</i>		0,75	0,72	0,78	0,85	0,84
<i>2AB+Art+Ay</i>		0,88	0,84			
<i>Sum6</i>		0,88 ^a	0,81		0,81	
<i>WSum6</i>			0,86		0,86	
<i>DV+DR</i>		0,79	0,72		0,74	
<i>INC+FAB</i>		0,82	0,89		0,92	
<i>ALOG</i>			0,90		0,93	
Mediação						
<i>XA%</i>			0,89			
<i>WDA%</i>			0,92			
<i>Popular</i>		0,73	0,83	0,84	0,81	0,78
<i>X+%</i>		0,80	0,86	0,84	0,87	0,88
<i>Xu%</i>			0,85		0,89	
<i>X-%</i>		0,86	0,92		0,88	
Processamento						
<i>Zf</i>		0,83	0,85	0,81	0,89	0,88
Interpessoal						
<i>COP</i>			0,81		0,88	
<i>AG</i>			0,82		0,81	
<i>Isolate/R</i>			0,84		0,83	
Auto-percepção						
<i>3r + (2)/R</i>		0,87	0,89	0,85	0,90	0,91
<i>Fr + rF</i>		0,78	0,82		0,89	
<i>FD</i>		0,83	0,88		0,90	
<i>MOR</i>			0,71		0,83	

Grupo 1 = Exner, Armbruster e Viglione (1978); Exner (1978, 1986, 1993, 2003); Exner e Weiner (1995); Viglione e Hilsenroth (2001).

Grupo 2 = Exner (1986, 1993, 1999a, 2003); Viglione e Hilsenroth (2001). Grupo 3 = Exner (1978). Grupo 4 = Exner (1986, 1993, 1999a, 2003); Viglione e Hilsenroth (2001); Grupo 5 = Exner (1978, 1980).

^a No estudo 1, Sum6 refere-se à soma dos 5 códigos especiais críticos que existiam na época.

	Correlações menores do que 0,50
	Correlações entre 0,50 e 0,74
	Correlações iguais a 0,75 ou maiores

Tabela 11 – Estabilidade temporal em 5 amostras de adultos norte-americanos pacientes

		Grupo				
N		1	2	3	4	5
Intervalo		30	35	25	35	20
Participantes		180 dias	90 dias	30 dias	3 dias	10 dias
		Pacientes neuróticos em psicoterapia longa	Pacientes com depressão reativa em psicoterapia breve	Pacientes em lista de espera	Pacientes adultos deprimidos	Esquizofrênicos internados
Variáveis						
Complexidade, estilo, controles						
<i>R</i>		0,64	0,83	0,83	0,77	0,89
<i>Lambda</i>		0,48	0,64	0,65	0,82	0,76
<i>EB</i>						
	<i>M</i>	0,68	0,82	0,82	0,75	0,77
	<i>WSumC</i>	0,38	0,49	0,68	0,68	0,81
<i>EA</i>		0,70	0,84	0,77	0,71	0,82
<i>FM</i>		0,49	0,73	0,74	0,28	0,75
<i>m</i>					0,84	
<i>SumC'</i>					0,77	
<i>SumV</i>					0,82	
<i>SumT</i>					0,86	
<i>SumY</i>					0,69	
<i>eb</i>						
	<i>FM+m</i>	0,41	0,70	0,62		0,73
	<i>SumSH</i>	0,42	0,28	0,49	0,74	0,72
<i>es</i>		0,61	0,76	0,67	0,69	0,71
Afeto						
<i>FC</i>		0,39	0,51	0,71	0,70	0,84
<i>CF</i>					0,51	
<i>CF+C</i>		0,47	0,44	0,57	0,74	0,89
<i>C+Cn</i>					0,50	
<i>Afr</i>		0,73	0,89	0,86	0,57	0,79
Ideação						
<i>a</i>		0,81	0,87	0,78	0,71	0,86
<i>p</i>		0,71	0,79	0,84	0,51	0,82
<i>Sum6</i>					0,84 ^a	
Mediação						
<i>Popular</i>		0,76	0,81	0,69	0,76	0,78
<i>X+%</i>		0,81	0,84	0,83	0,82	0,92
Processamento						
<i>Zf</i>		0,81	0,86	0,86		0,80
Interpessoal						
<i>Isolate/R</i>						
Auto-percepção						
<i>3r+(2)/R</i>		0,73	0,91	0,91	0,74	0,75
<i>FD</i>					0,76	
<i>MOR</i>					0,87	

Grupo 1 = Exner (1978). Grupo 2 = Exner (1978). Grupo 3 = Exner (1978). Grupo 4 = Haller e Exner (1985); Viglione e Hilsenroth (2001). Grupo 5 = Exner (1978).

^a no momento da publicação existiam somente 5 códigos especiais críticos.

	Correlações menores do que 0,50
	Correlações entre 0,50 e 0,74
	Correlações iguais a 0,75 ou maiores

Estudos também foram realizados por outros pesquisadores que não fazem parte do grupo de Exner.

Schwartz, Mebane e Malony (1990) aplicaram o Rorschach a 24 estudantes universitários (college students) surdos. Metade dos participantes respondeu ao teste por escrito e ao reteste usando a linguagem de sinais; a outra metade respondeu ao teste usando a linguagem de sinais e ao reteste por escrito. O reteste foi realizado aproximadamente uma semana após o teste. As correlações de teste-reteste encontradas foram menores do que os relatados por Exner: as 25 variáveis estudadas apresentaram correlações com média = 0,51, desvio-padrão = 0,25 e mediana = 0,62. Devemos lembrar que este não é um estudo simples de teste-reteste, porque existe mudança na forma de aplicação utilizada no teste e no reteste, fator que pode ser responsável pelas correlações menores. De fato, os autores notaram maior facilidade para os participantes responderem utilizando a linguagem de sinais; quando responderam por escrito, forneceram respostas mais breves, que frequentemente não continham palavras-chave que sugerissem o uso de certos determinantes; os protocolos obtidos por escrito apresentaram menos respostas de sombreado do que os protocolos com as respostas fornecidas pela linguagem de sinais.

Adair e Wagner (1992) estudaram 50 pacientes ambulatoriais com esquizofrenia. Encontraram baixa estabilidade temporal para os códigos especiais críticos (*DV*, *DR*, *INC*, *FAB*, *ALOG*) e para sua soma ponderada (*WSum6*): 0,50, 0,37, 0,26, 0,38, 0,16 e 0,385 respectivamente. Os autores apontam certas limitações deste estudo: alguns pacientes encontravam-se em psicoterapia, o que pode ter contribuído para mudanças e, conseqüentemente, estabilidade temporal mais baixa; a amostra foi composta por pacientes ambulatoriais, portanto casos mais leves e que apresentam menos falhas de pensamento do que pacientes mais graves – assim, a menor variabilidade dos escores pode ter contribuído para a baixa estabilidade temporal; a amostra também incluiu protocolos com menos de 14 respostas, considerados problemáticos por serem pouco fidedignos e, conseqüentemente inválidos (ver adiante a discussão sobre a fidedignidade dos protocolos com poucas respostas); a baixa estabilidade temporal também pode ser atribuída a oscilações associadas à esquizofrenia. Viglione e Hilsenroth (2001) comentam que a amostra deste estudo é bastante heterogênea, com pacientes esquizofrênicos com diferentes cursos da doença, medicação, hospitalizações e psicoterapia, e os intervalos entre teste e reteste foram também bastante

variados: de 1 a 18 anos. Todos estes fatores podem ter contribuído para as baixas correlações de teste-reteste encontradas.

Erstad (1996) investigou a menor frequência de movimento humano e respostas de cor em idosos. Seu grupo alvo foi composto por 33 indivíduos com mais de 60 anos. Utilizou também um grupo de comparação, composto por 33 adultos com idades de 18 a 55 anos. Parte da pesquisa envolvia a fidedignidade de teste-reteste e o autor obteve 17 protocolos de reteste no grupo alvo e 11 protocolos de reteste no grupo de comparação. Os resultados deste estudo são apresentados na Tabela 12. As correlações de teste-reteste encontradas foram baixas: para as 26 correlações informadas temos: no grupo de idosos, média = 0,60, desvio-padrão = 0,23 e mediana = 0,68; no grupo de adultos com até 55 anos de idade, média = 0,54, desvio-padrão = 0,22 e mediana = 0,54. As distribuições de valores das variáveis do SCR no grupo de comparação foram semelhante a dos adultos não-pacientes da amostra normativa norte-americana, portanto as menores correlações não decorrem de diferenças nas distribuições dos valores das variáveis

Tabela 12 – Estabilidade temporal em duas amostras de adultos norte-americanos

Grupo	> 60 anos	18-55 anos	Grupo	> 60 anos	18-55 anos
N	17	11	N	17	11
<i>R</i>	0,90	0,82	<i>WSumC</i>	0,75	0,53
<i>Popular</i>	0,77	0,36	<i>SumT</i>	0,70	0,92
<i>Zf</i>	0,70	0,67	<i>SumC'</i>	0,67	0,02
<i>F</i>	0,80	0,58	<i>SumY</i>	0,40	0,39
<i>M</i>	0,75	0,56	<i>SumV</i>	0,57	0,85
<i>FM</i>	0,68	0,40	<i>Lambda</i>	0,46	0,23
<i>M</i>	0,45	0,73	<i>X+%</i>	0,63	0,47
<i>a</i>	0,72	0,49	<i>Afr</i>	0,48	0,79
<i>p</i>	0,75	0,64	<i>3r+(2)/R</i>	0,60	0,68
<i>FC</i>	0,08	0,81	<i>EA</i>	0,73	0,44
<i>CF</i>	0,42	0,19	<i>es</i>	0,80	0,68
<i>C</i>	0,95	0,44	<i>D-score</i>	0,22	0,54
<i>CF+C</i>	0,62	0,39	<i>AdjD</i>	0,02	0,53

Fonte: Erstad (1996)

Erstad (1996) avaliou também a consistência das categorias interpretativas. Os resultados, apresentados na Tabela 13, foram mais baixos do que os de Exner (1991)⁴⁵.

Tabela 13 – Frequência e porcentagem de participantes com proporções estáveis em uma amostra de 8 participantes norte-americanos

Proporção	f (%)	Proporção	f (%)
<i>EB</i>	5 (63%)	<i>Lambda</i>	7 (88%)
<i>eb</i>	4 (50%)	<i>X+%</i>	5 (63%)
<i>D-score</i>	2 (25%)	<i>X-%</i>	6 (75%)
<i>AdjD</i>	7 (88%)	<i>Zd</i>	5 (63%)
<i>a : p</i>	5 (63%)	<i>W : D</i>	3 (38%)
<i>Ma : Mp</i>	7 (88%)	<i>W : M</i>	5 (63%)
<i>FC : CF+C</i>	4 (50%)	<i>HumanContent</i>	5 (63%)
<i>Afr</i>	6 (75%)	<i>Isolate/R</i>	5 (63%)
<i>EGO</i>	4 (50%)	<i>SumT</i>	7 (88%)

Fonte: Erstad (1996)

O autor afirma:

[...] este pesquisador notou que os participantes variaram amplamente em termos da riqueza dos protocolos fornecidos mesmo quando apresentaram mais do que 13 respostas. Parece não haver um padrão nesta variabilidade de uma sessão de testagem para a outra. Em alguns casos, o primeiro Rorschach mostrou-se defendido e escasso. No entanto, o reteste foi uma experiência de maior envolvimento e o Rorschach também mostrou o efeito deste envolvimento. Em outros casos, a primeira testagem foi uma experiência na qual os participantes estiveram mais envolvidos com o processo, enquanto que o reteste não se mostrou estruturalmente rico nem tão revelador num sentido qualitativo. (ERSTAD, 1996, p. 13).

Viglione e Hilsenroth (2001) criticam este estudo Erstad pelo número reduzido de participantes e pelo intervalo entre teste e reteste não ser especificado.

⁴⁵ EXNER, J. E. JR. **The Rorschach: A Comprehensive System**. Vol. 2: Interpretation, 2nd edition. New York: Willey, 1991.

O estudo de Exner (1991) mencionado por Erstad é o mesmo publicado em Exner (1988) e que será apresentado na parte em que discutiremos a estabilidade temporal e os protocolos com poucas respostas.

Saether e Berg (1999)⁴⁶ realizaram uma replicação do estudo de Haller e Exner (1985), com 66 indivíduos. O reteste foi realizado após de 1 a 17 dias e nele os participantes receberam instruções para fornecer respostas diferentes das do teste. Os autores encontraram mudanças importantes entre os resultados do teste e do reteste (mediana das correlações = 0,25). Comentando este estudo, Grønnerød (2003a) menciona que os autores também observaram que solicitar respostas diferentes enfatiza a tarefa de solucionar problemas, na qual testandos com mais instrução obtêm mais sucesso do que testandos com menos instrução.

Garb et al. (2001) comentam que o Sistema Compreensivo apresenta informações de fidedignidade de teste-reteste para apenas 40% das variáveis e que estudos de pesquisadores que não os do grupo de Exner encontraram correlações menores. Afirmam que não se poderia concluir que a maioria das variáveis apresenta excelente consistência temporal, como alegado por proponentes do SCR.

Estes críticos do Rorschach têm certa razão (ver Tabelas 10 e 11): algumas poucas variáveis são apresentadas em todos os estudos, enquanto que as demais são apresentadas em alguns estudos, mas não em outros – o que dificulta as comparações dos resultados de diferentes estudos. Além disso, algumas variáveis do Sumário Estrutural não têm suas correlações de teste-reteste informadas em nenhum dos estudos.

Meyer e Archer (2001) respondem à observação de Garb et al. (2001) de que os estudos de Exner encontraram maior estabilidade temporal do que os estudos de outros pesquisadores. Afirmam que estes últimos estudos apresentam alguns problemas metodológicos, tais como a falta de informações precisas sobre os intervalos entre o teste e o reteste, e sobre a ocorrência de intervenções entre as testagens, o que dificulta a comparação dos resultados.

Meyer e Archer (2001) compararam os dados de seis amostras de adultos com dados completos e intervalos entre o teste e o reteste de até um mês. As pesquisas são bastante diversificadas em termos da metodologia, mas todas apresentam o problema de serem amostras de tamanho relativamente pequeno. Três amostras são estudos de Exner e as outras

⁴⁶ As informações mencionadas provêm do resumo em inglês, que os autores gentilmente nos enviaram.

três são de outros grupos de pesquisadores: 1) Schwartz, Mebane e Malony (1990); 2) e 3) as duas amostras correspondem ao estudo de Erstad (1996), sendo a primeira de indivíduos com até 55 anos, e a segunda de indivíduos com 60 anos ou mais; 4) e 5) as duas amostras correspondem ao estudo de Haller e Exner (1985), sendo que uma delas recebeu as instruções originais e a outra recebeu instruções para fornecer respostas diferentes no reteste; 6) amostra de Exner (1993), composta por 35 não-pacientes retestados após 3 semanas. A Tabela 14 apresenta os resultados de 23 variáveis informadas nestas seis amostras.

Tabela 14 – Estabilidade temporal em 6 amostras norte-americanas

Grupo	1	2	3	4	5	6	Média	DP
N	24	11	17	25	25	35	Ponderada	
Reteste	7 dias	24 dias	?	4 dias	4 dias	21 dias		
<i>R</i>	0,70	0,93	0,90	0,72	0,77	0,84	0,80	0,10
<i>M</i>	0,66	0,75	0,75	0,78	0,75	0,83	0,76	0,06
<i>FM</i>	0,62	0,55	0,68	0,72	0,28	0,72	0,60	0,17
<i>m</i>	0,64	0,73	0,45	0,34	0,84	0,34	0,53	0,21
<i>a</i>	0,64	0,53	0,72	0,78	0,71	0,87	0,74	0,12
<i>p</i>	0,67	0,75	0,75	0,68	0,51	0,85	0,70	0,11
<i>FC</i>	0,61	0,80	0,08	0,79	0,70	0,92	0,69	0,30
<i>CF</i>	0,33	0,31	0,42	0,33	0,51	0,68	0,46	0,14
<i>C</i>	0,28	0,44	0,95	0,43	0,50	0,59	0,52	0,23
<i>CF+C</i>	0,26	0,48	0,62	0,62	0,74	0,83	0,62	0,20
<i>WSumC</i>	0,40	0,62	0,75	0,69	0,68	0,83	0,67	0,15
<i>SumT</i>	-0,07	0,92	0,70	0,84	0,86	0,96	0,70	0,39
<i>SumC'</i>	0,46	0,15	0,67	0,69	0,77	0,67	0,61	0,23
<i>SumY</i>	0,40	0,07	0,40	0,57	0,69	0,41	0,46	0,21
<i>SumV</i>	0,00	0,82	0,57	0,86	0,82	0,89	0,67	0,34
<i>F</i>	0,80	0,80	0,80	0,69	0,78	0,76	0,77	0,04
<i>Lambda</i>	0,12	0,26	0,46	0,40	0,82	0,76	0,52	0,28
<i>Popular</i>	0,76	0,40	0,77	0,37	0,76	0,81	0,67	0,20
<i>X+%</i>	0,83	0,57	0,63	0,72	0,82	0,87	0,77	0,12
<i>Afr</i>	0,24	0,78	0,48	0,67	0,57	0,85	0,61	0,22
<i>EGO</i>	0,79	0,74	0,60	0,70	0,74	0,90	0,76	0,10
<i>EA</i>	0,64	0,61	0,73	0,76	0,71	0,84	0,73	0,08
<i>es</i>	0,66	0,63	0,80	0,78	0,69	0,59	0,68	0,08
Média	0,50	0,60	0,64	0,65	0,70	0,77	0,66	0,18
DP	0,26	0,24	0,19	0,16	0,14	0,16	0,10	0,09
Mediana	0,62	0,62	0,68	0,69	0,74	0,83	0,69	0,17

Grupo 1 = Schwartz, Mebane e Malony (1990); grupo 2 = adultos não-pacientes com até 55 anos de idade (Erstad, 1996); grupo 3 = adultos não-pacientes com 60 anos ou mais (Erstad, 1996); grupo 4 = grupo experimental de Haller e Exner (1985); grupo 5 = grupo controle de Haller e Exner (1985); grupo 6 = Exner (1993).

	Correlações menores do que 0,50
	Correlações entre 0,50 e 0,74
	Correlações iguais a 0,75 ou maiores

Fonte: Meyer e Archer (2001).

De modo geral, as amostras de Exner produziram correlações um pouco maiores. Os coeficientes variaram consideravelmente de amostra para amostra, algumas vezes sendo maiores nas pesquisas de Exner, outras vezes nas pesquisas dos outros autores. Esta variação nas correlações está associada ao erro de amostragem, pois amostras de tamanho pequeno fornecem estimativas não generalizáveis, principalmente para variáveis infreqüentes (MEYER; ARCHER, 2001).

De fato, notamos que a maioria dos estudos de estabilidade temporal utilizou amostras de tamanho pequeno (aproximadamente 30 participantes). Partindo das correlações obtidas numa amostra, podemos determinar intervalos de confiança, que indicam os valores entre os quais provavelmente se situam as correlações na população. Uma correlação $r = 0,70$, obtida numa amostra de 30 participantes, apresenta como intervalo de confiança (com 95% de chance de acerto) valores de $r = 0,45$ a $r = 0,85$. Mesmo numa amostra de 100 participantes, a mesma correlação $r = 0,70$ apresenta como intervalo de confiança (com 95% de chance de acerto) valores de 0,58 a 0,79.

Efetivamente, para estimar de modo seguro a estabilidade temporal, uma pesquisa demandaria uma amostra com número bem maior do que 100 participantes. Discutindo as pesquisas sobre estabilidade temporal das medidas de personalidade, Watson (2004) sugere a utilização de amostras de no mínimo 300 participantes. Evidentemente, tal número pode ser conseguido em pesquisas com inventários de personalidade – instrumentos em que o participante responde às perguntas por conta própria, sem necessitar da presença de um aplicador –, mas dificilmente será obtido em pesquisas com o Rorschach.

De qualquer modo, é importante atentar para este problema do tamanho das amostras e a imprecisão das estimativas de estabilidade temporal. De fato, com amostras de tamanho pequeno, as diferenças entre as estimativas de estabilidade temporal podem ser mera decorrência do acaso na seleção das amostras. Apesar das limitações decorrentes do uso de amostras pequenas, a realização de estudos de estabilidade temporal com amostras deste tamanho justifica-se pelas características do Rorschach – instrumento que exige aplicação individual, assim demandando muito tempo ou muitos aplicadores para a coleta de amostras maiores – e também porque os diferentes estudos podem posteriormente ser reunidos numa metanálise.

Grønnerød (2003a) realizou uma metanálise de estudos sobre a estabilidade temporal do Rorschach. Seu levantamento bibliográfico coletou artigos de 1921 a 2002, incluindo artigos sobre seguimento em psicoterapia, pois alguns destes estudos incluíram grupos-controle, apresentando então dados de teste-reteste. A estabilidade temporal encontrada foi alta para todos os estudos, incluindo os que não utilizaram o Sistema Compreensivo: os níveis de estabilidade temporal ponderada combinada para um período de reteste médio de um pouco mais de 3 anos variaram de 0,68 a 0,73 nos diferentes conjuntos de dados analisados, sem considerar os moderadores sistematicamente relacionados à estabilidade. O Sistema Compreensivo mostrou maior estabilidade do que os outros sistemas e correlações mais semelhantes nas diferentes amostras do que as correlações entre as amostras de outros sistemas. Das 39 variáveis do Sistema Compreensivo analisadas, somente 7 apresentaram correlações menores do que 0,70, sendo que 3 delas são índices que incluem sombreado difuso e movimento inanimado; a estabilidade temporal combinada de 36 variáveis do Sistema Compreensivo foi $r = 0,79$. Grønnerød (2003a, p. 286) comenta ainda que “problemas metodológicos e restrições, tais como o tamanho e as características da amostra, são em muitos casos a razão mais importante para os baixos níveis de estabilidade [encontrados nas pesquisas]”. Em relação a pacientes versus não-pacientes, poder-se-ia pensar que as flutuações dos sintomas nos vários tipos de psicopatologia levariam a menor estabilidade, mas os estudos mostraram que isto não ocorreu em proporção importante.

O autor observa que as correlações de teste-reteste são comparações de grupos, mas outra questão refere-se à estabilidade de cada indivíduo – pode ocorrer uma significativa variabilidade nos resultados de um indivíduo, do teste para o reteste, variabilidade esta que não se percebe ao analisar o grupo. Por outro lado, existem as análises de Exner em relação à direcionalidade dos índices e à consistência das categorias interpretativas, que indicam que poucos participantes mudam em termos da configuração das variáveis centrais na interpretação do Rorschach – mas mais estudos deste tipo são necessários, com mais variáveis do Sumário Estrutural.

Grønnerød (2006) publicou uma revisão de sua metanálise sobre a estabilidade temporal do Rorschach, pois identificou problemas metodológicos. Utilizando uma metodologia mais adequada, os resultados foram bastante semelhantes aos originais. A

revisão também levou em conta o estudo realizado pelo autor sobre mudanças devidas à psicoterapia avaliadas pelo Rorschach (Grønnerød, 2003b) e chegou à seguinte conclusão:

[...] como os modelos de regressão são psicometricamente mais robustos do que anteriormente, a convergência dos resultados de estabilidade e os resultados do estudo sobre psicoterapia fortalecem a evidência relativa à estabilidade temporal do método de Rorschach, bem como de seu uso como um instrumento de análise da mudança devida à psicoterapia. Resultados dos dois conjuntos de dados claramente mostram que, com suficiente atenção às questões metodológicas, o método de Rorschach é tão estável quanto outros métodos de avaliação e pode captar as mudanças de personalidade tão bem quanto outros métodos”. (GRØNNERØD, 2006, p. 225).

Sultan et al. (2006) investigaram a estabilidade temporal das variáveis do Sistema Compreensivo do Rorschach em 75 adultos não-pacientes franceses. Os participantes foram indivíduos do estudo normativo francês e que concordaram em realizar o reteste após 3 meses. Além do Rorschach, foi aplicado o Questionário de Saúde Geral de Goldberg (General Health Questionnaire – GHQ), e participantes com pontuações indicando problemas de saúde mental foram excluídos da amostra.

Os autores calcularam correlações produto-momento de Pearson e correlações de postos de Spearman para 47 variáveis principais do SCR e também para o total do Índice de Depressão (*DEPI*) e da Constelação de Suicídio (*S-CON*). Os resultados deste estudo são apresentados na Tabela 15. Obtiveram correlações com valores bem menores do que as das pesquisas norte-americanas: média = 0,53, desvio-padrão = 0,15 e mediana = 0,51. Apenas nove variáveis mostraram correlações acima de 0,70.

Tabela 15 – Estabilidade temporal numa amostra de 75 adultos não-pacientes franceses

Variável	<i>r</i>	Variável	<i>r</i>	Variável	<i>r</i>
<i>R</i>	0,75	<i>SumT</i>	0,56	<i>Adjes</i>	0,46
<i>Popular</i>	0,54	<i>SumC'</i>	0,38	<i>D-score</i>	0,34
<i>Zf</i>	0,76	<i>SumY</i>	0,17	<i>XA%</i>	0,49
<i>Zd</i>	0,46	<i>SumV</i>	0,46	<i>WDA%</i>	0,49
<i>F</i>	0,70	<i>SumSH</i>	0,42	<i>X+%</i>	0,45
<i>M</i>	0,76	<i>FD</i>	0,51	<i>X-%</i>	0,55
<i>FM</i>	0,48	<i>Fr+rF</i>	0,65	<i>Xu%</i>	0,51
<i>m</i>	0,47	(2)	0,77	<i>Afr</i>	0,57
<i>FM+m</i>	0,50	<i>DV+DR</i>	0,26	<i>3r+(2)/R</i>	0,78
<i>a</i>	0,61	<i>INC+FAB</i>	0,61	<i>Sum6</i>	0,50
<i>p</i>	0,55	<i>COP</i>	0,38	<i>WSum6</i>	0,56
<i>FC</i>	0,61	<i>AG</i>	0,45	<i>Blends</i>	0,63
<i>CF</i>	0,47	<i>MOR</i>	0,45	<i>2AB+Art+Ay</i>	0,53
<i>C</i>	0,08	<i>Lambda</i>	0,72	<i>Isolate/R</i>	0,67
<i>CF+C</i>	0,55	<i>EA</i>	0,77	<i>DEPI</i>	0,28
<i>WSumC</i>	0,69	<i>es</i>	0,46	<i>S-CON</i>	0,47
<i>S</i>	0,70				

Fonte: Sultan et al. (2006)

Sultan et al. (2006) investigaram também a influência que a assimetria da distribuição de valores das variáveis pode desempenhar nas correlações, comparando as correlações obtidas com as variáveis originais versus as correlações obtidas com as variáveis transformadas de modo a apresentarem uma distribuição menos assimétrica⁴⁷. A transformação das variáveis contribuiu para gerar correlações maiores para algumas variáveis. Portanto, algumas correlações baixas estão associadas à distribuição assimétrica das variáveis. A assimetria também pode ser responsável pela diferença nos valores das correlações de Pearson em comparação com as correlações de postos de Spearman.

Investigaram também a direcionalidade das proporções e a consistência das categorias interpretativas (ver Tabelas 16 e 17). Fornecem as seguintes estatísticas: porcentagem de casos consistentes, Φ (Phi), V de Cramer e kappa⁴⁸. Encontraram resultados que indicam menor estabilidade do que os reportados nas pesquisas de Exner.

⁴⁷ A questão de a assimetria interferir nas correlações será discutida a seguir, na seção que aborda as questões estatísticas que interferem com as estimativas de estabilidade temporal.

⁴⁸ Phi é uma medida de correlação entre duas variáveis categóricas numa tabela 2x2 e seus valores variam de 0 (correlação nula) a 1 (correlação perfeita) (STATSOFT, 2007). O V de Cramer é um coeficiente de correlação para dados categóricos, interpretado da mesma forma que o r de Pearson (DANCEY; REIDY, 2006).

Tabela 16-- Direcionalidade das proporções numa amostra de 75 adultos não-pacientes franceses

Teste	Reteste			%	V	kappa
	E	S	D			
<i>EB</i>						
	E	20	5	1	76,9	
	S	8	14	10	43,8	0,43***
	D	2	5	10	58,8	0,38***
<i>EA : es</i>						
	E	5	3	7	33,3	
	S	2	4	5	36,4	0,21
	D	4	14	31	63,3	0,16
<i>eb</i>						
	E	16	11	7	47,1	
	S	8	9	6	39,1	0,25
	D	2	6	10	55,6	0,20*
<i>a : p</i>						
	E	20	6	2	71,4	
	S	5	7	8	35,0	0,35**
	D	6	11	10	37,0	0,24**
<i>FC : CF+C</i>						
	E	11	8	4	47,8	
	S	9	23	4	63,9	0,21
	D	3	10	3	18,8	0,17*

E = direcionalidade à esquerda; S = sem direcionalidade; D = direcionalidade à direita

* p < 0,05 ** p < 0,01 ***p < 0,001.

Fonte: Sultan et al. (2006)

Tabela 17 – Consistência das categorias interpretativas das variáveis do SCR numa amostra de 75 adultos não-pacientes franceses

Categoria teste	no	Reteste			%	V	kappa
		Primeira categoria	Segunda categoria	Terceira categoria			
<i>EB</i> introversivo	16		5	1	72,7		
<i>EB</i> ambigüal	7		25	5	67,6	0,46***	0,45***
<i>EB</i> extratensivo	2		6	8	50,0		
<i>EB</i> introversivo [#]	8		4	1	61,5		
<i>EB</i> ambigüal [#]	3		11	4	61,1	0,38*	0,34**
<i>EB</i> extratensivo [#]	1		5	5	45,5		
<i>D-score</i> < 0	22		12	1	62,9		
<i>D-score</i> = 0	12		16	4	50,0	0,26*	0,20*
<i>D-score</i> > 0	1		5	2	25,0		
<i>Zd</i> < -3,0	12		12	0	50,0		
<i>Zd</i> = -3,0 a 3,0	6		29	7	69,0	0,37***	0,29**
<i>Zd</i> > + 3,0	0		5	4	44,4		
<i>T</i> = 0	24		7	3	70,6		
<i>T</i> = 1	9		8	7	33,3	0,31**	0,27**
<i>T</i> > 1	4		5	8	47,1		
<i>EGO</i> < 0,33	30		12	1	69,8		
0,33 ≤ <i>EGO</i> ≤ 0,44	5		9	3	52,9	0,56***	0,48***
<i>EGO</i> > 0,44	1		2	12	80,0		

[#] somente participantes com Lambda <1,00

* p < 0,05 ** p < 0,01 ***p < 0,001.

Fonte: Sultan et al. (2006)

As Constelações são conjuntos de sinais do SCR. O Índice de Percepção e Pensamento (Perceptual-Thinking Index – *PTI*) identifica problemas na percepção e pensamento. O *PTI* substituiu o Índice de Esquizofrenia (Schizophrenia Index – *SCZI*), anteriormente usado, visando à identificação de esquizofrenia. O Índice de Depressão (Depression Index – *DEPI*) visa à identificação de perturbação emocional e cognitiva semelhante à encontrada em pacientes com depressão. O Índice de Déficit de Coping ou Índice de Déficit Relacional (Coping Deficit Index – *CDI*)⁴⁹ visa identificar indivíduos que experimentam dificuldade para lidar com as situações cotidianas, principalmente aquelas que envolvem relacionamento interpessoal. O Índice de Hipervigilância (Hypervigilance Index – *HVI*) identifica presença de desconfiança, retração à proximidade física e emocional e observação atenta do ambiente interpessoal a fim de se precaver contra eventuais danos antecipados ao se relacionar com os outros. O Índice de Estilo Obsessivo (Obsessive Style Index – *OBS*) identifica características de perfeccionismo e preocupação em ser correto.

Sultan et al. (2006) analisaram a consistência para as constelações consideradas em termos de seu status positivo ou negativo. (O índice *PTI* não foi incluído nesta análise porque não tem ponto de corte, sendo uma medida dimensional, na qual valores mais altos indicam maior perturbação da percepção e pensamento; em lugar do *PTI*, usaram o *Índice de Esquizofrenia – SCZI*). Analisaram também os índices *Lambda* e *reflexos* segundo uma categorização dicotômica, ou seja, com duas categorias interpretativas. Os resultados são apresentados na Tabela 18.

⁴⁹ O Coping Deficit Index foi traduzido para o português como Índice de Déficit Relacional. Optamos por indicar seu significado de déficit de recursos de *coping*. Esta palavra vem sendo traduzida para o português como *enfrentamento*, ou mantida no original, *coping*. Coping indica recursos para enfrentar as situações e ser eficiente em resolver problemas.

Tabela 18 – Consistência das categorias para 8 índices do SCR numa amostra de 75 adultos não-pacientes franceses retestados após 3 meses

Índices no teste	Reteste		% consistentes	Φ	kappa
	Positivo	Negativo			
<i>S-CON</i>					
Positiva	0	2	0,0		
Negativa	0	73	100,0	0,00	0,00
<i>HVI</i>					
Positivo	4	6	40,0		
Negativo	7	58	89,2	0,28*	0,28*
<i>OBS</i>					
Positivo	0	0	–		
Negativo	2	73	97,3	0,00	0,00
<i>SCZI</i>					
Positivo	5	5	50,0		
Negativo	9	56	86,2	0,31**	0,31**
<i>DEPI</i>					
Positivo	17	14	54,8		
Negativo	15	29	65,9	0,21	0,21
<i>CDI</i>					
Positivo	10	10	50,0		
Negativo	11	44	80,0	0,30*	0,30*
Lambda					
$\geq 1,00$	11	11	50,0		
$< 1,00$	11	42	79,2	0,29*	0,29*
Reflexos					
> 0	13	8	61,9		
$= 0$	3	51	94,4	0,62***	0,61***

* $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$

Fonte: Sultan et al. (2006)

Sultan et al. (2006, p. 340) comentam:

Em relação aos índices dicotômicos, as porcentagens de concordância sugerem que a ausência de um índice positivo é uma característica bastante estável, especificamente para S-CON, OBS, HVI, SCZI, CDI, e em menor grau para DEPI, enquanto que a presença de um índice positivo é mais provável de mudar entre as testagens, com aproximadamente 50% dos escores inicialmente positivos mudando para abaixo do ponto de corte crítico no reteste. Isto é consistente com a observação comum de que o poder preditivo positivo é menor do que o poder preditivo negativo (ou a

porcentagem de consistência para os índices negativos) quando um índice positivo apresenta uma baixa taxa de ocorrência (Streiner, 2003)⁵⁰.

A baixa taxa de ocorrência das *Constelações* positivas ocorreu nesta amostra de indivíduos não-pacientes. Em amostras de pacientes, as taxas de ocorrência devem ser maiores, e possivelmente haverá maior estabilidade para os índices positivos.

Investigaram também a influência do engajamento na tarefa e do estresse situacional na estabilidade temporal das variáveis do SCR. O engajamento na tarefa (Task Engagement - *TE*) corresponde ao fator principal encontrado na análise fatorial do Rorschach realizada por Meyer (1992a) e replicado por Lindgren e Carlsson (2002). É calculado a partir de escores padronizados, e tem como fórmula:

$$\begin{aligned} TE = & (0,436 * \text{Color-shading}) + (0,372 * FY) + (0,325 * FC') + (0,3 * FC) + \\ & (0,3 * CF+C) + (0,29 * m) + (0,29 * R) + (0,27 * S) + (0,24 * FM) + (0,22 * FV) + \\ & (0,21 * W) + (0,19 * MOR) + (0,18 * M) - (0,24 * \text{Lambda}) \end{aligned}$$

Sultan et al. (2006) encontraram que tanto o nível de engajamento na tarefa, quanto mudanças no engajamento de uma testagem para a outra, desempenham papel na estabilidade temporal de algumas variáveis. Maior engajamento na tarefa mostrou-se associado à maior estabilidade temporal para *m*, *FM+m*, *a*, *FC*, *SumC'*, *SumV*, *SumSH*, *Fr+rF*, *INC+FAB*, *COP*, *EA*, *es*, *Adjes*, *EGO* e *Blends*. Menor engajamento na tarefa mostrou-se associado a maior estabilidade para *F*. Menor mudança no engajamento na tarefa de uma testagem à outra se mostrou associada a maior estabilidade para *R*, *F*, *FM* e *p*.

Como o número de respostas é um componente importante do engajamento na tarefa, também investigaram se *R* seria um moderador da estabilidade temporal. Maior *R* mostrou-se associado a maior estabilidade temporal para *FC* e *SumV*. Menor mudança em *R*, de uma testagem à outra, mostrou-se associada a maior estabilidade *Popular*, *Zf*, *m*, *CF+C*, *SumC'*, *MOR*, *EA*, *es* e *Blends*.

⁵⁰ STREINER, D. L. Diagnosing tests: Using and misusing diagnostic and screening tests. **Journal of Personality Assessment**, v. 81, p. 209-219, 2003.

O estresse situacional foi avaliado pela pontuação no Questionário de Saúde Geral de Goldberg. Encontraram que maior mudança no nível de estresse situacional, de uma testagem à outra, mostrou-se associada a menor estabilidade para *Zd*, *m*, *FM+m*, *a*, *C*, *CF+C*, *WSumC* e *es*, e maior estabilidade para *FD*.

Em sua pesquisa, Sultan et al. (2006) utilizaram diferentes examinadores, e nenhum deles testou o mesmo participante nas duas testagens. Investigaram os resultados de cada examinador e também compararam pares de examinadores. Encontraram que alguns examinadores associam-se a menores números de respostas, menor engajamento na tarefa e valores menores de *EA*, bem como estabilidade temporal mais baixa. Comentam:

O impacto do engajamento na tarefa imediatamente coloca a questão dos fatores determinantes do engajamento, em particular com referência aos fatores relacionados ao contexto e ao examinador. De fato, como esperávamos, nossos resultados mostram que cada par de examinadores cujos resultados indicaram baixa estabilidade incluía um examinador que eliciou menor engajamento. [...] Embora não se tenha detectado qualquer erro de aplicação, é provável que um efeito devido ao examinador tenha desempenhado um papel aqui. Contudo, este efeito parece ser bastante limitado, porque os níveis médios de estabilidade não mudaram notavelmente quando os protocolos dos dois examinadores menos estáveis foram excluídos. (SULTAN et. al., 2006, p. 345).

Resumindo seus resultados, afirmam:

[...] embora nossos resultados tenham limitações [...], fornecem alguns argumentos em suporte: (a) da importância de emoções negativas situacionais na estabilidade de variáveis selecionadas do Sistema Compreensivo, (b) do papel do engajamento na tarefa em algumas importantes características pessoais que podem se expressar no Rorschach, (c) da importância da variação na responsividade em numerosas variáveis centrais, e (d) do provável efeito induzido pela atitude do examinador e contexto na estabilidade” (SULTAN et. al., 2006, p. 345).

Sultan et al. (2006) oferecem algumas explicações para a menor estabilidade temporal encontrada em seu estudo, em comparação com os estudos norte-americanos: 1) as distribuições de valores das variáveis; 2) a baixa fidedignidade inter-codificadores de algumas variáveis; 3) a baixa taxa de ocorrência de algumas variáveis, ocasionando restrição de faixa de valores; 4) mudanças ao fazer o reteste, pois neste os participantes já conhecem o

Rorschach; 5) mudanças nas características durante o intervalo de 3 meses – os participantes mostraram-se menos defensivos no reteste; 6) características da amostra: o nível de engajamento na tarefa é semelhante ao da amostra norte-americana, mas esta amostra francesa apresentou maior complexidade, com mais indicadores de sofrimento emocional, menos recursos e menor adequação perceptiva; a presença de maior sofrimento emocional pode levar os participantes a mudar mais do teste para o reteste. Por fim, apontam a necessidade de mais pesquisas, com amostras maiores.

Em nossa opinião, estas explicações sobre a menor estabilidade encontrada são plausíveis. No entanto, caberia questionar se alguns fatores apontados também não estariam presentes nas amostras de Exner, que apresentaram maior estabilidade temporal.

* * *

Ao final desta revisão dos estudos de estabilidade temporal do SCR, encontramos um quadro no qual as pesquisas originais do Sistema Compreensivo, realizadas por Exner e seus colaboradores, apresentaram estimativas mais altas de estabilidade temporal do que as encontradas por outros pesquisadores norte-americanos e também na recente pesquisa francesa de Sultan et al. (2006).

Os estudos de Exner infelizmente não apresentam informações sobre como os participantes foram recrutados e suas características sócio-demográficas, nem estatísticas descritivas dos resultados por eles obtidos. Diferenças nos resultados de estabilidade temporal das diferentes pesquisas podem estar relacionadas à composição das amostras e às distribuições de valores das variáveis do SCR, mas não contamos com estas importantes informações que ajudariam a compreender a maior estabilidade temporal obtida nas pesquisas de Exner.

A metanálise da estabilidade temporal do Rorschach (Grønnerød, 2003a) também encontrou estabilidade alta para o Rorschach como um todo. Além deste importante estudo, encontramos outros dois que ajudam a situar a estabilidade do Rorschach em comparação com outros instrumentos de avaliação:

Meyer (2004) revisou as metanálises de fidedignidade de teste-reteste na literatura psicológica. Para intervalos de até 12 meses, as correlações ponderadas variaram de 0,38 a 0,92, com a maioria dos instrumentos apresentando correlações na faixa de 0,70 e 0,80. A estabilidade temporal do Rorschach seria semelhante à destes outros instrumentos em psicologia e medicina.

Watson (2004) realizou uma revisão da fidedignidade de teste-reteste dos instrumentos de avaliação psicológica. Para 23 estudos com intervalos de 2 a 4 meses, a média das correlações mínimas foi de 0,63 e a média das correlações máximas foi de 0,79.

5.5.3. A estabilidade temporal e os protocolos com poucas respostas

Exner (1988) afirma que suspeitava que protocolos com poucas respostas prejudicassem os resultados das pesquisas. Por exemplo, o Índice de Esquizofrenia [*SCZI*] e a Constelação de Suicídio [*S-CON*] mostravam-se úteis na identificação da esquizofrenia e do risco de suicídio, mas muitos falso-negativos (testandos com a condição e não identificados pelo índice) apresentavam protocolos com poucas respostas. Isto levou a uma revisão de 10 estudos sobre a validade de variáveis do SCR:

Os dados de cada estudo foram analisados novamente duas vezes, utilizando-se os mesmos procedimentos estatísticos empregados no estudo original. Na primeira nova análise, todos os protocolos com menos do que 13 respostas foram excluídos das amostras. Os resultados foram marcados. Quando os estudos foram re-analizados com os protocolos com 12 respostas ou menos excluídos [das amostras], os níveis de significância que haviam sido detectados para os achados positivos nas análises originais melhoraram substancialmente. Quando os protocolos com 13 ou menos respostas foram excluídos [das amostras] [...], os achados melhoraram significativamente ainda mais. (EXNER, 1988, p. 641)

Exner (1988) investigou a estabilidade temporal de dois grupos de protocolos: num dos grupos, um dos protocolos apresentava menos de 14 respostas; no outro grupo, todos os protocolos apresentavam ao menos 15 respostas. Os resultados são apresentados na Tabela 19. O grupo dos protocolos com menos de 14 respostas apresentou menor estabilidade

temporal. Desde então o Sistema Compreensivo exige que os protocolos tenham pelo menos 14 respostas para serem interpretados.

Tabela 19 – Estabilidade temporal em duas amostras de 36 participantes retestados após de 4 a 30 dias

Variável	R < 14 em uma das testagens	R > 14 em ambas testagens
<i>R</i>	0,36	0,87
<i>Popular</i>	0,62	0,89
<i>Zf</i>	0,53	0,82
<i>F</i>	0,48	0,79
<i>M</i>	0,44	0,87
<i>FM</i>	0,28	0,74
<i>m</i>	0,36	0,28
<i>a</i>	0,41	0,88
<i>p</i>	0,22	0,83
<i>FC</i>	0,34	0,89
<i>CF</i>	0,49	0,71
<i>C</i>	0,33	0,59
<i>CF+C</i>	0,38	0,84
<i>WSumC</i>	0,27	0,86
<i>SumT</i>	0,54	0,91
<i>SumC'</i>	0,28	0,78
<i>SumY</i>	0,09	0,40
<i>SumV</i>	0,62	0,89
<i>Lambda</i>	0,46	0,83
<i>X+%</i>	0,81	0,92
<i>Afr</i>	0,49	0,89
<i>3r+(2)/R</i>	0,51	0,87
<i>EA</i>	0,44	0,86
<i>es</i>	0,23	0,71
<i>D</i>	0,76	0,92
<i>AdjD</i>	0,61	0,93
<i>Média</i>	0,44	0,80
<i>Desvio-padrão</i>	0,17	0,16

Fonte: Exner (1988)

Exner (1988) apresenta também dados da consistência das categorias interpretativas para os grupos com menos de 14 respostas e com mais de 14 respostas. Os resultados são apresentados na Tabela 20. Protocolos com menos de 14 respostas e $\text{Lambda} < 0,85$ apresentaram boa consistência das categorias interpretativas para algumas variáveis, enquanto que, para os protocolos curtos e com $\text{Lambda} > 0,99$, a consistência das categorias interpretativas foi baixa. De qualquer modo, os protocolos com menos do que 14 respostas apresentaram baixa consistência das categorias interpretativas para algumas variáveis, enquanto que os protocolos com mais de 14 respostas apresentaram boa

consistência das categorias interpretativas para todas as variáveis. Assim sendo, protocolos com menos de 14 respostas devem ser considerados pouco fidedignos e, conseqüentemente, inválidos.

Tabela 20 – Frequências e porcentagens para 3 grupos mostrando o número de pares de protocolos com valores na mesma categoria interpretativa em ambas testagens

	R < 14				R > 14	
	Lambda < 0,85 nos dois protocolos		Lambda > 0,99 no protocolo curto			
	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
<i>EB</i>	8	100	11	39	33	92
<i>eb</i>	7	88	13	46	35	97
<i>D-score</i>	8	100	21	75	35	97
<i>AdjD</i>	8	100	16	57	36	100
<i>a : p</i>	8	100	15	53	34	94
<i>Ma : Mp</i>	8	100	13	46	33	92
<i>FC : CF+C</i>	8	100	9	32	36	100
<i>Afr</i>	5	63	7	25	35	97
<i>3r+(2)</i>	6	75	11	39	35	97
<i>Lambda</i>	7	88	6	21	34	94
<i>X+%</i>	8	100	24	86	36	100
<i>X-%</i>	5	63	18	64	32	89
<i>Zd</i>	7	88	13	46	32	89
<i>W : D</i>	3	38	7	25	30	83
<i>W : M</i>	2	25	6	21	31	86
<i>HumanContent</i>	4	50	11	39	33	92
<i>Isolate/R</i>	6	75	20	71	35	97

Fonte: Exner (1988)

O procedimento de aplicação do Sistema Compreensivo indica que, no caso de o testando fornecer menos de 14 respostas, o teste deve ser interrompido e o examinador deve dizer

Agora você sabe como se faz. Mas temos um problema. Você não deu respostas suficientes para que eu possa fazer a avaliação do teste. Então vamos voltar aos cartões novamente e desta vez eu gostaria que você me desse mais respostas. Se você quiser, pode incluir as mesmas respostas que já deu, mas procure dar mais respostas desta vez. (EXNER, 1999b, p. 18-19)

Não se trata de obter respostas complementares, mas sim de uma nova aplicação do teste. Em nossa experiência, notamos que este procedimento muitas vezes não funciona a contento, impressão que é confirmada por Dean et al. (2007, p. 157):

Informalmente, os clínicos reclamam do procedimento de reaplicação, mencionando o tempo excessivo necessário, a frustração e confusão que o examinador e o testando experimentam, e a validade questionável dos resultados quando a fase de respostas é repetida, as respostas originais não são fornecidas novamente e então não são codificadas. Uma reação comum, mesmo em testandos cooperativos, é não repetir os perceptos originais ou repeti-los de modo resumido. O testando pode pensar que, como o examinador já anotou o primeiro conjunto de respostas, estas não precisam ser repetidas ou que ele pode fazer referência a elas sem os detalhes fornecidos durante a fase de respostas inicial. Por exemplo, durante a aplicação inicial, uma participante pode dizer ‘duas pessoas jantando’. Na re-aplicação, ela pode dizer ‘Ah sim, aquelas pessoas’. O examinador fica em dúvida se deve incluir o conteúdo descritivo da fase de respostas inicial [primeira aplicação] ou considerar somente a fala da fase de respostas da reaplicação.

O problema também foi identificado por Erstad (1996) em sua pesquisa com idosos. Uma das preocupações deste pesquisador era a fidedignidade de teste-reteste dos resultados, principalmente no caso de protocolos com poucas respostas, que o autor previu e de fato encontrou em sua amostra. Erstad notou que, utilizando as orientações do Sistema Compreensivo – dizer ao testando que voltariam aos cartões e que ele deveria fornecer mais respostas, podendo utilizar as respostas já fornecidas, se quisesse –, os participantes não compreendiam claramente que deveriam repetir as respostas fornecidas anteriormente para que estas fossem incluídas no protocolo da reaplicação; alguns participantes até disseram que pensavam que as respostas da primeira aplicação seriam automaticamente incluídas. Assim sendo, o examinador modificou as instruções fornecidas, incluindo esta orientação: “Se você quiser que as respostas que você forneceu antes sejam incluídas, você deve dizê-las novamente”. Este procedimento modificado parece ter resolvido a dúvida dos testandos sobre as respostas fornecidas inicialmente.

No entanto,

[...] é opinião do investigador de que este encorajamento [pedido de mais respostas na reaplicação] teve variados graus de sucesso. Para alguns sujeitos as respostas [na reaplicação] foram melhores [...]. Em outras palavras, não apenas houve um aumento no número de respostas; também houve uma melhora notável na riqueza das respostas. Para outros sujeitos, houve um aumento no número de respostas, mas essencialmente o protocolo foi tão defendido quanto o fornecido previamente. Nos piores casos, não houve nem um aumento no número de respostas, nem na riqueza do material das respostas. (ERSTAD, 1996, p. 135)

Erstad (1996) afirma que o procedimento de reaplicar o teste encorajando o testando a fornecer mais respostas nunca foi testado em termos de sua eficiência em melhorar a fidedignidade do protocolo. O autor, então, investigou esta questão em sua pesquisa.

O procedimento de encorajamento e reaplicação imediata foi experimentado com participantes do grupo de idosos, mas não se mostrou proveitoso. Num dos casos, com 11 respostas, a reaplicação imediata resultou novamente em 11 respostas; no reteste, o participante forneceu 13 respostas e não foi feita reaplicação. Em outro caso, o participante demonstrou hostilidade frente à testagem e o examinador decidiu não realizar o encorajamento e a reaplicação. Na verdade, “[...] em mais do que um caso, o examinador não tentou encorajar o participante idoso a fornecer outro conjunto de respostas, porque parecia improvável que isto produzisse resultados úteis” (ERSTAD, 1996, p. 106).

O estudo de Erstad apresenta, então, dados sobre os efeitos do encorajamento apenas para o grupo de adultos de 18 a 55 anos de idade. Frente a um protocolo com menos de 14 respostas, o examinador encorajou o participante a fornecer mais respostas e reaplicou o Rorschach imediatamente. Realizou o inquérito deste novo protocolo e, posteriormente, o inquérito das respostas fornecidas na primeira aplicação. Os participantes realizaram reteste após um período de poucos dias.

Este estudo contou com apenas 11 protocolos, sendo que o procedimento de encorajamento e reaplicação imediata foi realizado com 3 participantes. O autor estudou a fidedignidade de teste-reteste para as categorias interpretativas, comparando os resultados de casos cujos protocolos continham pelo menos 14 respostas obtidas sem encorajamento, protocolos com menos de 14 respostas e protocolos com mais de 14 respostas obtidos com encorajamento. Os níveis de fidedignidade de teste-reteste para as categorias interpretativas foram semelhantes nestas três situações, e todos mais baixos do que os reportados nas pesquisas de Exner. Assim sendo, Erstad concluiu que o procedimento de solicitar mais respostas e reaplicar o Rorschach imediatamente não contribui para melhorar a fidedignidade de teste-reteste.

Frente ao número pequeno de participantes, seria mais sensato dizer que o estudo apresentou limitações e não permite concluir se o procedimento de encorajamento e

reaplicação imediata é eficiente – o que é diferente de concluir que este procedimento é ineficiente.

Janson e Erdberg (1998, apud Grønnerød, 2003a)⁵¹ reportaram alta estabilidade para os resultados em um grupo de crianças com protocolos com poucas respostas e sugeriram que, para certos indivíduos, fornecer poucas respostas seria um estilo de resposta específico e estável.

A partir de sua metanálise sobre a estabilidade temporal do Rorschach, Grønnerød (2003a) concluiu que a questão dos protocolos com poucas respostas permanece aberta, necessitando de mais pesquisas.

A questão de o número de respostas fornecidas interferir nos escores do Rorschach gerou muito debate (ver: Cronbach, 1949; Fiske e Baughman, 1953; Lipgar, 1992; Meyer, 1992b; Exner, 1992; Kinder, 1992). Exner (1978) afirmou que, com o procedimento de aplicação do Sistema Compreensivo, o número de respostas é menos variado do que em outros sistemas. Posteriormente, Exner (1992) mostrou que *R* apresenta correlações baixas com os outros escores do SCR, ou seja, a interferência de *R* nos resultados seria mínima.

McGuire et al. (1995) afirmam que a questão da interferência do número de respostas nos escores é mais complexa do que até então foi conceituada, porque as respostas diferem em termos de sua complexidade: alguns indivíduos, por exemplo, fornecem protocolos com poucas respostas, mas muitas delas são respostas complexas; outros indivíduos fornecem muitas respostas, mas a maioria delas é simples.

De modo contrário a Exner, Meyer (1992b) destacou a interferência de *R* nos escores do Rorschach. Posteriormente, esta discussão tornou-se mais complexa, ao considerar os estilos de resposta ao Rorschach e o engajamento na tarefa.

⁵¹ JANSON, H.; ERDBERG, P. **Recoding Rorschachs from a longitudinal study into the Comprehensive System: Intercoder agreement and temporal stability**. Relatório não publicado. Stockholm: Stockholm University, 1998.

5.5.4. Os estilos de resposta, o engajamento na tarefa e a estabilidade temporal

Meyer (1997a, p. 301) afirma:

[...] os escores do Rorschach são bastante dependentes da habilidade dos pacientes de espontaneamente se engajarem na tarefa e articularem as respostas. Dificuldades cognitivas e emocionais que interferem com a articulação verbal, espontaneidade comportamental ou engajamento na tarefa comprometem a acurácia dos escores [...].

Refletindo sobre a freqüente discordância entre os resultados do Rorschach e do MMPI, Meyer (1996) afirma que estas divergências em parte decorrem de diferentes métodos de avaliação fornecerem perspectivas únicas dos mesmos fenômenos. Além disso, os testandos podem apresentar estilos de resposta diferentes no Rorschach e no MMPI e estes estilos de resposta moderam as correlações entre os resultados dos dois testes (MEYER, 1997a).

Meyer (1997a, p. 306) define os estilos de resposta ao Rorschach:

Alguns pacientes mostram-se situacionalmente bastante constrictos ou desconfiados sobre quais informações serão deles obtidas. Apresentam poucas respostas, mínimo investimento na tarefa, resistência ao inquérito, e mínima articulação dos determinantes (estilo 1-R). Outros apresentam comportamentos manifestos semelhantes, embora a causa seja muito diferente, provindo de recursos internos genuinamente limitados ou empobrecidos e [de uma experiência de] desconcerto com o caráter abstrato e a complexidade da tarefa (estilo 2-R). Outros pacientes são articulados, esforçam-se para trabalhar cooperativamente com o examinador, mantêm um engajamento ativo com a tarefa e espontaneamente fornecem respostas, articulam os determinantes e sintetizam [combinam, relacionam] áreas da mancha de acordo com suas predileções internas (estilo 3-R). Ainda outros se esforçam conscientemente para ser dramáticos, completos, criativos ou chocantes em suas respostas ao teste. Conseqüentemente, eles parecem altamente engajados na tarefa e fornecem muitas respostas estruturalmente complexas e sintetizadas [áreas da mancha relacionadas] (estilo 4-R). Finalmente, alguns outros pacientes também parecem altamente engajados na tarefa, produzindo protocolos longos e complexos. Contudo, seu engajamento provém de uma propensão inconsciente a descarregar estados internos caóticos, como se a tarefa abrisse seus frágeis limites psicológicos e liberasse uma torrente de percepções carregadas de emoções (estilo 5-R).

Meyer (1999) enfatiza que estes estilos de resposta podem ser tanto manipulações deliberadas quanto manifestações da estrutura intrínseca do caráter, então os denomina estilos de resposta/caráter (*response-character styles*).

Para definir os estilos de resposta, Meyer (1997a) considerou o fator principal obtido em sua análise fatorial do Rorschach (apresentada em Meyer, 1992a). Este fator foi inicialmente nomeado *articulação das respostas* (*response articulation*); posteriormente, Meyer (1997a) afirmou que *engajamento nas respostas* (*response engagement*) ou *complexidade das respostas* (*response complexity*) seriam nomes mais adequados ao fator. Sultan et al. (2006) o denominaram *engajamento na tarefa*. Os estilos 1 e 2 apresentam níveis baixos de engajamento na tarefa, o estilo 3 apresenta níveis médios de engajamento na tarefa e os estilos 3 e 4 apresentam níveis altos de engajamento na tarefa.

Meyer (1999) afirma que um modo mais simples e direto para definir os estilos de resposta seria utilizar *R* e *Lambda*. Os estilos 1 e 2 correspondem a *R* baixo e *Lambda* alto; o estilo 3 corresponde a *R* médio e *Lambda* médio; os estilos 4 e 5 correspondem a *R* alto e *Lambda* baixo.

Esta contribuição importantíssima de Gregory Meyer ao entendimento dos resultados do Rorschach certamente inspirou Sultan et al. (2006) a investigarem o papel que o engajamento na tarefa desempenha na estabilidade temporal do SCR. Encontraram que tanto o nível de engajamento na tarefa, quanto mudanças no engajamento de uma testagem à outra, moderam a estabilidade temporal de algumas variáveis do SCR.

5.5.5. Fatores situacionais interferindo nos resultados dos estudos de teste-reteste

Nas décadas de 1940 e 1950 foram realizadas muitas pesquisas sobre a interferência de fatores situacionais nos resultados dos testes projetivos. Estas pesquisas foram revistas por Joseph Masling em seu artigo **The influence of situational and interpersonal variables in projective testing**, publicado em 1960. As pesquisas revisadas são variadas: algumas utilizaram modificações nas instruções, solicitando que os testandos buscassem fornecer respostas com determinados tipos de localizações, determinantes ou

conteúdos; outras solicitaram que os testandos buscassem produzir determinada impressão, positiva ou negativa, no examinador; outras examinaram os resultados obtidos por testandos submetidos a situações estressantes, sob efeito de drogas ou sob efeito de hipnose; outras verificaram a influência dos estímulos na sala de espera sobre os resultados; outras investigaram a influência de características físicas dos examinadores nos resultados por eles obtidos; outras investigaram a diferença nos resultados obtidos por examinadores “calorosos” versus “frios” (encontrando que os examinadores mais calorosos obtinham respostas em maior número, mais criativas e com mais indicadores de facilidade para se relacionar). Masling concluiu que os fatores situacionais e interpessoais influenciam os testes projetivos e buscou compreender esta influência:

Existem evidências consideráveis de que os sujeitos, numa situação não-estruturada utilizarão todas as dicas disponíveis para completar as tarefas a eles designadas. O sujeito na situação de testagem projetiva não apenas utilizará aquelas dicas fornecidas pela mancha de tinta ou figura, mas também aquelas fornecidas por seus sentimentos sobre o examinador, aquelas fornecidas por suas necessidades, atitudes e medos, aquelas contidas nas instruções, na sala, no conhecimento prévio sobre o teste, e aquelas dicas fornecidas conscientemente ou inconscientemente pelo examinador. (MASLING, 1960, p. 81)

Exner (1978, 2003) também comenta sobre a interferência de fatores situacionais nos resultados do Rorschach.

Em primeiro lugar, pesquisas indicavam que diferentes examinadores obtinham resultados diferentes.

Outras pesquisas importantes foram as que encontraram que comportamentos verbais e não-verbais do examinador podem interferir nos resultados. Wickes (1956) encontrou que o examinador dizer “bom” (good) ou “excelente” (fine), ou sorrir ou balançar a cabeça em sinal de assentimento, quando o testando fornecia respostas de conteúdo humano, produziu um aumento na frequência dessas respostas – ou seja, as respostas de conteúdo humano foram reforçadas pela aprovação do examinador. Pesquisas semelhantes foram realizadas por Gross (1959), Simkins (1960), Dinoff (1960), Magnussen (1960) e Hersen e Greaves (1971), com resultados semelhantes.

Masling (1965) investigou o efeito que as expectativas dos examinadores podem ter nos resultados dos protocolos por eles obtidos. Instruiu 7 estudantes de psicologia dizendo que examinadores experientes obtinham mais respostas de conteúdo humano do que respostas de conteúdo animal; a outros 7 estudantes foi dito que examinadores experientes obtinham mais respostas de conteúdo animal do que respostas de conteúdo humano. Cada estudante testou dois participantes e as sessões de aplicação foram gravadas. Os dois grupos de examinadores diferiram entre si na direção esperada, obtendo mais respostas do conteúdo que supunham ser obtido por examinadores mais experientes. Analisando as gravações, Masling não encontrou evidências de reforçamento verbal, então concluiu que as diferenças nos resultados decorreram de reforçamento não-verbal das respostas.

Exner (2003) menciona um estudo⁵² de replicação do estudo de Masling (1965). Como no estudo original, os estudantes foram informados de que examinadores experientes obtinham mais respostas de conteúdo humano/animal, mas nesta replicação foram utilizados quatro grupos de examinadores: em dois grupos, os examinadores aplicaram o Rorschach sentados na frente dos testandos; nos outros dois grupos, os examinadores sentaram-se ao lado dos testandos. Cada examinador testou 3 participantes e as aplicações foram gravadas em vídeo. Os resultados foram semelhantes aos do estudo de Masling somente para os examinadores que aplicaram o Rorschach sentados em frente aos testandos e a gravação em vídeo mostrou que o examinador emitia comportamentos não-verbais que reforçavam a resposta que o examinador supunha ser obtida por examinadores experientes.

Vemos então que comportamentos verbais e não-verbais do examinador podem reforçar determinados tipos de respostas. Isto pode ser um dos fatores responsáveis pelas diferenças nos resultados obtidos por diferentes examinadores (EXNER, 2003). Desde sua formulação em 1974, o procedimento de aplicação Sistema Compreensivo do Rorschach foi escolhido de modo a tentar diminuir esta interferência do examinador nos resultados:

[...] o sujeito e o examinador sentarem lado a lado, assim reduzindo o potencial de reforçamento não-verbal, e [o procedimento de] minimizar a complexidade das

⁵² EXNER, J. E. Jr.; LEURA, A. V. ; GEORGE, L. M. **A replication of the Masling study using four groups of new examiners with two seating arrangements and video examination.** Workshops Study No. 256. Estudo não publicado. Rorschach Workshops, 1976.

instruções, assim reduzindo o potencial para um ‘predisposição’ (‘set’) direta, tende a reduzir a variância no número de respostas e limita o potencial de influência do examinador” (Exner, 1978, p. 39-40).

Na literatura mais recente, não encontramos mais esta discussão sobre as interferências da situação de testagem e do examinador nos resultados. No entanto, Meyer (1997a, 1999) discute a questão dos estilos de resposta, que remete ao modo como o testando responde à situação de testagem, e que tem implicação direta nos resultados.

Na pesquisa de Sultan et al. (2006) alguns aplicadores obtiveram protocolos com níveis menores de *TE*, *R* e *EA*. Como os participantes foram testados por aplicadores diferentes no teste e no reteste, estas diferenças entre os aplicadores, embora produzissem alterações pequenas nos resultados, contribuíram para diminuir a estabilidade temporal.

Algumas pesquisas de Exner e seus colaboradores investigaram outro aspecto da relação entre testando e examinador, e as interferências nos resultados: o caso de o testando conhecer o examinador.

Examinador conhecido versus desconhecido

Dois estudos do Sistema Compreensivo investigaram os efeitos de o testando ser avaliado por uma pessoa conhecida: Exner, Armbruster e Mittman (1978) investigaram os efeitos de terapeutas testarem seus pacientes, e Leura e Exner (1978, apud Exner, 2003)⁵³ investigaram os protocolos de alunos do ensino médio (high school) testados por seus professores. Nestes dois estudos, algumas variáveis do Rorschach alcançaram níveis muito mais altos quando os participantes foram testados pelos examinadores conhecidos. Aparentemente, então, quando o indivíduo se sente emocionalmente ou intelectualmente próximo ao examinador, tende a fornecer mais respostas e a se revelar mais (EXNER, 2003).

⁵³ LEURA, A. V.; EXNER, J. E. **Structural differences in the records of adolescents as a function of being tested by one’s own teacher**. Rorschach Workshops Study No. 265. Estudo não publicado. Rorschach Workshops, 1978.

Erstad (1996), em seu estudo sobre as respostas de idosos ao Rorschach, testou participantes conhecidos e desconhecidos. Os participantes que conheciam o examinador eram membros de uma congregação religiosa e o examinador, no caso o próprio Duane Erstad, era o pastor desta congregação. Os participantes que não conheciam o examinador eram parentes ou conhecidos dos membros da congregação. O autor levantou os resultados para *R*, *M*, *WSumC*, *SumColor*, *Blends*, *W*, *Dd*, *Sx*, *P* e *sombreado excluindo textura* para os participantes que conheciam o examinador e os comparou com as normas do SCR, encontrando diferenças apenas para *Dd* e *sombreado excluindo textura* (este último na direção oposta à prevista). Infelizmente, Erstad não comparou os resultados destes participantes com os daqueles que não conheciam o examinador, o que seria um modo melhor de investigar a influência de ser testado por um examinador conhecido.

Erstad (1996) Comparou também a estabilidade temporal dos protocolos de participantes que conheciam o examinador com a de participantes que não conheciam o examinador. Os resultados são apresentados na Tabela 21. As estimativas de estabilidade temporal não foram diferentes entre os grupos, tanto comparando os valores médios dos coeficientes, quanto comparando o número de coeficientes maiores num grupo em relação ao outro.

Considerando o número pequeno de participantes em cada subgrupo e lembrando que as estimativas de estabilidade temporal em amostras pequenas são pouco precisas, seria mais sensato dizer que o estudo não teve poder suficiente para responder à questão sobre a eventual interferência de os participantes conhecerem o examinador.

Tabela 21 – Estabilidade temporal de 26 variáveis do SCR em duas amostras norte-americanas

Grupo Examinador	Participantes com mais de 60 anos			18 a 55 anos
	Desconhecido	Conhecido	Total	
N	6	11	17	11
<i>R</i>	0,92	0,90	0,90	0,82
<i>P</i>	0,57	0,90	0,77	0,36
<i>Zf</i>	0,86	0,79	0,70	0,67
<i>F</i>	0,94	0,78	0,80	0,58
<i>M</i>	0,82	0,50	0,75	0,56
<i>FM</i>	0,69	0,67	0,68	0,40
<i>m</i>	-0,10	0,69	0,45	0,73
<i>a</i>	0,75	0,71	0,72	0,49
<i>p</i>	0,36	0,89	0,75	0,64
<i>FC</i>	-0,09	0,23	0,08	0,81
<i>CF</i>	0,46	0,52	0,42	0,19
<i>C</i>	0,63	1,00	0,95	0,44
<i>CF+C</i>	0,08	0,74	0,62	0,39
<i>WSumC</i>	0,18	0,88	0,75	0,53
<i>SumT</i>	0,62	0,83	0,70	0,92
<i>SumC'</i>	0,71	0,66	0,67	0,02
<i>SumY</i>	0,40	0,50	0,40	0,39
<i>SumV</i>	0,71	0,52	0,57	0,85
<i>Lambda</i>	0,61	0,35	0,46	0,23
<i>X+%</i>	0,85	0,48	0,63	0,47
<i>Afr</i>	0,16	0,54	0,48	0,79
<i>3r+(2)/R</i>	0,68	0,61	0,60	0,68
<i>EA</i>	0,56	0,88	0,73	0,44
<i>es</i>	0,90	0,70	0,80	0,68
<i>D-score</i>	0,00	0,34	0,22	0,54
<i>AdjD</i>	-0,50	0,29	0,02	0,53
Media	0,49	0,65	0,60	0,54
DP	0,38	0,21	0,23	0,22
Mediana	0,62	0,68	0,68	0,54

Fonte: Erstad (1996)

5.5.6 Questões estatísticas que interferem com as estimativas de estabilidade temporal

Os estudos com delineamento do tipo teste-reteste costumam utilizar como medida a correlação. Assim sendo, agora iremos apresentar alguns aspectos estatísticos que interferem nas correlações.

A correlação e o problema da restrição da faixa de valores

A correlação informa sobre o quanto duas variáveis estão relacionadas. Outra definição seria dizer que a correlação informa sobre o quanto uma variável muda em função da outra. Vê-se, portanto, que correlação implica haver variação dos valores das variáveis

(NUNNALLY, 1975). Nos casos em que há pouca variação dos valores de uma ou das duas variáveis, o coeficiente de correlação pode apresentar valores baixos. Este é o problema de restrição da faixa de valores: quando uma variável apresenta poucos valores, esta menor variabilidade costuma gerar coeficientes de correlação baixos.

Anastasi e Urbina (2000) exemplificam o problema de a variabilidade dos valores interferir diretamente nas correlações. A Figura 1, adaptada da Figura 4.5 apresentada na página 101 do livro **Testagem Psicológica**, apresenta um diagrama de dispersão de valores⁵⁴ e expõe como, restringindo o intervalo de valores, o coeficiente de correlação diminui.

Figura 1: O efeito de um intervalo restrito sobre um coeficiente de correlação

Fonte: Anastasi e Urbina (2000), p. 101 [modificado]

⁵⁴ No diagrama de dispersão de valores, cada ponto representa um caso. Sua posição no gráfico indica o valor no eixo horizontal (valor da primeira variável correlacionada) e o valor no eixo vertical (valor da segunda variável correlacionada). Na Figura 1, em vez de pontos utilizamos barras verticais inclinadas, para representar os casos.

O exame do digrama de dispersão hipotético apresentado [...] ilustrará melhor como os coeficientes de correlação dependem da variabilidade, ou da extensão das diferenças individuais, dentro do grupo. Este diagrama de dispersão mostra uma alta correlação positiva no grupo inteiro, heterogêneo, uma vez que os registros estão estreitamente agrupados na diagonal que vai do canto inferior esquerdo para o campo superior direito. Se, agora, nós considerarmos apenas o subgrupo situado dentro do pequeno retângulo na porção superior direita do diagrama, fica aparente que a correlação entre as duas variáveis é próxima de zero. (ANASTASI; URBINA, 2000, p. 100)

Muitas variáveis do SCR apresentam distribuições com poucos valores e predomínio de valores baixos. Por exemplo, *SumC'* pode apresentar, em amostras de adultos não-pacientes, uma distribuição semelhante à da Figura 2. A distribuição apresenta assimetria positiva, ou seja, predominam valores baixos. Exner denomina este tipo de distribuição *curva J* (o formato é o de um J em espelho). Este formato da distribuição dos valores pode diminuir as correlações de teste-reteste.

Figura 2: Distribuição dos valores de SumC' numa amostra de adultos não-pacientes

A correlação e o problema das distribuições de valores assimétricas

Dunlap, Burke e Greer (1994) demonstraram que a assimetria tem um efeito supressor nas correlações. Afirmam:

“[...] a correlação máxima ocorre quando ambas as variáveis não apresentam assimetria. A correlação é um pouco menor quando ambas as variáveis são assimétricas na mesma direção, mas a correlação é dramaticamente menor quando a assimetria [das variáveis] ocorre em direções opostas” (DUNLAP; BURKE; GREER, 1995, p. 367).

Meyer e Archer (2001) comentam que variáveis do Rorschach com faixa restrita de valores e assimetria podem produzir correlações de teste-reteste artificialmente baixas, que não revelam a verdadeira magnitude da associação. Recomendam então que se analise a estabilidade temporal observando o número de participantes que permanecem na mesma faixa interpretativa no teste e no reteste, em vez de calcular a correlação.

6. PESQUISAS BRASILEIRAS COM O SCR

Algumas pesquisas realizadas no Brasil com o Sistema Compreensivo do Rorschach detiveram-se no estudo de grupos clínicos: Nascimento (1993) estudou pacientes bipolares, Guntert (1996) estudou crianças com nódulo vocal, Bocchi (1998) estudou pacientes com queixa de disfunção da articulação temporomandibular, Silva Neto (1999) estudou jogadores patológicos, Semer (1999) estudou crianças enuréticas, Duarte (2000) estudou mães de crianças autistas, Gazire (2000) estudou pacientes com síndrome do intestino irritável, Rosenthal (2000, 2007) estudou pacientes com transtorno obsessivo-compulsivo, Santoantonio (2001) estudou adolescentes com lúpus, Godinho (2003) estudou pacientes com alopecia areata, Farah (2004) estudou pacientes com transtorno do pânico, Antúnez (2004) estudou portadores de transtorno obsessivo-compulsivo, Passos (2005) estudou pacientes obesos com transtorno de compulsão alimentar periódica, Brunori (2006) estudou portadores de transtorno invasivo do desenvolvimento e Hisatugo (2007) estudou pacientes ambulatoriais depressivos.

Outras pesquisas estudaram indivíduos não-pacientes: Campagna (2003) estudou aspectos da organização da identidade feminina no início da adolescência, Rocha Jr. (2004) estudou religiosidade e personalidade de universitários e Silva (2005) avaliou estudantes de psicologia, buscando relações entre características de personalidade e desempenho nos estágios em psicoterapia. Na área da Psicologia do Trânsito, Lamounier (2005) estudou motoristas infratores e não-infratores. Na área jurídica, Durat Jr. (2006) estudou adolescentes infratores.

Castro (2006) realizou um levantamento das pesquisas com o SCR apresentadas nos eventos da Associação Brasileira de Rorschach e outros métodos projetivos (ASBRo). Encontrou um aumento no número de trabalhos com este sistema. Comenta que

O aumento da quantidade de trabalhos, embora discreto, é significativo, se considerarmos que no período estudado [1995 a 2004], houve a apresentação do Sistema Compreensivo para a comunidade científica brasileira, que absorveu essa estratégia de trabalho com o método de Rorschach de maneira gradual, uma vez que houve a necessidade de estudos e de formação específica, além da necessidade de pesquisas normativas brasileiras. (CASTRO, 2006, p. 116)

A maioria dos trabalhos apresentados nos eventos da ASBRo foram casos clínicos, seguidos de estudos empíricos. Predominaram trabalhos relacionados ao estudo de psicopatologia. (CASTRO, 2006)

Notamos então o predomínio de pesquisas que utilizaram o Sistema Compreensivo do Rorschach para investigar as características de personalidade de grupos clínicos. A maioria destes estudos são, então, pesquisas com o método de Rorschach. Contudo, também são necessárias pesquisas brasileiras sobre o método de Rorschach, ou seja, estudos que investiguem as características psicométricas deste instrumento.

No processo de adaptação do Sistema Compreensivo do Rorschach ao Brasil, aspecto importante é o levantamento de normas brasileiras. Nascimento (2002, 2004, 2006; Nascimento et al., 2000, 2004) realizou pesquisa normativa, inicialmente com 200 indivíduos adultos não-pacientes da cidade de São Paulo. A amostra foi composta por pessoas com idade entre 17 e 65 anos, de ambos os sexos, de todos os estados civis e graus de escolaridade, de diferentes níveis sócio-econômicos, e que não estavam em tratamento psiquiátrico ou psicológico. Foram encontradas diferenças importantes entre os resultados desta amostra brasileira e os resultados da amostra normativa norte-americana do Sistema Compreensivo. Mais recentemente (Nascimento, 2007), a amostra foi ampliada com indivíduos do interior do Estado de São Paulo.

O SCR vem sendo utilizado no Brasil tendo como embasamento as pesquisas norte-americanas sobre fidedignidade e validade, supondo-se que este instrumento de avaliação da personalidade funcione de modo semelhante com brasileiros. Evidentemente, existe a necessidade de pesquisas brasileiras sobre sua fidedignidade e validade, pois estas não devem ser supostas, mas sim demonstradas empiricamente.

A pesquisa normativa e algumas das pesquisas com grupos clínicos investigaram a fidedignidade inter-codificadores. A Tabela 22 apresenta uma síntese das estimativas obtidas, quando informadas nos trabalhos. Nas pesquisas, estão informados os valores de fidedignidade inter-codificadores para códigos do SCR; na Tabela 22, apresentamos a média dos valores de fidedignidade dos códigos, por segmentos da codificação. De modo geral, os resultados encontrados foram adequados, indicando que diferentes avaliadores codificaram as

respostas de modo semelhante. Isto confirma que os critérios de codificação do Sistema Compreensivo são claros, permitindo uma classificação objetiva das respostas.

Tabela 22 – Valores médios de kappa para a fidedignidade inter-codificadores em pesquisas brasileiras com o SCR

	Loc	DQ	Det	FQ	Pair	Cont	P	Z	SS	Média
Silva Neto (1999)	0,92	0,86	0,84	0,64	0,86	0,80	0,97	0,84	0,48	0,80
Duarte (2000)	0,70	0,80	0,72	0,64	0,92	0,74	0,84	0,82	0,53	0,75
Santo Antonio (2001)	0,85	0,73	0,84	0,67	0,82	0,83	0,82	0,91*	0,47	0,75
Antúñez (2004)	0,87	0,52	0,77	0,56	0,89	0,70	0,90	0,89	0,52	0,74
Lamounier (2005)	0,79	0,67	0,76	0,76	0,79	0,74	0,91	0,70	0,52	0,74
Durat Jr. (2006)	0,81	0,70	0,88	0,78	0,79	0,81**	0,77	–	0,68***	0,78
Média	0,82	0,71	0,80	0,68	0,85	0,77	0,87	0,83	0,53	0,76

Loc = Localizações, DQ = Qualidade Evolutiva (Developmental Quality), Det = Determinantes, FQ = Qualidade Formal (Formal Quality), Pair = pares, Cont = Conteúdos, P = Popular; Z = notas Z, SS = Códigos Especiais (Special Scores).

* ICC. ** média da fidedignidade para os conteúdos humanos, animais e anatomia. *** média da fidedignidade para os códigos COP, AG e MOR.

Outro aspecto importante da fidedignidade refere-se à estabilidade temporal dos resultados. Além da necessidade de pesquisas brasileiras sobre fidedignidade dos testes psicológicos utilizados no país, uma pesquisa sobre estabilidade temporal do SCR justifica-se porque esta questão ainda é controversa: existem diferenças importantes entre os resultados obtidos por Exner e seus colaboradores, e os resultados obtidos por outros grupos de pesquisadores. Cabe ainda destacar a pesquisa recente de Sultan et al. (2006) com adultos franceses, que encontrou resultados mais baixos de estabilidade temporal do que os resultados norte-americanos, colocando a questão de uma eventual interferência de diferenças culturais (anglo-saxões versus latinos) nos resultados e no próprio funcionamento do SCR nestas diferentes populações.

7. JUSTIFICATIVA

Considerando, então,

- A importância de se conhecer a fidedignidade de qualquer teste psicológico;
- A inexistência de pesquisas sobre a estabilidade temporal do Sistema Compreensivo do Rorschach com indivíduos brasileiros;
- A importância de estudos de replicação sobre a estabilidade temporal do SCR;

Esta pesquisa visa estudar a estabilidade temporal das variáveis do Sistema Compreensivo do Rorschach numa amostra de indivíduos da cidade de São Paulo.

8. OBJETIVO

Investigar a estabilidade temporal das variáveis do SCR.

Este estudo é de natureza exploratória. Isto significa que não formulamos hipóteses sobre os resultados esperados nesta pesquisa e que realizamos análises adicionais a partir dos resultados obtidos.

9. MÉTODO

9.1. Amostra

Um dos grupos mais importantes para se estudar a estabilidade temporal é o de adultos não-pacientes, porque se espera que este grupo apresente estabilidade em seu funcionamento psicológico e, conseqüentemente, estabilidade nos resultados do Rorschach. Assim sendo, optamos por estudar adultos não-pacientes.

9.1.1. Recrutamento dos participantes

A maior dificuldade para se realizar uma pesquisa como esta é recrutar os participantes. Além de os brasileiros estarem pouco acostumados com situações de pesquisa, e muito menos ainda com situações de realização de testes psicológicos, nesta pesquisa deveriam realizar o Rorschach duas vezes, por se tratar de um estudo com delineamento de teste-reteste. Assim sendo, pensamos que um atrativo para recrutar os participantes seria oferecer os resultados da avaliação. Também optamos por realizar a pesquisa num local em que teríamos maior facilidade para divulgá-la: uma universidade particular da cidade de São Paulo⁵⁵.

Uma explicação sobre a pesquisa e convite para a participação voluntária foi enviada a todos os funcionários da universidade por e-mail. Além disso, foi marcada uma reunião com os funcionários da limpeza, na qual a pesquisa foi a eles apresentada pela coordenadora de recursos humanos, acompanhada pelo pesquisador. Também foi aberta a possibilidade de participação de outras pessoas, externas à universidade – os próprios funcionários poderiam divulgar a pesquisa para familiares e conhecidos e convidá-los a participar. Os interessados deveriam entrar em contato com a secretaria da Clínica de Psicologia da universidade e agendar horário para a primeira entrevista.

⁵⁵ Visando preservar o sigilo dos participantes, optamos por omitir o nome desta universidade.

9.2. Instrumentos

- Entrevista para coleta de dados sócio-demográficos.
- Método de Rorschach (Sistema Compreensivo).

9.3. Procedimento

9.3.1. Entrevista e aplicações do método de Rorschach

Os indivíduos dispostos a participar da pesquisa deveriam agendar uma primeira entrevista. Nesta, foram novamente informados sobre: a) o objetivo da pesquisa: investigar se os resultados de um teste de personalidade se mantêm ou mudam com o teste sendo realizado em dois momentos diferentes; b) o modo como seria realizada: o teste seria aplicado duas vezes; após a segunda aplicação, haveria uma entrevista para a comunicação dos resultados; c) o caráter voluntário da participação. Estando informados e de acordo, os interessados deveriam assinar o Termo de Consentimento Livre e Informado⁵⁶.

No dia da primeira entrevista, foram coletadas informações sócio-demográficas, ocupacionais, sobre o estado de saúde e eventuais tratamentos psiquiátricos ou psicológicos. Neste primeiro encontro também ocorreu a primeira aplicação do método de Rorschach. Esta aplicação foi realizada pelo autor desta pesquisa, seguindo-se as instruções do Sistema Compreensivo (Exner, 1999b):

[Trecho suprimido na versão online]

⁵⁶ O projeto desta pesquisa e o Termo de Consentimento Livre e Informado foram aprovados pelo Comitê de Ética do Instituto de Psicologia da Universidade de São Paulo. O Termo de Consentimento Livre e Informado encontra-se no Apêndice F.

[Trecho suprimido na versão online]

O reteste foi previsto para ocorrer após entre 3 e 4 meses. Estas segundas aplicações do Rorschach também foram realizadas pelo autor desta pesquisa.

9.3.2. Codificação dos protocolos do Rorschach e cálculo da fidedignidade inter-codificadores

Todos os protocolos obtidos foram codificados pelo autor da pesquisa. Foram sorteados, 10 protocolos do teste e 10 protocolos do reteste, que foram então codificados por um outro avaliador, sem conhecimento das codificações realizadas pelo primeiro avaliador. Estas duas classificações independentes foram então comparadas, avaliando-se a fidedignidade inter-codificadores. Para o cálculo da fidedignidade inter-codificadores, utilizamos o software **Rorschach Research Utilities (RRU)** versão 1.0.2, desenvolvido por Harald Janson. As codificações das respostas, agrupadas por protocolo e por avaliador, são introduzidas no programa, que realiza os cálculos das estatísticas de fidedignidade inter-codificadores.

Também confrontamos as respostas do teste com as do reteste e classificamos se cada resposta era igual a ou diferente às respostas fornecida na outra testagem. O critério para considerar duas respostas como sendo iguais é o seguinte:

“[...] uma mesma resposta [resposta igual] foi definida como aquela com conteúdo *idêntico* ao fornecido no teste, usando a mesma ou quase a mesma área de localização (conteúdo idêntico significando o mesmo conteúdo específico em vez de a mesma categoria de codificação de conteúdo). Em outras palavras, uma resposta de dois homens brigando, no teste, usando a área popular do cartão III, seria considerada a mesma resposta que dois homens trabalhando à noite, usando a mesma área, no reteste, mesmo que a segunda resposta incluía um determinante de cor acromática ou de sombreado-difuso. De modo contrário, se a resposta no teste 2 [reteste], usando a mesma área, fossem duas mulheres dançando, a resposta seria considerada diferente. De modo semelhante, um morcego, fornecido ao cartão I como um todo, seria considerado a mesma resposta que um morcego à noite, fornecida à mesma área no teste 2, enquanto uma borboleta, usando a área W [resposta global, em que se utiliza a mancha toda], seria considerada uma resposta diferente no reteste em comparação com o teste”. (EXNER; WEINER, 1995, p. 37)

Sorteamos 5 conjuntos de protocolos de teste e reteste para que suas respostas fossem classificadas, por um avaliador independente, em termos de serem iguais ou diferentes, procedimento que serviu para avaliarmos a fidedignidade da codificação de *respostas iguais*.

A partir da classificação das respostas como iguais ou diferentes, calculamos a porcentagem de respostas iguais para cada um dos participantes.

9.3.3. Seleção das variáveis a serem investigadas

Para o estudo da estabilidade temporal, selecionamos 59 variáveis centrais do SCR: *R, Lambda, F, M, WSumC, EA, FM, m, SumC', SumT, SumV, SumY, FM+m, SumSH, es, Adjes, D-score, AdjD, FC, CF, C, CF+C, Afr, Blends, S, COP, AG, GHR, PHR, HumanContent, Isolate/R, (2), EGO, Fr+rF, FD, MOR, H, (H)+Hd+(Hd), Zf, Zd, DQ+, DQv, XA%, WDA%, X+%, X-%, Xu%, Popular, a, p, Intell, DV+DR, INC+FAB, Sum6, WSum6, PTI, DEPI, CDI e S-CON*.

Também selecionamos:

- 6 proporções para estudar sua estabilidade em termos de direcionalidade: *EB, EA : es, eb, a : p, FC : CF+C e H : (H)+Hd+(Hd)*.
- 7 variáveis para estudo da consistência das categorias interpretativas: *Lambda, estilo passivo, reflexos, EB, D-score, Zd e Textura*.
- 6 *Constelações* para estudo da estabilidade em termos de seu status positivo ou negativo (*SCZI, DEPI, CDI, S-CON, HVI e OBS*) e também o *Índice de Percepção e pensamento [PTI]*, que foi analisado em termos dicotômicos ($PTI \leq 3$ versus $PTI > 3$).

9.3.4. Cálculos das variáveis do SCR e dos resultados do estudo da estabilidade temporal

As respostas dos protocolos da pesquisa foram codificadas pelo pesquisador e estas codificações foram inseridas no programa de computador **Rorschach Interpretive Assistance Program for Windows (RIAP) versão 5.0**. Este software checa eventuais falhas na codificação – por exemplo, omissão de algum elemento de codificação, inconsistências entre os elementos de codificação de uma mesma resposta –, realiza os cálculos das variáveis e fornece hipóteses interpretativas a partir da análise dos resultados obtidos comparados com as normas norte-americanas. As codificações inseridas no **RIAP** foram checadas a fim de se evitar eventuais erros na entrada dos dados.

Por meio do **RIAP**, obtivemos os valores das variáveis do Sumário Estrutural para cada um dos protocolos. Estes valores foram então inseridos manualmente no programa

Statistical Package for Social Sciences (SPSS). O **RIAP** também exporta algumas variáveis para o **SPSS** e, neste programa, realizamos os cálculos das demais variáveis do SCR. Contávamos então com dois bancos de dados da pesquisa: um deles inserido manualmente pelo pesquisador e o outro com os dados exportados pelo **RIAP** e com cálculos realizados no **SPSS**. Confrontamos os dois bancos de dados, de modo a checar erros na entrada destas informações. Deste modo, detectamos erros na inserção manual dos dados, e problemas decorrentes de o **SPSS** não ter realizado arredondamento ao calcular as variáveis do SCR. Para corrigir estas discrepâncias detectadas, utilizamos os Sumários Estruturais gerados pelo **RIAP** e realizamos as correções manualmente. Com todo este processo, garantimos que os dados fossem corretamente inseridos no **SPSS**.

Em relação às variáveis do SCR, levantamos as seguintes informações:

- 1) Estatísticas descritivas das variáveis do SCR
- 2) Comparação das médias do teste com as médias do reteste, utilizando o teste t de student para amostras pareadas.
- 3) Correlações produto-momento de Pearson e correlações de postos de Spearman, que informam sobre a estabilidade temporal das variáveis do SCR.

Por fim, investigamos a direcionalidade das proporções, a consistência das categorias interpretativas e a estabilidade das constelações consideradas em seu status positivo ou negativo.

10. RESULTADOS

10.1. Características da amostra obtida

Prontificaram-se a participar da pesquisa 61 indivíduos. Contudo, 29 participantes não foram incluídos, pois:

- Cinco indivíduos realizaram o teste, mas não o reteste.
- Cinco indivíduos estavam em tratamento psiquiátrico.
- Um indivíduo estava em psicoterapia.
- Um indivíduo realizou o reteste após 5 meses.
- Seis indivíduos conheciam o examinador. Pesquisas indicam que conhecer o examinador pode modificar os resultados do Rorschach. (Destes seis indivíduos que conheciam o examinador, três apresentaram protocolos com menos do que 14 respostas.)
- Três indivíduos forneceram muitas respostas no teste e seguimos o procedimento de retirar o cartão após a quinta resposta (Exner, 1999b, p. 18); somente depois da aplicação, compreendemos que este procedimento não seria indicado no nosso estudo. No reteste, não retiramos os cartões após a quinta resposta, obtendo protocolos ainda mais extensos. Como utilizamos procedimentos diferentes no teste e no reteste, decidimos excluir estes protocolos da análise.
- Seis indivíduos apresentaram menos de 14 respostas em uma ou ambas testagens (além deles, outros três participantes, que conheciam o examinador, forneceram protocolos com menos de 14 respostas e, então, foram excluídos da amostra). Protocolos com menos de 14 respostas são considerados inválidos. Estes protocolos foram analisados em outra parte desta pesquisa.
- Um indivíduo, no início do reteste, conversou com o aplicador, expressando sofrimento e solicitando ajuda. Neste caso, julgamos que esta situação de maior envolvimento pessoal interferiu exageradamente com a situação de testagem, enviesando os resultados, e decidimos excluir este protocolo da amostra.

- Um indivíduo respondeu ao reteste rapidamente, o que julgamos ser decorrência de desinteresse. Este modo de responder provavelmente prejudicou a validade dos resultados, e decidimos excluir este protocolo da amostra.

Obtivemos então 32 participantes compondo nossa amostra.

As informações sócio-demográficas foram obtidas no momento da primeira testagem.

Dos 32 participantes, 25 (78%) eram funcionários da universidade onde se realizou a pesquisa e 7 (22%) não eram funcionários, mas foram informados sobre a pesquisa por meio de algum funcionário e quiseram participar; dentre estes 7 participantes que não eram funcionários, 2 eram alunos da universidade. Nenhum dos participantes estava em tratamento psiquiátrico ou em psicoterapia no momento da pesquisa, mas 10 participantes (31%) informaram já ter realizado psicoterapia em outro momento.

Dos 32 participantes, 24 (75%) eram mulheres e 8 (25%) eram homens, com idades entre 19 e 58 anos (média = 33,22, desvio-padrão = 11,22, mediana = 30,50). Dezesesseis participantes (50%) eram solteiros, 11 (34%) casados ou amasiados, 4 (13%) separados e 1 (3%) viúvo. O nível de instrução variou de ensino fundamental incompleto até pós-graduação completa (anos de instrução: média = 13,34, desvio-padrão = 4,33, mediana = 13,00). A renda familiar variou de R\$ 505 até R\$ 12 000 (média = R\$ 3 683, desvio-padrão = R\$ 2 986, mediana = R\$ 3 000). A partir da renda familiar informada pelo participante, encontramos a classe do Critério de Classificação Econômica Brasil (Associação Brasileira de Empresas de Pesquisa, 2003) com nível de renda médio mais próximo. Em nossa amostra, 6 participantes (19%) pertenciam à classe C, 13 (41%) à classe B e 13 (41%) à classe A, ou seja, predominaram indivíduos das classes A e B. A Tabela 23 resume as características sócio-demográficas dos participantes da nossa amostra.

Tabela 23 – Características sócio-demográficas dos participantes da nossa amostra

Variável			
Idade (em anos)	Mínimo = 19, Máximo = 58, Média = 33,22, DP = 11,22, Mediana = 30,50		
Anos de Instrução	Mínimo = 5, Máximo = 23, Média = 13,34, DP = 4,33, Mediana = 13,00		
Estado Civil	Solteiros = 16 (50%) Casados/Amasiados = 11 (34%) Separados = 4 (13%) Viúvos = 1 (3%)		
Nível sócio- econômico		<i>f</i>	%
	A	13	41
	B	13	41
	C	6	19

O intervalo entre as testagens variou de 90 a 130 dias, com média = 104 dias, desvio-padrão = 11 dias e mediana = 104 dias.

10.2. Porcentagem de respostas iguais

As porcentagens de respostas iguais nos protocolos de teste e reteste variaram de 12% a 77%, com média = 47%, desvio-padrão = 16% e mediana = 50%.

Sorteamos 5 conjuntos de protocolos de teste e reteste para cálculo da fidedignidade da classificação das respostas iguais. Um avaliador com formação em Psicologia e conhecimentos de graduação sobre o método de Rorschach, mas sem experiência de utilização do instrumento, leu o critério para classificar respostas iguais e realizou treino de classificação de 1 conjunto de protocolos de teste e reteste, com supervisão do autor da pesquisa. Posteriormente, realizou, de modo independente, a classificação das respostas dos 5 conjuntos de protocolos sorteados. A porcentagem de concordância entre a classificação do autor e a do avaliador independente, para um total de 226 respostas dos 5 conjuntos de protocolos de teste e reteste, foi de 83%, com kappa = 0,66, o que corresponde a um nível de concordância substancial.

Observamos que algumas respostas com diversos elementos relacionados (por exemplo: “duas pessoas carregando um cesto e uma borboleta voando no meio delas”) foram fornecidas numa das testagens e, na outra testagem, os elementos apareceram em respostas separadas (“duas pessoas”, “uma borboleta”); classificamos casos como este como sendo respostas diferentes. Outras vezes, uma resposta foi fornecida de modo mais simples numa das testagens (por exemplo, “dois tigres”) e, na outra testagem, a resposta fornecida continha mais elementos (“dois tigres escalando um montanha”); casos como estes foram classificados às vezes como sendo respostas iguais – quando os elementos acrescentados foram julgados como pouco importantes para a resposta como um todo –, às vezes como sendo respostas diferentes – quando os elementos acrescentados foram julgados como importantes para a resposta como um todo. O critério para classificação de respostas iguais, fornecido por Exner e Weiner (1995) deveria especificar melhor como proceder em situações como essas; a falta de orientações precisas sobre como proceder nesses casos pode ter sido um fator responsável por parte das discordâncias entre os avaliadores na classificação das respostas como iguais ou diferentes.

Encontramos informações sobre a porcentagem de respostas iguais no teste e reteste apenas nos estudos do SCR em que houve reteste após intervalo de poucos dias, solicitando-se respostas diferentes: 1) Exner (1980) encontrou que crianças instruídas a fornecer respostas diferentes no reteste apresentaram 14% de respostas iguais, enquanto que crianças cujo reteste foi realizado com as instruções originais forneceram 86% de respostas iguais; 2) Haller e Exner (1985) encontraram que os adultos instruídos a fornecer respostas diferentes no reteste forneceram aproximadamente 1/3 de respostas iguais, enquanto os participantes que realizaram o reteste com as instruções originais forneceram aproximadamente 2/3 de respostas iguais. As porcentagens de respostas iguais nestes dois estudos foram então maiores do que a encontrada na nossa pesquisa. O intervalo de poucos dias entre o teste e o reteste nas pesquisas mencionadas deve ter contribuído para a maior porcentagem de respostas iguais. Talvez, também, diferenças no entendimento do critério para codificar as respostas como iguais tenham levado Exner e seus colaboradores a adotar uma convenção de codificar como iguais respostas que nós codificaríamos como diferentes – o que pode, em parte, explicar a porcentagem mais baixa de respostas iguais encontradas na nossa pesquisa.

Janson e Stattin (1994, apud Grønnerød, 2003)⁵⁷ reportaram uma recorrência de 45% das respostas numa amostra de 20 participantes testados aos 18 e aos 35 anos de idade. Uma parte da amostra de participantes, aqueles com vida estável, apresentou 54% de respostas repetidas, enquanto que o restante dos participantes apresentou 35% de respostas repetidas. A porcentagem de respostas iguais em nosso estudo foi semelhante à obtida neste estudo.

10.3. Fidedignidade inter-codificadores

10.3.1 Fidedignidade inter-codificadores ao nível das respostas

Dez protocolos do teste e dez protocolos do reteste foram sorteados e codificados por um avaliador independente. Estes 20 protocolos continham um total de 437 respostas.

Primeiramente, informamos a fidedignidade inter-codificadores para as decisões dicotômicas de codificação (assinalar ou não assinalar o código), que corresponde também à fidedignidade para os códigos individuais (Tabela 24). Esta análise foi denominada *agreement for single category*, por Janson (2003). Cabe lembrar que, neste caso, os resultados de *iota* são idênticos aos resultados de *kappa*.

Para a localização, a porcentagem de concordância foi maior que 94% para todos os códigos e os valores de *iota* variaram de 0,81 para *Dd* a 0,98 para *W*, com média = 0,90 e mediana = 0,91

Para a qualidade evolutiva, a porcentagem de concordância foi maior do que 88% para todos os códigos e os valores de *iota* variaram de 0,00 para *DQv/+* a 0,76 para *DQ+*, com média = 0,45 e mediana = 0,52.

⁵⁷ JANSON, H.; STAT TIN, H. Recurring objects in Rorschach records over a period of about 30 years: How often do people report the same thing again? **Rorschachiana 19: Yearbook of the International Rorschach Society**, p. 156-170. I. B. Weiner (editor). Göttingen, Alemanha: Hogrefe e Huber, 1994.

Para os determinantes, a porcentagem de concordância foi maior que 90% para todos os códigos e os valores de iota variaram de 0,00 para *C'F*, *TF* e *VF* a 0,83 para *M*, com média = 0,52 e mediana = 0,55.

Para a qualidade formal, a porcentagem de concordância foi maior do que 80% e os valores de iota variaram de 0,00 para *FQ+* a 0,79 para *FQo*, com média = 0,48 e mediana = 0,58.

Para os pares, a porcentagem de concordância foi de 92%, com $\tau = 0,82$.

Para os conteúdos, a porcentagem de concordância foi maior do que 96% e os valores de iota variaram de 0,00 para *Ex* a 1,00 para *Bl* e *Ge* com média = 0,71 e mediana = 0,78.

Para Popular, a porcentagem de concordância foi de 93%, com $\tau = 0,77$.

Para a atividade organizativa, houve concordância de 90% em termos de assinalar ou não assinalar notas *Z*, com $\tau = 0,79$.

Para os 6 códigos especiais críticos (*DV*, *DR*, *INCOM*, *FABCOM*, *ALOG*, *CONTAM*), a porcentagem de concordância foi maior do que 96% e os valores de iota variaram de -0,01 para *DVI* e *DR1* a 0,56 para *FABI*, com média = 0,15 e mediana = 0,00. Calculando a concordância para os códigos não levando em consideração as diferenças entre nível 1 e nível 2, melhora apenas a concordância de *INCOM*, com $\tau = 0,71$.

Para os demais códigos especiais assinalados (*PSV*, *AB*, *AG*, *COP*, *MOR*, *PER*, *GHR* e *PHR*), a porcentagem de concordância foi maior do que 94%, com valores de iota de 0,00 para *PSV* a 0,77 para *GHR*, com média = 0,55 e mediana = 0,58.

Considerando todos os códigos especiais em conjunto, os valores de iota apresentaram média = 0,36 e mediana = 0,53.

Tabela 24 – fidedignidade inter-codificadores para decisões dicotômicas de codificação

Códigos	BR	PA	Iota
Localização			
<i>W</i>	0,2883	0,9908	0,9777
<i>D</i>	0,5538	0,9542	0,9074
<i>Dd</i>	0,1579	0,9497	0,8109
<i>S</i>	0,0801	0,9863	0,9069
Qualidade Evolutiva			
<i>DQ+</i>	0,2082	0,9222	0,7643
<i>DQo</i>	0,7231	0,8902	0,7271
<i>DQv/+</i>	0,0149	0,9703	-0,0042
<i>DQv</i>	0,0538	0,9291	0,3177
Determinantes			
<i>F</i>	0,4600	0,9085	0,8162
<i>M</i> (ausente/presente)	0,1316	0,9611	0,8299
<i>Ma</i>	0,0412	0,9588	0,4810
<i>Mp</i>	0,0904	0,9474	0,6801
<i>Map</i>	0,0000	1,0000	–
<i>FM</i> (ausente/presente)	0,1602	0,9497	0,8129
<i>FMa</i>	0,0686	0,9588	0,6792
<i>FMp</i>	0,0915	0,9314	0,5898
<i>FMap</i>	0,0000	1,000	–
<i>m</i> (ausente/presente)	0,0423	0,9840	0,8025
<i>ma</i>	0,0252	0,9817	0,6277
<i>mp</i>	0,0172	0,9748	0,2544
<i>map</i>	0,0000	1,0000	–
<i>Cor cromática</i> (ausente/presente)	0,1487	0,9817	0,9278
<i>FC</i>	0,0767	0,9382	0,5645
<i>CF</i>	0,0595	0,9314	0,3934
<i>C</i>	0,0126	0,9886	0,5401
<i>Cn</i>	0,0000	1,0000	–
<i>Cor acromática</i> (ausente/presente)	0,0481	0,9817	0,8000
<i>FC'</i>	0,0435	0,9817	0,7799
<i>C'F</i>	0,0046	0,9908	-0,0034
<i>C'</i>	0,0000	1,000	–
<i>Textura</i> (ausente/presente)	0,0126	0,9931	0,7238
<i>FT</i>	0,0114	0,9954	0,7978
<i>TF</i>	0,0011	0,9977	0,0000
<i>T</i>	0,0000	1,0000	–
<i>Vista</i> (ausente/presente)	0,0183	0,9863	-0,6181
<i>FV</i>	0,0137	0,9863	0,4948
<i>VF</i>	0,0046	0,9908	0,0000
<i>V</i>	0,0000	1,0000	–
<i>Sombreado Difuso</i> (ausente/presente)	0,0561	0,9474	0,5042
<i>FY</i>	0,0423	0,9382	0,2468
<i>YF</i>	0,0137	0,9817	0,3282
<i>Y</i>	0,0000	1,0000	–
<i>FD</i>	0,0252	0,9725	0,4415
<i>Reflexos</i> (ausente/presente)	0,0183	0,9954	0,8727
<i>Fr</i>	0,0149	0,9931	0,7658
<i>rF</i>	0,0034	0,9977	0,6656

BR: taxa de ocorrência (Base Rate)

PA: proporção de concordância (Proportion of Agreement)

Tabela 24 – fidedignidade inter-codificadores para decisões dicotômicas de codificação (continua)

Códigos	BR	PA	Iota
Qualidade Formal			
<i>FQ+</i>	0,0011	0,9977	0,0000
<i>FQo</i>	0,4691	0,8947	0,7887
<i>FQu</i>	0,3261	0,8146	0,5783
<i>FQ-</i>	0,1865	0,8970	0,6607
<i>FQnone</i>	0,0172	0,9794	0,3918
Pares			
(2)	0,3032	0,9245	0,8223
Conteúdos			
<i>H</i>	0,0675	0,9840	0,8728
(<i>H</i>)	0,0320	0,9863	0,7787
<i>Hd</i>	0,0081	0,9748	0,8433
(<i>Hd</i>)	0,0297	0,9863	0,7622
<i>Hx</i>	0,0423	0,9657	0,5797
<i>A</i>	0,3753	0,9817	0,9610
(<i>A</i>)	0,0092	0,9954	0,7478
<i>Ad</i>	0,1018	0,9657	0,8125
(<i>Ad</i>)	0,0137	0,9954	0,8310
<i>An</i>	0,0767	0,9794	0,8545
<i>Art</i>	0,0389	0,9771	0,6940
<i>Ay</i>	0,0069	0,9908	0,3292
<i>Bl</i>	0,0046	1,0000	1,0000
<i>Bt</i>	0,0767	0,9886	0,9192
<i>Cg</i>	0,0366	0,9908	0,8703
<i>Cl</i>	0,0126	0,9977	0,9079
<i>Ex</i>	0,0023	0,9954	0,0000
<i>Fi</i>	0,0172	0,9977	0,9322
<i>Fd</i>	0,0126	0,9886	0,5409
<i>Ge</i>	0,0069	1,0000	1,0000
<i>Hh</i>	0,0149	0,9794	0,3023
<i>Ls</i>	0,0744	0,9703	0,7840
<i>Na</i>	0,0195	0,9657	0,1051
<i>Sc</i>	0,0263	0,9794	0,5986
<i>Sx</i>	0,0252	0,9863	0,7208
<i>Xy</i>	0,0080	0,9977	0,8560
<i>Id</i>	0,0286	0,9748	0,5471
Populares			
<i>P</i>	0,1854	0,9314	0,7728
Atividade Organizativa			
Assinalar uma nota Z	0,3947	0,9016	0,7941

BR: taxa de ocorrência (Base Rate)

PA: proporção de concordância (Proportion of Agreement)

Tabela 24 – fidedignidade inter-codificadores para decisões dicotômicas de codificação (conclusão)

Códigos	BR	PA	Iota
Códigos Especiais			
<i>DV</i> (ausente/presente)	0,0103	0,9794	-0,0072
<i>DV1</i>	0,0103	0,9794	-0,0072
<i>DV2</i>	0,0000	1,0000	–
<i>DR</i> (ausente/presente)	0,0172	0,9657	-0,0155
<i>DR1</i>	0,0149	0,9703	-0,0143
<i>DR2</i>	0,0023	0,9954	0,0000
<i>INC</i> (ausente/presente)	0,0286	0,9840	0,7119
<i>INC1</i>	0,0252	0,9771	0,5337
<i>INC2</i>	0,0034	0,9931	0,0000
<i>FAB</i> (ausente/presente)	0,0183	0,9817	0,4908
<i>FAB1</i>	0,0160	0,9863	0,5645
<i>FAB2</i>	0,0023	0,9954	0,0000
<i>ALOG</i>	0,0000	1,0000	–
<i>CONTAM</i>	0,0000	1,0000	–
<i>PSV</i>	0,0023	0,9954	-0,0023
<i>AB</i>	0,0069	0,9954	0,6646
<i>AG</i>	0,0195	0,9840	0,5815
<i>COP</i>	0,0172	0,9840	0,5261
<i>MOR</i>	0,0389	0,9680	0,5718
<i>PER</i>	0,0172	0,9840	0,5255
<i>CP</i>	0,0000	1,0000	–
<i>GHR</i>	0,1178	0,9519	0,7691
<i>PHR</i>	0,1144	0,9451	0,7290

BR: taxa de ocorrência (Base Rate)

PA: proporção de concordância (Proportion of Agreement)

Para a interpretação dos valores de iota, seguimos os parâmetros adotados por Acklin et al. (2000) para a interpretação do kappa: valores menores do que 0,61 indicam fidedignidade inaceitável, valores de 0,61 a 0,80 indicam fidedignidade substancial, e valores maiores do que 0,80 indicam fidedignidade excelente.

Considerando as 104 decisões de codificação calculadas pelo programa **RRU**, encontramos: fidedignidade excelente para 25 decisões de codificação (24%), substancial para outras 25 decisões de codificação (24%), inaceitável para 43 decisões de codificação (41%) e não pode ser calculada para 11 decisões de codificação (11%).

Em seguida, apresentamos a concordância para as decisões policotômicas de codificação (Tabela 25). Janson (2003) denomina esta análise como *agreement for multi-category coding decisions*. Os valores de iota indicam concordância substancial ou excelente

para a maioria das decisões de codificação, com exceção de vista, sombreado difuso e os códigos especiais críticos.

Tabela 25 – Fidedignidade inter-codificadores para decisões policotômicas de codificação

Códigos	PA	iota
Localização (W, D, Dd)	0,9474	0,9101
DQ (+, o, v/+, v)	0,8558	0,6671
M (*, Ma, Mp, Map)	0,9336	0,7191
FM (*, FMa, FMp, FMap)	0,9199	0,7164
m (*, a, p, ap)	0,9703	0,6373
Cor (*, FC, CF, C, Cn)	0,9199	0,6997
Cor Acromática (*, FC', C'F, C')	0,9771	0,7511
Textura (*, FT, TF, T)	0,9931	0,7242
Vista (*, FV, VF, V)	0,9817	0,4935
Sombreado Difuso (*, FY, YF, Y)	0,9336	0,3839
Reflexos (*, Fr, rF)	0,9931	0,8096
FQ (+, o, u, -, none)	0,7918	0,6740
Z	0,8009	0,6758
DV (*, DV1, DV2)	0,9794	-0,0072
DR (*, DR1, DR2)	0,9771	0,5897
INC (*, INC1, INC2)	0,9657	-0,0139
FAB (*, FAB1, FAB2)	0,9817	0,4919

PA = proporção de concordância (Proportion of Agreement)

* Indica decisão de não assinalar o código

Em seguida, apresentamos a concordância para os segmentos da codificação (Tabela 26). Janson (2003) denomina esta análise *agreement for coding segments*. Os valores de iota encontrados indicam concordância substancial ou excelente para todos os segmentos, exceto para os códigos especiais.

Tabela 26 – Fidedignidade inter-codificadores para os segmentos da codificação

Segmento	PA	Iota
Localizações e espaço branco (2 variáveis)	0,9668	0,9095
Determinantes (11 variáveis)	0,9549	0,7098
Conteúdos (27 variáveis)	0,9850	0,8086
Códigos Especiais (14 variáveis)	0,9868	0,4627

Por fim, temos a concordância geral para as codificações do SCR, que Janson (2003) denomina *agreement for whole responses*. Nesta pesquisa, a porcentagem de concordância foi de 0,9687, com iota = 0,7473, que se situa na faixa de concordância substancial.

10.3.2. Fidedignidade inter-codificadores ao nível do protocolo

Nesta análise, investigamos o grau de concordância dos resultados do Sumário Estrutural. Para o seu modelo de análise de dois fatores (two way), o programa **Rorschach Research Utilities** fornece a estatística *iota* e ICC (2,1). Os resultados são apresentados na Tabela 27.

Tabela 27 – Fidedignidade inter-codificadores para as variáveis do Sumário Estrutural

Agrupamento	Iota	ICC (2,1)
Seção Principal		
<i>R</i>	1,0000	1,0000
<i>Lambda</i>	0,7048	0,7153
<i>F</i>	0,9054	0,9097
<i>M</i>	0,7547	0,7640
<i>WSumC</i>	0,8860	0,8911
<i>EA</i>	0,8798	0,8851
<i>FM</i>	0,9273	0,9307
<i>m</i>	0,8627	0,8687
<i>FM+m</i>	0,9067	0,9110
<i>SumC'</i>	0,8814	0,8866
<i>SumV</i>	0,6954	0,7062
<i>SumT</i>	0,7500	0,7595
<i>SumY</i>	0,2600	0,2700
<i>SumSH</i>	0,8571	0,8633
<i>es</i>	0,9031	0,9075
<i>Adjes</i>	0,9383	0,9412
<i>D-score</i>	0,6975	0,7082
<i>AdjD</i>	0,6652	0,6765
Afeto		
<i>FC</i>	0,7943	0,8025
<i>CF</i>	0,4654	0,4782
<i>C</i>	0,5098	0,5226
<i>CF + C</i>	0,6272	0,6391
<i>Afr</i>	1,0000	1,0000
<i>Blends</i>	0,8528	0,8592
<i>S</i>	0,8795	0,8848
Relações Interpessoais		
<i>COP</i>	0,4167	0,4292
<i>AG</i>	0,3448	0,3565
<i>GHR</i>	0,5783	0,5908
<i>PHR</i>	0,7196	0,7299
<i>HumanContent</i>	N/C	N/C
<i>Isolate/R</i>	0,8434	0,8500

N/C = não calculado pelo programa RRU

Tabela 27 – Fidedignidade inter-codificadores para as variáveis do Sumário Estrutural (conclusão)

Agrupamento	Iota	ICC (2,1)
Autopercepção		
(2)	0,7843	0,7929
<i>EGO</i>	0,5981	0,6104
<i>Fr + rF</i>	0,9074	0,9116
<i>FD</i>	0,2857	0,2963
<i>MOR</i>	0,6267	0,6386
<i>H</i>	0,8478	0,8543
<i>(H) + Hd + (Hd)</i>	0,9286	0,9320
Processamento		
<i>Zf</i>	0,8091	0,8169
<i>Zd</i>	-0,0078	-0,0082
<i>DQ+</i>	0,7958	0,8040
<i>DQv</i>	0,2953	0,3060
Mediação		
<i>XA%</i>	0,5220	0,5348
<i>WDA%</i>	0,3987	0,4111
<i>X+</i>	0,7762	0,7850
<i>Xu%</i>	0,3576	0,3695
<i>X-%</i>	0,6226	0,6346
<i>Popular</i>	0,7260	0,7361
Ideação		
<i>a</i>	0,6756	0,6867
<i>p</i>	0,7186	0,7289
<i>2AB+Art+Ay</i>	0,6988	0,7095
<i>Sum6</i>	0,6350	0,6468
<i>WSum6</i>	0,7434	0,7531
Constelações		
<i>PTI</i>	0,3987	0,4111
<i>DEPI</i>	0,5714	0,5839
<i>CDI</i>	0,6000	0,6122
<i>S-CON</i>	0,6856	0,6966
<i>SCZI</i>	0,7000	0,7107

Dentre as 57 variáveis do Sumário Estrutural, sobre as quais temos informação dos valores de *iota*, 19 (33%) apresentaram fidedignidade inter-codificadores excelente e 22 (39%) apresentaram fidedignidade substancial. Contudo, 16 variáveis (28%) apresentaram fidedignidade inadequada: *SumY*, *CF*, *C*, *COP*, *AG*, *GHR*, *EGO*, *FD*, *Zd*, *DQv*, *XA%*, *WDA%*, *Xu%*, *PTI*, *DEPI* e *CDI*.

10.3.3 Comentário sobre os resultados do estudo da fidedignidade inter-codificadores

Os resultados da fidedignidade inter-codificadores indicaram concordância substancial para a maioria das variáveis do SCR. Contudo, algumas variáveis apresentaram fidedignidade insatisfatória. Em parte, a baixa fidedignidade para algumas codificações está associada às baixas taxas de ocorrência dos códigos assinalados. A Figura 3 mostra a fidedignidade das codificações (valores de *iota*) em função da taxa de ocorrência dos códigos. Observamos que a concordância para os códigos com baixa taxa de ocorrência é bastante variada e, à medida que aumenta a taxa de ocorrência, há uma tendência no sentido de aumento da concordância. Quando a taxa de ocorrência atinge 10%, os valores de *iota* são quase todos maiores do que 0,60, indicando concordância substancial.

Observamos também que existem pontos fora desta tendência de aumento da fidedignidade à medida que aumenta a taxa de ocorrência. No gráfico, correspondem a “picos invertidos”, todos eles relativos à codificação da qualidade formal: *FQ-* apresentou taxa de ocorrência = 19% e *iota* = 0,66; *FQu* apresentou taxa de ocorrência = 33% e *iota* = 0,58. Devemos lembrar que a codificação da qualidade formal é realizada consultando-se a Tabela de Qualidade Formal do Sistema Compreensivo, construída a partir de respostas de norte-americanos. Quando a resposta encontra-se listada na Tabela, o código é diretamente assinalado. Quando a resposta não se encontra na Tabela, nem é possível aproximar a resposta fornecida a alguma resposta da Tabela, o avaliador deve decidir se a resposta pode ser vista rápida e facilmente (codificada *FQu*) ou não (codificada *FQ-*). Suspeitamos que a menor concordância encontrada para *FQu* e *FQ-* decorra destas respostas que não estão listadas na Tabela e cuja qualidade formal assinalada depende do julgamento do avaliador. Este resultado sugere também que a Tabela de Qualidade Formal norte-americana talvez não seja tão adequada para brasileiros. A menor fidedignidade das codificações de *FQu* e *FQ-* refletiu-se na menor fidedignidade de *XA%*, *WDA%*, *Xu%* e *X-%*.

Figura 3 – Níveis de fidedignidade das codificações em função da taxa de ocorrência dos códigos

A baixa taxa de ocorrência dos códigos parece ser um dos fatores responsáveis pela baixa fidedignidade inter-codificadores obtida para os escores de *SumY*, *CF*, *C*, *COP*, *AG*, *FD* e *DQv*. A baixa fidedignidade dos escores de *GHR* é de difícil compreensão: ao nível das respostas, a codificação de *GHR* apresentou concordância substancial ($\kappa = 0,77$), porém, ao nível do protocolo, a concordância diminuiu para $\kappa = 0,59$. Algo semelhante ocorreu para *EGO*: ao nível das respostas, pares e reflexos apresentaram concordância excelente; porém, estas variáveis compondo o Índice de Egocentricidade resultaram em baixa fidedignidade dos escores de *EGO*. A menor concordância na codificação da qualidade formal refletiu-se em baixa fidedignidade dos escores de *XA%*, *WDA%* e *Xu%*, e também prejudicou a fidedignidade do Índice de Percepção e Pensamento [*PTI*], que envolve os índices de qualidade formal. A menor fidedignidade do Índice de egocentricidade [*EGO*] e de sombreado difuso [*SumY*] também deve ter contribuído para a fidedignidade insuficiente do Índice de Depressão [*DEPI*], do qual estas variáveis participam. A menor fidedignidade de *FC*, *CF*, *COP* e *AG* deve ter contribuído para a menor fidedignidade do Índice de Déficit Relacional [*CDI*], do qual participam.

McDowell e Acklin (1996) destacam fatores que influenciam a fidedignidade inter-codificadores: treino dos avaliadores, semelhança em termos de experiências de codificação, população da qual os protocolos foram obtidos, vieses e flutuações do avaliador ao codificar

as respostas, características das respostas avaliadas e métodos estatísticos de quantificação da concordância das codificações.

Nesta pesquisa, os codificadores foram pessoas devidamente treinadas na codificação do Rorschach: o pesquisador foi o tradutor do **Manual de Classificação do Rorschach para o Sistema Compreensivo** (Exner, 1999b), realizou sua pesquisa de mestrado com o Rorschach e leciona no curso de especialização no método de Rorschach promovido pela Associação Brasileira de Rorschach (ASBRo), sendo o responsável pelo ensino das codificações das respostas; o segundo codificador participou da pesquisa normativa com indivíduos do interior de São Paulo, na qual aplicou o Rorschach, codificou os protocolos e discutiu dúvidas de codificação com outros membros da equipe de psicólogos responsáveis pela coleta dos dados.

O nível de complexidade das respostas utilizadas na investigação da fidedignidade das codificações parece ser semelhante ao da amostra normativa de adultos não-pacientes da cidade de São Paulo (considerando que nossa amostra apresentou R maior do que a amostra normativa, mas também valores maiores de Λ). A complexidade das respostas de nossa amostra provavelmente é menor do que a dos estudos com grupos clínicos e certamente é menor do que as das pesquisas norte-americanas, considerando que os valores de R são maiores e os valores de Λ são menores nos Estados Unidos. A menor complexidade das respostas em nossa amostra deveria contribuir para uma maior concordância, pois respostas menos complexas são mais fáceis de serem codificadas, gerando menos dúvidas e discordâncias. Porém, os resultados da fidedignidade das codificações foram menores em nossa pesquisa do que nas pesquisas norte-americanas e do que em outras pesquisas brasileiras. Por outro lado, para algumas respostas, o menor nível de complexidade pode refletir uma falta de informações mais completas para uma codificação adequada – seja porque o aplicador falhou em aprofundar o inquérito, seja porque o participante falhou em explicar melhor suas respostas –, o que pode ter contribuído para diminuir a fidedignidade das codificações.

A fidedignidade inter-codificadores para os códigos especiais críticos foi menor do que a de outros aspectos da codificação das respostas. A baixa taxa de ocorrência destes

códigos pode ter contribuído para a menor fidedignidade, assim como eventuais deficiências dos critérios de codificação.

Considerando as médias de fidedignidade dos códigos agrupados em segmentos da codificação, os resultados foram menores na nossa pesquisa do que em outros estudos brasileiros com o SCR; apenas os códigos de localização apresentaram melhor fidedignidade em nossa amostra (Tabela 28).

Tabela 28 – Valores médios de kappa para a fidedignidade inter-codificadores em pesquisas brasileiras com o SCR e na nossa pesquisa

	Loc	DQ	Det	FQ	Pairs	Cont	P	Z	SS	Média
Silva Neto (1999)	0,92	0,86	0,84	0,64	0,86	0,80	0,97	0,84	0,48	0,80
Duarte (2000)	0,70	0,80	0,72	0,64	0,92	0,74	0,84	0,82	0,53	0,75
Santo Antonio (2001)	0,85	0,73	0,84	0,67	0,82	0,83	0,82	0,91*	0,47	0,75
Antúñez (2004)	0,87	0,52	0,77	0,56	0,89	0,70	0,90	0,89	0,52	0,74
Lamounier (2005)	0,79	0,67	0,76	0,76	0,79	0,74	0,91	0,70	0,52	0,74
Durat Júnior (2006)	0,81	0,70	0,88	0,78	0,79	0,81**	0,77	–	0,68***	0,78
Média	0,82	0,71	0,80	0,68	0,85	0,77	0,87	0,83	0,53	0,76
Nossa pesquisa	0,90	0,45	0,52	0,48	0,82	0,71	0,77	0,79	0,36	0,64

Loc = Localizações, DQ = Qualidade Evolutiva (Developmental Quality), Det = Determinantes, FQ = Qualidade Formal (Formal Quality), Pairs = pares, Cont = Conteúdos, P = Popular; Z = notas Z, SS = Códigos Especiais (Special Scores)

* ICC. ** média da fidedignidade para os conteúdos humanos, animais e anatomia. *** média da fidedignidade para os códigos COP, AG e MOR.

Um dos problemas nesta comparação é usar médias. Os segmentos da codificação *localização*, *qualidade evolutiva* e *qualidade formal*, apresentam apenas 4 códigos. Baixa fidedignidade de um destes códigos pode rebaixar bastante a média do segmento. Isto ocorreu em nossa amostra para a qualidade evolutiva e para a qualidade formal. Descontando este problema, ainda encontramos, em nossa amostra, menor fidedignidade da codificação dos determinantes e das respostas populares do que nos outros estudos brasileiros. A baixa fidedignidade para os códigos especiais foi um achado de todos os estudos brasileiros e também ocorreu na nossa amostra.

Chamou nossa atenção o nível de concordância obtido para a codificação das respostas populares ($\kappa = 0,77$). A codificação de uma resposta como popular ou não-popular é uma decisão simples e direta, devendo haver fidedignidade inter-codificadores quase perfeita. Investigamos as respostas em que o código *Popular* foi assinalado pelos dois avaliadores,

identificando 30 respostas em que houve discordância na codificação de *Popular*. A análise das respostas em que houve discordância na codificação de *Popular* encontra-se no Apêndice G.

O segundo codificador cometeu 17 erros de omissão (ou seja, deixou de codificar *Popular* quando deveria ter codificado) e 9 erros de comissão (ou seja, codificou *Popular* quando não deveria ter codificado) – a maioria destes erros ocorreu porque o avaliador não notou que a resposta foi vista com o cartão invertido, caso em que não é *Popular*. O primeiro codificador cometeu um erro de comissão. Três respostas foram consideradas como discordâncias legítimas, ou seja, são respostas para as quais o critério de codificação não permite decidir se o código deve ou não ser assinalado. Duas destas discordâncias legítimas envolveram animais (camaleão e lagarto) no cartão VIII. O critério de codificação define como popular, na área D1 do cartão VIII, uma figura animal inteira, comumente da variedade canina, felina ou roedor (EXNER, 2003). A dúvida é se estes animais que não são comumente vistos (*FQu*) devem ser codificados como populares. A outra discordância legítima ocorreu para uma resposta ao cartão VI, em que foi visto um animal. Neste cartão, a resposta popular é pele de animal, sendo que frequentemente a pele está incluída na descrição de um animal inteiro. A decisão de codificar *P* nestas respostas baseia-se em a pele ser mencionada ou estar claramente implicada na resposta (EXNER, 2003).

Notamos então que o segundo codificador cometeu muitos erros de omissão e também erros de comissão, por não observar que a resposta havia sido fornecida com o cartão invertido. O número elevado de erros na codificação das respostas populares por parte do segundo avaliador sugere que ele codificou as respostas com menor cuidado do que o necessário. Assim sendo, possivelmente ocorreram também erros na codificação de outros aspectos das respostas, o que ajudaria a explicar a menor fidedignidade das codificações encontrada na nossa pesquisa.

10.4. Estatísticas descritivas das variáveis do SCR

As estatísticas descritivas das variáveis do Rorschach selecionadas para esta pesquisa, no teste e no reteste, encontram-se apresentadas nas Tabelas 29 e 30, respectivamente.

Tabela 29 – Estatísticas descritivas para as variáveis do SCR no teste

Variável	Média	DP	Min	Max	Freq	Mediana	Moda	Assimetria	Curtose
<i>R</i>	22,94	7,31	14	42	32	22,00	–	0,92	0,52
<i>Lambda</i>	1,04	0,56	0,35	2,83	32	0,94	–	1,16	1,95
<i>F</i>	11,13	5,23	4	26	32	9,00	9	1,04	0,83
<i>M</i>	3,66	2,16	0	9	30	4,00	4	0,23	-0,28
<i>WSumC</i>	3,28	2,47	0,0	11,5	28	3,00	3,5	1,14	2,56
<i>EA</i>	6,94	3,01	1,5	12,5	32	6,75	–	0,04	-0,93
<i>FM</i>	3,16	1,89	0	8	30	3,00	3	0,44	0,09
<i>m</i>	0,75	1,19	0	4	12	0,00	0	1,50	1,02
<i>FM+m</i>	3,91	2,51	0	11	30	3,50	3	0,78	0,73
<i>SumC'</i>	0,78	1,13	0	3	13	0,00	0	1,18	-0,11
<i>SumT</i>	0,31	0,74	0	3	6	0,00	0	2,50	5,84
<i>SumV</i>	0,44	0,67	0	2	11	0,00	0	1,28	0,49
<i>SumY</i>	1,25	1,37	0	4	19	1,00	0	0,81	-0,61
<i>SumSH</i>	2,78	2,31	0	9	27	3,00	1	0,72	0,01
<i>es</i>	6,69	4,06	1	20	32	6,00	3	1,29	2,57
<i>Adjes</i>	5,66	3,00	1	15	32	5,50	3	1,01	1,62
<i>D-score</i>	-0,03	1,15	-3	4	32	0,00	0	0,74	5,49
<i>AdjD</i>	0,28	1,05	-2	4	32	0,00	0	1,33	4,46
<i>FC</i>	1,38	1,54	0	5	21	1,00	1	1,13	0,22
<i>CF</i>	2,03	1,68	0	5	24	2,00	0	0,30	-1,12
<i>C</i>	0,38	0,79	0	4	9	0,00	0	3,34	13,94
<i>CF+C</i>	2,41	1,98	0	8	26	2,00	–	0,83	0,56
<i>Afr</i>	0,52	0,20	0,25	1,08	32	0,50	–	1,17	1,48
<i>Blends</i>	2,56	2,69	0	9	23	2,00	0	1,03	0,25
<i>S</i>	2,78	3,09	0	13	26	2,00	2	2,10	4,85
<i>COP</i>	0,84	1,08	0	4	17	1,00	0	1,47	1,66
<i>AG</i>	0,44	0,72	0	2	10	0,00	0	1,36	0,46
<i>GHR</i>	2,94	1,76	0	7	29	3,00	2	0,29	-0,18
<i>PHR</i>	3,31	2,73	0	11	26	3,50	–	0,73	0,43
<i>Human Content</i>	5,63	3,08	0	15	31	6,00	7	0,65	1,44
<i>Isolate/R</i>	0,24	0,15	0,00	0,60	31	0,22	0,28	0,86	0,45
<i>(2)</i>	7,78	3,49	2	16	32	7,00	6	0,99	0,60
<i>3r+(2)/R</i>	0,38	0,10	0,27	0,76	32	0,355	0,33	0,78	4,62
<i>Fr+rF</i>	0,34	0,79	0	3	6	0,00	0	2,25	4,13
<i>FD</i>	0,56	0,80	0	2	12	0,00	0	0,99	-0,66
<i>MOR</i>	1,47	1,72	0	6	19	1,00	0	1,15	0,50
<i>H</i>	2,00	1,85	0	6	26	1,00	1	0,91	-0,31
<i>(H)+Hd+(Hd)</i>	3,63	2,46	0	9	31	3,00	1	0,38	-0,86
<i>Zf</i>	10,34	3,69	3	18	32	10,00	10	0,04	-0,31
<i>Zd</i>	0,77	4,32	-9,5	11,5	32	0,50	1,5	0,05	1,28
<i>DQ+</i>	5,41	2,73	1	11	32	5,50	3	0,36	-0,94
<i>DQv</i>	1,38	1,43	0	5	19	1,00	0	0,69	-0,43
<i>XA%</i>	0,79	0,09	0,63	0,95	32	0,79	–	0,10	-0,82
<i>WDA%</i>	0,82	0,09	0,63	1,00	32	0,815	0,71	-0,05	-0,78
<i>Popular</i>	4,19	1,77	1	7	32	4,00	–	0,03	-1,05
<i>X+%</i>	0,47	0,11	0,21	0,76	32	0,47	0,47	0,22	0,79
<i>Xu%</i>	0,32	0,11	0,13	0,50	32	0,305	0,44	0,04	-1,00
<i>X-%</i>	0,20	0,09	0,00	0,36	31	0,21	0,21	-0,37	-0,25
<i>a</i>	3,78	3,01	0	13	31	3,00	2	1,64	2,63
<i>p</i>	3,81	2,69	1	10	32	3,00	1	0,87	-0,08
<i>2AB+Art+Ay</i>	2,25	2,93	0	15	25	1,00	1	2,93	11,20
<i>DV+DR</i>	1,09	1,35	0	6	18	1,00	0	1,74	4,21
<i>INC+FAB</i>	1,53	1,55	0	6	21	1,00	0	0,92	0,55
<i>Sum6</i>	2,72	2,47	0	9	27	2	1	1,09	0,72
<i>WSum6</i>	8,50	9,35	0	41	27	5	0	1,75	3,54
<i>PTI</i>	0,56	1,13	0	5	9	0,00	0	2,52	7,06
<i>DEPI</i>	3,78	1,24	2	6	32	4,00	3	0,23	-0,94
<i>CDI</i>	3,09	1,15	1	5	32	3,00	3	-0,33	-0,40
<i>S-CON</i>	4,44	1,46	2	8	32	4,00	4	0,63	0,15
<i>SCZI</i>	1,66	1,21	0	6	30	1,00	1	1,78	4,37

Tabela 30 – Estatísticas descritivas para as variáveis do SCR no reteste

Variável	Média	DP	Min	Max	Freq	Mediana	Moda	Assimetria	Curtose
<i>R</i>	23,00	5,70	14	40	32	22,00	22	0,64	1,05
<i>Lambda</i>	1,09	0,59	0,37	3,00	32	1,00	1,00	1,63	3,46
<i>F</i>	11,38	4,14	5	21	32	10,50	10	0,58	-0,16
<i>M</i>	3,63	2,64	0	10	31	3,00	1	0,83	-0,07
<i>WSumC</i>	2,98	2,46	0,0	9,0	30	2,50	1,0	1,02	0,19
<i>EA</i>	6,61	3,56	1,5	15,0	32	6,25	–	0,63	-0,28
<i>FM</i>	3,63	1,83	0	7	31	4,00	4	0,16	-0,44
<i>m</i>	0,84	1,05	0	4	17	1,00	0	1,40	1,68
<i>FM+m</i>	4,47	2,00	0	10	31	4,00	–	0,46	1,08
<i>SumC'</i>	0,75	1,59	0	8	12	0,00	0	3,54	14,53
<i>SumT</i>	0,34	0,65	0	2	8	0,00	0	1,74	1,82
<i>SumV</i>	0,31	0,69	0	3	7	0,00	0	2,60	7,17
<i>SumY</i>	1,06	1,16	0	4	19	1,00	0	0,92	-0,09
<i>SumSH</i>	2,47	2,27	0	11	27	2,00	2	1,84	5,23
<i>es</i>	6,94	3,21	1	17	32	6,00	6	1,23	2,08
<i>Adjes</i>	6,16	2,52	1	14	32	6,00	6	1,06	2,30
<i>D-score</i>	0,03	1,06	-2	3	32	0,00	0	0,63	1,37
<i>AdjD</i>	0,25	1,41	-1	3	32	0,00	0	1,41	2,44
<i>FC</i>	1,50	1,44	0	5	23	1,50	1	0,97	0,42
<i>CF</i>	1,91	1,71	0	6	25	1,00	1	0,82	-0,19
<i>C</i>	0,22	0,49	0	2	6	0,00	0	2,26	4,77
<i>CF+C</i>	2,13	1,95	0	7	27	1,00	1	1,16	0,70
<i>Afr</i>	0,52	0,25	0,13	1,38	32	0,515	–	1,39	3,69
<i>Blends</i>	2,97	2,36	0	10	29	2,00	2	1,15	1,18
<i>S</i>	2,59	2,72	0	15	29	2,00	2	3,26	13,92
<i>COP</i>	0,97	1,40	0	6	15	0,00	0	1,93	4,42
<i>AG</i>	0,56	1,05	0	4	10	0,00	0	2,08	3,80
<i>GHR</i>	2,88	1,50	0	6	31	3,00	–	0,53	0,10
<i>PHR</i>	3,63	2,87	0	9	31	3,00	1	0,56	-1,30
<i>HumanContent</i>	5,53	2,93	1	13	32	5,00	4	0,60	-0,08
<i>Isolate/R</i>	0,21	0,11	0,00	0,42	30	0,205	–	0,04	-0,25
<i>(2)</i>	8,34	3,00	2	16	32	8,50	9	0,29	0,71
<i>3r+(2)/R</i>	0,42	0,09	0,27	0,58	32	0,41	0,38	0,20	-1,05
<i>Fr+rF</i>	0,41	0,84	0	3	8	0,00	0	2,25	4,52
<i>FD</i>	0,44	0,56	0	2	13	0,00	0	0,83	-0,28
<i>MOR</i>	1,00	1,19	0	5	18	1,00	0	1,47	2,67
<i>H</i>	2,00	1,97	0	8	27	2,00	1,50	1,74	2,83
<i>(H)+Hd+(Hd)</i>	3,53	2,10	0	9	31	3,00	–	0,61	-0,02
<i>Zf</i>	10,16	3,63	3	20	32	10,50	11	0,43	0,63
<i>Zd</i>	0,16	3,29	-6,0	8,0	32	-0,25	-0,5	0,16	0,05
<i>DQ+</i>	5,47	3,31	0	12	31	4,50	3	0,54	-0,78
<i>DQv</i>	1,09	1,40	0	5	17	1,00	0	1,33	1,04
<i>XA%</i>	0,78	0,10	0,54	0,96	32	0,78	0,73	-0,32	0,60
<i>WDA%</i>	0,81	0,10	0,53	0,96	32	0,82	0,71	-0,74	0,87
<i>Popular</i>	4,41	1,58	1	8	32	4,00	–	0,31	-0,06
<i>X+%</i>	0,47	0,11	0,21	0,73	32	0,47	0,47	0,22	0,79
<i>Xu</i>	0,32	0,11	0,11	0,50	32	0,325	0,29	-0,39	-0,81
<i>X-%</i>	0,21	0,09	0,04	0,46	32	0,22	0,27	0,43	1,07
<i>a</i>	3,78	3,32	1	12	32	3,00	1	1,31	0,74
<i>p</i>	4,38	2,56	0	10	30	4,00	4	0,27	-0,01
<i>2AB+Art+Ay</i>	1,91	2,96	0	16	24	1,00	–	3,83	17,29
<i>DV+DR</i>	1,13	0,98	0	3	21	1,00	–	0,18	-1,20
<i>INC+FAB</i>	1,38	1,52	0	6	22	1,00	1	1,49	2,10
<i>Sum6</i>	2,59	2,01	0	7	27	2,00	3	0,88	0,39
<i>WSum6</i>	8,22	9,18	0	38	27	6,00	8	2,02	3,82
<i>PTI</i>	0,69	1,15	0	3	10	0,00	0	1,35	0,15
<i>DEPI</i>	3,28	1,35	1	6	32	3,00	3	0,37	-0,03
<i>CDI</i>	3,09	1,33	1	5	32	3	2	0,17	-1,21
<i>S-CON</i>	4,41	1,39	2	7	32	4,00	–	0,44	-0,81
<i>SCZI</i>	1,78	1,21	0	4	29	1,50	1	1,78	4,37

Comparamos os resultados obtidos em nossa amostra com os encontrados na amostra normativa de 200 adultos não-pacientes da cidade de São Paulo, apresentados em Nascimento (2002):

Em nossa amostra, encontramos, para *R*: média = $22,94 \pm 7,31$ e mediana = 22,00 no teste, média = $23,00 \pm 5,70$ e mediana = 22,00 no reteste, em comparação com média = $20,40 \pm 6,84$ e mediana = 18,00 na amostra normativa⁵⁸.

Encontramos, para *Lambda*: média = $1,04 \pm 0,56$ e mediana = 0,94 no teste, média = $1,09 \pm 0,59$ e mediana = 1,00 no reteste, em comparação com média = $1,05 \pm 1,08$ e mediana = 0,72 na amostra normativa⁵⁹.

Segundo Exner (2003), valores de *Lambda* maiores do que 0,99 são considerados altos, caracterizando o estilo evitativo. Pessoas com este resultado lidaram com a situação do teste de modo excessivamente simplificado, o que pode indicar defesa frente à testagem ou uma preferência por lidar com as situações evitando maior envolvimento. Em nossa amostra, $Lambda \geq 1,00$ ocorreu em 16 participantes (50%) no teste e em 17 participantes (53%) no reteste. A proporção de participantes com $Lambda \geq 1,00$ em nossa amostra foi maior do que a encontrada na amostra normativa de 200 indivíduos da cidade de São Paulo, na qual 73 participantes (36,5%) apresentaram este resultado⁶⁰.

Vemos então que os participantes da nossa amostra forneceram mais respostas e uma proporção de respostas de forma pura maior do que a dos participantes da amostra normativa. Nascimento (2004) observou que, quanto maior o nível de instrução, maior o número de respostas e menor a proporção de forma pura. Os valores de *R* encontrados em nossa amostra podem ser compreendidos como associados ao maior nível de instrução de nossos participantes em comparação com os indivíduos da amostra normativa. No entanto, o

⁵⁸ Para apresentar os resultados, adotamos a seguinte notação: média $\pm$ desvio-padrão.

⁵⁹ Cabe observar que *Lambda*, assim como diversas variáveis do SCR não seguem uma distribuição semelhante à curva normal.

⁶⁰ Informação fornecida pela autora da pesquisa normativa, Profa. Dra. Regina Gattas Fernandes do Nascimento, e utilizada na preparação do artigo de Silva Neto (1999).

resultado de *Lambda* não seguiu o padrão descrito por Nascimento (2004), segundo o qual, quanto maior o nível de instrução menor o *Lambda*. Cabe observar que amostras de tamanho pequeno e com certas peculiaridades – no caso, nossa amostra foi composta por voluntários que quiseram participar da pesquisa – costumam apresentar diferenças em relação à amostra normativa.

[Trecho suprimido na versão online]

Em nossa amostra, no teste, encontramos: 10 participantes (31%) introversivos, 5 (16%) extratensivos e 17 (53%) ambíguais; no reteste: 12 participantes (38%) introversivos, 7 (22%) extratensivos e 13 (41%) ambíguais.

Nossa amostra foi composta por indivíduos não-pacientes e não avaliamos a eventual presença de transtorno mental ou sofrimento significativo. De qualquer modo, alguns índices do SCR podem sugerir a presença de problemas psicológicos.

Dois participantes (6%) no teste e cinco participantes (16%) no reteste apresentaram $PTI \geq 3$, resultado que indica dificuldades na percepção e pensamento.

Dez participantes (31%) no teste e cinco participantes (16%) no reteste apresentaram $DEPI \geq 5$, resultado que indica vulnerabilidade a perturbação afetiva. Um único participante (com $DEPI = 6$ no teste e no reteste) apresentou $S-CON = 8$ no teste, resultado que sugere risco de suicídio; no reteste, apresentou $S-CON = 6$.

O índice *CDI* é considerado positivo quando atinge valores iguais ou maiores do que 4, indicando, então, déficit de recursos de coping. *CDI* ≥ 4 ocorreu em 12 participantes (38%) no teste e em 12 participantes (38%) no reteste.

O *Índice de Hipervigilância* mostrou-se positivo para 4 participantes (13%) no teste e para 2 participantes (6%) no reteste. Índice de hipervigilância positivo indica que o indivíduo observa o ambiente atentamente e evita a proximidade física e emocional com as pessoas, devido a receio de vir a ser prejudicado pelos outros.

O *Índice de Estilo Obsessivo* mostrou-se negativo para todos os participantes da nossa amostra. Este índice avalia características de perfeccionismo e preocupação em lidar com as situações de modo correto e convencional.

A Tabela 31 apresenta a ocorrência de constelações positivas em nossa amostra (teste e reteste), na amostra de adultos não-pacientes da cidade de São Paulo (Nascimento, 2006) e na amostra normativa norte-americana (Exner, 2003).

Tabela 31 – Frequências e porcentagens de participantes com constelações positivas em nossa amostra (teste e reteste), na amostra normativa da cidade de São Paulo e na amostra normativa norte-americana

Amostra	1	2	3	4
País	Brasil	Brasil	Brasil	E.U.A
N	32	32	200	600
<i>PTI</i> ≥ 3	2 (6%)	5 (16%)	39 (20%)	1 (0%)
<i>DEPI</i> ≥ 5	10 (31%)	5 (16%)	47 (24%)	30 (5%)
<i>CDI</i> ≥ 4	12 (38%)	12 (38%)	108 (54%)	23 (4%)
<i>S-CON</i> ≥ 8	1 (3%)	0 (0%)	8 (4%)	0 (0%)
<i>HVI</i> positivo	4 (11%)	2 (6%)	18 (9%)	18 (3%)
<i>OBS</i> positivo	0 (0%)	0 (0%)	0 (0%)	8 (1%)

Amostras: 1 = Nossa amostra (teste); 2 = Nossa amostra (reteste);
3 = Amostra normativa da cidade de São Paulo (Nascimento, 2006);
4 = Amostra normativa norte-americana (Exner, 2003)

No teste, nossa amostra apresentou uma proporção significativamente maior de casos com *PTI*, *DEPI*, *CDI* e *HVI* positivos do que na amostra norte-americana (*S-CON* positivo, com uma proporção significativamente maior de casos em nossa amostra, quase alcançou significância estatística: $X^2 = 18,780$, $p = 0,051$ pelo teste exato de Fisher). A proporção de casos com constelações positivas não foi significativamente diferente da encontrada na amostra normativa da cidade de São Paulo (*PTI* e *CDI* positivos, com proporções menores de participantes em nossa amostra, quase alcançaram significância estatística: *PTI*: $X^2 = 3,130$ $p = 0,077$; *CDI*: $X^2 = 3,008$, $p = 0,083$).

No reteste, nossa amostra apresentou uma proporção significativamente maior de casos com *PTI*, *DEPI* e *CDI* positivos do que na amostra norte-americana e proporções não significativamente diferentes de participantes com constelações positivas à encontrada na amostra da cidade de São Paulo (*CDI* positivo, com proporção menor de participantes em nossa amostra, quase alcançou significância estatística: $X^2 = 3,008$, $p = 0,083$).

Conforme comentamos em outro trabalho (Silva Neto, 1999), a proporção elevada de adultos paulistanos não-pacientes com *DEPI* e *CDI* positivos coloca dúvidas sobre a validade destes índices no Brasil. Em relação ao índice *OBS*, embora em amostras de não-pacientes seja esperada uma proporção baixa de participantes com estilo obsessivo, a ausência de casos positivos tanto em nossa amostra quanto na pesquisa normativa, coloca sob suspeita a validade deste índice para identificar, no Brasil, as características de perfeccionismo e preocupação em lidar com as situações de modo correto e convencional.

Em nossa amostra, para *EA*, temos: média = $6,94 \pm 3,01$ no teste e média = $6,61 \pm 3,56$ no reteste, em comparação com média = $4,96 \pm 3,49$ na amostra normativa. *EA* refere-se aos recursos para lidar com os problemas cotidianos e diminuir as tensões experimentadas. A presença de mais recursos em nossa amostra provavelmente está associada ao maior nível de instrução, seguindo o mesmo padrão encontrado por Nascimento (2004).

Encontramos, para *es*: média = $6,69 \pm 4,06$ no teste, média = $6,94 \pm 3,21$ no reteste, em comparação com média = $8,57 \pm 5,01$ na amostra normativa. O índice *es* refere-se às tensões experimentadas.

Portanto, nossos participantes indicaram possuir mais recursos [*EA*] e experimentar menos tensões [*es*] do que os participantes da amostra normativa. [Trecho suprimido na versão online]

Para *FM+m*, encontramos: média = $3,91 \pm 2,51$ no teste, média = $4,47 \pm 2,00$ no reteste, e média = $4,83 \pm 2,94$ na amostra normativa; os valores são semelhantes ou um pouco mais baixos na nossa amostra. Para *SumSH*, encontramos: média = $2,78 \pm 2,31$ no teste, média = $2,47 \pm 2,27$ no reteste, e média = $3,74 \pm 3,42$ na amostra normativa; os valores são mais baixos na nossa amostra, aparentemente indicando menor sofrimento emocional em nossos participantes. As variáveis *SumC'*, *SumT*, *SumV* e *SumY* apresentaram distribuições semelhantes no teste, no reteste e na amostra normativa,

predominando valores 0 e 1; apenas o total de cor acromática e sombreado [*SumSH*] mostrou-se claramente menor em nossa amostra em comparação com a amostra normativa.

A variável *D-score* é obtida a partir da diferença *EA - es*, sendo, portanto, uma confrontação dos recursos [*EA*] com as tensões [*es*]. Encontramos, em nossa amostra, para *D-score*: média = $-0,03 \pm 1,15$ no teste, média = $0,03 \pm 1,06$ no reteste, em comparação com média = $-1,07 \pm 1,57$ na amostra normativa, indicando maior capacidade de controle e tolerância ao estresse na nossa amostra.

Para *WSumC*, encontramos: média = $3,28 \pm 2,47$ no teste, média = $2,98 \pm 2,46$ no reteste, em comparação com média = $2,11 \pm 1,78$ na amostra normativa. Isto indica que nossos participantes possuem maior capacidade de experimentar e expressar afetos do que os paulistanos da amostra normativa. Encontramos, para o quociente afetivo [*Afr*]: média = $0,52 \pm 0,20$ no teste, média = $0,52 \pm 0,25$ no reteste, e média = $0,56 \pm 0,26$ na amostra normativa; os valores foram semelhantes para este índice que indica a responsividade às situações afetivas.

Para *Zf*, encontramos: média = $10,34 \pm 3,69$ no teste, média = $10,16 \pm 3,63$ no reteste e média = $10,16 \pm 4,16$ na amostra normativa. Assim, o esforço de organização perceptiva foi semelhante nestas amostras. Para *Zd*, encontramos: média = $0,77 \pm 4,32$ no teste, média = $0,16 \pm 3,29$ no reteste e média = $-0,52 \pm 4,77$ na amostra normativa, ou seja, nossos participantes foram um pouco mais eficientes em organizar as informações do que os participantes da amostra normativa. Para *DQ+*, encontramos: média = $5,41 \pm 2,73$ no teste, $5,47 \pm 3,31$ no reteste e $4,76 \pm 3,00$ na amostra normativa, ou seja, nossos participantes estabeleceram mais relações entre os dados percebidos do que os participantes da amostra normativa. Estas (pequenas) diferenças em relação ao processamento das informações talvez estejam associadas ao maior nível de instrução nos participantes de nossa amostra, em comparação com a amostra normativa.

Os índices de adequação perceptiva mostraram valores semelhantes ou um pouco melhores em nossa amostra do que na amostra normativa. Para *XA%* (percepção adequada em geral), encontramos: média = $0,79 \pm 0,09$ no teste, média = $0,78 \pm 0,10$ no reteste, e média = $0,74 \pm 0,13$ na amostra normativa. Para *WDA%* (percepção adequada dos aspectos mais comuns do ambiente), encontramos: média = $0,82 \pm 0,09$ no teste, média = $0,81 \pm 0,10$ no

reteste, e média = $0,78 \pm 0,14$ na amostra normativa. Para *X-%* (percepção distorcida), encontramos: média = $0,20 \pm 0,09$ no teste, média = $0,21 \pm 0,09$ no reteste, e média = $0,25 \pm 0,12$ na amostra normativa. O número de respostas populares, indicando percepção convencional, obteve média = $4,19 \pm 1,77$ no teste, média = $4,41 \pm 1,58$ no reteste e média = $4,08 \pm 1,79$ na amostra normativa.

Encontramos menos falhas no pensamento em nossa amostra, em comparação com a amostra normativa. Para *Sum6*, encontramos: média = $2,72 \pm 2,47$ no teste, média = $2,59 \pm 2,01$ no reteste, e média = $3,75 \pm 3,04$ na amostra normativa. Para *WSum6*, encontramos: média = $8,50 \pm 9,35$ no teste, média = $8,22 \pm 9,18$ no reteste, e média = $10,63 \pm 11,08$ na amostra normativa. Comparando as médias dos códigos especiais críticos, encontramos médias maiores para *DV*, *INCOM* e *DR* e *ALOG* na amostra normativa, e médias maiores para *FABCOM* em nossa amostra; os códigos especiais de nível 2 apresentaram médias baixas em ambas as amostras. [Trecho suprimido na versão online]

Talvez este achado esteja relacionado às diferenças no nível de instrução, com mais instrução contribuindo para uma melhor comunicação e organização do pensamento.

Em nossa amostra, os conteúdos humanos apresentaram: média = $5,63 \pm 3,08$ no teste e média = $5,53 \pm 2,93$ no reteste, em comparação com média = $5,39 \pm 3,87$ na amostra normativa. [Trecho suprimido na versão online]

Para *COP*, encontramos: média = $0,84 \pm 1,08$ no teste, média = $0,97 \pm 1,40$ no reteste e média = $0,63 \pm 0,94$ na amostra normativa. Para *AG*: média = $0,44 \pm 0,72$ no teste, média = $0,56 \pm 1,05$ no reteste e média = $0,27 \pm 0,66$ na amostra normativa. Os valores de *COP* e *AG* foram um pouco maiores em nossa amostra, indicando [Trecho suprimido na versão online]

A comparação de outras variáveis mostrou-se difícil porque muitas delas apresentaram poucos valores e predominaram escores baixos. A impressão ao observar as estatísticas descritivas é que houve grande semelhança dos valores encontrados em nossa amostra e na amostra normativa.

10.5. Comparações entre os resultados de teste e reteste

Realizamos uma comparação entre os resultados do teste e do reteste, utilizando o teste t de student para amostras pareadas. Os resultados são apresentados na Tabela 32.

Tabela 32 – Comparação entre os resultados do teste e do reteste

Variável	t	p	d	Variável	t	p	d
<i>R</i>	-0,087	0,931	-0,01	<i>Isolate/R</i>	1,602	0,119	0,23
<i>Lambda</i>	-0,684	0,499	-0,09	<i>(2)</i>	-1,392	0,174	-0,42
<i>F</i>	-0,392	0,698	-0,05	<i>3r+(2)/R</i>	-2,189	0,036	-0,17
<i>M</i>	0,089	0,929	0,01	<i>Fr+rF</i>	-1,000	0,325	-0,09
<i>WSumC</i>	1,049	0,302	0,12	<i>FD</i>	1,000	0,325	0,18
<i>EA</i>	0,727	0,472	0,10	<i>MOR</i>	1,906	0,066	0,32
<i>FM</i>	-1,765	0,087	-0,25	<i>H</i>	0,000	1,000	0,00
<i>m</i>	-0,432	0,669	-0,08	<i>(H)+Hd+(Hd)</i>	0,281	0,781	0,04
<i>FM+m</i>	-1,579	0,124	-0,25	<i>Zf</i>	0,355	0,725	0,05
<i>SumC'</i>	0,133	0,895	0,02	<i>Zd</i>	0,731	0,470	0,16
<i>SumT</i>	-0,297	0,768	-0,04	<i>DQ+</i>	-0,146	0,885	-0,02
<i>SumV</i>	0,892	0,379	0,19	<i>DQv</i>	1,509	0,141	0,20
<i>SumY</i>	0,797	0,431	0,15	<i>XA%</i>	0,293	0,772	0,11
<i>SumSH</i>	0,781	0,441	0,14	<i>WDA%</i>	0,280	0,781	0,11
<i>es</i>	-0,459	0,649	-0,07	<i>Popular</i>	-0,865	0,394	-0,13
<i>Adjes</i>	-1,245	0,222	-0,18	<i>X+%</i>	0,056	0,955	0,00
<i>D-score</i>	-0,263	0,794	-0,05	<i>Xu%</i>	0,110	0,914	0,00
<i>AdjD</i>	0,147	0,884	0,02	<i>X-%</i>	-0,800	0,430	-0,11
<i>FC</i>	-0,571	0,572	-0,12	<i>A</i>	0,000	1,000	0,00
<i>CF</i>	0,528	0,601	0,07	<i>P</i>	-1,328	0,194	-0,22
<i>C</i>	1,305	0,201	0,25	<i>2AB+Art+Ay</i>	1,187	0,244	0,12
<i>CF+C</i>	1,121	0,271	0,14	<i>DV+DR</i>	-0,122	0,904	-0,03
<i>Afr</i>	0,099	0,922	0,00	<i>INC+FAB</i>	0,645	0,524	0,10
<i>Blends</i>	-1,141	0,263	-0,16	<i>Sum6</i>	0,416	0,680	0,06
<i>S</i>	0,561	0,579	0,07	<i>WSum6</i>	0,333	0,741	0,03
<i>COP</i>	-0,725	0,474	-0,10	<i>PTI</i>	-0,725	0,474	-0,11
<i>AG</i>	-1,072	0,292	-0,14	<i>DEPI</i>	2,184	0,037	0,39
<i>GHR</i>	0,218	0,829	0,04	<i>CDI</i>	0,000	1,000	0,00
<i>PHR</i>	-0,801	0,429	-0,11	<i>SCON</i>	0,112	0,911	0,00
<i>Human Content</i>	0,267	0,792	0,03				

t = valor do teste t de student; p = probabilidade de encontrar o valor de t obtido caso o valor de t na população seja igual à zero; d = magnitude do efeito

Devido ao grande número de testes t realizados, adotamos como nível de significância estatística $p < 0,02$. Não encontramos diferenças estatisticamente significantes entre os resultados no teste e no reteste, mas o *Índice de Egocentricidade* (com escores menores no teste) e o *Índice de Depressão* (com escores maiores no teste) quase alcançaram a significância estatística.

A estatística d de Cohen, apresentada na Tabela 32, é uma medida da magnitude do efeito, ou do tamanho da diferença entre as médias no teste e no reteste. Corresponde à diferença das médias dividida pela média dos desvios-padrão das amostras comparadas (na nossa pesquisa, o teste e o reteste). Segundo Dancey e Reidy (2006), tomando como base Joseph Cohen (**Statistical Power for Behavioral Sciences**. New York: Academic Press, 1988), $d = 0,20$ indica efeito pequeno, $d = 0,50$ indica efeito médio e $d = 0,80$ indica efeito grande. Em nossa amostra, as diferenças entre os resultados do teste e do reteste foram de magnitude pequena, com o valor mais alto sendo $d = 0,42$, para o Índice de Depressão [DEPI].

10.6. Estabilidade temporal do SCR

10.6.1. Correlações de teste-reteste

Na tabela 33 apresentamos as correlações produto-momento de Pearson e as correlações de postos de Spearman entre os escores no teste e no reteste para as variáveis do SCR investigadas na pesquisa.

As correlações produto-momento de Pearson para as 59 variáveis principais do SCR variaram de 0,25 a 0,91, com média = 0,61, desvio-padrão = 0,16 e mediana = 0,66. Dezenove correlações foram iguais ou maiores do que 0,70 e dez correlações foram iguais ou maiores do que 0,75.

As correlações de postos de Spearman para as 59 variáveis principais do SCR variaram de 0,12 a 0,86, com média = 0,59, desvio-padrão = 0,16 e mediana = 0,62. Dezesesseis correlações foram iguais ou maiores do que 0,70 e oito correlações iguais foram maiores do que 0,75.

Tabela 33 – Correlações produto-momento de Pearson e correlações de postos de Spearman, entre os escores no teste e no reteste

Variável	r	rho	Variável	r	rho
<i>R</i>	0,83	0,83	<i>Isolate/R</i>	0,76	0,74
<i>Lambda</i>	0,70	0,55	<i>(2)</i>	0,76	0,65
<i>F</i>	0,73	0,70	<i>3r+(2)/R</i>	0,35	0,25
<i>M</i>	0,68	0,70	<i>Fr+rF</i>	0,91	0,86
<i>WSumC</i>	0,79	0,75	<i>FD</i>	0,51	0,47
<i>EA</i>	0,71	0,79	<i>MOR</i>	0,60	0,67
<i>FM</i>	0,67	0,66	<i>H</i>	0,73	0,74
<i>m</i>	0,41	0,34	<i>(H)+Hd+(Hd)</i>	0,67	0,73
<i>FM+m</i>	0,62	0,57	<i>Zf</i>	0,67	0,64
<i>SumC'</i>	0,56	0,57	<i>Zd</i>	0,25	0,37
<i>SumT</i>	0,64	0,52	<i>DQ+</i>	0,69	0,71
<i>SumV</i>	0,32	0,42	<i>DQv</i>	0,72	0,64
<i>SumY</i>	0,46	0,50	<i>XA%</i>	0,65	0,65
<i>SumSH</i>	0,51	0,45	<i>WDA%</i>	0,43	0,48
<i>es</i>	0,66	0,58	<i>Popular</i>	0,64	0,60
<i>Adjes</i>	0,67	0,60	<i>X+%</i>	0,49	0,42
<i>D-score</i>	0,27	0,29	<i>Xu%</i>	0,29	0,29
<i>AdjD</i>	0,30	0,33	<i>X-%</i>	0,62	0,65
<i>FC</i>	0,66	0,60	<i>a</i>	0,72	0,63
<i>CF</i>	0,69	0,69	<i>p</i>	0,59	0,62
<i>C</i>	0,53	0,12	<i>2AB+Art+Ay</i>	0,85	0,56
<i>CF+C</i>	0,74	0,69	<i>DV+DR</i>	0,26	0,34
<i>Afr</i>	0,64	0,77	<i>INC+FAB</i>	0,60	0,72
<i>Blends</i>	0,69	0,79	<i>Sum6</i>	0,73	0,67
<i>S</i>	0,80	0,60	<i>WSum6</i>	0,87	0,79
<i>COP</i>	0,72	0,75	<i>PTI</i>	0,64	0,58
<i>AG</i>	0,78	0,57	<i>DEPI</i>	0,50	0,46
<i>GHR</i>	0,51	0,50	<i>CDI</i>	0,63	0,64
<i>PHR</i>	0,69	0,68	<i>SCON</i>	0,39	0,39
<i>Human Content</i>	0,78	0,74			

Para as 59 variáveis principais do SCR, 33 apresentaram valores maiores da correlação de Pearson, 21 da correlação de postos de Spearman e 5 variáveis apresentaram valores iguais dos dois coeficientes. De modo geral, então, a correlação de Pearson resultou em valores maiores do que a correlação de postos de Spearman.

Dez variáveis apresentaram estabilidade temporal alta ($r \geq 0,75$): *R*, *WSumC*, *S*, *AG*, *HumanContent*, *Isolate/R*, *(2)*, *Fr+rF*, *2AB+Art+Ay* e *WSum6*. Outras nove variáveis apresentaram estabilidade de moderada a alta ($0,70 \leq r < 0,75$): *Lambda*, *F*, *EA*, *CF+C*, *COP*, *H*, *DQv*, *a* e *Sum6*. Outras 28 variáveis apresentaram estabilidade moderada ($0,50 \leq r < 0,70$): *M*, *FM*, *FM+m*, *SumC'*, *SumT*, *SumSH*, *es*, *Adjes*, *FC*, *CF*, *C*, *Afr*, *Blends*, *GHR*, *PHR*, *FD*, *MOR*, *(H)+Hd+(Hd)*, *Zf*, *DQ+*, *XA%*, *Popular*, *X-%*, *p*, *INC+FAB*, *PTI*, *DEPI* e *CDI*. A

estabilidade temporal de movimento inanimado [*m*] e sombreado difuso [*SumY*] costuma ser baixa, pois se entende que estas variáveis estão associadas a estresse situacional. Em nossa pesquisa, *m* e *SumY* apresentaram estabilidade temporal baixa. Contudo, além destas duas variáveis, outras 10 apresentaram baixa estabilidade temporal ($r < 0,50$): *SumV*, *D-score*, *AdjD*, $3r+(2)/R$, *Zd*, *WDA%*, *X+%*, *Xu%*, *DV+DR* e *S-CON*.

Em comunicação pessoal com Cato Grønnerød, autor da metanálise sobre estabilidade temporal do Rorschach, fomos informados de que a estabilidade prevista para um período de reteste de 104 dias seria $r = 0,76$. Esta estimativa foi obtida a partir de todas as pesquisas com o Rorschach, incluindo outros sistemas além do SCR. A média das correlações em nossa amostra foi, então, menor do que a obtida a partir da metanálise sobre estabilidade temporal do Rorschach.

Uma das dificuldades na comparação com as pesquisas realizadas é o fato de estas não apresentarem informações sobre todas as variáveis do SCR, sendo única exceção a pesquisa de Sultan et al. (2006), que as apresenta em apêndice. Optamos então por comparar nossos resultados com: 1) os resultados da amostra de 50 adultos não-pacientes norte-americanos retestados após 1 ano (Exner, 1986, 2003); 2) os resultados da amostra de 35 adultos não-pacientes norte-americanos retestados após 3 semanas (Exner, 1986, 2003) e 3) os resultados da amostra de 75 adultos não-pacientes franceses retestados após 3 meses (Sultan et al., 2006) – estas são as amostras com maior número de variáveis para as quais a estabilidade temporal é informada. A Tabela 34 apresenta os resultados na nossa amostra e nestas outras três pesquisas.

Tabela 34 – Estabilidade temporal das variáveis do SCR em nossa amostra, em duas amostras norte-americanas e numa amostra francesa

Amostra	Nossa amostra	Exner (1986)	Exner (1986)	Sultan et al. (2006)
N	32	50	35	75
Intervalo	3 meses	1 ano	3 semanas	3 meses
<i>R</i>	0,83	0,86	0,84	0,75
<i>Lambda</i>	0,70	0,78	0,76	0,72
<i>F</i>	0,73	0,74	0,76	0,70
<i>M</i>	0,68	0,84	0,83	0,76
<i>WSumC</i>	0,79	0,82	0,83	0,69
<i>EA</i>	0,71	0,83	0,84	0,77
<i>FM</i>	0,67	0,77	0,72	0,48
<i>m</i>	0,41	0,26	0,34	0,47
<i>FM+m</i>	0,62	–	–	0,50
<i>SumC'</i>	0,56	0,73	0,67	0,38
<i>SumT</i>	0,64	0,91	0,96	0,56
<i>SumV</i>	0,32	0,87	0,89	0,46
<i>SumY</i>	0,46	0,31	0,41	0,17
<i>SumSH</i>	0,51	–	0,71	0,42
<i>es</i>	0,66	0,64	0,59	0,46
<i>Adjes</i>	0,67	0,82	0,79	0,46
<i>D-score</i>	0,27	0,91	0,88	0,34
<i>AdjD</i>	0,30	–	–	0,38
<i>FC</i>	0,66	0,86	0,92	0,61
<i>CF</i>	0,69	0,58	0,68	0,47
<i>C</i>	0,53	0,56	0,59	0,08
<i>CF+C</i>	0,74	0,81	0,83	0,55
<i>Afr</i>	0,64	0,82	0,85	0,57
<i>Blends</i>	0,69	0,62	0,71	0,63
<i>S</i>	0,80	–	–	0,70
<i>COP</i>	0,72	0,81	0,88	0,38
<i>AG</i>	0,78	0,82	0,81	0,45
<i>GHR</i>	0,51	–	–	0,55
<i>PHR</i>	0,69	–	–	0,70
<i>HumanContent</i>	0,78	–	–	0,68
<i>Isolate/R</i>	0,76	0,84	0,83	0,67
<i>(2)</i>	0,76	0,81	0,83	0,77
<i>EGO</i>	0,35	0,89	0,90	0,78
<i>Fr+rF</i>	0,91	0,82	0,89	0,65
<i>FD</i>	0,51	0,88	0,90	0,51
<i>MOR</i>	0,60	0,71	0,83	0,62
<i>H</i>	0,73	–	–	0,67
<i>(H)+Hd+Hd</i>	0,67	–	–	0,58
<i>Zf</i>	0,67	0,85	0,89	0,76
<i>Zd</i>	0,25	–	–	0,46
<i>DQ+</i>	0,69	–	–	0,70
<i>DQv</i>	0,72	–	–	0,55
<i>XA%</i>	0,65	0,89	–	0,49
<i>WDA%</i>	0,43	0,92	–	0,45
<i>Popular</i>	0,64	0,83	0,81	0,54
<i>X+%</i>	0,49	0,86	0,87	0,55
<i>Xu%</i>	0,29	0,85	0,89	0,32
<i>X-%</i>	0,60	0,92	0,88	0,51
<i>a</i>	0,72	0,83	0,87	0,61
<i>p</i>	0,59	0,72	0,85	0,55
<i>2AB+Art+Ay</i>	0,85	0,84	–	0,53
<i>DV+DR</i>	0,26	0,72	0,74	0,26
<i>INC+FAB</i>	0,60	0,89	0,92	0,61
<i>Sum6</i>	0,73	0,81	0,81	0,50
<i>Wsum6</i>	0,87	0,86	0,86	0,56
<i>PTI</i>	0,64	–	–	0,48
<i>DEPI</i>	0,50	–	–	0,28
<i>CDI</i>	0,63	–	–	0,58
<i>S-CON</i>	0,39	–	–	0,47

Correlações menores do que 0,50
Correlações entre 0,50 e 0,74
Correlações iguais a 0,75 ou maiores

Comparando a estabilidade obtida na nossa amostra com a da amostra de 50 adultos norte-americanos retestados após um ano (Exner, 1986, 2003), encontramos que a média das correlações de 44 variáveis foi $0,62 \pm 0,17$ na nossa amostra e $0,78 \pm 0,14$ na amostra norte-americana. Trinta e seis correlações foram maiores na amostra norte-americana. A diferença entre as médias de estabilidade temporal nas amostras foi de magnitude grande ($d = -1,03$).

Comparando a estabilidade obtida na nossa amostra com a da amostra de 35 adultos norte-americanos retestados após 3 semanas (Exner, 1986, 2003), a média das correlações de 41 variáveis foi $0,62 \pm 0,16$ na nossa amostra e $0,80 \pm 0,13$ na amostra norte-americana. Trinta e cinco correlações foram maiores na amostra norte-americana. A diferença entre as médias de estabilidade nas amostras foi de magnitude grande ($d = -1,24$).

Comparando a estabilidade obtida na nossa amostra com a da amostra de 75 adultos franceses retestados após 3 meses (Sultan et al., 2006), encontramos que a média das correlações de 59 variáveis foi $0,61 \pm 0,16$ na nossa amostra e $0,54 \pm 0,15$ na amostra francesa. Trinta e sete correlações foram maiores na nossa amostra. A diferença entre as médias de estabilidade temporal nas amostras foi de magnitude moderada ($d = 0,45$).

A partir das pesquisas do Sistema Compreensivo, estabeleceu-se que sombreado difuso e movimento inanimado são variáveis que apresentam baixa estabilidade temporal ($r = 0,20$ a $r = 0,40$), por estarem associadas a características do tipo estado; os índices dos quais movimento inanimado e sombreado difuso participam ($FM+m$, $SumSH$ e es) apresentam correlações na faixa de $r = 0,60$; outras variáveis com correlações na faixa de $r = 0,60$ (FM , $SumC'$, CF , $Blends$) são consideradas predominantemente traço, mas com um componente estado; as demais variáveis apresentam correlações altas, iguais ou maiores do que 0,75, sendo consideradas associadas a traços de personalidade.

Para as variáveis que apresentaram alta estabilidade temporal ($r \geq 0,75$) nas pesquisas de Exner, nossa amostra somente apresentou estabilidade semelhante para R , $WSumC$, S , AG , $HumanContent$, $Isolate/R$, (2), $Fr+rF$, $2AB+Art+Ay$ e $WSum6$. Movimento inanimado e sombreado difuso apresentaram correlações baixas em quase todas as pesquisas e também em nossa amostra; os índices dos quais estas variáveis participam ($FM+m$, $SumSH$, es) também apresentaram estabilidade semelhante na nossa amostra e nas pesquisas de Exner

(ou um pouco mais baixa na nossa amostra). Para outras variáveis que apresentaram correlações próximas de 0,60 nas amostras de Exner (*FM*, *SumC'*, *CF*, *Blends*), nossa amostra apresentou correlações semelhantes. Algumas variáveis merecem ser destacadas, por terem apresentado correlações muito mais baixas na nossa amostra: *SumV*, *D-score*, *AdjD*, $3r+(2)/R$, *WDA%*, *X+%*, *Xu%*, *DV+DR* e *INC+FAB*.

SumV apresentou estabilidade temporal baixa ($r = 0,32$) em nossa amostra. A baixa fidedignidade das codificações de *Vista* pode ter contribuído para a baixa estabilidade temporal encontrada. A baixa taxa de ocorrência desta variável faz com que a assimetria e a restrição da faixa de valores também possam ter contribuído para diminuir a estabilidade temporal.

Nas amostras de Exner, as correlações de teste-reteste para *SumV* situam-se na faixa de 0,80, ou seja, esta variável apresenta boa estabilidade. Contudo, Weiner (1998a) comenta que estas correlações altas se devem à maioria dos indivíduos não-pacientes apresentarem $SumV = 0$ em ambas testagens. [Trecho suprimido na versão online]

Tendo encontrado este resultado, resolvermos investigar a estabilidade de *SumV* em termos de categorias interpretativas. Os resultados são apresentados na Tabela 35. Vinte e quatro participantes (75%) mantiveram-se na mesma categoria interpretativa no teste e no reteste. A ausência de respostas de vista foi uma característica mais estável, que tendeu a se manter entre diferentes testagens; a presença de vista foi menos estável, ou seja, aparecendo numa das testagens, não necessariamente ocorreu na outra.

Tabela 35 – Consistência das categorias interpretativas de SumV

Teste <i>SumV</i>	Reteste <i>SumV</i>		% consistentes	Φ	k
	= 0	> 0			
= 0	19	2	90%		
> 0	6	5	45%	0,41*	0,39*

* $p < 0,05$

A variável *D-score* apresentou baixa estabilidade temporal ($r = 0,27$) em nossa amostra, enquanto apresentou estabilidade alta ($r \approx 0,90$) nas amostras de Exner⁶¹. Esta variável corresponde à diferença *EA - es*, que então é convertida em *D-score* utilizando-se uma tabela de valores escalonados. Portanto, *D-score* depende de *EA* e *es*. Estas variáveis apresentaram estabilidade temporal bem mais alta: $r = 0,71$ e $r = 0,66$, respectivamente. Curiosamente, apesar de as variáveis componentes apresentarem um nível moderado de estabilidade, a variável *D-score* apresentou baixa estabilidade temporal. A variável *D-score* será também analisada em termos da consistência das categorias interpretativas (ver adiante).

O Índice de egocentricidade [$3r+(2)/R$] apresentou estabilidade temporal bem mais baixa em nossa amostra ($r = 0,35$), em comparação com a alta estabilidade ($r \approx 0,90$) nas pesquisas de Exner. Este achado é de difícil compreensão, principalmente porque pares e respostas de reflexo, variáveis que compõem o *Índice de Egocentricidade*, apresentaram alta estabilidade ($r = 0,76$ e $r = 0,91$, respectivamente)

As variáveis relacionadas à adequação da percepção – ou seja, os índices que envolvem a qualidade formal: *XA%*, *WDA%*, *X+%*, *Xu%* e *X-%* – também apresentaram estabilidade mais baixa em nossa amostra, em comparação com as amostras de Exner. Algumas destas variáveis apresentaram baixa fidedignidade das codificações, o que pode ter contribuído para a baixa estabilidade temporal.

Investigamos a consistência temporal das categorias interpretativas das principais variáveis relativas à adequação da percepção. Os resultados são apresentados na Tabela 36.

⁶¹ As correlações para *AdjD* não são informadas nas amostras de Exner.

Para $XA\%$, 29 participantes (83%) mantiveram-se na mesma categoria interpretativa – a maioria com $XA\% \geq 0,70$ em ambas testagens. Para $WDA\%$, 26 participantes (74%) mantiveram-se na mesma categoria interpretativa – a maioria com $WDA\% \geq 0,75$. Para $X-\%$, 28 participantes (80%) mantiveram-se na mesma categoria interpretativa. Em termos de proporção de participantes na mesma categoria interpretativa, portanto, os resultados são melhores. A concentração de escores numa determinada faixa de valores (problemas de restrição da faixa de valores e assimetria) pode ter contribuído para a menor estabilidade temporal encontrada.

Tabela 36 – Consistência das categorias interpretativas de $XA\%$, $WDA\%$ e $X-\%$

Teste	Reteste		% consistentes	Φ	k
	1ª categoria	2ª categoria			
$XA\%$	< 0,70	2	4	33%	
	$\geq 0,70$	2	27	93%	0,31
$WDA\%$	< 0,75	4	7	36%	
	$\geq 0,75$	2	22	92%	0,35*
$X-\%$	< 0,30	28	3	90%	
	$\geq 0,30$	4	0	0%	-0,11

* $p < 0,05$

Os *códigos especiais críticos*, agrupados nas variáveis $DV+DR$ e $INC+FAB$, também apresentaram menor estabilidade temporal em nossa amostra do que nas amostras de Exner. Os adultos não-pacientes norte-americanos apresentam valores baixos para os códigos especiais críticos, então a boa estabilidade se deve a ausência (ou escores baixos) destes códigos nos seus protocolos. Já nos participantes da nossa amostra, ocorreram mais falhas de pensamento, que recebem estes códigos especiais críticos, havendo também maior instabilidade no aparecimento destas falhas em diferentes momentos. Estes códigos, agrupados em $DV+DR$ e $INC+FAB$ apresentam baixa taxa de ocorrência e valores baixos, fatores que podem ter contribuído para a baixa estabilidade temporal (problemas de restrição da faixa de valores e assimetria). A variável $Sum6$ engloba todos os códigos especiais críticos ($DV + DR + INC + FAB + ALOG$), então apresenta valores mais altos, e apresentou uma maior estabilidade temporal ($r = 0,73$).

Outras comparações podem ser feitas, no caso entre a estabilidade temporal na nossa amostra e nas pesquisas norte-americanas com pacientes. A Tabela 37 apresenta a estabilidade temporal em nossa amostra de 32 adultos não-pacientes retestados após 3 meses, na amostra de 25 pacientes norte-americanos em lista de espera, retestados após 30 dias [quarto estudo de Exner (1978)] e na amostra de 35 pacientes norte-americanos com depressão reativa em psicoterapia breve, retestados após 90 dias [sexto estudo de Exner (1978)].

Tabela 37 – Estabilidade temporal de 18 variáveis do SCR em nossa amostra e em duas amostras norte-americanas de pacientes psiquiátricos

País	Brasil	EUA	EUA
N	32	25	35
Grupo	Não-pacientes	Lista de espera	Depressão reativa
Reteste	90 dias	30 dias	90 dias
R	0,83	0,83	0,83
Lambda	0,70	0,65	0,64
F	0,73	0,57	0,77
M	0,68	0,82	0,82
WSumC	0,79	0,68	0,49
EA	0,71	0,77	0,84
FM	0,67	0,74	0,73
FM+m	0,62	0,62	0,70
SumSH	0,51	0,49	0,28
es	0,66	0,67	0,76
FC	0,66	0,71	0,51
CF+C	0,74	0,57	0,44
Afr	0,64	0,86	0,89
3r + (2)	0,35	0,91	0,91
Zf	0,67	0,86	0,86
Popular	0,64	0,69	0,81
X+%	0,49	0,83	0,84
a	0,72	0,78	0,87
p	0,61	0,84	0,79
Média	0,65	0,73	0,73
Desvio-padrão	0,11	0,12	0,17
Faixas de valores			
≥ 0,75	2	9	12
0,50 a 0,74	15	9	4
< 0,50	2	1	3

	Correlações menores do que 0,50
	Correlações entre 0,50 e 0,74
	Correlações iguais a 0,75 ou maiores

Para as 19 correlações comparadas, nossa amostra apresentou média = $0,65 \pm 0,11$, a primeira amostra norte-americana, média = $0,73 \pm 0,12$ e a segunda amostra norte-americana, média = $0,73 \pm 0,17$. Nossa amostra apresentou apenas 8 correlações maiores do que as da primeira amostra norte-americana e 6 correlações maiores do que as da segunda amostra norte-americana. A diferença entre as médias de estabilidade temporal em nossa amostra e na primeira amostra norte-americana foi de magnitude moderada ($d = -0,70$), assim como a diferença em relação à segunda amostra norte-americana ($d = -0,57$).

Outra comparação contrasta os resultados da nossa amostra com os de três estudos norte-americanos realizados por pesquisadores que não fazem parte do grupo de Exner (Tabela 38). As médias da estabilidade temporal em nossa amostra e na amostra de Schwartz, Mebane e Malony (1990) apresentaram uma diferença de magnitude moderada ($d = 0,60$). As diferenças entre as médias de estabilidade temporal em nossa amostra e nas amostras de Erstad (1996) foram de magnitude pequena ($d \approx 0,11$).

Vemos então que nossos resultados foram menores do que os das pesquisas de Exner (tanto com não-pacientes, quanto com pacientes), menores do que as estimativas obtidas a partir da metanálise sobre estabilidade temporal do Rorschach (Grønnerød, 2003) semelhantes aos de Erstad (1996) e maiores do que os da pesquisa francesa (Sultan et al., 2006).

Tabela 38 – Estabilidade temporal em nossa amostra e em 3 amostras norte-americanas

Grupo	1	2	3	4
N	32	24	11	17
Reteste	90 dias	7 dias	24 dias	dias
<i>R</i>	0,83	0,70	0,93	0,90
<i>Lambda</i>	0,70	0,12	0,26	0,46
<i>F</i>	0,73	0,80	0,80	0,80
<i>M</i>	0,68	0,66	0,75	0,75
<i>WSumC</i>	0,79	0,40	0,62	0,75
<i>EA</i>	0,71	0,64	0,61	0,73
<i>FM</i>	0,67	0,62	0,55	0,68
<i>m</i>	0,41	0,64	0,73	0,45
<i>SumC'</i>	0,56	0,46	0,15	0,67
<i>SumT</i>	0,64	-0,07	0,92	0,70
<i>SumV</i>	0,32	0,00	0,82	0,57
<i>SumY</i>	0,46	0,40	0,07	0,40
<i>es</i>	0,66	0,66	0,63	0,80
<i>FC</i>	0,66	0,61	0,80	0,08
<i>CF</i>	0,69	0,33	0,31	0,42
<i>C</i>	0,53	0,28	0,44	0,95
<i>CF+C</i>	0,74	0,26	0,48	0,62
<i>Afr</i>	0,64	0,24	0,78	0,48
<i>EGO</i>	0,35	0,79	0,74	0,60
<i>X+%</i>	0,49	0,83	0,57	0,63
<i>Popular</i>	0,64	0,76	0,40	0,77
<i>a</i>	0,72	0,64	0,53	0,72
<i>p</i>	0,59	0,67	0,75	0,75
Média	0,62	0,50	0,60	0,64
DP	0,14	0,26	0,24	0,19

Faixa	Número de correlações			
≥ 0,75	2	4	8	8
0,50 a 0,74	16	9	8	9
<0,50	5	10	7	6

Grupo 1 = nossa amostra

Grupo 2 = Schwartz, Mebane e Malony (1990)

Grupo 3 = adultos não-pacientes com até 55 anos de idade (Erstad, 1996)

Grupo 4 = adultos não-pacientes com 60 anos ou mais (Erstad, 1996)

Correlações menores do que 0,50
Correlações entre 0,50 e 0,74
Correlações iguais a 0,75 ou maiores

No SCR, os tipos de vivência introversivo e extratensivo são considerados estilos definidos de resposta, ou seja, indicam preferências bem estabelecidas no modo de lidar com as situações e resolver problemas: o introversivo prefere pensar antes de agir, o extratensivo prefere solucionar os problemas por tentativa e erro, sendo bastante influenciado pelos afetos. O tipo de vivência ambigüal é concebido como indicando que o indivíduo oscila entre as duas abordagens (lidar com as situações pensando ou sentindo), não tendo um estilo de resposta definido.

Exner (2003, p. 316) comenta que, separando os 100 participantes do estudo de teste-reteste com intervalo de 3 anos (Exner, Armbruster e Viglione, 1978) em termos de estilos de resposta do *EB*, encontrou que os ambigüais apresentaram menor estabilidade temporal dos resultados no Rorschach. Assim sendo, a maior proporção de ambigüais em nossa amostra pode ser um dos fatores que contribuiu para a menor estabilidade temporal encontrada. Investigamos esta questão comparando as correlações de postos para as variáveis do SCR selecionando os participantes com mesmo estilo do *EB* no teste e no reteste. Encontramos 7 introversivos, 4 extratensivos e 10 ambigüais. Devido ao número pequeno de extratensivos, os reunimos com os introversivos – deste modo, formamos dois subgrupos: um deles com participantes com o mesmo estilo definido (introversivo ou extratensivo) no teste e no reteste; o outro com participantes ambigüais em ambas testagens. Os resultados são apresentados na Tabela 39. A estabilidade temporal foi um pouco maior no grupo de introversivos e extratensivos. A diferença entre as médias da estabilidade temporal nos dois subgrupos foi de magnitude pequena ($d = 0,20$). Assim sendo, a maior proporção de ambigüais em nossa amostra pode ter contribuído para a menor estabilidade temporal encontrada, embora esta contribuição pareça ser pequena.

Tabela 39 – Correlações de postos de Spearman entre os escores no teste e no reteste para grupos separados pelo tipo de vivência

Variável	I-E	A	Variável	I-E	A
<i>R</i>	0,81	0,92	$3r+(2)/R$	0,42	0,24
<i>Lambda</i>	0,22	0,84	$Fr+rF$	1,00	0,71
<i>F</i>	0,70	0,80	<i>FD</i>	0,12	0,65
<i>M</i>	0,86	0,74	<i>MOR</i>	0,78	0,73
<i>WSumC</i>	0,82	0,92	<i>H</i>	0,94	0,28
<i>EA</i>	0,76	0,86	$(H)+Hd+(Hd)$	0,70	0,36
<i>FM</i>	0,58	0,39	<i>Zf</i>	0,49	0,78
<i>m</i>	0,69	0,03	<i>Zd</i>	0,24	0,07
$FM+m$	0,59	0,19	<i>DQ+</i>	0,79	0,75
<i>SumC'</i>	0,62	0,48	<i>DQv</i>	0,72	0,87
<i>SumT</i>	–	0,59	<i>XA%</i>	0,31	0,77
<i>SumV</i>	0,67	0,16	<i>WDA%</i>	0,26	0,74
<i>SumY</i>	0,60	0,64	<i>Popular</i>	0,48	0,69
<i>SumSH</i>	0,47	0,34	<i>X+%</i>	-0,48	0,72
<i>es</i>	0,78	0,23	<i>Xu%</i>	0,48	0,19
<i>Adjes</i>	0,76	0,33	<i>X-%</i>	-0,04	0,75
<i>D-score</i>	0,30	0,13	<i>a</i>	0,73	0,52
<i>AdjD</i>	-0,25	0,46	<i>p</i>	0,72	0,61
<i>FC</i>	0,70	0,08	$2AB+Art+Ay$	0,49	0,62
<i>CF</i>	0,83	0,50	$DV+DR$	0,06	0,75
<i>C</i>	0,32	-0,41	$INC+FAB$	0,79	0,61
$CF+C$	0,76	0,74	<i>Sum6</i>	0,25	0,67
<i>Afr</i>	0,56	0,71	<i>WSum6</i>	0,73	0,57
<i>Blends</i>	0,79	0,93	<i>PTI</i>	0,61	0,36
<i>S</i>	0,58	0,45	<i>DEPI</i>	0,80	0,01
<i>COP</i>	0,93	0,76	<i>CDI</i>	0,45	0,79
<i>AG</i>	0,72	0,39	<i>SCON</i>	0,23	0,03
<i>GHR</i>	0,60	0,41			
<i>PHR</i>	0,73	0,29	Média	0,57	0,51
<i>Human Content</i>	0,74	0,56	Desvio-padrão	0,30	0,29
<i>Isolate/R</i>	0,93	0,51	Mediana	0,68	0,57
<i>(2)</i>	0,90	0,46			

I-E = Introversivos e Extratensivos

A = Ambíguais

10.6.2. Direcionalidade das proporções e consistência das categorias interpretativas

Em seguida, investigamos a direcionalidade das proporções do Sumário Estrutural (Tabela 40). Consideramos direcionalidade quando um dos lados da proporção foi maior do que o outro em mais de um ponto. *EB* mostrou-se estável para 22 participantes (69%); *EA* : *es*, para 16 participantes (50%); *eb*, para 19 participantes (59%); *a* : *p*, para 19 participantes (59%); *FC* : $CF+C$, para 22 participantes (69%); *H* : $(H)+Hd+(Hd)$, para 21 participantes (66%).

Tabela 40 – Direcionalidade de 6 proporções do SCR no teste e no reteste na nossa amostra

Teste	Reteste			%	V	kappa
	À esquerda	Sem direcionalidade	À direita			
EB						
À esquerda	9	3	1	69%	0,51**	0,48**
Sem direcionalidade	3	6	2	55%		
À direita	1	1	7	63%		
EA : es						
À esquerda	7	3	3	54%	0,36	0,26*
Sem direcionalidade	3	2	5	20%		
À direita	0	2	7	78%		
eb						
À esquerda	12	1	1	86%	0,33	0,32*
Sem direcionalidade	6	6	1	46%		
À direita	2	2	1	20%		
a : p						
À esquerda	6	4	2	50%	0,44*	0,38**
Sem direcionalidade	1	3	4	38%		
À direita	1	1	10	83%		
FC : CF+C						
À esquerda	1	2	0	33%	0,42*	0,43**
Sem direcionalidade	1	15	1	88%		
À direita	1	5	6	50%		
H: (H)+Hd+(Hd)						
À esquerda	2	2	0	50%	0,49**	0,41**
Sem direcionalidade	1	7	4	58%		
À direita	0	4	12	75%		

* p < 0,05; ** p < 0,01

Em relação à direcionalidade das proporções, comparando os resultados da nossa amostra com os resultados da amostra norte-americana de 100 adultos não-pacientes retestados após 3 anos (Exner, Armbruster e Viglione, 1978), encontramos, na nossa amostra, estabilidade significativamente menor para *EB* e *EA : es*; a estabilidade das demais proporções investigadas nas duas pesquisas não foi significativamente diferente (mas *a : p* quase alcançou significância estatística: $X^2 = 3,321$, $p = 0,068$). Os resultados encontram-se resumidos na Tabela 41.

Tabela 41 – Frequência e porcentagem de participantes com proporções estáveis na nossa amostra e numa amostra norte-americana

País	Brasil	E.U.A
Amostra	32 não-pacientes	100 não-pacientes
Reteste	3 meses	3 anos
EB*	22 (69%)	86 (86%)
EA : es**	16 (50%)	82 (82%)
eb	19 (59%)	58 (58%)
a: p	19 (59%)	76 (76%)
FC : CF+C	22 (69%)	78 (78%)

* $p < 0,05$; ** $p < 0,01$

A Tabela 42 compara os resultados da nossa amostra com os resultados da amostra francesa de 75 adultos não-pacientes retestados após 3 meses (Sultan et al., 2006). Encontramos que a estabilidade não foi significativamente diferente para as 5 proporções analisadas nos dois estudos (mas $FC : CF+C$, com estabilidade maior em nossa amostra, quase alcançou significância estatística: $X^2 = 3,419$, $p = 0,064$).

Tabela 42 – Frequência e porcentagem de participantes com proporções estáveis na nossa amostra e numa amostra francesa

País	Brasil	França
Amostra	32 não-pacientes	75 não-pacientes
Reteste	3 meses	3 meses
EB	22 (69%)	44 (59%)
EA : es	16 (50%)	40 (53%)
eb	19 (59%)	35 (47%)
a: p	19 (59%)	37 (49%)
FC : CF+C	22 (69%)	37 (49%)

Continuando, investigamos a consistência das categorias interpretativas dos índices *Lambda*, estilo passivo e respostas de reflexo (Tabela 43). Para *Lambda*, 23 participantes (72%) mantiveram-se na mesma faixa de valores em ambas as testagens. Para o estilo passivo, 24 participantes (75%) mantiveram-se na mesma categoria interpretativa, no teste e no reteste. Para as respostas de reflexo, 30 participantes (94%) mantiveram-se na mesma categoria interpretativa.

Tabela 43 – Consistência das categorias interpretativas de Lambda, estilo passivo e respostas de reflexos em nossa amostra

Índices no teste	Reteste		%	Φ	kappa
	1ª categoria	2ª categoria			
Lambda					
< 1,00	11	5	69%	0,44*	0,44*
≥ 1,00	4	12	75%		
Estilo passivo					
Ausente	14	6	70%	0,52**	0,50**
Presente	2	10	83%		
Reflexos					
= 0	24	2	92%	0,83**	0,82**
> 0	0	6	100%		

* $p < 0,05$; ** $p < 0,01$

A consistência das categorias interpretativas de *Lambda*, estilo passivo e reflexos, na nossa amostra, não foi significativamente diferente da encontrada na amostra francesa de Sultan et al. (2006), conforme apresentado na Tabela 44.

Tabela 44 – Frequência e porcentagem de participantes com proporções estáveis em nossa amostra e numa amostra francesa

País	Brasil	França
Amostra	32 não-pacientes	75 não-pacientes
Reteste	3 meses	3 meses
Lambda	23 (72%)	53 (71%)
Estilo passivo	24 (75%)	48 (64%)
Reflexos	30 (94%)	64 (85%)

O *EB* também pode ser analisado em termos de tipos de vivência: introversivo, extratensivo ou ambigüal. Vinte e um participantes (66%) apresentaram o mesmo tipo de vivência no teste e no reteste (Tabela 45).

Tabela 45 – Consistência das categorias do EB em nossa amostra

Teste	Reteste			%	V	kappa
	Introversivo	Ambigual	Extratensivo	consistentes		
EB						
Introversivo	7	3	0	70%	0,53**	0,46**
Ambigual	4	10	3	59%		
Extratensivo	1	0	4	80%		

Em seguida, comparamos a consistência das categorias interpretativas para *Lambda*, *a : p*, *EB* e *FC : CF+C*, na nossa amostra, na amostra francesa (Sultan et al., 2006) e em três amostras norte-americanas: 35 adultos não-pacientes retestados após 3 semanas (Exner, 1999a), 50 adultos não-pacientes retestados após 1 ano (Exner, 1999a) e 8 idosos (Erstad, 1996). A Tabela 46 resume os resultados nas 5 amostras.

Tabela 46 – Frequência e porcentagem de participantes com consistência das categorias em nossa amostra, em três amostras norte-americanas e numa amostra francesa

País	Brasil	E.U.A	EUA	EUA	França
N	32	35	50	8	75
Reteste	90 dias	3 semanas	1 ano	8 dias	90 dias
<i>Lambda</i>	23 (72%)	33 (94%)*	49 (98%)**	7 (88%)	53 (71%)
<i>a : p</i>	24 (75%)	35 (100%)**	49 (98%)**	5 (63%)	48 (64%)
<i>EB^a</i>	11 (92%)	35 (100%)	49 (98%)	–	24 (89%)
<i>FC : CF+C</i>	22 (69%)	34 (97%)**	48 (96%)**	4 (50%)	37 (49%)

^a ambíguos foram excluídos.

* $p < 0,05$ ** $p < 0,01$

Nas amostras de Exner, os ambíguos foram excluídos da análise. Considerando então apenas os introversivos e extratensivos, encontrou que todos os 29 participantes (100%) retestados após 3 semanas e 39 dos 40 participantes (98%) retestados após 1 ano apresentaram o mesmo estilo do *EB* no teste e no reteste. Se excluirmos os ambíguos da nossa amostra, sobram 12 participantes, sendo que 11 deles (92%) apresentaram o mesmo estilo do *EB* no teste e no reteste. Para a amostra francesa (Sultan et al., 2006), excluindo os ambíguos, sobram 27 participantes, sendo que 24 deles (89%) apresentaram o mesmo estilo do *EB* no teste e no reteste. Nesta análise, a consistência das categorias interpretativas do *EB* não foi significativamente diferente nestas três amostras. Contudo, devemos lembrar que a proporção de ambíguos, excluídos nesta análise, foi menor nos norte-americanos do que nos brasileiros e nos franceses.

A consistência das categorias interpretativas em nossa amostra foi significativamente menor do que nas amostras norte-americanas de Exner, para *Lambda*, *a : p* e *FC : CF+C*, mas não foi significativamente diferente da encontrada na amostra de idosos de Erstad (1996) e na amostra francesa (Sultan et al., 2006).

O SCR destaca situações em que os resultados do *EB* são pouco confiáveis: quando *M* ou *WSumC* apresentam valores iguais à zero. Estes resultados não são esperados e estariam associadas a estados de estresse. Assim sendo, decidimos verificar os participantes com *M* ou *WSumC* iguais à zero em nossa amostra. Um participante apresentou *M* = 0 em ambas testagens, e outro participante apresentou *M* = 0 no teste e *M* = 1 no reteste. Um participante apresentou *WSumC* = 0,0 em ambas testagens, outros 3 participantes apresentaram *WSumC* = 0,0 apenas no teste e um outro participante apresentou *WSumC* = 0 apenas no reteste. Excluindo estes 7 participantes, sobram 25 com *EB* válido para indicar os estilos de resposta. Dentre estes 25 participantes, 15 (60%) apresentaram o mesmo tipo de vivência no teste e no reteste (Tabela 47).

Tabela 47 – Consistência das categorias do tipo de vivência na nossa amostra, excluindo-se os casos com *M* = 0 ou *WSumC* = 0 no teste ou no reteste

Teste	Reteste			% consistentes	V	kappa
	Introversivo	Ambigual	Extratensivo			
EB						
Introversivo	3	2	0	60%	0,40*	0,35*
Ambigual	4	9	3	56%		
Extratensivo	1	0	3	75%		

* $p < 0,05$ ** $p < 0,01$

Exner (2003) indica que *EB* deve ser interpretado em conjunto com *Lambda*. Quando encontramos $Lambda \geq 1,00$, identificamos o estilo evitativo, que predomina sobre os estilos identificados pelo *EB*. Assim sendo, teríamos quatro diferentes tipos: evitativo, introversivo, extratensivo e ambigual. Seguindo esta classificação, encontramos em nossa amostra 18 participantes (56%) com mesmo estilo no teste e no reteste (Tabela 48).

Tabela 48 – Consistência das categorias do tipo de vivência no teste e no reteste, considerando *EB* e *Lambda*

Teste	Reteste				% consistentes	V	kappa
	Evitativo	Introversivo	Ambigual	Extratensivo			
Tipo de vivência							
Evitativo	12	3	0	1	75%	0,41	0,33*
Introversivo	0	2	2	0	50%		
Ambigual	4	1	3	1	33%		
Extratensivo	1	1	0	1	33%		

* $p < 0,01$

Eliminando de nossa amostra os casos com $\Lambda \geq 1,00$ no teste ou no reteste, sobram 11 casos, sendo que 6 deles (55%) apresentaram o mesmo tipo de vivência (avaliado pelo *EB*) no teste e no reteste (Tabela 49). A porcentagem de casos que permaneceram na mesma categoria em ambas testagens não foi significativamente diferente da encontrada na amostra francesa de Sultan et al. (2006): 24 participantes (57%) dentre os 42 com $\Lambda < 1,00$.

Tabela 49 – Consistência das categorias do tipo de vivência para 11 participantes da nossa amostra com $\Lambda < 1,00$ no teste e no reteste

Teste	Reteste			%	V	Kappa
	Introversivo	Ambigual	Extratensivo	consistentes		
EB						
Introversivo	2	2	0	50%	0,41	0,28
Ambigual	1	3	1	60%		
Extratensivo	1	0	1	50%		

Eliminando da análise casos com *EB* com valores baixos ($EA < 4,0$, lembrando que *EA* corresponde à soma dos dois lados do *EB*) no teste ou no reteste, sobram 22 casos, sendo 13 deles (59%) consistentes, em termos das categorias do *EB* no teste e no reteste (Tabela 50).

Tabela 50 – Consistência das categorias do tipo de vivência para os casos com $EA \geq 4,0$ em nossa amostra

Teste	Reteste			%	V	kappa
	Introversivo	Ambigual	Extratensivo	consistentes		
EB						
Introversivo	6	2	0	75%	0,50*	0,38**
Ambigual	4	3	2	33%		
Extratensivo	1	0	4	80%		

* $p < 0,05$ ** $p < 0,01$

Para o tipo de vivência (*EB* considerado em termos de categorias interpretativas), as porcentagens de participantes com mesmo tipo no teste e no reteste foram semelhantes nas várias situações investigadas: a) considerando todos os protocolos (66%); b) eliminando os casos com $M = 0$ ou $WSumC = 0,0$ (60%); c) classificando o tipo de vivência em quatro categorias, de modo a incluir como uma delas o estilo evitativo (56%); d) eliminando os protocolos com $\Lambda \geq 1,00$ (55%); e) eliminando os casos com $EA < 4,0$ (59%).

Continuando com a análise da consistência das categorias interpretativas, investigamos a consistência para as variáveis *D-score*, *Zd* e *Textura* (Tabela 51). Permaneceram na mesma categoria interpretativa quatorze participantes (44%) para *D-score*, 16 participantes (50%) para *Zd* e 25 participantes (78%) para *Textura*.

Tabela 51 – Consistência das categorias para D-score, Zd e Textura em nossa amostra

Categoria no teste	Reteste			%	V	kappa
	Primeira categoria	Segunda categoria	Terceira categoria	consistentes		
<i>D-score</i> < 0	2	3	0	60%	0,23	-0,02
<i>D-score</i> = 0	6	11	6	48%		
<i>D-score</i> > 0	0	3	1	25%		
<i>Zd</i> < -3,0	0	2	1	0%	0,28	-0,01
<i>Zd</i> = -3,0 a 3,0	6	14	2	64%		
<i>Zd</i> > + 3,0	0	5	2	29%		
<i>T</i> = 0	22	3	1	85%	0,50*	0,40*
<i>T</i> = 1	2	1	0	33%		
<i>T</i> > 1	0	1	2	67%		

* $p < 0,01$

Comparamos a consistência das categorias interpretativas de *D-score*, *Zd* e *Textura* encontrada na nossa amostra, na amostra de Exner (1988) e na amostra francesa de Sultan et al. (2006), conforme apresentado na Tabela 52. Em relação à amostra francesa, nossa amostra apresentou consistência significativamente maior para *Textura*. Por outro lado, apresentou consistência de *D-score* e *Zd* significativamente menor do que a da amostra norte-americana.

Tabela 52 – Frequência e porcentagem de participantes na mesma categoria interpretativa no teste e no reteste para D-score, Zd e Textura na nossa amostra, numa amostra norte-americana e numa amostra francesa

País	Brasil	E.U.A	França
Amostra	32 não-pacientes	36 participantes	75 não-pacientes
Reteste	3 meses	4 a 30 dias	3 meses
<i>D-score</i>	14 (44%)	35 (97%)**	40 (53%)
<i>Zd</i>	16 (50%)	32 (89%)**	45 (60%)
<i>Textura</i>	25 (78%)	–	40 (53%)*

* $p < 0,05$ ** $p < 0,01$

Por fim, investigamos a estabilidade das constelações.

O Índice de percepção e pensamento [*PTI*] apresenta valores de zero a cinco. É interpretado de modo dimensional, com valores maiores indicando maior prejuízo. Apesar de não possuir ponto de corte, valores maiores do que 3 são considerados um indício de importante prejuízo na percepção e pensamento. Na Tabela 53, apresentamos a análise do *PTI* com seus valores divididos em duas categorias: $PTI \leq 3$ e $PTI > 3$. Encontramos que 31 participantes (97%) mantiveram-se com $PTI \leq 3$ e um participante passou de $PTI \geq 3$ no teste para $PTI < 3$ no reteste.

Tabela 53 – Frequência de participantes em duas categorias interpretativas do índice PTI

Teste	Reteste PTI		% consistentes		Φ	k
	> 3	≤ 3	> 3	≤ 3		
PTI	> 3	0	1	0%	–	–
	≤ 3	0	31	100%		

As demais constelações foram analisadas em termos de seu status positivo ou negativo, com os resultados apresentados na Tabela 54.

Tabela 54 – Frequência de participantes com constelações positivas e negativas na nossa amostra

Índices no teste		Índices no reteste		% consistentes	Φ	k
		Positivo	Negativo			
SCZI						
	Positivo	2	0	100%	0,60**	0,53**
	Negativo	3	27	90%		
DEPI						
	Positivo	4	6	40%	0,45**	0,41**
	Negativo	1	21	95%		
CDI						
	Positivo	10	2	83%	0,73**	0,73**
	Negativo	2	18	90%		
S-CON						
	Positiva	0	1	0%	–	–
	Negativa	0	31	100%		
HVI						
	Positivo	1	3	25%	0,29	0,27
	Negativo	1	27	96%		
OBS						
	Positivo	0	0	0%	–	–
	Negativo	0	35	100%		

* $p < 0,05$ ** $p \leq 0,01$

O Índice de Esquizofrenia [*SCZI*] apresenta valores de zero a seis e $SCZI \geq 4$ é considerado positivo. O *Índice de Esquizofrenia* mostrou-se estável, em termos de seu status positivo ou negativo, para 29 dos 32 participantes (91%).

O Índice de depressão [*DEPI*] apresenta valores de zero a sete e $DEPI \geq 5$ é considerado positivo. O *Índice de Depressão* mostrou-se estável em termos de seu status positivo ou negativo para 29 dos 32 participantes (91%).

O Índice de Déficit Relacional [*CDI*] apresenta valores de zero a cinco e $CDI \geq 4$ é considerado positivo. O *Índice de Déficit Relacional* mostrou-se estável em termos de seu status positivo ou negativo para 28 dos 32 participantes (88%).

A Constelação de Suicídio [*S-CON*] apresenta valores de zero a doze e $S-CON \geq 8$ é considerada positiva. A *Constelação de Suicídio* mostrou-se estável em termos de seu status negativo ou positivo para 31 dos 32 participantes (97%).

O *Índice de Hipervigilância* mostrou-se estável em termos de seu status negativo ou positivo para 28 dos 32 participantes (88%).

Todos os 32 participantes (100%) apresentaram *Índice de Estilo Obsessivo* [*OBS*] negativo no teste e no reteste.

Para todas as constelações, exceto *CDI*, a ausência de um índice positivo no teste foi uma característica bastante estável, enquanto que um índice positivo no teste mais provavelmente mudou para negativo no reteste. Este resultado é esperado, pois o Poder Preditivo Positivo (ou a porcentagem de consistência para os índices positivos) é menor do que o Poder Preditivo Negativo (ou a porcentagem de consistência para os índices negativos) quando um índice positivo apresenta uma baixa taxa de ocorrência (STEINER, 2003), tal como ocorreu em nossa amostra.

No teste, o *CDI* foi positivo para 12 dos 32 participantes, sendo que 10 deles mantiveram-se com *CDI* positivo no reteste (Poder Preditivo Positivo = 83%); foi negativo para 20 dos 32 participantes, sendo que 18 deles mantiveram-se com *CDI* negativo no reteste (Poder Preditivo Negativo = 90%).

Comparando a consistência das constelações em nossa amostra com a encontrada na amostra francesa (Sultan et al., 2006), apenas *DEPI* mostrou-se significativamente mais estável em nossa amostra (Tabela 55).

Tabela 55 – Frequência e porcentagem de participantes com constelações consistentes em nossa amostra e numa amostra francesa

País	Brasil	França
Amostra	32 adultos	75 adultos
Reteste	90 dias	90 dias
SCZI	29 (91%)	61 (81%)
DEPI	29 (91%)*	46 (61%)
CDI	28 (88%)	54 (72%)
S-CON	31 (97%)	73 (97%)
HVI	28 (88%)	62 (82%)
OBS	32 (100%)	73 (97%)

* $p < 0,01$

10.7. Problemas estatísticos interferindo nas estimativas da estabilidade temporal

Um dos fatores que pode diminuir a estabilidade temporal é a restrição da faixa de valores. Infelizmente, as pesquisas de Exner sobre estabilidade temporal não apresentam estatísticas descritivas das variáveis. No entanto, contamos com as estatísticas descritivas da amostra normativa do SCR, apresentadas em Exner (2003). Decidimos comparar os desvios-padrão das variáveis em nossa amostra (utilizando a média dos desvios-padrão do teste e do reteste) com os desvios-padrão das variáveis na amostra normativa norte-americana. Os resultados são apresentados na Tabela 56.

Quarenta e quatro desvios-padrão foram maiores em nossa amostra. Parece então que nossa amostra apresentou faixas de valores mais amplas do que as das amostras de Exner (supondo-se que estas apresentem desvios-padrão semelhantes aos da amostra normativa). Assim sendo, a estabilidade temporal mais baixa encontrada em nossa amostra não pode ser explicada como decorrência de faixas de valores mais restritas.

Tabela 56 – Desvios-padrão de 50 variáveis do SCR em nossa amostra (média do teste e do reteste) e na amostra normativa norte-americana

Variável	Nossa amostra	Normativa EUA	Variável	Nossa amostra	Normativa EUA
<i>R</i>	6,51	4,40	<i>GHR</i>	1,63	1,78
<i>Lambda</i>	0,58	0,31	<i>PHR</i>	2,80	1,46
<i>F</i>	4,69	2,83	<i>Human Content</i>	3,01	1,75
<i>M</i>	2,40	1,95	<i>Isolate/R</i>	0,13	0,09
<i>WSumC</i>	2,47	1,78	<i>(2)</i>	3,25	2,18
<i>EA</i>	3,29	2,38	<i>3r+(2)/R</i>	0,10	0,09
<i>FM</i>	1,86	1,31	<i>Fr+rF</i>	0,82	0,43
<i>m</i>	1,12	0,99	<i>FD</i>	0,68	0,94
<i>FM+m</i>	2,26	1,70	<i>MOR</i>	1,46	0,89
<i>SumC'</i>	1,36	1,16	<i>H</i>	1,91	1,71
<i>SumT</i>	0,70	0,61	<i>Zf</i>	3,66	2,78
<i>SumV</i>	0,68	0,61	<i>Zd</i>	3,81	2,98
<i>SumY</i>	1,27	0,96	<i>DQ+</i>	3,02	2,23
<i>SumSH</i>	2,29	2,09	<i>DQv</i>	1,42	1,26
<i>es</i>	3,64	2,99	<i>XA%</i>	0,10	0,06
<i>D-score</i>	1,11	0,97	<i>WDA%</i>	0,10	0,94
<i>AdjD</i>	1,23	0,82	<i>Popular</i>	1,68	1,39
<i>FC</i>	1,52	1,88	<i>X+%</i>	0,11	0,09
<i>CF</i>	1,70	1,31	<i>Xu%</i>	0,10	0,07
<i>C</i>	0,64	0,37	<i>X-%</i>	0,09	0,05
<i>Afr</i>	0,23	0,16	<i>a</i>	3,17	2,23
<i>Blends</i>	2,53	2,08	<i>p</i>	2,63	1,64
<i>S</i>	2,91	1,28	<i>2AB+Art+Ay</i>	2,95	1,48
<i>COP</i>	1,24	1,38	<i>Sum6</i>	2,24	1,47
<i>AG</i>	0,89	1,15	<i>WSum6</i>	9,27	4,08

Outro fator que pode resultar em correlações mais baixas é a assimetria das variáveis. Quando a assimetria é maior do que 1,00, os dados não seguem uma distribuição semelhante à curva normal (DANCEY; REIDY, 2006). Comparamos a assimetria das variáveis do SCR em nossa amostra, no teste e no reteste, com a assimetria das variáveis na amostra normativa do Sistema Compreensivo. Os resultados são apresentados na Tabela 57.

Tabela 57 – Assimetria de 50 variáveis do SCR em nossa amostra e na amostra normativa norte-americana

Amostra	Nossa amostra (teste)	Nossa amostra (reteste)	Normativa EUA
<i>R</i>	0,92	0,64	0,86
<i>Lambda</i>	1,16	1,63	2,27
<i>F</i>	1,04	0,58	0,92
<i>M</i>	0,23	0,83	0,48
<i>WSumC</i>	1,14	1,02	0,11
<i>EA</i>	0,04	0,63	-0,04
<i>FM</i>	0,44	0,16	0,15
<i>m</i>	1,50	1,40	0,62
<i>FM+m</i>	0,78	0,46	0,20
<i>SumC'</i>	1,18	3,54	1,41
<i>SumT</i>	2,50	1,74	0,83
<i>SumV</i>	1,28	2,60	2,71
<i>SumY</i>	0,81	0,92	3,53
<i>SumSH</i>	0,72	1,84	2,54
<i>es</i>	1,29	1,23	1,43
<i>D-score</i>	0,74	0,63	-3,06
<i>AdjD</i>	1,33	1,41	-0,88
<i>FC</i>	1,54	0,97	0,29
<i>CF</i>	0,30	0,82	3,76
<i>C</i>	3,34	2,26	12,15
<i>Afr</i>	1,17	1,39	0,35
<i>Blends</i>	1,03	1,15	0,00
<i>S</i>	2,10	3,26	1,99
<i>COP</i>	1,47	1,93	0,25
<i>AG</i>	1,36	2,08	1,02
<i>GHR</i>	0,29	0,53	0,36
<i>PHR</i>	0,73	0,56	1,25
<i>HumanContent</i>	0,65	0,60	0,59
<i>Isolate/R</i>	0,86	0,04	0,51
<i>(2)</i>	0,99	0,29	0,29
<i>3r+(2)/R</i>	0,78	0,20	0,47
<i>Fr+fr</i>	2,25	2,25	4,98
<i>FD</i>	0,99	0,83	0,84
<i>MOR</i>	1,15	1,47	1,01
<i>H</i>	0,91	1,74	0,97
<i>Zf</i>	0,04	0,43	0,87
<i>Zd</i>	0,05	0,16	0,31
<i>DQ+</i>	0,36	0,54	0,53
<i>DQv</i>	0,69	1,33	1,35
<i>XA%</i>	0,10	-0,32	-1,34
<i>WDA%</i>	-0,05	-0,74	-1,42
<i>Popular</i>	0,03	0,31	-0,09
<i>X+%</i>	0,22	0,22	-0,86
<i>Xu</i>	0,04	-0,39	0,54
<i>X-%</i>	-0,37	0,43	1,41
<i>a</i>	1,64	1,31	0,32
<i>p</i>	0,87	0,27	0,57
<i>2AB+Art+Ay</i>	2,93	3,83	1,27
<i>Sum6</i>	1,09	0,88	0,80
<i>WSum6</i>	1,75	2,02	1,42

Para as 50 variáveis comparadas:

- 19 variáveis apresentaram assimetria < 1 em nossa amostra no teste, no reteste, e na amostra normativa norte-americana;
- 14 variáveis apresentaram assimetria > 1 em nossa amostra no teste e/ou no reteste e na amostra normativa norte-americana;
- 11 variáveis apresentaram assimetria > 1 em nossa amostra no teste e/ou no reteste, mas não na amostra normativa norte americana;
- 6 variáveis apresentaram assimetria > 1 na amostra norte-americana, mas não em nossa amostra.

Nossa amostra, no teste, apresentou 31 assimetrias maiores do que as da amostra norte-americana; no reteste, apresentou 29 assimetrias maiores. Assim sendo, parece que a assimetria não é um fator importante para explicar as correlações mais baixas em nossa amostra.

Algumas variáveis do Rorschach apresentam valores baixos, com muitos testandos obtendo escores iguais à zero. Isto pode contribuir para restrição da faixa de valores e assimetria, que costumam diminuir a estabilidade temporal. Comparamos a proporção de casos com escores iguais à zero em nossa amostra (média do teste e do reteste) com a proporção na amostra normativa norte-americana. Para encontrar o número de casos com escores iguais à zero, subtraímos do número total de participantes a frequência de casos com escores maiores do que zero, informada nas estatísticas descritivas. Esta análise foi realizada apenas para as variáveis que costuma apresentar valores baixos, com escores iguais à zero sendo bastante frequentes. Os resultados são apresentados na Tabela 58.

Nossa amostra apresentou uma proporção de casos com escores iguais à zero significativamente maior do que a da amostra norte-americana para as variáveis *m*, *SumC'*, *SumT*, *FC*, *CF*, *COP*, *AG*, *FD* e *H*. Para estas variáveis, então, a distribuição de valores com maior frequência de escores igual à zero em nossa amostra pode ter contribuído para a menor estabilidade temporal. Nossa amostra apresentou também uma proporção significativamente menor de casos com escores iguais a zero para *Fr+rF* e *DQv* do que a amostra norte-americana. Isto pode ter contribuído para a estabilidade temporal maior de *Fr+rF* em nossa amostra (nenhuma das pesquisas de Exner informa a estabilidade temporal de *DQv*).

Tabela 58 – Frequência e porcentagem de casos com escores iguais à zero para 19 variáveis do SCR em nossa amostra (média do teste e reteste) e na amostra normativa norte-americana

Variável	Nossa amostra		Normativa EUA	
	f	%	f	%
<i>m</i>	17	53	142**	24
<i>SumC'</i>	19	59	110**	18
<i>SumT</i>	25	78	110**	18
<i>SumV</i>	23	72	476	79
<i>SumY</i>	13	41	338	56
<i>FC</i>	10	31	20**	3
<i>CF</i>	7	22	36**	6
<i>C</i>	24	75	539	90
<i>S</i>	4	13	86	14
<i>COP</i>	16	50	102**	17
<i>AG</i>	22	69	220**	37
<i>Fr+rF</i>	25	78	552**	92
<i>FD</i>	19	59	144**	24
<i>MOR</i>	13	41	279	47
<i>H</i>	5	16	5**	1
<i>DQv</i>	14	44	407**	68
<i>2AB+Art+Ay</i>	7	22	151	25
<i>Sum6</i>	5	16	104	17
<i>WSum6</i>	5	16	104	17

* $p < 0,05$; ** $p < 0,01$

Em amostras relativamente pequenas como a nossa, escores extremos (outliers) podem enviesar as estimativas de estabilidade temporal. Para identificar os escores extremos, realizamos análises de regressão para as variáveis do SCR, tomando como variável dependente os escores no reteste e como variável dependente os escores no teste. Analisamos os resíduos, identificando os casos cujos resíduos se situaram a mais do que dois desvios-padrão abaixo ou acima da média dos resíduos. Com esta análise, identificamos valores extremos para 9 variáveis. Para cada uma destas variáveis, recalculamos a estabilidade temporal, excluindo os participantes com escores extremos (Tabela 59). Notamos que, para todas as variáveis, a exclusão dos escores extremos fez com que as estimativas de estabilidade temporal aumentassem, às vezes em magnitude importante (por exemplo, *Afr* apresentou $r = 0,64$, mas, excluindo um participante, a correlação passou a ser $r = 0,84$).

Tabela 59 – Estabilidade temporal das variáveis do SCR na amostra original e retirando-se os escores extremos (outliers)

Amostra	Original	Sem extremos	#
<i>SumC'</i>	0,56	0,74	3
<i>SumT</i>	0,64	0,83	3
<i>SumV</i>	0,32	0,38	3
<i>C</i>	0,53	0,72	7
<i>Afr</i>	0,64	0,84	1
<i>AG</i>	0,78	0,89	5
<i>a</i>	0,72	0,83	1
<i>INC+FAB</i>	0,60	0,69	1
<i>PTI</i>	0,64	0,68	1

número de participantes com valores extremos

O procedimento de eliminar os escores extremos é controverso. Por um lado, gera estimativas da estabilidade temporal menos enviesadas pela presença de uns poucos escores fora do padrão geral. Por outro lado, a própria existência de escores extremos indica que alguns participantes apresentaram importantes mudanças de uma testagem à outra. Existe também o problema do critério a ser utilizado para identificar os escores extremos – nosso critério não se mostrou muito eficiente para a variável *C*, pois foram identificados 7 escores extremos e estes não devem ser tão frequentes; para outras variáveis talvez tenha deixado de identificar escores que enviesaram as estimativas da estabilidade temporal. O importante é assinalar que, para algumas variáveis, a presença de escores extremos pode ter diminuído a estabilidade temporal.

Sultan et al. (2006) levantaram a hipótese de que a taxa de ocorrência da variável influencia sua estabilidade temporal: variáveis com baixa taxa de ocorrência devem apresentar menor estabilidade do que variáveis com taxa de ocorrência maior. A baixa taxa de ocorrência se relaciona aos problemas da restrição da faixa de valores e assimetria.

Em nossa amostra, a relação entre a taxa de ocorrência da variável e a estabilidade temporal foi bastante variada (Figura 4). A correlação entre a taxa de ocorrência (média do teste e do reteste) e a estabilidade temporal foi $r = 0,30$, o que é uma associação de magnitude pequena. De qualquer modo, observamos esta *tendência* de que quanto maior a taxa de ocorrência, maior a estabilidade temporal.

Figura 4 – Estabilidade temporal em função da taxa de ocorrência para as variáveis do SCR em nossa amostra

10.8. A estabilidade temporal e o engajamento na tarefa

O nível de engajamento na tarefa em nossa amostra, considerando-se os valores de R e Λ , aparentemente foi semelhante ao encontrado na amostra normativa: nossa amostra apresentou R maior, mas Λ também maior, disso resultando um nível de engajamento aproximadamente semelhante. No entanto, o nível de engajamento na tarefa em nossa amostra foi menor do que na amostra normativa norte-americana (Exner, 2003) e na pesquisa francesa sobre estabilidade temporal (Sultan et al., 2006).

O engajamento na tarefa também pode ser avaliado pela variável Task Engagement [TE]. Para o cálculo de TE , utilizamos escores padronizados, ou seja, os escores originais são transformados, subtraindo a média e dividindo este resultado pelo desvio-padrão da amostra. Utilizamos os escores do teste e do reteste em conjunto para calcular as médias e desvios-padrão das variáveis componentes de TE . Então, calculamos o índice de engajamento na tarefa [Task Engagement – TE], no teste e no reteste, para os participantes da nossa amostra. Quatorze participantes apresentaram maior engajamento no teste; dezoito, no reteste.

Como o engajamento na tarefa [TE] é calculado a partir de escores padronizados – ou seja, TE é uma medida ancorada na amostra estudada –, isto impede comparações dos valores de TE entre amostras diferentes. Também é difícil estimar a magnitude das mudanças em TE de uma testagem à outra, na nossa amostra.

Encontramos um outro modo de avaliar o engajamento na tarefa. R corresponde ao total de respostas do protocolo; F corresponde às respostas mais simples. Assim, $R-F$ corresponde ao total de respostas mais complexas ou nas quais houve maior engajamento. Devemos observar que $R-F$ corresponde ao denominador do índice $Lambda$, que compara as respostas de forma pura com as respostas com outros determinantes [$Lambda = F/(R-F)$]. Os resultados de $R-F$, no teste e no reteste, são apresentados na Tabela 60.

Tabela 60 – Engajamento na tarefa (avaliado por R-F) no teste e no reteste

Participante	Teste	Reteste	Teste - Reteste
1	14	19	5
2	6	10	4
3	16	21	5
4	17	17	0
5	15	16	1
6	6	6	0
7	15	13	2
8	8	12	4
9	11	11	0
10	12	10	2
11	15	9	6
12	23	14	9
13	10	11	1
14	9	7	2
15	16	10	6
16	9	7	2
17	11	11	0
18	8	8	0
19	12	15	3
20	10	13	3
21	11	15	4
22	17	15	2
23	13	11	2
24	6	6	0
25	18	13	5
26	6	8	2
27	12	14	2
28	7	6	1
29	8	10	2
30	9	8	1
31	16	15	1
32	12	11	1
Média	11,81	11,63	2,44
DP	4,18	3,81	2,17
Mediana	11,50	11,00	2

O engajamento na tarefa, avaliado por $R-F$, aumentou para 12 participantes, permaneceu igual para 6 participantes e diminuiu para 14 participantes. Considerando o

módulo da diferença entre $R-F$ no teste e $R-F$ no reteste, encontramos que $2/3$ dos participantes apresentou diferenças de até 2 pontos de uma testagem à outra, o que pode ser considerada uma diferença “pequena”; porém outro $1/3$ dos participantes apresentou diferenças maiores do que 2 pontos, o que pode ser considerada uma diferença “grande”.

Na amostra normativa de Exner (2003), R apresentou média ≈ 22 e F apresentou média ≈ 8 . Assim, utilizando estes valores, encontramos $R-F \approx 14$, valor 2 pontos maior do que a média de $R-F$ em nossa amostra. Na amostra francesa de Sultan et al. (2006), R apresentou média ≈ 24 e F apresentou média ≈ 9 . Utilizando estes valores, encontramos $R-F \approx 15$, valor 3 pontos maior do que a média de $R-F$ em nossa amostra. Considerando que o número médio de respostas nestas três amostras foi semelhante ($R \approx 23$), e que $R-F$ diferiu aproximadamente 3 pontos entre nossa amostra e as outras duas, podemos dizer que o engajamento na tarefa foi apenas um pouco menor na nossa amostra, já que 3 pontos em 23 respostas correspondem a 13%.

10.9. Protocolos de participantes que conheciam o examinador

Seis participantes conheciam o examinador antes da aplicação do teste. Dois destes participantes conheceram o examinador por meio de conhecidos em comum e os outros quatro participantes conheceram o examinador por trabalharem na Clínica de Psicologia, onde se realizou a pesquisa.

Dos quatro participantes que conheciam o examinador por trabalharem na Clínica de Psicologia, 3 forneceram protocolos com menos de 14 respostas, em uma ou em ambas as testagens, e 1 participante forneceu protocolos com mais de 14 respostas nas duas testagens. Dos participantes com protocolos com menos de 14 respostas, dois apresentaram valores baixos de Λ . Este resultado (R baixo, Λ baixo) indica que se envolveram com a testagem, apresentando respostas complexas, mas forneceram poucas respostas – segundo um estilo que privilegia a qualidade à quantidade. O outro participante com $R < 14$ apresentou $\Lambda > 2,00$ (alto) em ambas testagens.

Dos dois participantes que conheceram o examinador por meio de conhecidos em comum, um apresentou R elevado e $Lambda$ dentro da média da amostra, em ambas testagens, e o outro apresentou R dentro da média e $Lambda$ abaixo da média, em ambas testagens. Assim sendo, ambos se engajaram na tarefa. Apenas 1 participante apresentou o padrão descrito por Exner (1978), de aumento no número de respostas associado ao testando conhecer o examinador.

É difícil avaliarmos o papel que conhecer o examinador pode ter desempenhado nos resultados. Pode ser que para alguns participantes, conhecer o examinador tenha contribuído para fornecerem respostas mais elaboradas, embora não tenha aumentado o número de respostas; para outros participantes, pode ter aumentado o número de respostas, mas sem melhorar sua complexidade; para outros, pode ter provocado defensividade, resultando em protocolos com R baixo e $Lambda$ alto.

Cabe notar que nas pesquisas de Exner, os examinadores eram terapeutas que testaram seus pacientes ou professores que testaram seus alunos, situação diferente da nossa, em que os participantes conheciam socialmente o examinador, sem ter contato mais freqüente com ele.

10.10. Protocolos com poucas respostas

Conforme informado anteriormente, 9 indivíduos que participaram da pesquisa apresentaram protocolos com menos de 14 respostas. Destes, 3 conheciam o examinador antes da testagem; os protocolos dos outros 6 participantes serão agora analisados.

Devido ao número pequeno de protocolos com menos do que 14 respostas, decidimos agrupá-los junto aos 32 participantes e recalcular as correlações. A Tabela 61 apresenta os resultados para a amostra original de 32 participantes e para a amostra de 38 participantes, que inclui os 6 protocolos com menos de 14 respostas.

Tabela 61 – Estabilidade temporal das variáveis do SCR para a amostra original de 32 participantes e para a amostra de 38 participantes, que inclui os 6 casos com menos de 14 respostas em uma ou ambas testagens

Amostra	Original	38	Amostra	Original	38
R	0,83	0,88	(2)	0,76	0,79
Lambda	0,70	0,71	3r+(2)/R	0,35	0,39
F	0,73	0,78	Fr+rF	0,91	0,89
M	0,68	0,69	FD	0,51	0,47
WSumC	0,79	0,78	MOR	0,60	0,57
EA	0,71	0,74	H	0,73	0,71
FM	0,67	0,63	(H)+Hd+(Hd)	0,67	0,73
m	0,41	0,43	Zf	0,67	0,65
FM+m	0,62	0,58	Zd	0,25	0,21
SumC'	0,56	0,57	DQ+	0,69	0,69
SumT	0,64	0,58	DQv	0,72	0,73
SumV	0,32	0,30	XA%	0,65	0,68
SumY	0,46	0,45	WDA%	0,43	0,42
SumSH	0,51	0,51	Popular	0,64	0,60
es	0,66	0,64	X+%	0,49	0,49
Adjes	0,67	0,61	Xu%	0,29	0,34
D-score	0,27	0,28	X-%	0,62	0,67
AdjD	0,30	0,34	a	0,72	0,71
FC	0,66	0,69	p	0,59	0,57
CF	0,69	0,66	2AB+Art+Ay	0,85	0,86
C	0,53	0,50	DV+DR	0,26	0,27
CF+C	0,74	0,73	INC+FAB	0,60	0,52
Afr	0,64	0,64	Sum6	0,73	0,66
Blends	0,69	0,70	WSum6	0,87	0,81
S	0,80	0,81	PTI	0,64	0,69
COP	0,72	0,72	DEPI	0,50	0,49
AG	0,78	0,69	CDI	0,63	0,62
GHR	0,51	0,50	SCON	0,39	0,31
PHR	0,69	0,72			
Human Content	0,78	0,81	Média	0,62	0,61
Isolate/R	0,76	0,76	Desvio-padrão	0,16	0,16

Para a amostra de 38 participantes, as correlações de teste-reteste variaram de 0,21 a 0,89, com média = $0,61 \pm 0,16$ e mediana = 0,66. Para a amostra original de 32 participantes, as correlações de teste-reteste variaram de 0,25 a 0,91, com média = $0,62 \pm 0,16$ e mediana = 0,66. Vinte e nove correlações foram maiores na amostra com 32 participantes, 6 correlações foram iguais nas duas amostras e 25 correlações foram maiores na amostra com 38 participantes. Aparentemente, então, os resultados foram semelhantes, ou seja, a inclusão dos protocolos com menos do que 14 respostas não diminuiu significativamente a estabilidade temporal das variáveis. Devido ao número pequeno de participantes com menos de 14 respostas, 6 casos, talvez não tenha sido possível identificar uma eventual diminuição na estabilidade temporal com a inclusão de protocolos com menos de 14 respostas.

Realizamos também comparações dos resultados no teste e no reteste para os 6 protocolos com menos do que 14 respostas. Esta análise encontra-se no Apêndice H [Suprimido na versão online]. De modo geral, os resultados, no teste e no reteste, para os protocolos com menos de 14 respostas, foram semelhantes para a maioria das variáveis. No entanto, para alguns destes protocolos, ocorreram mudanças importantes em variáveis centrais para a interpretação dos resultados.

11. DISCUSSÃO

Nesta pesquisa, encontramos fidedignidade inter-codificadores substancial ou excelente para a maioria dos códigos do SCR. Contudo, algumas codificações apresentaram fidedignidade insatisfatória. Em parte, a menor fidedignidade ocorreu para variáveis com baixa taxa de ocorrência, problema que já havia sido identificado por outros autores (Acklin et al., 2000; Meyer et al., 2002; Viglione e Taylor, 2003; McGrath, Pogge e Strokes, 2005; Sahly, 2006). Para códigos com baixa taxa de ocorrência, devemos considerar que as estimativas de fidedignidade são imprecisas, mais do que propriamente baixas.

É importante destacar que, considerando o conjunto dos estudos norte-americanos e brasileiros, a fidedignidade inter-codificadores do SCR mostra-se adequada: diferentes estudos encontraram níveis substanciais de fidedignidade para a maioria das variáveis, o que indica que os critérios de codificação são suficientemente claros para garantir uma codificação objetiva das respostas. Contudo, observamos que a fidedignidade inter-codificadores encontrada nos estudos brasileiros foi menor do que a reportada nos estudos norte-americanos. Um dos problemas parece ser a utilização da Tabela de Qualidade Formal do Sistema Compreensivo. Trata-se de uma tabela norte-americana, aparentemente não tão adequada para brasileiros: muitas respostas não são encontradas na Tabela, então sua qualidade formal é avaliada pelo examinador, o que parece diminuir a fidedignidade das codificações de qualidade formal. Também vem sendo mais difícil obter melhor concordância entre diferentes avaliadores brasileiros para os códigos especiais. Este é um aspecto mais difícil da codificação, tanto que, nas diversas pesquisas norte-americanas, a fidedignidade dos códigos especiais costuma ser menor do que a de outros segmentos. Talvez os critérios de codificação não sejam suficientemente claros ou ainda não tenham sido bem compreendidos pelos examinadores brasileiros. De qualquer modo, alguns critérios de codificação do SCR ainda podem ser aprimorados.

Erstad (1996) comenta que os exercícios para treino das codificações fornecidos no Workbook⁶² do SCR mostram que a codificação das respostas segue princípios lógicos e coerentes.

Por outro lado, parte da lógica da codificação do Sistema Compreensivo raramente ou nunca é apresentada em trabalhos impressos e encontra-se disponível apenas nos seminários oferecidos por Rorschach Workshops [cursos ministrados pela organização de pesquisas de Exner]. Contar com informações mais completas sobre a lógica das codificações ajudaria a tornar mais fácil o aprendizado das codificações do Sistema Compreensivo. Em alguns casos, esta informação pode ser o único modo de chegar a compreender as codificações realizadas por Rorschach Workshops [e apresentadas no Workbook]. Esta informação também ajudaria a assegurar que aqueles que querem usar o Sistema Compreensivo estão codificando tão corretamente quanto possível no presente estágio de desenvolvimento do Sistema Compreensivo. (ERSTAD, 1996, p. 177-178).

A publicação do livro **Rorschach Coding Solutions** (Viglione, 2002) parece ser uma contribuição importante no sentido de detalhar os critérios de codificação e esclarecer dúvidas.

Além dos critérios detalhados para a codificação das respostas, existe a necessidade de treinamento adequado dos codificadores para realizar codificações de modo correto. Um avaliador menos bem treinado pode cometer mais erros de codificação. O procedimento de treinamento das codificações do SCR, desenvolvido por Hilsenroth, Charnas e Zoldan (2007) parece ser uma iniciativa louvável e seria importante que os usuários brasileiros do SCR estudassem o material preparado para este treinamento.

Acreditamos que as codificações das respostas, por nós realizadas, estão corretas, apesar de eventuais erros que possam ter ocorrido [Meyer (1998) encontrou que erros na coleta, codificação e análise dos dados são freqüentes nas pesquisas, mas afetam os resultados numa proporção pequena]. Os resultados obtidos para a fidedignidade das codificações foram, de modo geral, satisfatórios, mas com valores menores do que os de outros estudos

⁶² Ver nota 15, à página 38.

brasileiros. Buscando hipóteses que expliquem este resultado, podemos pensar que houve uma codificação menos cuidadosa por parte do segundo avaliador.

Em relação à estabilidade temporal, dez variáveis apresentaram estabilidade alta, semelhante à encontrada nos estudos de Exner. Contudo, a maioria das variáveis apresentou níveis moderados de estabilidade e 10 variáveis – além de *m* e *SumY*, para as quais a baixa estabilidade é esperada – apresentaram baixa estabilidade temporal. De modo geral, a estabilidade temporal em nossa amostra foi menor do que nas pesquisas de Exner e do que a estimativa obtida a partir da metanálise sobre a estabilidade temporal do Rorschach (Grønnerød, 2003a), mas maior do que a obtida na amostra francesa de Sultan et al. (2006).

A menor fidedignidade das codificações deve ser considerada como um dos eventuais fatores que contribuíram para a menor estabilidade temporal encontrada nesta pesquisa. Uma menor fidedignidade das codificações limita a estabilidade temporal, porque, se os protocolos não são codificados de modo consistente, isto pode gerar maior variação entre os resultados do teste e do reteste, e consequentemente, menor estabilidade temporal.

A análise da direcionalidade das proporções e da consistência das categorias interpretativas também mostrou menor estabilidade temporal em nossa amostra do que nas pesquisas de Exner, mas semelhante à encontrada na pesquisa francesa de Sultan et al. (2006). Em nossa amostra, entre 44% e 70% dos participantes mantiveram-se nas mesmas categorias interpretativas no teste e no reteste. Para as constelações, consideradas em termos de seu status negativo ou positivo, a estabilidade temporal encontrada foi semelhante à da amostra francesa, com a maioria dos participantes mantendo o status negativo para as constelações em ambas testagens. O Índice de Déficit Relacional [*CDI*] apresentou mais casos positivos, mas a proporção de participantes que mantiveram o status (negativo ou positivo) deste índice nas duas testagens também foi alta (88%). Para as demais constelações, pelo menos 88% dos participantes apresentaram o mesmo status no teste e no reteste.

Aparentemente, a menor estabilidade temporal encontrada nesta pesquisa não pode ser explicada por problemas de restrição de faixa de valores e assimetria. No entanto, a distribuição dos valores, principalmente quando predominaram escores baixos ou iguais à zero, parece ter contribuído para a menor estabilidade temporal de algumas variáveis.

Encontramos também que escores extremos (outliers) ocorreram para 9 variáveis e diminuíram a estabilidade temporal.

O nível de engajamento na tarefa em nossa amostra aparentemente foi semelhante ao encontrado na amostra normativa da cidade de São Paulo, mas menor do que na amostra normativa norte-americana e na pesquisa francesa sobre estabilidade temporal. O menor nível de engajamento na tarefa pode ter contribuído para a menor estabilidade temporal.

Sultan et al. (2006) investigaram o papel que o engajamento na tarefa desempenha na estabilidade temporal das variáveis do SCR, encontrando que níveis maiores de engajamento na tarefa estavam associados a maior estabilidade temporal para algumas variáveis. Infelizmente, o procedimento utilizado naquele estudo não está descrito de modo claro no artigo (ou nós não conseguimos compreender a descrição) e a comunicação, por e-mail, com o primeiro autor, Serge Sultan, trouxe mais dúvidas do que esclarecimentos. Assim sendo, (ainda) não realizamos esta análise. O menor engajamento na tarefa em nossa amostra ter contribuído para a menor estabilidade mantém-se como uma hipótese a ser futuramente investigada. Pesquisas com amostras com variados níveis de engajamento na tarefa permitiriam investigar esta relação entre o engajamento na tarefa e a estabilidade temporal.

Sultan et al. (2006) comentam que participantes com menor nível de engajamento na tarefa podem mudar ao responder ao teste num segundo momento – o que parece ter ocorrido na nossa amostra. Por outro lado, participantes com excessivo engajamento na tarefa podem se expressar de modo diferente em diferentes momentos – o que parece ter ocorrido na amostra francesa. Um nível ótimo de engajamento na tarefa parece contribuir para uma melhor estabilidade dos resultados – o que parece ter ocorrido nas amostras de Exner.

Holzberg (1977) menciona dois fatores que podem modificar os resultados no reteste. Por um lado, alguns participantes podem apresentar baixa motivação para realizar o reteste. Por outro lado, ao realizar o reteste, os participantes já conhecem o material, o que pode facilitar a realização da tarefa. Analisando o engajamento na tarefa [*TE*], encontramos que 14 participantes apresentaram maior engajamento no teste, e 18 participantes, no reteste. Aparentemente, para a maioria dos participantes, a mudança no engajamento na tarefa, de uma testagem à outra, foi pequena.

Na nossa pesquisa, a estabilidade temporal calculada incluindo-se os protocolos com menos de 14 respostas na amostra não foi significativamente diferente da obtida com apenas os protocolos com pelo menos 14 respostas. Porém, esta análise foi realizada com um número pequeno de protocolos com poucas respostas, o que pode ter dificultado detectar eventuais diferenças que existam.

Os protocolos com menos de 14 respostas foram compreendidos por Exner como sendo inválidos, primeiramente porque a estabilidade temporal dos resultados mostrou-se bem menor nestes protocolos. Em segundo lugar, há o problema da representatividade dos resultados nos protocolos curtos: poucas respostas podem gerar resultados que não representam suficientemente o funcionamento psicológico do indivíduo. Este problema da representatividade dos resultados mostra-se de modo mais claro para os protocolos com poucas respostas, mas, na verdade, permeia qualquer testagem com o Rorschach.

Meyer (1997a) enfatiza que os resultados no Rorschach dependem do modo como o testando aborda a situação de testagem. Um baixo engajamento na tarefa pode decorrer tanto de uma atitude de defesa frente à testagem, quanto expressar uma personalidade com poucos recursos; um alto engajamento na tarefa pode decorrer tanto de uma atitude de exageração ou envolvimento excessivo com a testagem, quanto expressar uma personalidade mais complexa ou perturbada. O Rorschach parece funcionar melhor e fornecer resultados mais representativos quando os participantes apresentam um nível ótimo de engajamento na tarefa, expresso por um número de respostas entre 18 e 27 e valores de *Lambda* entre 0,30 e 0,99, segundo os parâmetros norte-americanos.

Agora discutiremos fatores que podem explicar os resultados do Rorschach obtidos em nossa amostra. Primeiramente, comentaremos as características dos participantes.

Nossa amostra foi composta por participantes voluntários. Weiner (1995) comenta que participantes voluntários podem não se preocupar em censurar algumas respostas, assim fornecendo protocolos com mais respostas e/ou escores desviantes. Isto aparentemente não ocorreu em nossa amostra, cujos participantes forneceram mais respostas do que os participantes da amostra normativa da cidade de São Paulo, mas também apresentaram uma proporção maior de respostas de forma pura – indicando evitação de maior envolvimento – e

menos respostas com códigos especiais críticos (estes códigos estão associados a falhas no pensamento e podem aumentar caso o testando não se preocupe em censurar respostas).

Perguntamos aos participantes por que eles quiseram participar da pesquisa e os motivos alegados foram variados: 1) “quero me conhecer melhor”; 2) “a ... [secretária] perguntou se eu poderia fazer e eu vim”; 3) “minha amiga contou que fez e achou interessante e eu fiquei curiosa para ver como é”; 4) “quero ajudar na pesquisa”. Estas informações não foram coletadas de modo sistemático, mas sim como parte da conversa informal que ocorreu antes da aplicação do teste. Temos a impressão de que querer conhecer os resultados foi um dos principais motivos alegados. Portanto, receber os resultados da avaliação da personalidade serviu como atrativo para a participação, conforme havíamos planejado.

Nossa amostra foi composta por adultos não-pacientes, ou seja, indivíduos que aparentemente não apresentavam problemas psicológicos graves ao ponto de impedirem ou prejudicarem significativamente o desempenho de uma profissão ou os estudos. Contudo, não temos garantias de que todos apresentam boa condição de saúde mental. Aliás, temos a informação de que alguns deles pensavam em procurar psicoterapia e de que outros já haviam realizado psicoterapia em outro momento de suas vidas. Talvez nossa amostra contenha um número importante de participantes que experimentam sofrimento e/ou apresentam demanda por psicoterapia. A menor estabilidade temporal encontrada pode estar associada a esta característica da amostra. Exner (1978) encontrou menor estabilidade temporal em pacientes ambulatoriais em lista de espera, em comparação com que a estabilidade encontrada em amostras de adultos não-pacientes.

Tivemos a impressão de que receber os resultados da avaliação foi proveitoso para muitos participantes. Para aqueles que referiram estar pensando em procurar psicoterapia, receber uma confirmação de suas dificuldades ou de seu sofrimento, a partir dos resultados do Rorschach, parece ter reforçado a demanda por um tratamento. Para todos estes participantes, afirmamos que uma psicoterapia poderia ser benéfica e nos prontificamos a fornecer indicações de serviços em que pudessem ser atendidos.

Outra característica da amostra a ser destacada é a predominância de mulheres. A pesquisa normativa norte-americana não encontrou diferenças entre os resultados de homens e de mulheres e nenhum dos estudos de estabilidade temporal menciona qualquer diferença na

estabilidade em função do gênero dos participantes. Apesar da lenda popular que afirma que as mulheres são mais instáveis do que os homens, em nossa opinião, se a predominância de mulheres desempenhou algum papel para a menor estabilidade encontrada em nossa amostra, este papel foi pequeno.

Utilizando os resultados do Rorschach para caracterizar nossos participantes, encontramos que eles experimentavam menos tensões e contavam com mais recursos para lidar com as situações, o que fazia com que estivessem menos sobrecarregados; esforçaram-se para organizar as informações em proporção semelhante, mas estabeleceram mais relações entre os dados captados e foram mais eficientes na organização das informações; perceberam as situações de modo adequado e apresentaram menos falhas no pensamento do que os participantes da amostra normativa de Nascimento (2002). Consideramos que muitas destas diferenças estariam associadas ao maior nível de instrução dos participantes da nossa amostra. Não parece haver relação entre estas características da amostra e a menor estabilidade temporal.

Em resumo, pensamos que as características dos nossos participantes (voluntários, alguns participantes com demanda de psicoterapia, predomínio de mulheres) não foram o principal fator responsável pela menor estabilidade encontrada em nossa amostra. De qualquer modo, as particularidades da amostra, incluindo-se aqui também o predomínio de indivíduos das classes sociais A e B, limitam a generalização dos resultados. Aspecto importante a ser destacado é o número relativamente pequeno de participantes, que pode ter contribuído para gerar estimativas imprecisas da estabilidade temporal.

Infelizmente, não temos informações sobre como os participantes das pesquisas de Exner foram recrutados – fator que interfere na motivação para realizar o teste e que afeta os resultados –, nem sobre como foram preparados para a situação de testagem. Talvez estes aspectos possam, ao menos em parte, explicar a maior estabilidade temporal obtida nas pesquisas originais do SCR.

Segundo nosso conhecimento, o único artigo de Exner em que encontramos uma descrição mais detalhada sobre o procedimento de recrutamento dos participantes e características dos examinadores é o recente informe sobre a nova pesquisa normativa norte-americana para o SCR (Exner, 2007). Neste artigo, somos informados de que os

examinadores foram psicólogos que receberam treinamento no SCR por meio da realização de cursos oferecidos por Rorschach Workshops. Estes examinadores recrutaram os participantes em empresas, utilizando uma carta com informações sobre a pesquisa: o projeto visava uma nova padronização do Rorschach e envolvia aplicar o teste em centenas de pessoas que não são e nem nunca foram pacientes psiquiátricos. Como atrativo, os participantes poderiam escolher uma instituição de caridade para a qual seriam doados 25 dólares em seu nome. Nenhum participante recebeu os resultados do teste. As aplicações do Rorschach ocorreram no momento e local conveniente para os participantes. No encontro para a realização do Rorschach, os examinadores descreveram novamente a pesquisa, pediram que os participantes assinassem um termo de consentimento e então aplicaram o teste. Eventualmente, os participantes dos estudos de estabilidade temporal de Exner também foram recrutados desta forma, e isto pode ter desempenhado algum papel na composição das amostras (indivíduos que se interessam em participar de uma pesquisa, tendo como benefício uma doação de dinheiro, em seu nome, para uma instituição de caridade) e na estabilidade temporal obtida.

Agora comentaremos os resultados obtidos no Rorschach nas duas testagens.

É importante notar que, tanto em nossa amostra, quanto na amostra normativa de adultos da cidade de São Paulo, os valores de *Lambda* foram maiores do que na amostra normativa norte-americana. Podemos formular quatro hipóteses sobre o significado da maior proporção de respostas de forma pura nos paulistanos: 1) o modo como o Rorschach foi aplicado resultou em protocolos com mais respostas simples; 2) os paulistanos têm dificuldade para explicar sua percepção, do que resulta uma proporção maior de respostas codificadas com o determinante forma pura; 3) os paulistanos apresentam um funcionamento psicológico simples; 4) os resultados de *Lambda* estão associados a aspectos da situação de testagem e ao modo como os participantes lidam com esta situação.

A primeira hipótese formulada postula que o modo como foram realizadas as aplicações do Rorschach resultou em escores mais elevados de *Lambda* nos paulistanos.

Lis et al. (2007) relatam a pesquisa normativa do SCR realizada na Itália. Os aplicadores foram estudantes de psicologia, que receberam treinamento intensivo na aplicação e codificação das respostas do SCR. Num primeiro momento da coleta de dados, os resultados obtidos apresentaram valores altos para *Lambda*. A equipe de pesquisadores, então, realizou

curso com a Prof^a. Andronikoff, rorschachista com ampla experiência com o SCR, sendo inclusive credenciada por Rorschach Workshops para oferecer cursos de treinamento. Após este novo treinamento, a equipe coletou novos protocolos, que apresentaram escores de *Lambda* semelhantes aos da amostra normativa norte-americana. Os autores comentam:

Estes resultados destacam a importância do treinamento estruturado no Sistema Compreensivo para além do estudo e discussão dos livros de Exner e do Workbook. Notamos que os examinadores apresentavam diferentes níveis de entendimento destas fontes de informação, incluindo um uso superficial das orientações, que não seguia as regras e o uso padronizado do Sistema Compreensivo. Também acreditamos que certas características metodológicas específicas, tais como a ativa cooperação do cliente, a natureza relacional dos encorajamentos [prompting] e outros aspectos referentes à natureza interativa da aplicação e do inquérito são particularmente difíceis de serem aplicados sem treinamento formal. (LIS et al., 2007, p. S198-S199).

Pode ser que as orientações para a aplicação do Rorschach, principalmente o que se refere ao inquérito das respostas, não venham sendo compreendidas e/ou aplicadas, no Brasil, de modo inteiramente correto para o Sistema Compreensivo. Haveria a necessidade de os usuários brasileiros realizarem treinamento e supervisão com instrutores mais diretamente ligados ao Sistema Compreensivo, credenciados por Rorschach Workshops.

A segunda hipótese levantada para explicar os escores mais altos de *Lambda* nos paulistanos foi a de dificuldade dos participantes em explicar sua percepção. Esta dificuldade pode ser uma decorrência do menor nível de instrução da nossa população, já que Nascimento (2004) encontrou que quanto menor o nível de instrução, maiores os valores de *Lambda*. Contudo, em nossa amostra, a maioria dos participantes possuía pelo menos o ensino médio completo, ou seja, para os nossos participantes não podemos atribuir os escores altos de *Lambda* ao menor nível de instrução. Uma dificuldade para explicar suas percepções tampouco parece plausível.

A terceira hipótese, de que os paulistanos apresentam um funcionamento psicológico simples (ou mais simples do que o dos norte-americanos) também não parece muito convincente para os participantes da nossa amostra, que foi composta por indivíduos predominantemente das classes A e B, com razoável nível de instrução e moradores de uma

cidade que oferece a complexidade de vida de uma grande metrópole – o que deve contribuir para uma maior complexidade psicológica.

Em nossa opinião, a quarta hipótese – os escores mais altos de *Lambda* estão associados à situação de testagem e ao modo como os participantes lidaram esta situação – seria a mais plausível. Discutiremos, então, aspectos relacionados à situação de testagem.

O reteste ter sido realizado pelo mesmo examinador que aplicou o teste é um fator a ser considerado. Nos estudos de Exner e no estudo de Sultan et al. (2006), as aplicações do Rorschach, no teste e no reteste, foram realizadas por examinadores diferentes. Provavelmente, utilizar examinadores diferentes no teste e no reteste traz maior variabilidade na interação examinador-testando, então a interferência da situação de testagem e da relação interpessoal nos resultados deve ser maior – o que contribuiria para diminuir a estabilidade temporal. Não temos como avaliar o papel que a utilização do mesmo examinador no teste e no reteste desempenhou nos resultados da nossa pesquisa. A impressão é que cada participante reagiu a esta situação de modo individual: talvez alguns participantes se sentiram mais à vontade no reteste, porque já conheciam o examinador; talvez outros participantes se sentiram menos à vontade com o examinador no reteste.

Pode ser que, no reteste, alguns participantes estivessem preocupados com o examinador já ter os dados da primeira testagem, eventualmente imaginando que ele tivesse identificado problemas no participante – devemos lembrar que o psicólogo frequentemente é considerado o profissional que trata de pessoas com problemas e que a situação de avaliação, particularmente com o Rorschach, pode despertar no testando temores de ser invadido e julgado (CHABERT, 1983). Ocampo e Arzeno (1987, p. 317-318) descrevem como muitos indivíduos que realizam um processo psicodiagnóstico se sentem em relação à avaliação, principalmente quando não há devolução dos resultados:

[...] o psicólogo assumirá, para ele [indivíduo avaliado] desde o princípio, o papel de uma figura extremamente ameaçadora, pelo depósito maciço de tudo o que é persecutório que alberga em seu mundo interno. Ocupar-se-á, quase que exclusivamente, de controlá-lo, de mantê-lo à distância e de evitá-lo. Isto pode ser registrado como [...] bloqueios totais ou reiterados antes as pranchas que lhe são apresentadas [...]. Se a informação não é devolvida, intensificam-se as fantasias de doença, gravidade, incurabilidade, loucura, etc.

Por se tratar de uma situação de pesquisa – em vez de uma situação de psicodiagnóstico, que costuma ocorrer quando há alguma dificuldade que demanda a avaliação – provavelmente o impacto destas preocupações foi menor. De qualquer modo, pode ter desempenhado um papel na atitude de menor envolvimento no reteste, para alguns participantes.

Os estudos de Exner, Armbruster e Mitman (1978) e Leura e Exner (1978, apud Exner, 2003)⁶³ revelaram que os testandos ficaram mais à vontade e forneceram mais respostas quando testados por examinadores conhecidos. Devemos observar que os participantes destes estudos haviam tido uma convivência freqüente com os examinadores, que eram professores que testaram seus alunos ou terapeutas que testaram seus pacientes. Em nosso estudo, os participantes conheciam o examinador apenas por terem realizado com ele o teste. O número de respostas no teste e no reteste foi semelhante, o que indica que conhecer o examinador, por ter realizado com ele o teste, não teve o efeito de aumentar o número de respostas no reteste.

Seis participantes conheciam o examinador antes da aplicação do teste. Não tinham convivência assídua como a de pacientes com seus terapeutas ou de alunos com seus professores, mas sim a convivência de pessoas que se conheciam socialmente. Os resultados dos participantes que conheciam o examinador foram variados: um participante forneceu poucas respostas e predomínio de respostas simples, o que pode indicar que se defendeu frente à testagem; outros participantes forneceram poucas respostas, porém bastante elaboradas – ou seja, houve envolvimento com a testagem, mas não houve um aumento do número de respostas; um participante forneceu um número elevado de respostas, no teste e no reteste – neste caso, conhecer o examinador pode ter contribuído para aumentar o número de respostas.

Exner (2007) comparou os resultados obtidos por diferentes examinadores e encontrou diferenças estatisticamente significantes para $F\%$, Zf e $X\%$. Sultan et al. (2006) também encontraram diferenças nos resultados obtidos por diferentes examinadores: alguns

⁶³ LEURA, A. V.; EXNER, J. E. **Structural differences in the records of adolescents as a function of being tested by one's own teacher**. Workshops Study No. 265. Estudo não publicado. Rorschach Workshops, 1978.

examinadores associaram-se a menores números de respostas, menor engajamento na tarefa e valores menores de *EA*. Os valores mais altos de *Lambda* em nossa amostra podem ter sido decorrência do examinador que realizou as testagens, no caso, o autor desta pesquisa. Fatores relacionados às características físicas e psicológicas do examinador podem ter desempenhado algum papel, assim como a situação de testagem considerada em um contexto mais amplo.

Comentando a maior proporção de respostas de forma pura nos indivíduos de países de cultura latina, Vinet (2000, p. 360) considera esse resultado decorrente do modo como os indivíduos lidam com a situação de testagem:

A situação de testagem é comumente uma situação nova e inesperada para o examinando; ela provoca uma grande quantidade de ansiedade e estresse; e não existem regras estruturadas para lidar com isto. O examinador é percebido como uma pessoa com autoridade e poder sobre o sujeito, que pertence a um outro grupo. Neste contexto, o indivíduo irá provavelmente se comportar de um modo *seguro*, caracterizado por respeito, obediência e conformismo, mas também por baixa cooperação e baixa confiança interpessoal. Nestas circunstâncias, é fácil compreender os protocolos com valores altos de *Lambda* indicando defesa [...] Além disso, o sujeito não manifestará abertamente sentimentos de mal-estar porque a situação solicita comportamento correto e *simpatia* [...] O sujeito estará atento às necessidades do examinador, evitando deixar-lhe em uma situação difícil (grifos do autor).

Esta caracterização do modo de lidar com a testagem ajuda a compreender os resultados obtidos na pesquisa normativa com adultos da cidade de São Paulo (Nascimento, 2002) e também os resultados da nossa pesquisa. Nesta, a questão da autoridade do examinador pode ter desempenhado um papel ainda maior.

Muitos participantes souberam que este estudo estava sendo realizado na universidade porque havia sido aberta esta possibilidade ao pesquisador por uma das diretoras (e proprietária) da universidade, com quem ele mantinha convívio próximo. Assim sendo, é provável que o pesquisador (e examinador) tenha sido considerado alguém associado a uma figura de poder do local de trabalho dos participantes, e isto pode ter desempenhado o papel de fazer com que os participantes apresentassem um bom número de respostas, mas um predomínio de respostas de forma pura – respostas mais simples, que indicam um modo mais defendido de lidar com a tarefa.

Apesar de o procedimento de aplicação do SCR ter sido planejado de modo a diminuir a interferência do examinador nos resultados, parece que, ainda assim, fatores da situação de testagem podem, algumas vezes, desempenhar importante papel nos resultados obtidos. Isto parece ter ocorrido na nossa pesquisa.

Qualquer que seja o motivo para a maior proporção de forma pura encontrada em nossa amostra, o mais importante é notar que houve menor engajamento na tarefa do que nas pesquisas originais do SCR. Levantamos a hipótese de que a menor estabilidade temporal encontrada em nossa amostra está associada ao menor engajamento na tarefa.

Os resultados de estabilidade temporal encontrados na nossa pesquisa foram, de modo geral, menores do que os das pesquisas de Exner e do que a estimativa obtida com a metanálise dos estudos sobre estabilidade temporal do Rorschach. Devemos considerar que continuam prevalecendo os resultados das outras pesquisas, ou seja, é mais sensato continuar afirmando que o Rorschach é um instrumento que avalia características estáveis da personalidade de modo consistente, e tentar compreender os resultados (estabilidade temporal menor) obtidos em nossa amostra – como fizemos. A explicação mais simples para a menor estabilidade temporal encontrada em nossa amostra é considerar que ela ocorreu devido a, por mero acaso, termos obtido esta amostra em particular.

Entendemos que a menor fidedignidade encontrada em nossa amostra não deve, de modo algum, levar ao abandono da utilização do Rorschach como instrumento psicométrico, privilegiando-se uma avaliação mais qualitativa dos protocolos. No passado, frente a alguns resultados baixos em estudos sobre a fidedignidade do Rorschach, alguns autores afirmaram que este aspecto não seria importante, uma vez que a interpretação dependeria não dos escores isolados, mas sim da configuração total do protocolo. Frente a afirmações como estas, Jensen (1959, p. 111-112) é definitivo:

[O Rorschach ser interpretado em termos da configuração total do protocolo ...] de modo algum diminui a necessidade de fidedignidade; uma configuração é na verdade um conjunto de diferenças entre varias quantidades ou qualidades classificáveis. A fidedignidade destas quantidades ou classificações, então, obviamente afeta a fidedignidade da configuração total da qual são uma parte.

Um aspecto importante a ser destacado é a interferência do engajamento na tarefa na estabilidade temporal e, em sentido mais amplo, na própria validade do SCR. Conforme comentado por Meyer (1997a), menor engajamento na tarefa resulta em protocolos que muitas vezes não representam corretamente ou de modo mais completo a personalidade dos indivíduos avaliados. O modo como o Rorschach é aplicado, incluindo o modo como o teste é apresentado e a relação interpessoal entre o examinador e o testando, está diretamente ligado ao nível de engajamento na tarefa por parte do testando. Mais estudos são necessários sobre esta questão.

Recentemente, vêm sendo estudadas modificações no procedimento de aplicação do Rorschach. Dean et al. (2007) utilizaram um procedimento modificado de aplicação, segundo o qual, toda vez que o testando fornece apenas uma resposta e pretende devolver o cartão, é estimulado a fornecer mais respostas; e toda vez que fornece três respostas a um cartão, este cartão é recolhido. Este procedimento visa obter um número de respostas que não seja nem muito baixo, nem muito alto. Dean et al. (2007) compararam os resultados obtidos com a aplicação do Rorschach segundo as instruções originais do Sistema Compreensivo versus os resultados obtidos com esta aplicação modificada. Encontraram que os protocolos obtidos com o procedimento modificado apresentaram maior validade na identificação de transtornos do pensamento do que os protocolos obtidos com o procedimento original de aplicação. Sugeriram que mais pesquisas sejam realizadas para se compreender como procedimentos modificados de aplicação afetam a validade dos resultados. Seria interessante realizar pesquisas deste tipo com brasileiros, principalmente buscando encontrar fatores que promovam maior engajamento na tarefa, pois parece que níveis maiores de engajamento na tarefa geram protocolos mais completos, fidedignos e válidos.

Formulamos as hipóteses sobre a interferência da situação de testagem nos resultados de paulistanos e sobre o menor engajamento na tarefa contribuírem para uma menor estabilidade temporal. Se estas hipóteses forem verdadeiras, isso demandaria ainda mais cautela ao interpretar os resultados obtidos com o SCR, principalmente em protocolos com baixo engajamento na tarefa (identificados por *R* baixo e *Lambda* alto).

Cabe, por fim, enfatizar que os resultados em nossa amostra, apesar de não apresentarem níveis de estabilidade tão altos quantos os das pesquisas originais do SCR,

indicaram níveis moderados de estabilidade temporal para a maioria das variáveis, o que é positivo para reafirmar a qualidade do Sistema Compreensivo do Rorschach.

12. CONCLUSÕES

- Encontramos fidedignidade inter-codificadores substancial ($\kappa > 0,60$) para a maioria das variáveis do SCR em nossa pesquisa.
- A média das correlações de teste-reteste encontradas em nossa amostra foi $r = 0,61$, o que corresponde a um nível moderado de estabilidade temporal.
- A estabilidade temporal encontrada nesta pesquisa foi menor do que a reportada nos estudos originais do SCR e do que a estimativa obtida a partir da metanálise sobre estabilidade temporal do Rorschach, mas semelhante ou maior do que a encontrada em um estudo francês recente.
- Em nossa amostra, houve um nível de engajamento na tarefa semelhante ao da amostra normativa da cidade de São Paulo, mas menor do que nas pesquisas norte-americanas. Isto pode ter contribuído para a menor estabilidade temporal encontrada.
- A situação de testagem aparentemente desempenhou importante papel nos resultados.
- Os usuários brasileiros devem ter cuidado ao interpretar os resultados obtidos com o Sistema Compreensivo do Rorschach, principalmente frente aos protocolos com baixo nível de engajamento na tarefa (identificados de modo mais direto por R baixo e Λ alto).
- Mais pesquisas brasileiras com o SCR são necessárias, inclusive para investigar procedimentos que promovam maior engajamento na tarefa, assim contribuindo para maiores fidedignidade e validade dos resultados obtidos na testagem.

13. REFERÊNCIAS

ACKLIN, M. W.; MCDOWELL, C. J., 2nd; VERSCHELL, M. S.; CHAN, D. Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. **Journal of Personality Assessment**, v. 74, n. 1, p. 15-47, 2000.

ADAIR, H. E.; WAGNER, E. E. Stability of unusual verbalizations on the Rorschach for outpatients with schizophrenia. **Journal of Clinical Psychology**, v. 48, p. 250-256. 1992.

AMERICAN PSYCHOLOGICAL ASSOCIATION. **Standards for Educational and Psychological Tests**. Washington: American Psychological Association, 1974.

ANASTASI, A.; URBINA, S. **Testagem psicológica**. Porto Alegre, RS: ArtMed, 2000.

ANTÚNEZ, A. E. A. **Estudo de características afetivas em pacientes diagnosticados com transtorno obsessivo compulsivo e voluntários normais por meio do exame de Rorschach**. Tese de doutorado. Universidade Federal de São Paulo, 2004.

ARKES, H. R. Impediments to accurate clinical judgment and possible ways to minimize their impact. **Journal of Consulting and Clinical Psychology**, v. 49, n. 3, p. 323-330, 1981.

ARONOW, E.; RESNIKOFF, M.; MORELAND, K. The Rorschach: projective technique or psychometric test? **Journal of Personality Assessment**, v. 64, n. 2, p. 213-228, 1995.

ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA. **Critério de Classificação Econômica Brasil**. ABEP, 2003.

BARTKO, J. J. The intraclass correlation coefficient as a measure of reliability. **Psychological Reports**, v. 19, p. 3-11, 1966.

BOCCHI, E. A. **Características psicológicas de pacientes com queixa de disfunção da articulação temporomandibular**. Tese de mestrado. Universidade Federal de São Paulo, 1998.

BRUNONI, G. R. **Características de personalidade de portadores de transtorno invasivo do desenvolvimento (TID): um estudo pelo método de Rorschach**. Trabalho de conclusão de curso. Pontifícia Universidade Católica de São Paulo, 2006.

CAMPGANA, V. N. **Aspectos da organização da identidade feminina no início da adolescência**. Dissertação de mestrado. Universidade de São Paulo, 2003.

CASTRO, P. F. Histórico da produção científica do Sistema Compreensivo no Brasil a partir dos eventos da ASBRo. In: **Métodos projetivos: instrumentos atuais para investigação psicológica e da cultura**. SILVA NETO, N. A.; AMPARO, D. M. (organizadores). Brasília: ASBRo, 2006. p. 114-123.

CHABERT, C. **Le Rorschach em clinique adulte: interpretation psychanalytique**. Paris : Dunod, 1983.

COHEN, J. A power primer. **Psychological Bulletin**, v. 112, n.1, 155-159, 1992.

CONSELHO FEDERAL DE PSICOLOGIA. **O Conselho avalia a qualidade dos testes psicológicos**. Disponível em <http://www.pol.org.br>. Acessado em 04/11/2007.

CRONBACH, L. J. Statistical methods applied to Rorschach scores: a review. **Psychological Bulletin**, v. 46, p. 393-429, 1949.

CRONBACH, L. J. **Fundamentos da testagem psicológica**. Porto Alegre: Artes Medicas, 1996.

DANCEY, C. P.; REIDY, J. **Estatística sem matemática para Psicologia usando SPSS para Windows**. Porto Alegre: ArtMed, 2006.

DEAN, K. L.; VIGLIONE, D. J.; PERRY, W.; MEYER, G. A method to optimize the response range while maintaining Rorschach Comprehensive System Validity. **Journal of Personality Assessment**, v. 89, n. 2, p. 149-161, 2007.

DINNOFF, M. Subject awareness of examiner influence in a testing situation. **Journal of Consulting Psychology**, v. 24, p. 465, 1960.

DUARTE, C. S. **Características de personalidade de mães de crianças com diagnóstico de autismo infantil: um estudo comparativo**. Tese de doutorado. Universidade Federal de São Paulo, 2000.

DUNLAP, W. P.; CHEN, R.; GREER, T. Skew reduces test-retest reliability. **Journal of Applied Psychology**, v. 79, p. 310-313, 1994.

DURAT Jr., C. **O Rorschach na avaliação psicológica de adolescentes em conflito com a lei**. Dissertação de mestrado. Universidade São Francisco. Itatiba, 2006.

ERSTAD, D. **An investigation of older adults' less frequent human movement and color responses on the Rorschach**. Tese de doutorado. Marquette University, 1996.

EXNER, J. E. Jr. **The Rorschach Systems**. New York: Grune & Straton, 1969.

EXNER, J. E. Jr. **The Rorschach: a Comprehensive System**. New York: John Willey & Sons, 1974.

EXNER, J. E. Jr. **The Rorschach: a Comprehensive System. Volume 2: Interpretation**. New York: John Willey & Sons, 1978.

EXNER, J. E. Jr. But it's only an inkblot. **Journal of Personality Assessment**, v. 44, n. 6, p. 563-577, 1980.

EXNER, J. E. Jr. Rorschach assessment. In: **Clinical methods in psychology**. WEINER, I. (editor). New York: Wiley, 1983. p. 58-99.

EXNER, J. E. Jr. **The Rorschach: a Comprehensive System. Vol.1: Basic Foundations**. 2nd. ed. New York: John Willey & Sons, 1986.

EXNER, J. E. Jr. Problems with brief Rorschach protocols. **Journal of Personality Assessment**, v. 52, n. 4, p. 640-647, 1988.

EXNER, J. E. Jr. R in Rorschach research: a ghost revisited. **Journal of Personality Assessment**, v. 58, n.2, p. 245-251, 1992.

EXNER, J. E. Jr. **The Rorschach: a Comprehensive System. Volume 1: Basic Foundations**, 3rd. ed. New York: John Wiley & Sons, 1993.

EXNER, J. E. Jr. Critical bits and the Rorschach response process. **Journal of Personality Assessment**, v. 53, p. 520-536, 1996a.

EXNER, J. E. Jr. A comment on “the Comprehensive System for the Rorschach: A critical examination”. **Psychological Science**, v. 7, p. 11-13, 1996b.

EXNER, J. E. Jr. The Rorschach: Measurement concepts and issues of validity. In: **The new rules of measurement: What every psychologist and educator should know**.

EMBRETSON, S. E.; HERSHBERGER, S. L. (editores). Mahwah, NJ: Lawrence Erlbaum, 1999a. p. 159-183.

EXNER, J. E. Jr. **Manual de Classificação do Rorschach para o Sistema Compreensivo**. São Paulo: Casa do Psicólogo, 1999b.

EXNER, J. E. Jr. **The Rorschach: A Comprehensive System. Volume 1: Basic Foundations and Principles of Interpretation**. 4th ed. Hoboken, NJ: Wiley, 2003.

EXNER, J. E. Jr. A new U.S. adult nonpatients sample. **Journal of Personality Assessment**, v. 89, n. S1, p. S154-S158, 2007.

EXNER, J. E. Jr.; ARMBRUSTER, G.; MITTMAN, B. The Rorschach response process. **Journal of Personality Assessment**, v. 42, n. 1, p. 27-38, 1978.

EXNER, J. E. Jr.; ARMBRUSTER, G. L.; VIGLIONE, D. The temporal stability of some Rorschach features. **Journal of Personality Assessment**, v. 42, n. 5, p. 474-482, 1978.

EXNER, J. E. Jr.; THOMAS, E. A.; MASON, B. J. Children's Rorschachs: description and prediction. **Journal of Personality Assessment**, v. 49, p. 13-20, 1985.

EXNER, J. E. Jr.; WEINER, I. B. **The Rorschach: A Comprehensive System. Volume 3: Assessment of Children and Adolescents**. New York: Wiley, 1982.

EXNER, J. E. Jr.; WEINER, I. B. **The Rorschach: A comprehensive System. Volume 3: Assessment of children and adolescents.** 2nd. ed. New York: Wiley, 1995.

FACHEL, J. M. G.; CAMEY, S. Avaliação psicométrica: a qualidade das medidas e o entendimento dos dados. In: **Psicodiagnóstico V.** CUNHA, J. A. (editora). Porto Alegre, RS: ArtMed, 2000. p. 158-170.

FARAH, F. H. Z. **O transtorno do pânico no método de Rorschach e na escala de ansiedade de Beck.** Dissertação de mestrado. Universidade São Francisco, 2004.

FERREIRA, A. B. H. **Novo dicionário básico da língua portuguesa Folha/Aurélio.** Rio de Janeiro: Nova Fronteira, 1988.

FISKE, D. W.; BAUGHMAN, E. E. Relationships between Rorschach scoring categories and the total number of responses. **The Journal of Abnormal and Social Psychology**, v. 48, n.1, p. 25-32, 1953.

FRANK, L. K. Projective methods for the study of personality. In: **Handbook of projective techniques.** MURSTEIN, B. I. (editor). New York: Basic Books, 1965. [Publicado originalmente no Journal of Psychology, v. 8, p. 389-413, 1939.]

GARB, H. N.; WOOD, J. M.; NEZWORSKI, M. T.; GROVE, W. M.; STEJSKAL, W. J. Toward a resolution of the Rorschach controversy. **Psychological Assessment**, v. 13, n. 4, p. 433-448, 2001.

GAZIRE, P. **Estudo descritivo de aspectos da personalidade de pacientes com síndrome do intestino irritável.** Dissertação de mestrado. Universidade Federal de São Paulo, 2000.

GODINHO, S. M. **Estudo do manejo do estresse em pacientes acometidos por alopecia areata.** Dissertação de mestrado. Universidade Federal de São Paulo, 2003.

GRØNNERØD, C. Rorschach interrater agreement estimates: an empirical evaluation. **Scandinavian Journal of Psychology**, v. 40, p. 115-120, 1999.

GRØNNERØD, C. Temporal Stability in the Rorschach Method: A Meta-Analytic Review. **Journal of Personality Assessment**, v. 80, n. 3, p. 272-293, 2003a.

GRØNNERØD, C. Rorschach assessment of changes following psychotherapy: A meta-analytic review. **Journal of Personality Assessment**, v. 83, n. 3, p. 2565-276, 2003b.

GRØNNERØD, C. Reanalysis of the Grønnerød (2003) Rorschach temporal stability meta-analysis data set. **Journal of Personality Assessment**, v. 86, n. 2, p. 222-225, 2006.

GROSS, L. R. Effects of verbal and nonverbal reinforcement on the Rorschach. **Journal of Consulting Psychology**, v. 23, p. 66-68, 1959.

GUARNACCIA, V.; DILL, C. A.; SABATINO, S.; SOUTHWICK, S. Scoring accuracy using the comprehensive system for the Rorschach. **Journal of Personality Assessment**, v. 77, n. 3, p. 464-474, 2001.

GUNTERT, A. E. V. A. **Crianças com nódulo vocal: estudo da personalidade por meio da prova de Rorschach**. Tese de doutorado. Universidade Federal de São Paulo, 1996.

HALLER, N.; EXNER, J. E. Jr. The reliability of Rorschach variables for inpatients presenting symptoms of depression and/or helplessness. **Journal of Personality Assessment**, v. 49, n. 5, p. 516-521, 1985.

HERSEN, M.; GREAVES, S. T. Rorschach productivity as related to verbal reinforcement. **Journal of Personality Assessment**, v. 35, p. 436-441, 1971.

HILSENROTH, M.; CHARNAS, J. **Training manual for Rorschach interrater reliability**. The Derner Institute of Advanced Psychological Studies. Adelphi University, Garden City, NY: 2006. (Manuscrito não publicado).

HILSENROTH, M. J.; CHARNAS, J. W.; ZOLDAN, J. Criterion-based training for Rorschach scoring. **Training and Education in Professional Psychology**, v. 1, n. 2, p. 125-134, 2007.

HISATUGO, C. L. C. **Estudo do índice depressivo, DEPI, do Rorschach, Sistema Compreensivo, em amostra de pacientes depressivos atendidos em ambulatório público de São Paulo**. Tese de doutorado. Universidade Federal de São Paulo, 2007.

HOLZBERG, J. Reliability re-examined. In: **Rorschach Psychology**. M. A. RICKERS-OVSIANKINA (editora). New York: Wiley, 1977.

HOUAISS, A.; CARDIM I. **Novo dicionário Folha-Webster's Inglês/Português, Português/Inglês**. São Paulo: Publifolha, 1996.

JANSON, H. **Rorschach Research Utilities version 1.0.0 user's guide**. Harald Janson, 2003. (Manual do programa de computador)

JANSON, H.; OLSSON, U. A measure of agreement for interval or nominal multivariate observations. **Educational and Psychological Measurement**, v. 61, n. 2, p. 277-289, 2001.

JANSON, H.; OLSSON, U. A measure of agreement for interval or nominal multivariate observations by different sets of judges. **Educational and Psychological Measurement**, v. 64, n. 1, p. 62-70, 2004.

JENSEN, A. R. The reliability of projective techniques: Review of the literature. **Acta Psychologica Amsterdam**, v. 16, p. 108-136, 1959.

KINDER, B. N. The problems of R in clinical settings and in research: suggestion for the future. **Journal of Personality Assessment**, v. 58, n.2, p. 252-259, 1992.

KUBISZYN, T.; BORICH, G. **Educational testing and measurement**. Hoboken, NJ: John Wiley & Sons, 2003.

LAMOUNIER, R. **O Rorschach e a avaliação de motoristas: evidencia de validade**. Dissertação de mestrado. Universidade São Francisco, 2005.

LEVIN, J. **Estatística aplica às ciências humanas**. São Paulo, SP: Harper & Row do Brasil, 1985.

LINDGREN, T.; CARLSSON, A. M. Relationship between the Rorschach and the MMPI-2 in a Swedish population: A replication study of the effects of first-factor related test-interaction styles. **Journal of Personality Assessment**, v. 79, n. 2, p. 357-370, 2002.

LIPGAR, R. M. The problem of R in the Rorschach: the value of varying responses. **Journal of Personality Assessment**, v. 58, n.2, p. 223-230, 1992.

LIS, A.; PAROLIN, L.; CALVO, V.; ZENNARO, A.; MEYER, G. The impact of administration and inquiry on Rorschach Comprehensive System Protocols in a National Reference Sample. **Journal of Personality Assessment**, v. 89, n. S1, p. S193-S2000, 2007.

MAGNUSSEN, M. G. Verbal and nonverbal reinforcers in the Rorschach situation. **Journal of Clinical Psychology**, v. 16, p. 167-169, 1960.

MASLING, J. The influence of situational and interpersonal variables in projective testing. **Psychological Bulletin**, v. 57, n. 1, p. 65-85, 1960.

MASLING, J. Differential indoctrination of examiners and Rorschach responses. **Journal of Consulting Psychology**, v. 29, n..3 , p. 198-201, 1965.

MCDOWELL, C.; ACKLIN, M. W. Standardizing procedures for calculating Rorschach interrater reliability: conceptual and empirical foundations. **Journal of Personality Assessment**, v. 66, n. 2, p. 308-320, 1996.

MCGRATH, R. E.; POGGE, D. L.; STROKES, J. M. Field reliability of Comprehensive System Scoring in an adolescent inpatient sample. **Assessment**, v. 45, n. 5, p. 199-206, 2005.

MCGUIRE, H.; KINDER, B. N.; CURTISS, G.; VIGLIONE, D. J. Some special issues in data analysis. In: **Issues and methods in Rorschach research**. EXNER, J. E. Jr. (editor). Mahwah, NJ: Lawrence Erlbaum Associates, 1995. p. 227-250.

MEYER, G. J. The Rorschach factor structure: a contemporary investigation and historical review. **Journal of Personality Assessment**, v. 59, p. 117-136, 1992a.

MEYER, G. J. Response frequency problems in the Rorschach: clinical and research implications with suggestions for the future. **Journal of Personality Assessment**, v. 58, n.2, p. 231-244, 1992b.

MEYER, G. J. The impact of response frequency on the Rorschach Constellation Indices and on their validity with diagnostic and MMPI-2 criteria. **Journal of Personality Assessment**, v. 60, n. 1, p. 153-180, 1993.

MEYER, G. J. The Rorschach and MMPI: toward a more scientifically differentiated understanding of cross-method assessment. **Journal of Personality Assessment**, v. 67, n.3, p. 558-578, 1996.

MEYER, G. J. On the integration of personality assessment methods: the Rorschach and the MMPI. **Journal of Personality Assessment**, v. 68, n. 2, p. 297-330, 1997a.

MEYER, G. J. Assessing reliability: critical corrections for a critical examination of the Rorschach Comprehensive System. **Psychological Assessment**, v. 9, p. 480-489, 1997b.

MEYER, G. J. Error in research and assessment data with an erratum for Meyer (1993). **Journal of Personality Assessment**, v. 71, p. 195-211, 1998.

MEYER, G. J. The convergent validity of MMPI and Rorschach scales: an extension using profile scores to define response and character styles on both methods and a reexamination of simple Rorschach response frequency. **Journal of Personality Assessment**, v.72, n. 1, p. 1-35, 1999.

MEYER, G. J. The reliability and validity of the Rorschach and Thematic Apperception Test (TAT) compared to other psychological and medical procedures: An analysis of systematically gathered evidence. In: **Comprehensive Handbook of Personality Assessment**. HERSEN, M. (editor chefe). Volume 2: Personality Assessment. HILSENROTH, M.; SEGAL, D. (editores). Hoboken, NJ: John Wiley & Sons, 2004. p. 315-342.

MEYER, G. J. Advancing personality assessment terminology: Time to retire “objective” and “projective” as personality descriptors. **Journal of Personality Assessment**, v. 87, n. 3, p. 223-225, 2006.

MEYER, G. J. **Challenges in Rorschach-based Assessment**. Material do curso apresentado no III Congresso Brasileiro de Avaliação Psicológica e XII Conferência Internacional de Avaliação Psicológica, realizado em João Pessoa, em 2007.

MEYER, G. J.; ARCHER, R.P. The hard science of Rorschach research: what do we know and where do we go? **Psychological Assessment**, v. 13, n. 4, p. 486-502, 2001.

MEYER, G. J.; BAXTER, D.; EXNER, J. E. Jr.; FOWLER, J. C.; HILSENROTH, M. J.; PIERS, C. C.; RESNICK, J. An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. **Journal of Personality Assessment**, v. 78, n. 2, p. 219-274, 2002.

MEYER, G. J.; ERDBERG, P.; SHAFFER, T. Toward international normative reference data for the Comprehensive System. **Journal of Personality Assessment**, v. 89, n. 1, p. S201-S216, 2007.

MEYER, G. J.; MIHURA, J. L.; SMITH, B. L. The Interclinician Reliability of Rorschach Interpretation in Four Data Sets. **Journal of Personality Assessment**, v. 84, n. 3, p. 296-314, 2005.

NAKATA, L. M. **Interrater reliability and the Comprehensive System for the Rorschach: Clinical and non-clinical protocols**. Tese de doutorado. Pacific Graduate School of Psychology, 2000.

NASCIMENTO, R. S. G. F. **Bipolares: Um Estudo de Características de Personalidade através do Rorschach**. Tese de doutorado. Pontifícia Universidade Católica de São Paulo, 1993.

NASCIMENTO, R. S. G. F. Resultados de estudo normativo para o Sistema Compreensivo do Rorschach: um estudo para a cidade de São Paulo. **Psico-USF**, v. 7, n. 2, p. 127-141, 2002.

NASCIMENTO, R. S. G. F. The impact of educational and socioeconomic conditions on Rorschach data in a Brazilian nonpatient sample. **Rorschachiana**, v. 26, p. 45-62, 2004.

NASCIMENTO, R. S. G. F. Estudo normativo do Sistema Compreensivo do Rorschach para São Paulo: resultados dos índices PTI, SCZI, DEPI, CDI, HVI, OBS e S-CON. **Avaliação Psicológica**, 2006.

NASCIMENTO, R. S. G. F. Rorschach Comprehensive System data for a sample of 409 adults nonpatients from Brazil. **Journal of Personality Assessment**, v. 89, n. S1, p. S35-S41, 2007.

NASCIMENTO, R. S. G. F.; FREITAS, E. S.; MUCARBEL, F.; VELHO, L. M.; MOTOOKA, R. C. F. R.; INFANTE, T. C. Um estudo normativo do Rorschach para cidades do estado de São Paulo: resultados parciais. In: **Técnicas projetivas: produtividade em pesquisa**. VAZ, C. E.; GRAEFF, R. F. (editores). Porto Alegre: Casa do Psicólogo e SBRO, 2004.

NASCIMENTO, R. S. G. F.; HISATUGO, C. C.; MUCARBEL, F. C.; MATTOS, A. P.; MORI NETO, A.; CALVARO, P. E. S. Um estudo das médias de qualidade formal em um grupo de sujeitos não-pacientes de São Paulo, conforme o Sistema Compreensivo de Exner. In: **Livro dos Anais do II Congresso Nacional da Sociedade Brasileira de Rorschach e outros métodos projetivos** VAZ, C. E. (editor). Porto Alegre: AGE editora, 2000. p. 40-47.

NUNNALLY, J. C. **Psychometric Theory**. New York: McGraw-Hill, 1967.

OCAMPO, M. L. S.; ARZENO, M. E. Devolução de informação no processo psicodiagnóstico. In: **O processo psicodiagnóstico e as técnicas projetivas**. OCAMPO, M. L. S.; ARZENO, M. E. G.; Piccolo, E. G. e colaboradores. 5ª edição. São Paulo: Martins Fontes, 1987. p. 315-332.

PARKER, K. C. H.; HANSON, R. K.; HUNSLEY, J. MMPI, Rorschach and WAIS: a meta-analytic comparison of reliability, stability, and validity. **Psychological Bulletin**, v. 103. n. 3, p. 367-373, 1988.

PASQUALI, L. (organizador). **Técnicas de Exame Psicológico – TEP: manual**. São Paulo: Casa do Psicólogo/Conselho Federal de Psicologia, 2001.

PASSOS, T. C. B. M. **Psicodinâmica depressiva em obesos com transtorno de compulsão alimentar periódica: estudo com o método de Rorschach**. Dissertação de mestrado.

Universidade Federal de São Paulo, 2005

PIOTROWSKI, C. The status of Exner's Comprehensive System in contemporary research.

Perceptual and Motor Skills, v. 82, p. 1341-1342, 1996.

PIOTROWSKI, Z. A. **Perceptanalysis: The Rorschach method fundamentally reworked, expanded and systematized**. Philadelphia: Ex Libris, 1974.

ROCHA Jr., A. **A influência da religião na vida dos adolescentes universitários da grande São Paulo: um estudo à luz do método de Rorschach**. Tese de doutorado. Universidade Metodista de São Paulo, 2004.

RORSCHACH, H. **Psicodiagnóstico**. São Paulo: Mestre Jou, 1967 [1921].

ROSENTHAL, M. C. P. **Transtorno obsessivo-compulsivo e personalidade**. Dissertação de mestrado. Pontifícia Universidade Católica de São Paulo, 2000.

ROSENTHAL, M. C. P. **“Assim é se lhe parece” – Transtorno obsessivo-compulsivo: estudo de variáveis cognitivo-afetivas de personalidade por meio do método de Rorschach**. Tese de doutorado. Pontifícia Universidade Católica de São Paulo, 2007.

SAETHER, L.; BERG, M. **“Inkblots again?!” Stability of Rorschach scores: An artifact of memory?** Dissertação de mestrado. University of Tromso, Noruega, 1999.

SAHLY, J. M. **Rorschach interrater reliability as assessed by iota statistics on protocol-level Comprehensive System variables for an international sample: meeting the Daubert standard**. Tese de doutorado. Alliant International University, 2006.

SANTOANTONIO, J. **Estudo de características da personalidade de adolescentes com diagnóstico de lúpus eritematoso sistêmico por meio do método de Rorschach**.

Dissertação de doutorado. Universidade Federal de São Paulo, 2001.

SCHAFER, R. **Psychoanalytic Interpretation in Rorschach testing**. New York: Grune & Stratton, 1954.

SCHWARTZ, N. S.; MEBANE, D. L.; MALONY, H. N. Effects of alternate modes of administration on Rorschach performance of deaf adults. **Journal of Personality Assessment**, v. 54, n. 3&4, p. 671-683, 1990.

SEMER, N. L. **Estudo da auto-estima de crianças enuréticas pelo método de Rorschach**. Tese de doutorado. Universidade Federal de São Paulo, 1999.

SHROUT, P. E.; FLIES, J. L. Intraclass correlations: uses in assessing rater reliability. **Psychological Bulletin**, v. 86, n. 2, p. 420-428, 1979.

SILVA, S. M. **Habilidades e desempenho terapêutico em estudantes de psicologia: utilizando o método de Rorschach**. Dissertação de mestrado. Universidade São Francisco, 2005.

SILVA NETO, A. C. P. **Características de personalidade de jogadores patológicos avaliados pelo método de Rorschach**. Dissertação de mestrado. Universidade Federal de São Paulo, 1999.

SILVA NETO, A. C. P. Usando corretamente o Sistema Compreensivo do Rorschach em indivíduos brasileiros. **Técnicas projetivas: produtividade e pesquisa** (Trabalhos completos do III Congresso Nacional da Sociedade Brasileira de Rorschach e Métodos Projetivos). VAZ, C. E.; GRAEFF, R. F. (organizadores). Porto Alegre: Sociedade Brasileira de Rorschach (SBRo), 2004.

SIMKINS, L. Examiner reinforcement and situational variables in a projective testing situation. **Journal of Consulting Psychology**, v. 24, p. 541-547, 1960.

STATSOFT. **Eletronic Statistics Textbook**. Disponível em <http://www.statsoft.com/textbook/stathome.html>. Acessado em 05/11/2007.

SULTAN, S.; ANDRONIKOF, A.; RÉVEILLÈRE, C.; LEMMEL, G. A Rorschach stability study in a nonpatient adult sample. **Journal of Personality Assessment**, v. 87, n. 3, p. 330-348, 2006.

TAVARES, M. Validade Clínica. **Psico-USF**, v. 8, n. 2, p. 125-136, 2003.

TRAUB, R. E.; ROWLEY, G. L. **Understanding reliability. Instructional Topics in Educational Measurement.** National Council on Measurement in Education, 1991.

VIGLIONE, D. J. Basic considerations regarding data analysis. In: **Issues and methods in Rorschach Research.** EXNER, J. E. Jr. (editor). p. 195-226. Mahwah, NJ: Lawrence Erlbaum Associates, 1995.

VIGLIONE, D. W. **Rorschach coding solutions: A reference guide for the Comprehensive System.** Donald Viglione, 2002.

VIGLIONE, D. J.; HILSENROTH, M. J. The Rorschach: facts, fictions, and future. **Psychological Assessment**, v. 13, n. 4, p. 452-471, 2001.

VIGLIONE, D. J.; TAYLOR, N. Empirical support for interrater reliability of Rorschach Comprehensive System. **Journal of Clinical Psychology**, v. 59, n. 1, p. 111-121, 2003.

VINET, E. The Rorschach Comprehensive System in Iberoamerica. In: DANA, R. H. (editor) **Handbook of Cross-cultural and multicultural personality assessment.** Hillsdale: Laurence Erlbaum Associates, 2000.

VOIGT, W. H.; DANA, R. H. Inter- and Intra-Scorer Rorschach Reliability. **Journal of Projective Techniques and Personality Assessment**, v. 28, n.1, p. 92-95, 1964.

WAGNER, E. E.; DAUBNEY, J. H. Test-retest reliability of the Rorschach on neurologically impaired subjects. **Journal of Personality Assessment**, v. 40, n. 6, p. 579-581, 1976.

WATSON, D. Stability versus change, dependability versus error: Issues in the assessment of personality over time. **Journal of Research in Personality**. v. 38, p. 319-350, 2004.

WEINER, I. B. Editor's note: Interscorer agreement in Rorschach research. **Journal of Personality Assessment**, v. 56, n. 1, p. 1, 1991.

WEINER, I. B. Methodological considerations in Rorschach research. **Psychological Assessment**, v. 7, n.3, p. 330-337, 1995.

WEINER, I. B. Some observations on the validity of the Rorschach Inkblot Method. **Psychological Assessment**, v.8, n.2, p. 206-213, 1996.

WEINER, I. B. **Principles of Rorschach Interpretation**. Mahwah, NJ: Erlbaum, 1998a.

WEINER, I. B. Teaching the Rorschach Comprehensive System. In: **Teaching and Learning Personality Assessment**. HANDLER, L.; HILSENROTH, M. (eds.). Mahwah, NJ: Erlbaum, 1998b.

WICKES, T. H. Examiner influence in a testing situation. **Journal of Consulting Psychology**, v. 20, p. 23-26, 1956.

WILLIAMSON, E. L. Internal consistency of the Rorschach ink blot test: An analysis of Exner's comprehensive system using Cronbach's alpha. **Tese de doutorado**. Brigham Young University, 1988.

WOOD, J. M.; NEZWORSKI, M. T., W. J. Thinking critically about the Comprehensive System for the Rorschach. A comment on Meyer (1997). **Psychological Assessment**, v. 9, p. 490-494, 1997.

WOOD, J. M.; NEZWORSKI, M. T.; LILIENFELD, S. O.; GARB, H. N. **What's wrong with the Rorschach**. San Francisco: Jossey-Bass, 2003.

WOOD, J. M.; NEZWORSKI, M. T.; STEJSKAL, W. J. The Comprehensive System for the Rorschach: a critical examination. **Psychological Science**, v. 7, p. 3-10, 1996a.

WOOD, J. M.; NEZWORSKI, M. T.; STEJSKAL, W. J. Thinking critically about the Comprehensive System for the Rorschach. A reply to Exner. **Psychological Science**, v. 7, p. 14-17, 1996b.

ZUBIN, J.; ERON, L. D.; SCHUMER, F. **An experimental approach to projective techniques**. New York: Wiley, 1965.

APÊNDICE A – CORRELAÇÃO

A exposição que se segue foi desenvolvida a partir de Levin (1985), Anastasi e Urbina (2000), Kubiszyn e Borich (2003) e Meyer (2007).

Muitas vezes, os cientistas querem investigar relações entre diferentes aspectos da realidade. Por exemplo, em Psicologia pode-se investigar a relação entre inteligência e rendimento mensal em adultos. Primeiramente, o pesquisador precisaria definir o que entende como sendo inteligência e como esta pode ser avaliada (uma definição possível seria capacidade de abstração e de estabelecer relações entre idéias, e esta seria avaliada por um teste psicológico, tal como as Matrizes Progressivas de Raven); também deveria definir o rendimento mensal, por exemplo, como sendo aquele declarado pelo indivíduo. Este tipo de estudo busca então medir duas variáveis e depois analisar sua relação. Os dados deste exemplo são apresentados na Tabela A1.

Tabela A1 – Inteligência e rendimento mensal de uma amostra de adultos

Inteligência (QI)	Rendimento mensal (em reais)	Inteligência (QI)	Rendimento mensal (em reais)
72	1 000	100	2200
75	900	100	2000
79	1200	101	2200
80	2 000	102	2500
81	1500	103	3000
85	1600	105	2600
89	1500	107	2200
90	1800	108	2900
91	1600	110	2600
92	2000	112	3200
94	2000	114	3000
96	3000	118	3500
98	1800	121	2900
98	2300	130	3300
99	2100	132	3500

Podemos também apresentar estes dados em um gráfico, no qual os valores de inteligência encontram-se no eixo horizontal (eixo x) e os valores de rendimento mensal encontram-se no eixo vertical (eixo y). Cada ponto no gráfico corresponde a um indivíduo da amostra. Este tipo de gráfico é chamado de diagrama de dispersão de valores (scatter-plot) e encontra-se apresentado na Figura A1.

Figura A1: Inteligência e renda mensal numa amostra de 30 indivíduos adultos

Os dados apresentam uma tendência geral: à medida que aumenta a inteligência, aumenta o rendimento mensal. Toda vez que, à medida que uma variável aumenta, a outra também aumenta, temos uma relação positiva.

Um outro exemplo seria avaliar a relação entre agressividade e estima por parte dos colegas, em crianças. Neste caso, a agressividade seria medida contando-se o número de vezes que uma criança bate em seus colegas durante o período de recreio, numa semana, e o grau de estima por parte dos colegas seria medido pedindo-se que cada criança avaliasse, numa escala de zero a dez pontos, o quanto gosta de cada um de seus colegas – a média das notas dadas a cada uma das crianças por seus colegas corresponderia ao escore de estima por parte dos colegas. Um gráfico dessa relação é apresentado na Figura A2. Além dos pontos, que representam cada uma das crianças, o gráfico apresenta uma linha que mostra o padrão de relação entre agressividade e estima pelos colegas.

Figura A2: Agressividade e estima por parte dos colegas numa amostra de 30 crianças

Neste caso, observa-se como tendência geral: quanto maior a agressividade, menor a estima por parte dos colegas. Toda vez que, à medida que uma variável aumenta, a outra diminui, temos uma relação negativa.

Nestes dois exemplos, a relação entre as variáveis apresenta um sentido constante: à medida que uma variável aumenta, ou a outra também aumenta, ou diminui. Entretanto, podem ocorrer situações em que o sentido da relação entre as variáveis muda ao longo do intervalo de valores. Como exemplo, temos a relação entre ansiedade e resolução de problemas, apresentada na Figura A3. A ansiedade poderia ser avaliada por um conjunto de medidas fisiológicas, tais como frequência cardíaca e condutividade da pele; a resolução de problemas poderia ser avaliada por meio de testes com tarefas de resolução de problemas (por exemplo, uma prova de execução de uma escala Wechsler de inteligência).

Figura A3: Ansiedade e resolução de problemas

Conforme vemos na parte mais à esquerda da figura, pessoas com baixa ansiedade (provavelmente pouco “preocupadas” com seu desempenho) apresentam baixa resolução de problemas; à medida que a ansiedade aumenta, aumentam os escores de resolução de problemas (pode-se dizer então que a ansiedade fornece certa “energia” para resolver problemas); contudo, a partir de certo ponto, a ansiedade aumenta e diminuem os escores de resolução de problemas, conforme vemos na parte mais à direita da figura (podemos dizer que a ansiedade torna-se excessiva, passando então a prejudicar a resolução de problemas). Então, para níveis baixos de ansiedade (parte esquerda da figura), a relação entre ansiedade e resolução problemas é positiva; para níveis altos de ansiedade (parte direita da figura), passa a ser uma relação negativa. Neste exemplo, explicamos a resolução de problemas como decorrência da ansiedade, estabelecendo então uma relação causal. No entanto, é importante sublinhar que a *correlação* mostra apenas uma relação entre variáveis, não permitindo determinar causalidade. Explicamos a resolução de problemas como decorrência da ansiedade, mas poderíamos também explicar a ansiedade como decorrência da resolução de problemas: existem dois tipos de pessoas com baixa resolução de problemas: as que experimentam baixa ansiedade (estão acostumadas com suas dificuldades para resolver problemas, então não sofrem) e as que experimentam alta ansiedade (sofrem com seu fracasso em resolver problemas); à medida que a resolução de problemas se aproxima de um nível médio, a ansiedade também se aproxima de um nível médio.

Quando a relação entre as variáveis segue apenas um sentido (positivo ou negativo), a relação é monotônica (por exemplo, as relações representadas nas Figuras A1 e A2); porém, quando a relação muda de sentido ao longo da faixa de valores, dizemos que a relação é não-monotônica.

Existem ainda outros termos para descrever as relações entre variáveis (ou os formatos das curvas nos gráficos). Algumas relações são lineares, quando o aumento em uma variável associa-se a um aumento proporcional na outra variável (por exemplo, a relação entre QI e renda, apresentada na Figura A1 é linear, com a forma da curva que representa a tendência geral sendo, então, uma linha reta); outras relações são não-lineares (por exemplo, a relação entre agressividade e estima por parte dos colegas, apresentada na Figura A2 – neste caso a relação é curvilínea).

Ao analisar a relação entre duas variáveis, primeiramente o pesquisador quer saber se existe um padrão discernível de relação. As várias curvas desenhadas nos gráficos apresentados justamente indicam alguns padrões de relação. Pode-se então descrever relações monotônicas (positivas ou negativas) ou não-monotônicas, lineares ou não-lineares, ou até mesmo encontrar variáveis que não apresentam um padrão claro de relação entre si, não sendo possível traçar uma curva que represente a tendência geral. A Figura A4 mostra uma situação em que não encontramos uma tendência geral.

Figura A4: Exemplo de gráfico de dispersão de valores, no qual não se identifica uma relação entre as variáveis

A Figura A5 é um diagrama de dispersão de valores que apresenta a relação entre tempo dedicado ao estudo e grau de retenção das informações. O tempo dedicado ao estudo corresponderia ao número de minutos que o aluno estudou durante uma semana e o grau de retenção das informações seria avaliado por uma prova com questões referentes ao conteúdo estudado. Nesta figura, desenhamos também a reta que representa a tendência geral.

A curva que representa a tendência geral apresenta então pontos em torno de si. Se todos os pontos caíssem exatamente sobre a curva, esta representaria perfeitamente a relação entre as variáveis. Contudo, o mais comum é encontrarmos pontos em torno da curva.

Figura A5: Tempo dedicado ao estudo e retenção de informações

A forma da curva já indica o modo como as variáveis se relacionam, a tendência geral. Determina também o sentido da relação (quando esta é monotônica). Contudo, o pesquisador costuma querer conhecer também a “força” da relação entre as variáveis, o quanto os valores de uma variável mudam em função da outra. Em linguagem técnica: ele quer saber o grau de associação entre as variáveis. Quanto mais os pontos encontram-se próximos da curva, mais “forte” é a relação entre as variáveis; quanto mais os pontos estão afastados da curva, menos “forte” é a relação entre as variáveis.

Em muitos casos, o modelo linear parece ser adequado para descrever a relação entre duas variáveis – isto é determinado observando-se o diagrama de dispersão, que deve mostrar pontos alinhados de tal modo que uma reta expresse relativamente bem a tendência geral da relação entre as variáveis.

A Figura A6 retoma a Figura 1 apresentada anteriormente, que mostra dados de inteligência e rendimento mensal. Pode-se traçar a curva que representa a tendência geral da relação, obtendo-se uma linha reta, chamada de linha de regressão. Como a tendência geral pode ser bem representada por uma linha reta, isto indica que a relação entre inteligência e rendimento mensal é linear.

Figura A6: Relação entre inteligência e rendimento mensal

No caso da relação entre inteligência e rendimento mensal, os pontos se distribuem um pouco “afastados” da linha de regressão. Se a relação entre inteligência e rendimento fosse perfeitamente linear, todos os pontos estariam exatamente sobre a linha de regressão. (Não encontramos em psicologia relações perfeitamente lineares, mas sim relações que se aproximam disso.)

No caso de relações lineares, os pontos do diagrama de dispersão encontram-se em torno da linha de regressão formando uma elipse. Quanto mais “fechada” a elipse (pontos mais próximos à linha de regressão) maior a associação entre as variáveis (Figura A7); quanto mais “aberta” a elipse (pontos mais afastados da linha de regressão) menor a associação entre as variáveis (Figura A8).

Figura A7: Exemplo de associação forte entre variáveis

Figura A8: Exemplo de associação fraca entre variáveis

Quanto mais os pontos estão próximos da linha de regressão, mais forte é a associação entre as variáveis; quanto mais os pontos estão afastados da linha de regressão, mais fraca é a associação entre as variáveis. Exemplificamos isto, para o caso de relações lineares, com a imagem das elipses. Contudo, seria conveniente medir o grau de associação, em vez de usar termos como “pontos mais próximos / pontos mais afastados” ou “elipse mais aberta / elipse mais fechada”. Apresentamos os exemplos de forma visual para facilitar a compreensão do leitor, mas existem também medidas para representar a “força” da relação entre as variáveis, com a conveniência de fornecerem uma quantificação do grau de associação. Justamente, o *coeficiente de correlação* descreve a associação entre duas variáveis, informando sobre o sentido da relação e o grau de associação entre as variáveis.

A medida mais utilizada de correlação é o *coeficiente de correlação produto-momento de Pearson*, representado por r . Este pressupõe que a relação entre as variáveis seja linear e que as variáveis sejam de nível intervalar.

Os níveis de mensuração são:

- 1) Nominal, quando as variáveis são categorias de classificação apenas, sem que se possa ordená-las em termos de valor. Por exemplo o sexo dos participantes de uma pesquisa é uma variável nominal, porque os participantes podem ser classificados em termos de serem do sexo masculino ou do sexo feminino;
- 2) Ordinal, quando se pode estabelecer uma ordem do tipo menor/maior. Por exemplo, a classe sócio-econômica é uma variável ordinal, pois os indivíduos

podem ser classificados em termos de pertencerem às classes A, B, C, D ou E, e estas classes podem ser colocadas numa ordem da classe mais baixa (E) à classe mais alta (A);

- 3) Intervalar, quando se pode estabelecer uma ordem do tipo menor/maior e se conhecer a distância entre os valores. Por exemplo, a renda mensal é uma variável intervalar. Se Pedro tem renda mensal de R\$ 1 000, João tem renda mensal de 2 000 e Maria tem renda mensal de R\$ 2 500, podemos dizer que a renda de João é duas vezes maior do que a de Pedro e que a renda de Maria é maior do que a de João em 500 reais.

O *coeficiente de correlação* apresenta um sinal (+ ou -) que indica o sentido da relação entre as variáveis (sentido positivo ou sentido negativo, respectivamente) e um valor, que indica o grau de associação entre as variáveis. O *coeficiente de correlação* varia de -1 a $+1$. O valor -1 indica uma correlação negativa perfeita, ou seja, todos os pontos caem sobre a linha de regressão e, à medida que uma variável aumenta de valor, a outra diminui de valor. O valor 0 indica ausência de relação entre as variáveis. O valor $+1$ indica uma correlação positiva perfeita, ou seja, todos os pontos caem sobre a linha de regressão e, à medida que uma variável aumenta de valor, a outra também aumenta de valor. Valores entre -1 e 0 indicam correlações parciais negativas e valores entre 0 e $+1$ indicam correlações parciais positivas (Figura A9).

Figura A9: valores do coeficiente de correlação e seu significado interpretativo

A maioria das correlações entre variáveis psicológicas apresenta valores bem menores do que os de uma correlação perfeita. Assim, os parâmetros para a avaliação do grau de associação entre as variáveis precisam ser estabelecidos de acordo com as características

próprias do campo de estudos – no caso, a Psicologia. Segundo Cohen (1992) correlações com valor de 0,10 indicam um grau de associação pequeno, mas não necessariamente trivial; correlações com valor de 0,30 indicam grau médio de associação, e correlações com valor de 0,50 indicam grau elevado de associação⁶⁴.

Uma vez determinado o coeficiente de correlação na amostra, permanece a questão de se saber se ele representa o resultado da população. Devemos lembrar que, partindo da população, podemos selecionar diferentes amostras e cada uma delas apresentará determinado coeficiente de correlação, com valor mais próximo ou mais distante do valor da correlação na população.

Devemos sempre lembrar que a correlação obtida na amostra é uma aproximação, mais próxima ou menos próxima, da correlação da população. Uma abordagem para se investigar a generalização dos resultados da amostra para a população consiste no estabelecimento de *intervalos de confiança*, ou seja, a partir do valor encontrado na amostra calcula-se a faixa de valores em que o valor na população mais provavelmente se situa.

Como exemplo dos intervalos de confiança, vamos comparar os resultados de SumC' em uma amostra de Exner e seus colaboradores e em nossa amostra:

Exner, Armbruster e Viglione (1978) encontraram, em sua amostra de 100 adultos não-pacientes retestados após 3 anos, para SumC', $r = 0,67$. Adotando um intervalo de confiança de 95%, encontramos que a correlação na população situa-se entre 0,55 e 0,80. Isto significa que, podemos afirmar, com 95% de chance de estarmos corretos, que SumC' apresenta correlação entre 0,55 e 0,80 na população.

Em nossa amostra de 32 participantes retestados após 3 meses encontramos, para SumC', $r = 0,56$. Adotando um intervalo de confiança de 95%, encontramos que a correlação na população situa-se entre 0,26 e 0,76.

⁶⁴ Este valores se aplicariam aos exemplos citados anteriormente, mas numa pesquisa sobre estabilidade temporal, espera-se que os resultados nas testagens sejam bastante semelhantes, resultando em correlações altas, iguais ou maiores do que 0,75.

Neste exemplo, vemos que, apesar de os coeficientes para as amostras serem diferentes (0,67 versus 0,56), os intervalos de confiança se sobrepõem, o que significa que, para esta variável, não podemos dizer que as correlações sejam diferentes para as populações.

Em vários casos (relação não-linear entre as variáveis, amostras de tamanho pequeno, nível de mensuração ordinal ou nominal em vez de intervalar) não se deve utilizar o *coeficiente de correlação de Pearson*, mas sim outros coeficientes (não-paramétricos). Para dados ordinais, uma das estatísticas mais utilizadas é o *coeficiente de correlação de ordem de postos de Spearman*, que informa sobre o grau em que as ordens dos valores de duas variáveis estão associadas.

Como um exemplo de utilização da correlação de postos de Spearman, teríamos uma pesquisa sobre a relação entre o estado de humor dos coletores de laranja e sua produção no trabalho. O estado de humor foi investigado pedindo-se que cada trabalhador assinalasse uma das seguintes categorias: 1- muito triste, 2- triste, 3- neutro, 4- alegre, 5- muito alegre. A produção no trabalho corresponde ao número de laranjas colhidas em um dia de trabalho de 6 horas, dado que também foi dividido em categorias: 1- até 299 laranjas, 2- 300 a 599 laranjas, 3- 600 a 999 laranjas, 4- 1 000 laranjas ou mais. É importante notar que o nível de mensuração destas variáveis é ordinal; os números apenas indicam uma ordem entre os diferentes valores, mas um valor igual a 4 não é duas vezes maior do que um valor igual a 2.

Os resultados encontrados são apresentados na Tabela A2. Os dados foram organizados em ordem crescente (da variável “humor”) para facilitar a visualização das relações. Notamos que, quanto mais triste o humor, menor a produção. A correlação de postos de Spearman para a amostra foi de 0,75.

Tabela A2: estado de humor relatado e produção de 30 coletores de laranja

Humor	Produção	Humor	Produção
1	1	3	2
1	2	3	3
1	2	3	3
2	3	4	3
2	3	4	4
2	2	4	4
2	3	4	4
3	2	4	4
3	3	4	4
3	2	4	4
3	2	4	4
3	3	5	4
3	3	5	3
3	3	5	4
3	4	5	4

Estado de humor: 1- muito triste, 2- triste, 3- neutro, 4- alegre, 5- muito alegre

Produção: 1- até 299 laranjas. 2- 300 a 599 laranjas. 3- 600 a 999 laranjas, 4- 1 000 laranjas ou mais.

APÊNDICE B – CÓDIGOS E VARIÁVEIS DO SCR

[Suprimido na versão online]

APÊNDICE C – PORCENTAGEM DE CONCORDÂNCIA E KAPPA COMO MEDIDAS DA FIDEDIGNIDADE INTER-CODIFICADORES

A porcentagem de concordância, usada como medida de fidedignidade inter-codificadores, não leva em conta a concordância obtida ao acaso.

Por exemplo, vamos analisar a concordância de dois avaliadores para um conjunto de respostas. As Tabelas C1 e C2 mostram duas possíveis situações de concordância entre estes dois avaliadores para a codificação movimento inanimado [*m*]:

Tabela C1 – Concordância entre dois codificadores

Codificador 2	Codificador 1		Total
	*	m	
*	85	5	90
m	5	5	10
Total	90	10	100

* indica decisão de não codificar m

Tabela C2 – Concordância entre dois codificadores

Codificador 2	Codificador 1		total
	*	m	
*	55	5	60
m	5	35	40
total	60	40	100

* indica decisão de não codificar m

Nas duas situações, a porcentagem de concordância é de 90%. Contudo, no primeiro caso, os avaliadores concordaram que 5 respostas apresentam o código, dentre as 15 para os quais o código em questão foi assinalado por ao menos um dos avaliadores, enquanto que, no segundo caso, os avaliadores concordaram que 35 respostas apresentam o código, dentre as 45 para os quais o código foi assinalado por ao menos um dos codificadores. Assim, a concordância para quando o código foi assinalado é maior no segundo caso do que no primeiro.

Se não existir concordância real entre os codificadores, as frequências esperadas para as células da tabela serão determinadas pelo acaso. Temos então, apresentadas nas Tabelas C3 e C4 as seguintes expectativas, calculadas pelo mesmo método utilizado para o cálculo do qui-quadrado (X^2):

Tabela C3 – Concordância esperada ao acaso

Codificador 2	Codificador 1		total
	*	m	
*	81	9	90
m	9	1	10
total	90	10	100

* indica decisão de não codificar m

Tabela C4 – Concordância esperada ao acaso

Codificador 2	Codificador 1		total
	*	m	
*	36	24	60
m	24	16	40
total	60	40	100

* indica decisão de não codificar m

Para a primeira situação, a porcentagem de concordância esperada no caso de não haver concordância real é de 82%; na segunda situação, é de 52%.

Comparando os valores obtidos com os valores que seriam esperados por mero acaso, notamos que, na primeira situação, a concordância obtida apresenta uma diferença pequena em relação ao que seria esperado por acaso (90% versus 82%), enquanto que a diferença é grande para a segunda situação (90% versus 52%). Assim, podemos afirmar com maior segurança que os valores obtidos na segunda situação expressam uma concordância que realmente existe entre os codificadores, em vez de eventualmente ter sido obtida por acaso.

O coeficiente kappa foi introduzido justamente para superar o problema da concordância devida ao acaso. Na fórmula do *kappa* temos a porcentagem de concordância observada (percentage of agreement – PA) e a porcentagem obtida ao acaso (percentage of chance – PC). A Tabela C5 mostra a matriz utilizada para o cálculo do kappa.

Tabela C5 – Exemplificação do cálculo do kappa

Codificador 2	Codificador 1		total
	*	código	
*	a	b	a + b
código	c	d	c + d
total	a + c	b + d	n

* indica decisão de não codificar o código

$$PA = \frac{(a + d)}{n}$$

$$PC = \frac{(a + c) * (a + b) + (b + d) * (c + d)}{n^2}$$

O coeficiente kappa é definido como a diferença entre a concordância observada e a concordância obtida ao acaso, expresso como uma fração da diferença máxima. Como o valor máximo para a concordância é igual a 1,00 (100%), temos:

$$k = \frac{PA - PC}{1 - PC}$$

Para os dados da Tabela C1, teríamos:

$$PA = 0,900 \quad (90\%)$$

$$PC = 0,820 \quad (82\%)$$

$$Kappa = (0,900 - 0,820)/(1,000 - 0,820) = 0,08/0,18 = 0,444$$

Para os dados da Tabela C2, teríamos:

$$PA = 0,900$$

$$PC = 0,520$$

$$Kappa = (0,900 - 0,520)/(1,000 - 0,520) = 0,380/0,480 = 0,792$$

Kappa varia de -1,00 a +1,00, sendo que valores maiores do que zero indicam que a concordância obtida é maior do que a esperada por mero acaso; kappa igual à zero indica que a concordância obtida é igual à esperada por mero acaso; valores de kappa menores do que zero indicam que a concordância obtida é menor do que a concordância esperada por mero acaso.

Na literatura, encontramos diversos parâmetros para a interpretação dos valores de kappa, propostos por diferentes autores e reunidos por Sahly (2006):

Landis e Kock (1977)⁶⁵ fornecem os seguintes parâmetros:

- Kappa < 0,00: concordância pobre (poor agreement);
- 0,00 a 0,20: concordância fraca (slight agreement);
- 0,21 a 0,40: concordância regular (fair agreement);
- 0,41 a 0,60: concordância moderada (moderate agreement);
- 0,61 a 0,80: concordância substancial (substantial agreement);
- 0,81 a 1,00: concordância quase perfeita (almost perfect agreement).

Segundo Fleiss (Fleiss, 1981; Fleiss e Cohen, 1973; Shrout e Fleiss, 1979)⁶⁶:

- Kappa < 0,40 = concordância pobre (poor)
- 0,40 – 0,59 = concordância regular (fair)
- 0,60 – 0,74 = concordância boa (good);
- Kappa > 0,74 = concordância excelente (excellent).

⁶⁵ Landis, J. R.; Kock, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159-174, 1977.

⁶⁶ Fleiss, J. **Statistical methods for rates and proportions**. 2nd ed.. New York: Wiley, 1981.

Fleiss, J.; Cohen, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. **Educational and Psychological Measurement**, v. 33, p. 613-619, 1973.

Shrout, P.; Fleiss, J. Intraclass correlations: Uses in assessing rater reliability. **Psychological Bulletin**, v. 86, p. 420-428, 1979.

Segundo Cicchetti, (1981, 1994)⁶⁷:

Kappa < 0,40 = concordância pobre (poor);

0,40 – 0,59 = concordância regular (fair);

0,60 – 0,74/0,79 = concordância boa (good);

Kappa > 0,75/0,79 = concordância excelente (excellent)

Kappa > 0,80 = quase perfeito (nearly perfect).

Outro elemento a ser informado é a taxa de ocorrência do código. Na primeira situação exemplificada, o código foi assinalado 10 vezes por cada um dos codificadores, apresentando então uma taxa de ocorrência igual a 20 em 200 (10%); no segundo exemplo, o código foi assinalado 40 vezes por cada um dos examinadores, apresentando então uma taxa de ocorrência igual a 80 em 200 (40%).

Tabela C6 – Exemplificação do cálculo da taxa de ocorrência

Codificador 2	Codificador 1		total
	*	código	
*	a	b	a + b
código	c	d	c + d
total	a + c	b + d	n

* indica decisão de não codificar o código

Taxa de ocorrência =

=
$$\frac{\text{ocorrência do código para o codificador 1} + \text{ocorrência do código para o codificador 2}}{\text{total de respostas codificadas pelos dois codificadores}}$$

ou seja, taxa de ocorrência =
$$\frac{(b + d) + (c + d)}{n + n}$$

⁶⁷ Cicchetti, D. V. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptative behavior. **American Journal of Mental Deficiency**, v. 86, p. 127-137, 1981.

Cicchetti, D. V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. **Psychological Assessment**, v. 6, p. 284-290, 1994.

APÊNDICE D – DIFERENTES MODOS DE AVALIAR A FIDEDIGNIDADE INTER-CODIFICADORES DO SCR

Podemos avaliar a fidedignidade das codificações ao nível das respostas. Neste nível de análise, podemos analisar a concordância para decisões de codificação dicotômicas, decisões de codificação policotômicas e por segmentos da codificação (ACKLIN et al., 2000).

Uma decisão dicotômica de codificação envolve atribuir ou não um código para uma resposta. Um exemplo de decisão dicotômica é codificar *S* ou não codificar *S*. (Esta decisão de codificação também corresponde à fidedignidade do código *S*.) A Tabela D1 apresenta uma situação hipotética em que se avalia a concordância para esta decisão de codificação.

Tabela D1 – Exemplo do cálculo da concordância para a decisão de codificar ou não codificar *S*

Codificador B	Codificador A		Total
	*	<i>S</i>	
*	78	2	80
<i>S</i>	3	17	20
total	81	19	100

* indica decisão de não codificar *S*

Para as 100 respostas que foram codificadas pelos dois avaliadores, eles concordaram em não codificar *S* para 78 respostas e em codificar *S* para 17 respostas – portanto a percentagem de concordância foi de 95 respostas dentre as 100 codificadas, ou seja, 0,95 ou 95%. *S* ocorreu 19 vezes para o codificador A e 20 vezes para o codificador B – portanto *S* ocorreu 39 vezes dentre as 200 codificações realizadas pelos dois avaliadores, ou seja, a taxa de ocorrência (Base Rate – BR) foi de $39/200 = 0,195$ ou 19,5%. Kappa é calculado como $(PA-PC)/(1-PC)$, onde PA é a percentagem de concordância (Percentage of Agreement) e PC é a percentagem de concordância que seria obtida por mero acaso (Percentage of Chance). Para os dados do exemplo, teríamos $PA = 0,950$, $PC = 0,686$ e $kappa = 0,841$.

Uma decisão policotômica de codificação envolve mais do que duas possibilidades de codificação. Um exemplo de decisão policotômica refere-se à decisão de codificar FC, CF, C, Cn ou não codificar o determinante cor cromática.. A Tabela D2 apresenta uma situação hipotética em que se avalia a concordância para esta decisão de codificação.

Tabela D2 – exemplo de cálculo do kappa para decisões policotômicas

codificador 2	codification 1					Total
	*	FC	CF	C	Cn	
*	70	2	1	1	0	74
FC	3	7	4	0	0	14
CF	1	2	5	2	0	10
C	0	0	1	1	0	2
Cn	0	0	0	0	0	0
total	74	11	11	4	0	100

Para o cálculo de decisões de codificação policotômicas, a fórmulas seria:

$$PA = \frac{\text{número de ocorrência de concordâncias}}{\text{número total de respostas}}$$

$$PC = \frac{\text{soma dos produtos das proporções marginais dos codificadores para cada categoria de codificação}}{n^2}$$

$$Kappa = \frac{PA - PC}{1 - PC}$$

No exemplo da Tabela D2:

$$PA = 0,830$$

$$PC = (74*74) + (11*14) + (11*10) + (4*2) + (0*0) / 100^2 = (5476 + 154 + 110 + 8 + 0) / 10000 = 5748 / 10000 = 0,5748$$

$$Kappa = (0,830 - 0,5748) / (1,000 - 0,5748) = 0,2552 / 0,4252 = 0,600$$

Os segmentos de codificação são os seguintes: Localização, Qualidade Evolutiva, Determinantes, Qualidade Formal, Pares, Conteúdos, Populares, Notas Z e Códigos Especiais. A concordância também pode ser calculada por segmento de codificação. Por exemplo, podemos avaliar a concordância para o segmento da Qualidade Formal (Tabela D3).

Tabela D3 – Exemplo do cálculo da concordância das codificações para o segmento da qualidade formal

codificador 2	codification 1					Total
	FQ+	FQo	FQu	FQ-	FQnone	
FQ+	0	0	0	0	0	0
FQo	1	58	1	1	0	61
FQu	0	2	18	1	0	21
FQ-	0	1	1	14	0	16
FQnone	0	0	1	0	1	1
Total	1	61	21	16	1	100

PA = 0,79

PC = 0,44

Kappa = 0,63

Podemos também avaliar a concordância ao nível do protocolo, ou seja, para as variáveis do Sumário Estrutural. Como exemplo, vamos avaliar a concordância para a variável *WSumC*, utilizando 2 avaliadores que codificaram as respostas de 10 protocolos. A Tabela D4 apresenta os escores de *WSumC* para cada protocolo e para cada um dos avaliadores.

Tabela D4 – Escores de WSumC para 10 protocolos codificados por 2 avaliadores

Protocolo	Codificador A	Codificador B
1	5,5	4,5
2	0,5	1,0
3	4,0	6,0
4	3,0	2,5
5	7,0	6,5
6	5,0	3,5
7	2,0	2,0
8	3,5	3,0
9	2,0	2,5
10	1,5	4,0

Para a avaliação da concordância ao nível do protocolo utiliza-se o Coeficiente de Correlação Intraclass (Intraclass Correlation Coefficient - ICC). O ICC baseia-se nos componentes da análise de variância (BARTKO, 1966). A interpretação do ICC é semelhante à do kappa (ACKLIN et al., 2000).

Podemos também trabalhar com as categorias interpretativas. Por exemplo, a variável *es* possui três faixas de valores: valores de acordo com o esperado (*es* de 7 a 11), abaixo do esperado (*es* < 6) e acima do esperado (*es* > 11). A Tabela D5 apresenta a avaliação da concordância entre dois codificadores para a variável *es* considerada em termos de categorias interpretativas.

Tabela D5 – Concordância entre dois codificadores para *es* analisado de acordo com as categorias interpretativas

Codificador B	Codificador A			Total
	<i>es</i> < 6	$7 \leq \textit{es} < 11$	<i>es</i> > 11	
<i>es</i> < 6	2	1	0	3
$7 < \textit{es} < 11$	1	11	1	13
<i>es</i> > 11	0	2	2	4
Total	3	14	3	20

APÊNDICE E – COEFICIENTE IOTA

Janson e Olsson (2001, p. 279) retomam definições importantes.

Tal como na literatura convencional (e.g. Shrout & Fleiss, 1979)⁶⁸, observadores ou avaliadores serão denominados juízes, e os objetos avaliados (pessoas, casos, respostas, eventos etc.) serão denominados alvos.

O Coeficiente iota, como medida da fidedignidade inter-codificadores, é mais diretamente apresentado em Janson e Olsson (2004, p. 63):

$$\iota = 1 - \frac{d_o}{d_e}$$

onde d_o , a discordância observada, representa a média, ao longo dos alvos e pares de juízes, das distâncias das pontuações entre os juízes para os mesmos alvos; e d_e , a discordância esperada, representa a média, ao longo dos alvos e pares de juízes, das distâncias das pontuações entre os juízes para quaisquer alvos. Com dados intervalares, a distância multivariada é definida como a distância Euclidiana, e com dados nominais, a distância multivariada é definida como o número de discordâncias ao longo das variáveis.

Para dados nominais (por exemplo, as decisões de codificação do Rorschach), iota é equivalente ao kappa. Para dois juízes e dados intervalares (por exemplo, a concordância do Rorschach ao nível do protocolo), iota é semelhante ao Coeficiente de Correlação Intraclass [ICC(2,1)], exceto por uma diferença pequena resultante do uso de $n-1$ em vez de n , para os graus de liberdade (JANSON; OLSSON, 2001).

Tal como inicialmente formulado (Janson e Olsson, 2001), iota seria um coeficiente a ser utilizado quando todos os diferentes juízes codificam todos os alvos. Este é o modelo de dois fatores (two way). Posteriormente, Janson e Olsson (2004), ampliaram iota para utilização com diferentes juízes, sem a necessidade de todos codificarem todos os alvos. Este é o modelo de um fator (one way).

⁶⁸ Shrout, P. E.; Fleiss, J. L. Intraclass correlations: Uses in assessing rater reliability. **Psychological Bulletin**, v. 86, p. 420-428, 1979.

Para a avaliação da fidedignidade inter-codificadores ao nível do protocolo, *iota* com modelo de dois fatores é semelhante ao ICC (2,1); *iota* com modelo de um fator não é semelhante a nenhuma outra medida (JANSON, 2003).

APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE E INFORMADO

Prezado Sr.(a),

Estamos realizando uma pesquisa sobre a estabilidade temporal dos resultados do método de Rorschach.

O método de Rorschach consiste em um conjunto de manchas de tinta, que a pessoa deve observar e dizer com o que se parecem. A partir das respostas fornecidas, podemos conhecer algumas características da personalidade, ou seja, algumas características do modo como a pessoa experimenta e lida com as situações.

Em nossa pesquisa estaremos realizando duas aplicações do método de Rorschach. Deste modo, veremos o quanto os resultados permanecem os mesmos ou variam em diferentes momentos. Este tipo de estudo é chamado de estudo da estabilidade temporal dos resultados.

Temos o compromisso ético de manter sigilo sobre suas informações pessoais, ou seja, os resultados da entrevista e da avaliação pelo método de Rorschach serão utilizados na pesquisa, mas você não será identificado.

Se for de seu interesse, os resultados da avaliação do método de Rorschach lhe serão comunicados. Para tanto, agendaremos uma entrevista, após a realização das duas aplicações do método de Rorschach para a presente pesquisa.

A participação na pesquisa é voluntária e você tem o direito de desistir a qualquer momento, sem ter que se justificar – basta comunicar ao pesquisador sua decisão de interromper a participação. A desistência não afetará de modo algum seu vínculo com a instituição em que trabalha.

Você pode pedir esclarecimentos sobre a pesquisa a qualquer momento, bastando entrar em contato com o pesquisador responsável, o psicólogo Antonio Carlos Pacheco e Silva Neto, no telefone 3842-4811.

Autorizo a utilização de meus resultados na pesquisa sobre estabilidade temporal do método de Rorschach

Nome: _____

R. G.: _____ Data: _____ Telefone : _____

Assinatura: _____

APÊNDICE G – ANÁLISE DAS DISCORDÂNCIAS NAS CODIFICAÇÕES DAS RESPOSTAS POPULARES

[Suprimido na versão online]

**APÊNDICE H – COMPARAÇÃO DOS RESULTADOS NO TESTE E NO
RETESTE DE 6 PROTOCOLOS COM MENOS DE 14 RESPOSTAS**

[Suprimido na versão online]