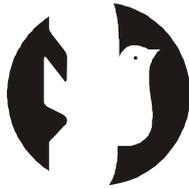


FÁBIO CAMILO DA SILVA



UNIVERSIDADE
SÃO FRANCISCO

EVIDÊNCIAS DE VALIDADE PELO PROCESSO DE RESPOSTA PARA O TESTE DE
RACIOCÍNIO INFERENCIAL (R_{In})

ORIENTADOR: PROF. DR. RICARDO PRIMI

ITATIBA - SP
2011

FÁBIO CAMILO DA SILVA



EVIDÊNCIAS DE VALIDADE PELO PROCESSO DE RESPOSTA PARA O TESTE DE
RACIOCÍNIO INFERENCIAL (RIn)

Dissertação apresentada ao Programa de
Pós-Graduação *Stricto Sensu* em Psicologia
da Universidade São Francisco para
obtenção do título de Mestre em Psicologia

ORIENTADOR: PROF. DR. RICARDO PRIMI

ITATIBA - SP
2011

153.94 Silva, Fabio Camilo da.
S58e Evidências de validade pelo processo de resposta para o
Teste de Raciocínio Inferencial (RIn). / Fabio Camilo da
Silva. -- Itatiba, 2011.
86 p.

Dissertação (mestrado) – Programa de Pós-Graduação
Stricto Sensu em Psicologia da Universidade São
Francisco.

Orientação de: Ricardo Primi.

1. Processo de resposta. 2. Teste de Raciocínio Inferencial
(RIn). 3. Fator g. I. Primi, Ricardo. II. Título.

Ficha catalográfica elaborada pelas bibliotecárias do Setor de
Processamento Técnico da Universidade São Francisco.

UNIVERSIDADE SÃO FRANCISCO

PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM PSICOLOGIA
MESTRADO

EVIDÊNCIAS DE VALIDADE PELO PROCESSO DE RESPOSTA PARA O TESTE DE
RACIOCÍNIO INFERENCIAL (RIn)

Autor: Fábio Camilo da Silva
Orientador: Prof. Dr. Ricardo Primi

Este exemplar corresponde à redação final da dissertação de mestrado defendida por Fábio Camilo da Silva e aprovada pela comissão examinadora.

Data: 13 / 12 / 2011



Prof. Dr. Ricardo Primi (Presidente)
Universidade São Francisco



Prof. Dr. Lucas de Francisco Carvalho
Universidade Presbiteriana Mackenzie



Prof. Dr. Fabián Javier Marín Rueda
Universidade São Francisco

ITATIBA - SP
2011

DEDICATÓRIA

À minha mãe Ivana que, ao me ensinar a ler, me abriu as portas para o mundo e à minha sobrinha Isabela que, ao nascer, me fez descobrir um novo mundo.

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus, N'aquele em que acredito, confio e que me abriu todas as portas necessárias para chegar até aqui.

Aos meus pais, Ivana e Camilo, que me deram sempre o melhor que puderam dentro de suas condições e à minha irmã Juliana, meu cunhado Cristiano e minha sobrinha Belinha, cuja existência dá sentido à minha vida.

Ao Dr. José Glauco Bardella, suas filhas Márcia e Mônica. A confiança e crença que depositaram em mim, além de toda ajuda material e profissional, foram cruciais para a concretização desse sonho.

Ao Prof. Dr. Fermino Fernandes Sisto, com quem tive momentos que variaram do amor ao ódio, às vezes concomitantemente, que muito me ensinou e a quem devo uma nova visão da vida acadêmica.

Ao Prof. Dr. Ricardo Primi, por me “adotar” enquanto orientador e contribuir com sua experiência e conhecimento.

À Prof. Dra. Ana Paula Porto Noronha por seu apoio e colaboração, tanto em sala de aula quanto como coordenadora do curso.

A todos os professores que participaram da minha formação. Em especial ao professor Dr. Armando Rocha Júnior, a quem devo meu ingresso na área de pesquisa ainda

na graduação. Aos professores Acácia, Claudette e Fabian, que sempre que foi preciso me apoiaram dentro e fora da sala de aula.

A todos com quem tive a oportunidade de compartilhar as aulas do mestrado e tornaram esses momentos inesquecíveis. Em especial à amiga Áurea Vianna por seu carinho, amizade e companheirismo.

Aos amigos da Vetor Editora que sempre acreditaram em mim, me deram força, incentivo, oraram e se mostraram dispostos a me ajudar dentro do que era possível a cada um. Dentre eles, Cristiano Esteves, Rodrigo Fonseca, Luciano Franzim, Tabata Cardoso, Cristiane Raymundo e Thais Fonseca.

À eterna amiga, irmã, espírito de luz, Isabel Piza.

A todas as pessoas que posso ter omitido, não propositalmente, neste agradecimento, mas que de alguma forma fizeram parte da minha vida nesses dois últimos anos.

RESUMO

Silva, F.C. (2011). *Evidências de validade pelo processo de resposta para o Teste de Raciocínio Inferencial (RIn)*. Dissertação de mestrado, Programa de Pós-Graduação *Strictu Sensu*, Universidade São Francisco: Itatiba.

O presente estudo teve por objetivo buscar evidências de validade pelo processo de resposta para o Teste de Raciocínio Inferencial (RIn), que avalia a inteligência ou fator *g* de Spearman. Para isso, todos os itens do teste foram decompostos, possibilitando a identificação de seus componentes (número de estímulos, agrupamentos, tamanho das matrizes, dentre outros) e das regras necessárias para respondê-los corretamente. Utilizou-se como fonte um banco de dados contendo os resultados de 503 pessoas que foram submetidas ao teste durante um processo de avaliação psicológica, das quais 78,9% eram do sexo masculino e 63,6% tinham Ensino Superior. A idade média dos participantes foi de 29,42 anos (mediana de 28 e moda de 27 anos) e o desvio-padrão de 8,596. O modelo criado pelo autor da pesquisa foi analisado por dois juízes e as correlações encontradas (mínima de $r = 0,640$ e máxima de $r = 0,994$) confirmaram sua adequação aos objetivos propostos. O índice de dificuldade, cuja correlação entre os dados da presente amostra e do manual do teste foi alta e significativa ($r = 0,926$; $p < 0,01$), foi a variável dependente a partir da qual todas as correlações foram efetuadas. Os componentes que evidenciaram uma correlação moderada com a dificuldade dos itens foram a quantidade de estímulos, seja no problema ($r = 0,491$; $p < 0,01$) ou na resposta ($r = 0,529$; $p < 0,01$), assim como seus respectivos agrupamentos ($r = 0,542$; $p < 0,01$ e $r = 0,440$; $p < 0,01$). Também foram encontradas correlações entre a dificuldade com as inferências horizontais ($r = 0,570$; $p < 0,01$) e com a quantidade de inferências possíveis ($r = -0,415$; $p < 0,01$). Uma análise de regressão indicou que o modelo composto pelas regras Identidade Horizontal, Identidade Vertical, Sequencia Horizontal e Quantidade de estímulos agrupados na resposta explicaria 69,9% da variância da dificuldade. Com exceção desta última, as demais contribuiriam para tornar os itens mais fáceis. A identificação destes elementos ou regras permite não apenas prever a dificuldade do item, como também pode servir de base para a criação de novos instrumentos similares. Sugere-se a realização de novos estudos que possam confirmar a influência das regras e componentes aqui encontrados sobre a dificuldade dos itens de outros instrumentos.

Palavras-chave: Teste de Raciocínio Inferencial (RIn); processo de resposta; inteligência; fator *g*; análise de regressão.

ABSTRACT

Silva, F. C. (2011). Validity Evidences based on response processes for the Teste de Raciocínio Inferencial (RIn). *Master's Degree Dissertation*, Program of Post-Graduation Strict Sense in Psychology, São Francisco University: Itatiba.

The aim of this study was to find validity evidences based on response processes for the Inferential Reasoning Test (RIn), which assesses intelligence or Spearman's g factor. For this, all the test items were decomposed allowing the identification of its components (number of stimuli, groups, size of the matrices, among others) and the necessary rules to answer them correctly. A source used was a database containing the tests results of 503 people who were submitted during a psychological evaluation process, of which 78,9% were male and 63,6% had higher education. The participants average age was 29,42 years (median 28 and mode 27 years) and the standard deviation of 8,596. The model created by the author of the study was analyzed by two judges and the correlations (among 0,640 and 0,994) confirmed its adequacy to the proposed objectives. The difficulty index, whose correlation between this data and manual test sample was high and significant ($r = 0,926$, $p < 0,01$), was the dependent variable from which all correlations were performed. The components that showed a moderate correlation with the difficulty of the items were: a number of stimuli, whether the problem ($r = 0,491$, $p < 0,01$) or response ($r = 0,529$, $p < 0,01$) as well as their respective groups ($r = 0,542$, $p < 0,01$ and $r = 0,440$, $p < 0,01$). It have also been found correlations between the difficulty and the horizontal inferences ($r = 0,570$, $p < 0,01$) and the number of possible inferences ($r = -0,415$, $p < 0,01$). A regression analysis indicated that the model composed by the rules: Horizontal Identity, Vertical Identity, Horizontal Sequence and Number of Grouped Stimuli in responding explain 69,9% of the variance of the difficulty. With exception of that last one, the others would help to make items easier. The identification of these elements or rules allows not only predict the difficulty of the item, but also can serve as a basis for creating new similar instruments. It is suggested that further studies may be conducted to confirm the influence of rules and components found on the difficulty of the items from other instruments.

Keywords: Inferential Reasoning Test (RIn); response process, intelligence, g factor, the regression analysis.

SUMÁRIO

DEDICATÓRIA.....	iv
AGRADECIMENTO	v
RESUMO	vii
ABSTRACT	viii
LISTA DE TABELAS	x
LISTA DE FIGURAS	xii
APRESENTAÇÃO.....	1
TESTES PSICOLÓGICOS.....	5
INTELIGÊNCIA.....	9
EVIDÊNCIAS DE VALIDADE.....	19
PROCESSO DE RESPOSTA E MODELOS TEÓRICOS	22
OBJETIVOS.....	38
MÉTODO.....	39
APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	50
CONSIDERAÇÕES FINAIS.....	73
REFERÊNCIAS	77

LISTA DE TABELAS

Tabela 1. Distribuição da amostra quanto ao sexo e escolaridade.....	39
Tabela 2. Taxonomia para análise dos itens.....	49
Tabela 3. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série A.....	50
Tabela 4. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série B.....	51
Tabela 5. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série C.....	52
Tabela 6. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série D.....	53
Tabela 7. Índices de <i>infit</i> e <i>outfit</i> dos itens do teste.....	54
Tabela 8. Correlação entre a análise de juízes quanto às regras identificadas em cada item.....	56
Tabela 9. Correlação entre a dificuldade dos itens e a localização da célula para resposta no problema.....	59
Tabela 10. Correlação entre a dificuldade dos itens com a quantidade de opções de resposta e com a localização da célula com a resposta correta.....	59
Tabela 11. Correlação entre a dificuldade dos itens e a quantidade de linhas, colunas ou tamanho da matriz.....	60
Tabela 12. Correlação entre a dificuldade dos itens e o tipo de estímulos presentes nos problemas e nas opções de resposta.....	61

Tabela 13. Correlação entre a dificuldade dos itens e a quantidade total de estímulos existentes nos problemas e nas opções de resposta.....	62
Tabela 14. Correlação entre a dificuldade dos itens e as regras horizontais e verticais.....	63
Tabela 15. Correlação entre a dificuldade dos itens e a quantidade de inferências.....	64
Tabela 16. Correlação entre a dificuldade dos itens e as regras horizontais e verticais.....	65
Tabela 17. Correlação entre as regras horizontais e verticais.....	66
Tabela 18. Estatística descritiva da análise de regressão.....	69
Tabela 19. Resumo dos modelos obtidos na análise de regressão.....	70
Tabela 20. Coeficientes de regressão, padronizados e teste <i>t</i> de Student.....	71

LISTA DE FIGURAS

Figura 1. Exemplo de item.....	34
Figura 2. Frequência dos participantes por idade.....	41
Figura 3. Item utilizado como modelo para descrição dos componentes.....	57
Figura 4 – Dendograma.....	66

APRESENTAÇÃO

Assim como outras ciências empíricas, a psicologia tem como meta descrever, prever e explicar os fenômenos de sua área de interesse, o que muitas vezes é feito a partir da utilização de testes psicológicos (Anastasi, 1961/1977). A palavra *teste* deriva do latim *testis* (Tyler, 1973) e tem o significado de um exame, prova, ensaio ou experiência. No campo da psicologia, vários autores fornecem definições para teste psicológico como uma forma objetiva de se avaliar determinadas características do comportamento (Braga, 1938/1998; Muñiz, 1996; Cronbach, 1996; Anastasi & Urbina, 2000; Urbina, 2007). Também é comum culturalmente se empregar a palavra *teste* com a finalidade de avaliar o conhecimento ou capacidade de uma pessoa. Nesse sentido, o conceito pode englobar um grande número de aspectos, aptidões específicas como atenção e memória, por exemplo, ou inteligência, dentre outras (Braga, 1938/1998).

Tão importante quanto definir o que é um teste psicológico, é entender o contexto e com quais objetivos deve ser utilizado num processo de avaliação psicológica (Werlang, Villemor-Amaral & Nascimento, 2010). Os testes podem ser aplicados com diferentes objetivos como, por exemplo, em diagnóstico de fracassos acadêmicos, orientação vocacional ou avaliação de candidatos para processos seletivos organizacionais. Especialmente nesta última situação, dificilmente existiria uma atividade para a qual não haja um instrumento psicológico que possa ser útil, seja no processo de admissão, transferência, promoção ou demissão (Anastasi, 1961/1977; Sisto, 1975). Independentemente do contexto em que venham a ser utilizados, uma questão primordial é

que as decisões que são tomadas por base em seus resultados, como num concurso público perícia judicial ou no contexto do trânsito, por exemplo, podem interferir significativamente na vida de um indivíduo. Configura-se, portanto, não apenas a necessidade de que os testes apresentem estudos que subsidiem tais decisões, como também de um adequado preparo e conhecimento por parte do profissional que os utiliza.

Durante muito tempo, a objetividade dos testes foi questionada pela falta de instrumentos adequadamente adaptados e normatizados à população local, o que acabou prejudicando a história da avaliação psicológica no Brasil (Sisto, 1975). Santos (2011) afirma que o princípio básico do código de ética dos psicólogos não era respeitado, pois a utilização de instrumentos sem normatização não assegurava a qualidade dos serviços prestados pelo psicólogo à época, muito menos a dignidade dos avaliados. Atualmente esse cenário tem outra configuração, uma vez que a resolução 02/2003 do Conselho Federal de Psicologia determinou parâmetros mínimos aceitáveis para que um teste psicológico seja utilizado. Entre outros aspectos que asseguram a qualidade de um instrumento psicológico, encontram-se as evidências de validade, que se referem a quanto um teste mede aquilo a que se propõe (Anastasi & Urbina, 2000). Diferentes fontes de evidências podem indicar a validade de um teste, dentre as quais se encontra a validade pelo processo de resposta, pela qual são analisados os processos cognitivos envolvidos na realização das tarefas exigidas por determinado instrumento (Primi, Muniz & Nunes, 2009).

A proposta deste estudo é levantar evidências de validade pelo processo de resposta para o Teste de Raciocínio Inferencial – RIn (Sisto, 2006), cujo construto avaliado é a inteligência ou fator *g* a partir do raciocínio inferencial. Por raciocínio inferencial entende-se o processo de manipular, a partir de regras inferenciais, dados que fazem parte de uma

base de conhecimento, com vistas a chegar a uma conclusão (Roisenberg, 1999). O Teste de Raciocínio Inferencial – RIn (Sisto, 2006) já possui estudos de validade baseada na estrutura interna e critério, mas não no processo de resposta. Na presente pesquisa pretende-se encontrar mais uma evidência de validade para o referido teste. Para isso, serão identificados os elementos e regras que compõem cada item e sua influência sobre o processo de resposta. A partir dessa identificação será proposto um modelo de avaliação dos processos cognitivos, modelo esse que pode contribuir para que estudos futuros sejam realizados com outros instrumentos similares e até mesmo auxiliar em sua criação.

Dentre as justificativas para a realização deste trabalho, pode-se citar uma crítica de Hutz (2011) quanto ao prazo em que os estudos de validade dos instrumentos são considerados e a sugestão de que seus manuais tivessem novas edições a cada cinco anos, indicando novas pesquisas sobre o construto abordado e estudos sobre a validade e utilização do instrumento em diferentes contextos ocorridos nesse período. A inteligência, por exemplo, é um dos construtos mais pesquisados na literatura psicológica e as diferentes formas de medi-la, avaliá-la e interpretar seus resultados, evidencia a importância da colocação de Hutz (2011). Coincidentemente, o instrumento aqui utilizado completa cinco anos de sua publicação no ano de 2011 e o presente estudo vai ao encontro dos argumentos do referido autor. Não se trata de desconsiderar os estudos realizados anteriormente com o instrumento, mas sim de atualizá-los e consolidar sua validade.

Considerando que a história da testagem psicológica está diretamente relacionada a dos testes de inteligência, será feito, inicialmente, um breve histórico da história dos testes e das principais teorias da inteligência, focando-se na do fator *g* de Spearman (1927), no qual o Teste de Raciocínio Inferencial – RIn (Sisto, 2006) é fundamentado. A seguir, serão

abordados o processo de criação de instrumentos psicológicos, uma vez que as análises realizadas são inerente à construção dos itens do teste. Também serão comentadas as evidências de validade, destacando-se o processo de resposta e, finalmente, os modelos teóricos que servirão de base para a adaptação necessária ao alcance dos objetivos propostos.

TESTES PSICOLÓGICOS

A história dos testes tem suas raízes na antiguidade. Encontram-se relatos de que por volta do ano 3.000 a.C. os testes já eram usados para a seleção de funcionários civis na China (Anastasi, 1961/1977; Anastasi & Urbina, 2000; Pasquali, 2003). Mais especificamente na literatura psicológica, Cattell, em um artigo de 1890, foi o primeiro a empregar a expressão “teste mental”, descrevendo uma série de testes que eram aplicados anualmente a estudantes universitários com o intuito de determinar seu nível intelectual (Anastasi & Urbina, 2000). Embora existam registros desde essa época, efetivamente esses instrumentos psicológicos podem ser rastreados a partir dos trabalhos de Galton, biólogo inglês e maior responsável pelo início do movimento de testagem (Anastasi & Urbina, 2000; Pasquali, 2003).

Pasquali (1999) considera que existem três grandes grupos de testes psicológicos, sendo eles os testes de inteligência, psicomotores e de personalidade. Já Anzieu (1981 citado por Werlang, Villemor-Amaral & Nascimento, 2010) afirma que as técnicas de avaliação eram classificadas na literatura como expressivas, projetivas ou psicométricas. Arias (1996), por sua vez, faz uma descrição mais completa da classificação dos testes, considerando aspectos que vão desde os materiais utilizados e tempo de aplicação, às normas, características e finalidade do instrumento. Considerando a abrangência da visão desta autora, alguns pontos dessa classificação dos testes serão abordados a seguir.

A primeira classificação seria quanto ao método, segundo a qual os testes dividem-se em psicométricos e projetivos. Nos primeiros, no qual se enquadra o RIn (Sisto, 2006) as

respostas são avaliadas de acordo com normas quantitativas e o resultado final é uma pontuação, sendo que seus elementos variam numericamente e independentemente. Já os segundos se referem a características gerais da personalidade, sendo avaliados por normas e critérios globais e qualitativos.

A segunda classificação aborda a finalidade da utilização do teste. Um mesmo instrumento pode servir a diferentes finalidades de acordo com o contexto de utilização como, por exemplo, investigação (provar hipóteses, analisar diferenças individuais, dentre outros) ou diagnóstico (qualidade ou grau de um determinado traço psicológico ou de uma conduta). Também podem ser divididos em testes de execução máxima (inteligência, rendimento, dentre outros) onde a pessoa utiliza sua capacidade para resolver problemas, ou de execução típica (avaliação de personalidade, interesse e atitudes), em que são descritas situações cotidianas e as respostas se baseiam no comportamento típico do sujeito nestas situações. Neste sentido, o RIn (Sisto, 2006) configura-se como um teste de investigação e execução máxima.

A área de comportamento abordada seria uma quarta possibilidade de classificação de um teste, considerando a) rendimento acadêmico ou profissional, b) inteligência e atitudes (traços de natureza cognitiva) ou c) personalidade, atitudes e interesses. Outras seis divisões são propostas pela autora, sendo elas a modalidade de aplicação (oral, papel e lápis, manipulação ou execução, dentre outras); a demanda exigida do sujeito: a velocidade e rapidez (há um controle rigoroso do tempo e as diferenças individuais são fundamentais para rapidez da resposta) ou potência da dificuldade (as diferenças se dão basicamente em função da qualidade de execução, sendo o tempo um fator menos determinante); o grau de aculturação do teste; a população a quem o teste é dirigido, o modelo estatístico em que se

baseia e o modelo conceitual (testes referidos ao critério, traço latente ou diferenças individuais).

As informações trazidas até aqui apontam classificações e formas de utilização dos testes. Entretanto, não é muito comum se ater à maneira como um teste foi criado, o que é a base para um trabalho como o aqui proposto, uma vez que é a partir das características dos itens do teste que se pretende encontrar a validade por meio do processo de resposta. Arias (1996) sugere que sejam seguidas uma série de fases na construção de um teste, a saber:

- a) Identificação da finalidade ou propósito em que as pontuações serão utilizadas: definir a população alvo (idades, conhecimentos prévios necessários, público geral ou grupos específicos, clínicos) e quais decisões serão tomadas por base em seus resultados (seleção ou promoção para determinada cargo, decisões relativas à escola, diagnóstico, dentre outras);
- b) Especificação das principais restrições operacionais, como tempo de aplicação ou compreensão das instruções;
- c) Definição do construto a partir de um modelo teórico;
- d) Criação dos itens do teste, pautados no referencial teórico adotado, revisões da literatura, observações empíricas, informações levantadas diretamente com o grupo a ser pesquisado, dentre outras;
- e) Especificação do formato do teste, definindo, por exemplo, se as respostas serão escolhidas dentre um grupo de opções ou construídas pelo próprio testado;

- f) Planejar a amostra a ser pesquisada no estudo piloto e realizar os estudos estatísticos dos dados obtidos nesse estudo, definindo a forma definitiva do teste;
- g) Definir os parâmetros desejados para os índices de dificuldade dos itens, validade e precisão;
- h) Planejar os procedimentos para a construção das normas do teste;
- i) Realizar os estudos psicométricos e, por fim
- j) Elaborar o manual do teste (Arias, 1996).

Além dos procedimentos para construção de um teste, a autora ressalta a importância de que os instrumentos disponíveis no mercado sejam avaliados quanto a diversos aspectos, como a adequação à população a que se pretende aplicar o instrumento, exemplos de protocolos analisados, índice de dificuldade e de alocação dos itens segundo esse critério, descrição dos estudos psicométricos, clareza das informações quanto à correção e fundamentação em uma teoria que embase o construto a que se pretende medir. No caso do RIn (Sisto, 2006) não é possível verificar se todos os processos citados por Arias (1996) foram utilizados, uma vez que não há informações quanto às fases D e F em seu manual. Quanto ao construto a ser medido – a inteligência – será apresentado um breve histórico nos próximos parágrafos, com ênfase na Teoria Bi-fatorial de Spearman (1927), a partir da qual o teste foi construído.

INTELIGÊNCIA

Brody (1992) considera marcantes para o estudo da inteligência a perspectiva psicométrica de Spearman que, em 1904, introduziu o construto de inteligência geral em seu artigo denominado *General Intelligence objectively determined and measured* e o estudo de Binet e Simon intitulado *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*. Almeida (1988) acrescentou à primeira citação outras duas teorias, a perspectiva cognitivista de Sternberg e a desenvolvimentalista de Piaget. Já Oliveira-Castro e Oliveira-Castro (2001) apontaram a dificuldade em se adotar exclusivamente uma abordagem psicométrica, cognitiva ou biológica, embora citem autores que encontraram pontos convergentes entre as diferentes teorias.

Quanto à sua utilização na psicologia, o conceito inteligência exerce uma função adverbial “na linguagem cotidiana, não indicando ações específicas, mas formas de realizar quase qualquer ação”, o que pode explicar as muitas definições e, inclusive, algumas incongruências encontradas na literatura (Oliveira-Castro & Oliveira-Castro, 2001, p.257). Primi (2002), ao realizar uma busca no banco de dados PsycInfo da *American Psychological Association* (APA) encontrou 18448 artigos com os termos *intelligence*, *intellectual* ou *ability*, concluindo que, apesar de possivelmente ser o construto psicométrico mais pesquisado na literatura, o número de pesquisas realizadas dificilmente permitiria um consenso sobre o assunto. Tal conclusão corrobora a afirmação de Lubinsky e Benbow (2000), segundo os quais nos últimos cem anos o debate sobre a inteligência e as formas de medi-la continuam e ainda não resultaram em uma definição plenamente satisfatória.

Um dos estudos mais conhecidos sobre o tema foi apresentado em 1921 no simpósio intitulado “*Intelligence and its Measurement*”, no qual quatorze especialistas deram suas definições acerca da natureza desse fenômeno (Sternberg, 1992). A palavra “adaptação” foi um dos termos mais utilizados pela maior parte dos pesquisadores, que conceituaram a inteligência como uma capacidade de aprender e adaptar-se (Sisto, 2006). Cornoldi (2006) chamou a atenção para o fato de que a definição de inteligência como capacidade de adaptação pode ter sido uma influência darwiniana, uma vez que, de acordo com a lei de seleção natural, uma maior capacidade adaptativa é preditiva da sobrevivência. Sua afirmação teve como base o fato de que Galton, primeiro pesquisador a sistematizar o estudo da inteligência, não apenas tinha laços familiares com Darwin, como também só fez sua primeira publicação alguns anos após o lançamento do livro “*A origem das espécies*”. Mais recentemente, em 1986, vinte e quatro psicólogos cognitivos acrescentaram às definições anteriores da inteligência a importância da metacognição (Sternberg & Detterman, 1986), ou seja, do conhecimento e controle que se tem sobre a própria cognição.

Um fato importante a ser considerado é o de que a inteligência pode contribuir para que uma pessoa se ajuste e busque a harmonia dentro do contexto em que está inserida, mas não é garantia de uma boa adaptação. Dessa maneira, algumas vezes o indivíduo pode precisar intervir ou alterar seu meio para que seja mais coerente com suas necessidades ou buscar um novo ambiente (Sternberg, 1984). Há, portanto, certa associação da capacidade de adaptação com a variável cultura, uma vez que diferentes períodos históricos ou populações com culturas distintas podem privilegiar determinadas habilidades e, conseqüentemente, o grau de inteligência estará relacionado ao quanto se possui da habilidade privilegiada, o que irá interferir na adaptação individual (Primi, 1998; Sisto,

2006). Sisto (2006), por exemplo, citou estudos de vários autores que abordaram essa questão, como Goodnow (1976), Irvine (1978) e Serpell (1982), dentre outros, em que uma mesma característica, como o silêncio ou tempo de execução de uma atividade, é percebida de forma contrária por diferentes povos.

A existência de definições diferentes para a inteligência, por vezes demasiadamente amplas ou contraditórias, contribuiu tanto para o interesse pelo tema fora do meio acadêmico quanto para a descrença de alguns cientistas (Cornoldi, 2006). Pesquisando ambos os grupos – leigos e cientistas – Miranda (2002, p.19) afirmou que para o senso comum a inteligência se refere a uma qualidade do indivíduo, enquanto para o meio científico é “entendida e abordada como um conceito plural, que marca o comportamento adaptativo”. Já Sternberg e cols. (1981 citado por Sternberg, 1992) encontraram uma correlação de 0,96 entre a visão de leigos e cientistas sobre as características de inteligência e de 0,85 sobre sua importância. Ainda assim, informaram que os dois grupos divergiram em relação ao aspecto motivação (citada pelos especialistas e não pelos leigos) e pela maior importância dada aos aspectos socioculturais pelos leigos.

Diferentemente do que ocorre com o estado de ânimo e sua influência na conduta diária, nem sempre as pessoas se atêm ao fato de que a inteligência está implícita em muitas ações cotidianas (Colom, 2008). Dentre os construtos avaliados por um teste psicológico a inteligência é, possivelmente, o mais importante dos atributos humanos e nenhum outro fator psicológico está relacionado a um número tão grande de fenômenos sociais como ela (Colom, 2002). Considerando os principais achados científicos acerca da inteligência, a questão principal das discussões está em torno da existência de um fator geral ou de fatores mais específicos, sendo o fator geral apenas um deles (Anastasi & Urbina, 2000; Sternberg,

1992; Sisto, 2006). Essa discussão tem como base os postulados de Spearman e sua Teoria Bi-Fatorial, de 1904.

Discordando da simples descrição da atividade inteligente e, observando se as atitudes intelectuais estariam correlacionadas e dependeriam de uma inteligência geral ou se, pelo contrário, eram independentes, Spearman postulou a Teoria dos Dois Fatores ou Teoria Bi-fatorial (Braga, 1938/1998). Sua proposta era conciliar e resolver de forma harmônica as diferenças entre as teorias monárquicas (que postulavam uma inteligência geral), oligárquicas (que defendiam várias habilidades diferentes) e anárquicas (que pressupunham múltiplas atitudes independentes) existentes à época (Spearman, 1927).

A base da teoria de Spearman foi a *psicologia das faculdades*. De acordo com essa teoria existem sete faculdades primárias – intelecto, sentido, memória, imaginação ou invenção, linguagem, atenção e movimento – que se referem a funções diferentes e, portanto, devem ser medidas separadamente. Na época de concepção dessa teoria, que estava bastante ligada à filosofia, o intuito era identificar aspectos comuns a todas as pessoas e não diferenciá-las, embora posteriormente tenham sido analisadas à luz da psicologia das diferenças individuais. A partir da medição individual de cada um dos traços ou faculdades era feita uma representação gráfica, pela qual os traços eram interligados. Pressupondo que poderia haver correlações entre essas medidas, Spearman correlacionou os resultados de diferentes testes para os quais encontrou coeficientes que não eram perfeitos ou nulos, mas oscilavam regularmente entre dois valores, o que permitia representá-los matematicamente. À essa regularidade encontrada nas correlações entre as diferentes habilidades mentais, denominou fator *g* (Sisto, 1975).

Além do fator *g*, comum a todas as habilidades, Spearman (1927) também definiu um fator específico para cada habilidade, denominado fator *s*. Os fatores específicos explicariam a parte da medida dos testes que não se refere ao fator *g*, uma vez que, como não há uma correlação perfeita entre as diferentes habilidades, não se pode dizer que um teste meça exclusivamente o fator *g* (Sisto, 2006). Apesar de todas as habilidades envolverem ambos os fatores, para algumas o fator *g* desempenha um papel principal, enquanto para outras é o fator *s* que o faz. Outros estudos também indicaram a existência de fatores que se encontram em grande parte de um conjunto de habilidades afins, denominados de *fatores de grupo* (Spearman, 1927).

O fator *g* não é um observável concreto, mas sim um valor, uma magnitude (Sisto, 1975). Quantitativamente, é um fator comum e fundamental a todas as funções cognitivas de um mesmo indivíduo, ao mesmo tempo em que sua magnitude é amplamente variável de um indivíduo a outro. Qualitativamente, é explicado como uma “energia subjacente e constante a todas as operações psíquicas” (Spearman, 1927, p.14).

O fator específico (*s*) varia, do ponto de vista quantitativo, tanto entre as habilidades de um mesmo indivíduo quanto de um indivíduo para outro e não depende ou se correlaciona com o fator *g* ou com outros fatores específicos. Já do ponto de vista qualitativo, é o instrumento pelo qual atua e opera a energia mental (Spearman, 1927) e está relacionado à particularidade de cada item do teste (Sisto, 2006).

Os fatores de grupo são comuns a muitas habilidades de um mesmo conjunto, como os fatores V (verbal), M (mecânico, espacial), N (numérico), P (perseverança), dentre outros. De uma forma geral, Spearman (1927) considerou a inteligência como sendo

determinada por um grande número de fatores específicos (um para cada função), um número limitado de fatores de grupo e um fator geral (fator *g*) comum a todas as funções.

Apesar de, aparentemente, Spearman ter encontrado uma conciliação entre as discrepâncias existentes à época, logo surgiram outras teorias que se levantaram em sua oposição, revivendo e prolongando as divergências que ele acreditava haver solucionado. Nesse contexto, nos primeiros 50 anos do século XX houve uma predominância de duas posições opostas sobre os estudos fatoriais da inteligência, a de Spearman e a de Thurstone. Possivelmente, uma das premissas dessa discussão foi o fato de que, mesmo para Spearman, o fator *g* não era o único determinante da inteligência (Sisto, 2006).

Thurstone (1938), maior opositor de Spearman, não acreditava na existência de um fator geral, mas sim de um conjunto de habilidades primárias ou básicas, definidas como V (compreensão verbal), W (fluência verbal), P (perceptual), N (numérica), S (espacial), M (memória) e I (raciocínio indutivo). Entretanto, o próprio autor admitiu posteriormente a semelhança entre sua teoria e a de Spearman, quando detectou uma alta correlação entre esses fatores e afirmou que o fator *g* seria o elemento comum às diversas capacidades (Thurstone & Thurstone, 1941). Tendo por base as observações das correlações entre as teorias de Spearman e Thurstone, Cattell (1971) observou a existência de dois grandes fatores, aos quais denominou *inteligência fluida* (*Gf*) e *inteligência cristalizada* (*Gc*).

Por inteligência fluida entendem-se as operações mentais que são realizadas por uma pessoa ao se deparar com uma situação nova e que não pode ser executada automaticamente (Flanagan & Ortiz, 2001). Para Almeida (1988) a degeneração das estruturas fisiológicas causada pelo avanço da idade leva ao declínio dessa capacidade,

ocorrendo o oposto com a inteligência cristalizada, que se refere “à extensão e profundidade dos conhecimentos adquiridos de uma determinada cultura” (Almeida & Primi, 2004, p.136), relacionando-se ao que foi aprendido anteriormente para operacionalizar o conhecimento. Tais definições explicariam o porquê de alguns idosos encontrarem dificuldade ao lidar com informações ou situações desconhecidas, não ocorrendo o mesmo àquelas que já fazem parte de seu repertório (Cornoldi, 2006). Horn (1991) acrescentou outras seis habilidades cognitivas ao modelo de Cattell, a saber, processamento visual, memória de curto prazo, armazenamento, recuperação a longo prazo e velocidade de processamento, rapidez para decisão correta e processamento auditivo.

Em 1993, Carrol procurou integrar as diferentes posições quanto à abordagem fatorial da inteligência e selecionou aproximadamente 460 tabelas contendo correlações entre testes de inteligência de pesquisas realizadas entre os anos de 1927 e 1987. A partir dos resultados da análise fatorial, concluiu que a inteligência e as habilidades mentais poderiam ser explicadas em três estratos (Sisto, 2006). A integração entre essa teoria e a *Gf-Gc* deu origem à Teoria de Cattell-Horn-Carrol (CHC), dentro de uma perspectiva multidimensional que contém dez fatores integrados a áreas do funcionamento cognitivo. Esses fatores são a Inteligência fluida (*Gf*), Inteligência cristalizada (*Gc*), Conhecimento quantitativo (*Gq*), Leitura e escrita (*Grw*), Memória de curto prazo (*GSM*), Processamento visual (*Gv*), Processamento auditivo (*Ga*), Capacidade e armazenamento e recuperação da memória de longo prazo (*Glr*), Velocidade de processamento (*Gs*) e Rapidez de decisão (*Gt*) (Primi, 2003).

Enquanto os fatores amplos ocupam o segundo de uma hierarquia de três níveis, no terceiro nível estão aproximadamente 70 fatores específicos que os subdividem. O fator *g*

de Spearman ocupa, sozinho, o Estrato III dessa teoria e explicou aproximadamente 50% da variância das cerca de mil medidas de inteligência estudadas, o que indica que todos os outros elementos da cadeia estariam relacionados a ele (Sisto, 2006). A teoria CHC é bastante aceita dentro da psicometria atual e considerada uma evolução do modelo teórico da inteligência (Primi, 2003). Além disso, é constantemente atualizada. Entretanto, apesar de ser o mais abrangente modelo sobre a inteligência, muitas das habilidades descritas provêm de outros modelos já existentes (Sisto, 2006).

Em comum, as teorias apresentadas postulam uma descrição da inteligência a partir de um modelo hierárquico, mas o fator *g*, postulado por Spearman, continua sendo o fator comum a todas as demais habilidades e no qual a maior parte dos testes de inteligência se baseiam, como no caso do RIn (Sisto, 2006). O raciocínio inferencial, a partir do qual o RIn (Sisto, 2006) foi construído, refere-se à maneira pela qual busca-se uma conclusão a partir da manipulação de dados conhecidos, utilizando-se o processo inferencial (Roisenberg, 1999). Tais conclusões podem ser corretas ou não, pois os dados que uma pessoa seleciona, suas interpretações e conclusões sofrem influência do conhecimento que essa pessoa tem do mundo. Também é verdade que a maneira como uma inferência é realizada pode tornar uma conclusão válida, mesmo que esta não seja verdadeira. Um professor pode, por exemplo, supor que seus alunos estudam apenas quando sabem que farão uma prova e, como dará uma prova em determinada data, inferir que naquele dia todos os seus alunos estudaram. Entretanto, alguns alunos podem não tê-lo feito, o que tornaria a conclusão errônea, apesar da coerência com a qual realizou sua inferência. Da mesma maneira, considerar que todas as frutas são doces, pelo fato de morango ser uma fruta e ser doce, colocaria os limões na mesma situação, o que não estaria correto.

É possível relacionar o raciocínio inferencial com o conceito de noegênese proposto por Spearman (1927) em alguns estudos que tinham por objetivo explicar o que seria o fator *g*. No termo noegénético estão envolvidos dois conceitos, o noético e o genético. O primeiro conceito refere-se a uma entidade intelectual, àquilo que pertence à mente, enquanto o segundo está relacionado à geração de um novo conteúdo mental e não apenas a uma mera reprodução (Sisto, 2006). Três leis ou processos foram postulados em relação à noegênese. A primeira, intitulada “*Lei da apreensão da própria experiência*” está relacionada a quanto o indivíduo é capaz de perceber sua própria mente, no sentido de não apenas sentir, se esforçar ou saber, mas ter consciência do que sente, de seus esforços e do que sabe (Spearman, 1927). No RIn (Sisto, 2006), segundo seu autor, os primeiros itens de cada série do teste cumprem esse papel, uma vez que servem para que o examinando compreenda o que será exigido nos itens seguintes.

As duas outras leis noegénéticas, diferentemente da primeira, estão presentes em todos os testes de inteligência e se referem à educação, sendo a segunda lei relacionada à educação de relações e a terceira à educação de correlatos. A segunda lei relaciona-se à maior ou menor capacidade que uma pessoa dispõe para estabelecer relações entre duas ou mais ideias, que podem ter sido originadas a partir de sua percepção ou de uma representação mnêmica. Um exemplo seria ver uma banana e uma maçã e classificá-las como frutas. Já a terceira lei refere-se à capacidade de criar uma nova ideia a partir de outra e de uma relação já existentes na mente (Spearman, 1927, p.156). Pode-se concluir que estas leis estão relacionadas à quantidade das relações que são estabelecidas, inferidas ou abstraídas pelos dados fornecidos (Sisto, 2006).

Além das três leis qualitativas ou noogenéticas apresentadas, Spearman (1927) também estudou a idade, a herança e o sexo. Quanto à idade, observou um aumento mais significativo do fator g até aproximadamente quinze ou dezesseis anos, além de certo declínio na senilidade. A herança foi abordada a partir da influência da educação e da nacionalidade ou da raça, mas os resultados não foram totalmente conclusivos, dadas as dificuldades em se controlar as variáveis. Ainda assim, obteve algumas conclusões, como uma possível maior influência da educação nas habilidades específicas do que no fator g . Em relação ao gênero, houve superioridade feminina em alguns aspectos e masculina em outros, mas em relação ao fator g as diferenças encontradas foram desprezíveis.

A amplitude das teorias acerca da inteligência permitiria um extenso discurso e debate sobre os pontos convergentes e divergentes, mas aprofundar-se nesse tema fugiria dos objetivos centrais deste texto. Dessa forma, apesar de bastante rica, essa discussão não será prolongada, uma vez que o Teste de Raciocínio Inferencial – RIn (Sisto, 2006) – foco deste estudo – tem por objetivo a avaliação do fator g segundo a definição de Spearman, da qual já foram apresentados os principais pontos. A seguir, serão abordadas as evidências de validade, especialmente pautadas no processo de resposta.

EVIDÊNCIAS DE VALIDADE

Pode-se inferir que a qualidade de um instrumento psicológico está diretamente relacionada às evidências de validade, ou seja, às confirmações de que o instrumento realmente mede aquilo à que se propõe (Anastasi & Urbina, 2000). A validade é o conceito central da psicometria e está relacionada à interpretabilidade dos escores de um teste (American Psychological Association, American Educational Research Association & National Council on Measurement in Education, 1999). Por psicometria entende-se “o conjunto de modelos formais que possibilitam a medição das variáveis psicológicas, centrando-se nas condições que permitem levar a cabo todo processo de medição em psicologia e estabelecer as bases para que estes processos se realizem de forma adequada” (Arias, 1996, p.21).

Anastasi (1961/1977) adverte que a multiplicidade de técnicas disponíveis para a obtenção da validade de um instrumento pode ser algo perigoso, pois um autor pode usar diferentes métodos para encontrar evidências de validade de seu instrumento e publicar apenas aquele que, até mesmo por acaso, tenha sido mais positivo. A omissão dos demais resultados pode, conseqüentemente, gerar uma impressão errônea acerca da interpretabilidade dos resultados. Essa afirmação destaca a importância de que, cada vez mais, sejam levantadas evidências de validade dos instrumentos psicológicos, mesmo daqueles que já tem alguns estudos desse tipo, como é o caso do RIn (Sisto, 2006), e é uma primeira justificativa para a realização do estudo proposto, correspondendo à necessidade de atualização dos instrumentos sugerida por Hutz (2011).

Pasquali (2003) considera possível fazer um delineamento da validade a partir de três períodos históricos, onde houve a predominância de um dos tipos do modelo trinitário proposto por Cronbach e Meehl (1955), a saber, conteúdo, critério e construto. No primeiro período, de 1900 a 1950, os testes eram considerados válidos ao apresentar uma correspondência entre seu conteúdo e o conteúdo das definições apresentadas pela teoria que lhe subsidiavam. É um tipo de validade comumente empregada na área educacional para verificar-se se o conteúdo de uma prova, por exemplo, engloba os conhecimentos da área específica a que se destina (Muniz, 2004).

Num segundo período, compreendido entre os anos de 1950 e 1970, a validade foi caracterizada por predizer precisamente comportamentos futuros ou em outras condições, comportamentos e condições estas que se tornariam os critérios de validade. Tal fato pode ser relacionado à influência do behaviorismo skinneriano nesse período e à conseqüente ênfase dada aos processos estatísticos em detrimento à teoria psicológica predominante no período anterior (Pasquali, 2003). Estudos abordando esse tipo de validade são essenciais especialmente quando os testes serão utilizados para tomada de decisões importantes, como no contexto de avaliação psicológica para, por exemplo, fins clínicos, organizacionais ou jurídicos (Muniz, 2004).

Um terceiro período, a partir do ano de 1970, teria como ponto de partida um impacto tardio do artigo de Cronbach e Meehl (1955) sobre o modelo trinitário da validade, onde passou a existir uma maior preocupação em se estabelecer a validade de construto. A partir das afirmações e constatações de Messick (1980) de que os demais tipos de validade (conteúdo e critério) eram contemplados pela validade de construto, houve a necessidade de reformulação dessa visão. O termo validade de construto passou então a ser aceito como

sinônimo de validade, o que não tornava lógica a existência de procedimentos específicos para defini-la. Conseqüentemente, a expressão “tipos de validade” foi substituída por “evidências de validade”, explicitando que não haveriam diferentes tipos de validade e sim formas distintas de encontrar evidências que sustentassem as interpretações feitas a partir dos escores dos testes (Anastasi & Urbina, 1997).

Os aspectos descritos anteriormente puderam ser acompanhados também pelas modificações nos *Standards for educational and psychological testing* (American Psychological Association, American Educational Research Association & National Council on Measurement in Education, 1999), normalmente conhecido como Padrões de testagem ou Padrões (Cronbach, 1996). O primeiro desses documentos foi publicado em 1954, no qual eram abordadas a validade de conteúdo, preditiva, concorrente e de construto. Outras versões foram publicadas em 1966, 1974, 1985 e 1999. Nas versões de 1974 e 1985 as evidências de validade preditiva e concorrente passaram a integrar a validade de critério. Na edição de 1999 pode ser percebida a influência das idéias de Messick (1989), quando a validade passou a ser definida como “o grau em que evidência e teoria sustentam as interpretações dos escores dos testes vinculados aos usos propostos dos testes” (APA, AERA & NCME, 1999, p. 9). Foram definidas cinco fontes de evidência de validade, baseadas no conteúdo, nas relações com outras variáveis, na estrutura interna, no processo de resposta e nas conseqüências de testagem. O presente trabalho será pautado nas evidências de validade com base no processo de resposta, a partir do qual “busca-se analisar a coerência entre as explicações teóricas e os dados empíricos” (Nunes & Primi, 2010).

PROCESSO DE RESPOSTA E MODELOS TEÓRICOS

Do ponto de vista metodológico, as evidências pelo processo de resposta partem da “análise de protocolos de resposta, entrevistas ou outros procedimentos que permitam a análise individualizada do par sujeito/item” (Cunha & Santos, 2009, p.551). As anotações feitas em rascunhos durante a realização do teste, o monitoramento dos movimentos dos olhos ou do tempo de resposta, bem como a comparação do desempenho em determinadas partes de um mesmo teste ou com outras variáveis podem ser monitoradas e fornecer importantes dados sobre o processo de resposta envolvido na resolução dos itens do teste. Também é preciso considerar a necessidade de haver uma consistência entre as avaliações feitas por diferentes juízes ou observadores, verificando se os processos adotados por eles não sofrem influência de fatores irrelevantes (APA, AERA & NCME, 1999).

Apesar de ser uma forma de evidência empregada pelos psicólogos cognitivos a testes de inteligência e ter surgido a partir da década de 1980, o processo de resposta ainda é pouco utilizado (Anastasi & Urbina, 1997). Embora a afirmação anterior já tenha alguns anos, no Brasil, onde o precursor desse tipo de estudo foi Primi (1995; 1998), ainda percebe-se a escassez de estudos desse tipo. Esse autor, em sua dissertação, criou uma prova experimental de raciocínio analógico geométrico, baseado nos princípios de organização perceptual da teoria da Gestalt e na abordagem do processamento da informação. Tal prova foi construída em função de três variáveis, a saber, número de elementos, de transformações e tipo de organização perceptual. Seu estudo partiu de uma revisão teórica acerca dos métodos de investigação da inteligência e das concepções oriundas dessa investigação, culminando na importância atribuída aos processos mentais

envolvidos na realização de determinada tarefa. Procurou, então, “explicar a complexidade dos problemas de raciocínio analógico geométrico presentes em testes de inteligência, bem como as diferenças individuais, analisando ocorrências particulares de interferências entre as fases do raciocínio (ou componentes) ligadas, por um lado, à percepção e, por outro, à abstração” (p.18). A partir dos resultados obtidos pela aplicação da Bateria de Provas de Raciocínio Diferencial, a BPRD (Almeida, 1992), Primi (1995) dividiu sua amostra em grupos com maior e menor habilidade. As respostas e o tempo de reação foram gravados pelo programa informatizado a partir do qual a prova de raciocínio foi aplicada. Percebeu-se que as pessoas mais rápidas ao responder às questões e que conseguiram trabalhar com um número maior de elementos eram aquelas cujo desempenho na BPRD (Almeida, 1992) havia sido melhor. A conclusão da pesquisa indicou que a dinâmica entre as diferenças individuais e a complexidade da estimulação interferem na complexidade da tarefa (Primi, 1995).

Já em sua tese, Primi (1998) desenvolveu um modelo para indicar fatores de complexidade em problemas de raciocínio analítico a partir da psicologia cognitiva e de estudos sobre os componentes de processamento desse tipo de raciocínio. O autor criou dois instrumentos diferentes, com a mesma estrutura e compostos, cada um, por trinta e dois itens de analogias com figuras geométricas em forma de matrizes. A construção dos instrumentos partiu da identificação de fatores que poderiam influenciar sua dificuldade, como o número de figuras geométricas e de regras, a natureza dessas regras e a organização perceptual. A amostra foi composta por 365 universitários de 17 a 22 anos, dos quais 68% pertenciam ao sexo feminino.

A partir da utilização do modelo criado por Primi (1998) foi possível prever entre 75% e 82% da dificuldade dos itens, a qual estava atrelada principalmente ao número de figuras e regras e à organização perceptual. O tempo de reação e a utilização de respostas de eliminação de alternativas foram influenciados pelos fatores de complexidade e os itens com informações irrelevantes ou maior número de informações consumiam mais tempo e levavam a um aumento da utilização de estratégias analíticas, respectivamente. Já os alunos com melhor desempenho foram mais sistemáticos tanto na distribuição do tempo quanto na utilização de recursos de eliminação diante dos itens mais complexos. Além disso, foram encontradas correlações com diversas variáveis externas como conhecimentos gerais, compreensão em leitura, raciocínio espacial, lógico-dedutivo e a Escala Avançada de Raven, cujas correlações foram de 0,55 a 0,81. Utilizando a Teoria de Resposta ao Item (TRI), mais especificamente o modelo de um parâmetro, o autor criou uma escala que indica a proficiência em problemas de distintos níveis de exigência da memória de curto prazo e do gerenciamento metacognitivo, que permite também a comparação do desempenho relativo quanto à persistência, flexibilidade e gerenciamento eficaz do tempo.

Mais recentemente, Rueda e Sisto (2008) realizaram um estudo com o objetivo de buscar evidências de validade para o Teste Pictórico de Memória. O referido teste mede a memória de curto prazo, pela apresentação por um minuto de um cartão contendo 51 desenhos, que a pessoa deve memorizar e depois tem dois minutos para recuperar as informações e escrever os nomes dos desenhos dos quais se recorda. A amostra da pesquisa foi composta por 220 participantes do sexo masculino e 291 do feminino. A idade variou de 10 a 60 anos, com média de 19,80 e o DP = 8,38. Em relação à escolaridade, cursavam desde o segundo ano do Ensino Fundamental ao Superior. Para levantar as evidências pelo

processo de resposta os autores partiram do pressuposto que os ambientes presentes no teste (céu, terra e água) gerariam níveis de dificuldade significativamente diferenciados, afetando a recuperação da informação. A partir da confirmação de tal premissa, concluíram que o teste apresentava as evidências desejadas. Apesar da conclusão dos autores, a metodologia empregada não parece adequada para se considerar os resultados obtidos como evidências baseadas no processo de resposta.

Outro estudo utilizando o processo de resposta foi publicado por Cunha e Santos (2009). O objetivo das autoras foi encontrar evidências de validade pelo processo de resposta para o teste de Cloze. Participaram da pesquisa 266 crianças, das quais 52,6% eram do sexo masculino e 47,4% do feminino. As idades variaram de 8 a 13 anos, com média de 9,42 anos e DP = 0,80. Alunos do quarto ano representaram 44,4% da amostra e do quinto ano 55,6%. O teste de Cloze utilizado apresentava evidências de validade de critério e estava adequado à escolaridade dos participantes. Tratava-se de um texto cujo quinto vocábulo de cada frase era omitido, o qual deveria ser substituído pela palavra correta. A lacuna onde essa palavra deveria ser encaixada era proporcional ao tamanho da palavra omitida. A metodologia empregada foi a análise dos erros cometidos no teste, considerando a divisão da amostra em dois grupos, crianças com as melhores e com as piores pontuações. Os resultados indicaram que as crianças do quarto ano cometeram mais erros lexicais e as do quinto ano mais erros semânticos. Com base na avaliação da homogeneidade da distribuição dos tipos de erros as autoras encontraram evidências de validade por meio do processo de resposta, embora tenham sugerido a ampliação dos estudos psicométricos do teste utilizado.

Os estudos citados anteriormente foram os únicos encontrados na literatura nacional utilizando o processo de resposta, o que confirma a afirmação de que estudos usando essa evidência não são comuns (Anastasi & Urbina, 1997). O RIn (Sisto, 2006) inclui-se nessa constatação, pois nenhum de seus cinco estudos de validade foi baseado no processo de resposta, o que se configura como segunda justificativa para a realização do presente trabalho, aumentando também a literatura sobre essa evidência. No que tange à forma de obtenção de evidências de validade pelo processo de resposta, Primi, Muniz e Nunes (2009) fazem uma descrição bastante abrangente. De acordo com esses autores, a partir do referencial teórico do construto são elaborados modelos explicativos do processo mental utilizado para responder aos itens do teste, assim como previsões sobre os parâmetros de comportamento, tendo por base as diferentes propriedades dos itens. Afirmam, ainda que o

modelo teórico inclui explicações sobre como as diferentes características dos itens afetam o processamento e, especialmente, como essas alterações processuais culminam em alterações nos parâmetros comportamentais – ocorrência de acertos, maior ou menor tempo de reação, dentre outros. Assim, o modelo teórico pode ser empiricamente testado comparando-se as observações dos padrões de resposta ou os padrões dos parâmetros comportamentais em diferentes itens verificando-se o quanto esses padrões correspondem às previsões feitas pelo modelo. Quanto mais correspondentes, mais confiança se tem no modelo teórico de interpretação sobre o que o teste mede (p.250).

Apesar de muitos teóricos terem demonstrado interesse quanto aos processos cognitivos envolvidos na capacidade humana de raciocinar e solucionar problemas (Almeida, 1988), não é comum estudar esses processos entre as pessoas que realizam testes,

o que constitui uma crítica dos psicólogos cognitivistas às teorias psicométricas (Lemos, 2007) e configura uma terceira justificativa para a realização desta pesquisa. Sternberg (1983) considerou que a metodologia empregada pelo modelo psicométrico tradicional limita a compreensão dos processos cognitivos implícitos, uma vez que normalmente procura prever resultados a partir da tentativa de se encontrar fatores subjacentes às habilidades humanas. Primi (1998) fez uma afirmação semelhante, quando disse que a abordagem psicométrica enfatiza a face estrutural da inteligência e, apesar de permitir a decomposição da capacidade cognitiva de uma pessoa em distintos fatores, não dá suporte a uma compreensão de como as capacidades cognitivas são utilizadas por uma pessoa ao ser confrontada com um problema.

Uma abordagem que pode sanar tais limitações é a psicologia cognitiva ou abordagem do processamento humano da informação, que surgiu após a segunda metade do século passado e a partir da qual a abordagem psicométrica deu lugar à descrição de quais processos cognitivos estariam relacionados a atividades cognitivas específicas (Primi, 1998). Dentre os diferentes métodos que compõem essa teoria, que busca entender a maneira como as pessoas processam informações, encontra-se a análise componencial (Sternberg, 1979).

A análise dos componentes dos itens dos testes foi utilizada por Sternberg (1983) para uma melhor compreensão da inteligência e das exigências que os testes fazem às pessoas que o respondem. Diferentemente da abordagem psicométrica, esse modelo procura explicar as diferenças entre as tarefas, o que, de acordo com Primi (1998) equivale indiretamente a explicar as funções cognitivas utilizadas pelos avaliados. Enquanto na abordagem fatorial são identificados os fatores estruturais das habilidades mentais, na teoria

componencial identificam-se os processos cognitivos associados a tais habilidades (Sternberg, 1984). Estudos posteriores apontaram diversos pontos positivos e negativos desse método, os quais não serão discutidos aqui, uma vez que será dada prioridade aos estudos realizados com figuras geométricas, por serem os estímulos utilizados pelo RIn (Sisto, 2006). Em contrapartida, serão relacionados os pontos em comum entre os estudos de Sternberg e as teorias apresentadas.

De acordo com Primi (1998), um dos primeiros trabalhos de processamento de analogias a partir de figuras geométricas foi desenvolvido por Mulholland, Pellegrino e Glaser (1980), a partir da integração de dois modelos existentes, o de Sternberg (1977) e o de Evans (1968). Tal integração originou um modelo composto por quatro componentes, sendo eles a) a comparação e decomposição de elementos, b) análise de transformações e gerações de regras, c) comparação de regras e d) resposta.

O primeiro componente está relacionado à identificação de elementos comuns entre os dois primeiros termos apresentados numa analogia do tipo $A:B // C:D$, a partir de sua comparação. Assemelha-se ao que Sternberg (1977) denominou codificação, embora nesse caso sejam contextualmente dependentes. Em um segundo momento são analisadas as transformações ocorridas em cada elemento para que seja possível elaborar uma regra que transforme A em B e C em D. As transformações identificadas passam a fazer parte da memória de curto prazo. Esse processo corresponde ao que Sternberg denominou inferência. Posteriormente, as regras das analogias A-B e C-D são comparadas pelo sujeito e, caso sejam semelhantes, indicam que a analogia é verdadeira, ocorrendo o oposto se forem distintas. Quanto ao número de elementos e de transformações, as variáveis manipuladas por Mulholland, Pellegrino e Glaser (1980) compunham até três elementos

para cada termo (A, B, C ou D) que poderiam ser de seis formas distintas (linhas, triângulo, círculo, cruz, retângulo ou pentágono) e comportar até três transformações (identidade, aumento de tamanho, rotação de 45° à direita, adição de um elemento ou remoção de sua metade). Diferentemente de Sternberg (1979) que decompunha os problemas, esse método tinha por base a suposição da influência de fatores experimentais (Primi, 1998).

Sintetizando, o modelo de Sternberg (1977) pressupunha uma representação mental de uma possível resposta que seria comparada às opções existentes, enquanto o modelo de Mulholland, Pellegrino e Glaser (1980) partia da comparação das regras criadas para os termos A e B e para os termos C e D. Considerando que um item teria quatro ou seis alternativas de resposta, como ocorre no RIn (Sisto, 2006), a estratégia utilizada seria comparar o termo C com as alternativas D1, D2, D3, D4, D5 e D6, inferir um regra para cada um dos pares possíveis (C-D1, C-D2, C-D3, C-D4, C-D5 e C-D6), comparar a regra A-B com cada uma das demais regras C-D e escolher a alternativa correta a partir da coincidência das regras, eliminando aquelas que não atendem a esse critério. Entretanto, no caso dos itens do RIn (Sisto, 2006) não é possível utilizar esse modelo tal como concebido sem realizar algumas alterações, uma vez que os itens não correspondem à analogias do tipo A-B // C-D.

Assim como Mulholland, Pellegrino e Glaser (1980, citados por Primi, 1998) propuseram um modelo para análise de testes compostos por figuras geométricas, Carpenter, Just e Shell (1990, citados por Primi, 1998) realizaram um estudo para testes de matrizes, utilizando a escala avançada das Matrizes Progressivas de Raven (Raven, 1962), já que haviam críticas quanto aos estudos anteriores envolverem testes de pouca complexidade. De acordo com esse modelo, existiriam três estágios para a resolução de

problemas, denominados de *análise perceptual*, *análise conceitual* e *geração de resposta e seleção da alternativa correta*.

Três etapas compõem a *análise perceptual*. A primeira, denominada *codificação*, relaciona-se à decomposição de termos, considerando os elementos e atributos que os formam. A segunda, chamada *estabelecimento de correspondências*, refere-se à necessidade de agrupar elementos regidos por uma mesma regra, no caso de itens que possuem mais de um elemento. Nesse caso, os sujeitos poderiam prever que formas semelhantes seriam governadas pelas mesmas regras (semelhança de nome) ou que poderia haver correspondência entre os elementos de um termo que foram deixados à parte enquanto os demais foram agrupados. Finalmente, seria realizada a comparação das correspondências encontradas para cada par de elementos, cujos resultados seriam armazenados na memória de curto prazo, servindo de base para a criação de regras conceituais (Carpenter, Just & Shell, 1990).

No segundo estágio, as regras seriam aplicadas à primeira linha da matriz, para em seguida ser realizada uma análise perceptual e conceitual da segunda linha, sendo possível, então, identificar a correspondência entre as regras dos elementos de ambas as linhas. A partir da criação de uma regra geral para explicar as transformações dos elementos dessas duas linhas, seria realizada a análise da terceira linha. No último estágio, “o sujeito usaria as regras inferidas e formularia uma instância para os elementos específicos encontrados nos dois primeiros termos da terceira linha” (Primi, 1998, p.103). Havendo mais de uma regra para os problemas, seria necessário encontrar correspondências entre os elementos da terceira linha com os da segunda e da primeira linha, regidos por uma mesma regra geral. Dessa maneira, seria possível identificar qual regra aplicar aos problemas da terceira linha e

assim encontrar uma resposta e selecionar a alternativa correspondente a ela. Diferentemente dos modelos anteriormente descritos, a proposta de Carpenter, Just e Shell (1990) é mais plausível de ser utilizada como base para o presente trabalho, dadas as características dos itens e do RIn (Sisto, 2006) como um todo.

Jacobs e Vandeventer (1972, citados por Primi, 1998) também criaram uma classificação para a análise das regras inferenciais de testes de raciocínio analógicos com figuras geométricas. As regras previstas por esse modelo eram *identidade* (repetição de um elemento sem mudanças), *forma* (mudança completa da forma), *sombreamento* (alteração completa ou parcial do sombreamento interno do elemento), *tamanho* (aumento ou diminuição de um elemento), *movimento no plano* (o elemento move-se em graus regulares), *séries* (séries progressivas ou não, com números ou aspectos geométricos), *adição de elemento* (um elemento adicionado ou um elemento removido do segundo termo), *adição através da combinação* (os elementos de duas colunas são adicionados e o resultado colocado em uma terceira coluna), *elementos de um conjunto* (elementos aparecem três vezes em uma matriz 3X3), *adição única* (dois termos são sobrepostos e elementos únicos são tratados de modo diferencial, ou seja, adicionados enquanto elementos comuns são removidos), *reflexão* (um elemento é o reflexo em espelho do outro) e *inversão* (o intercâmbio ou troca de algum elemento ou propriedade como forma, sombreamento e/ou posição).

Com exceção das regras de *reflexão* e *inversão*, Carpenter, Just e Shell (1990) reagruparam as demais dez categorias do modelo de Jacobs e Vandeventer (1972) em cinco regras. Chamaram de *constante em linhas* (um mesmo valor se mantém constante dentro das linhas, mas muda entre colunas) as regras de *identidade* e *forma*. *Sombreamento*,

tamanho, movimento no plano, séries e adição de elementos passaram a compor uma única regra, denominada *progressão quantitativa* (um incremento ou decremento quantitativo ocorre entre células adjacentes como atributos de tamanho, posição, número). *Adição através da combinação* foi denominada *adição ou subtração de elementos* (o elemento de uma coluna é adicionado, justaposto, superposto ou subtraído de um elemento de outra coluna para produzir um terceiro termo) e *elementos de um conjunto* como *distribuição de três valores* (três elementos ou três valores de um atributo categórico são distribuídos ao longo de uma linha). Finalmente, *adição única*, nesse modelo é chamada *distribuição de dois valores* (dois elementos ou dois valores de um atributo categórico são distribuídos nas três colunas, onde uma delas é nula).

As taxonomias apresentadas anteriormente servirão como subsídio para responder à primeira hipótese (H1) da qual parte este trabalho, a saber, de que os componentes dos itens do teste influenciam o processo de resposta. Em segundo lugar (H2), supõe-se que regras diferentes implicam distintos processos de resposta e, conseqüentemente, estão relacionadas com o grau de dificuldade dos itens, o que procurar-se-á responder a partir das correlações entre as regras presentes em cada item e o grau de dificuldade dos mesmos.

A H2 envolve não apenas as regras, mas também a dificuldade dos itens. De acordo com Primi (1998) existe uma relação direta entre o entendimento das diferenças individuais e os fatores que interferem no nível de dificuldade dos itens, pois, a partir da identificação de quais fatores permitem aos indivíduos mais hábeis resolverem adequadamente tanto itens fáceis quanto difíceis, pode-se criar testes que estejam melhor atrelados a significados psicológicos. Dessa forma, a partir dos componentes comuns aos itens, que foram

agrupados em ordem hierárquica de acordo com seu nível de dificuldade, pretende-se identificar quais fatores podem influenciar essa característica.

Partindo das duas hipóteses mencionadas, objetiva-se encontrar evidências de validade por meio do processo de resposta para o RIn (Sisto, 2006). Entretanto, será necessária a adaptação dos modelos apresentados anteriormente, uma vez que a estrutura dos itens do RIn (Sisto, 2006) não se enquadra totalmente em apenas um modelo, havendo regras que precisam ser modificadas e outras que não se aplicam ao teste. O modelo criado para a concretização deste trabalho será apresentado quando da descrição dos procedimentos.

Todos os problemas do teste são compostos por figuras geométricas (quadrado, triângulo e círculo), embora alguns itens não disponham dessas três formas. Além disso, há uma célula em branco, para a qual se espera que seja dada a resposta correta. Há itens com 4 e outros com 6 opções de resposta, havendo apenas uma correta. Para efeitos deste estudo, intitulou-se como “componentes” todas as variáveis que compõem os itens do teste, sejam nos problemas ou nas respostas. São esses componentes que permitirão a análise dos itens a partir dos níveis de dificuldade apresentados.

Uma vez descritos os modelos teóricos encontrados na literatura e a proposta de modelo a ser utilizado nesta pesquisa, será abordado de forma mais específica e prática o processo de resposta. Para isso, utilizar-se-á um dos problemas do RIn (Sisto, 2006), escolhido aleatoriamente (Figura 1).

No item utilizado como exemplo (Figura 1) estão presentes as três formas geométricas citadas e a célula para resposta ocupa a segunda coluna da segunda linha,

posição esta que se alterna em outros problemas. De acordo com Sisto (2006), para chegar à resposta correta é possível utilizar mais de uma inferência. Por esse motivo, os itens foram analisados tendo por base a possibilidade de existir mais de uma forma de respondê-los.

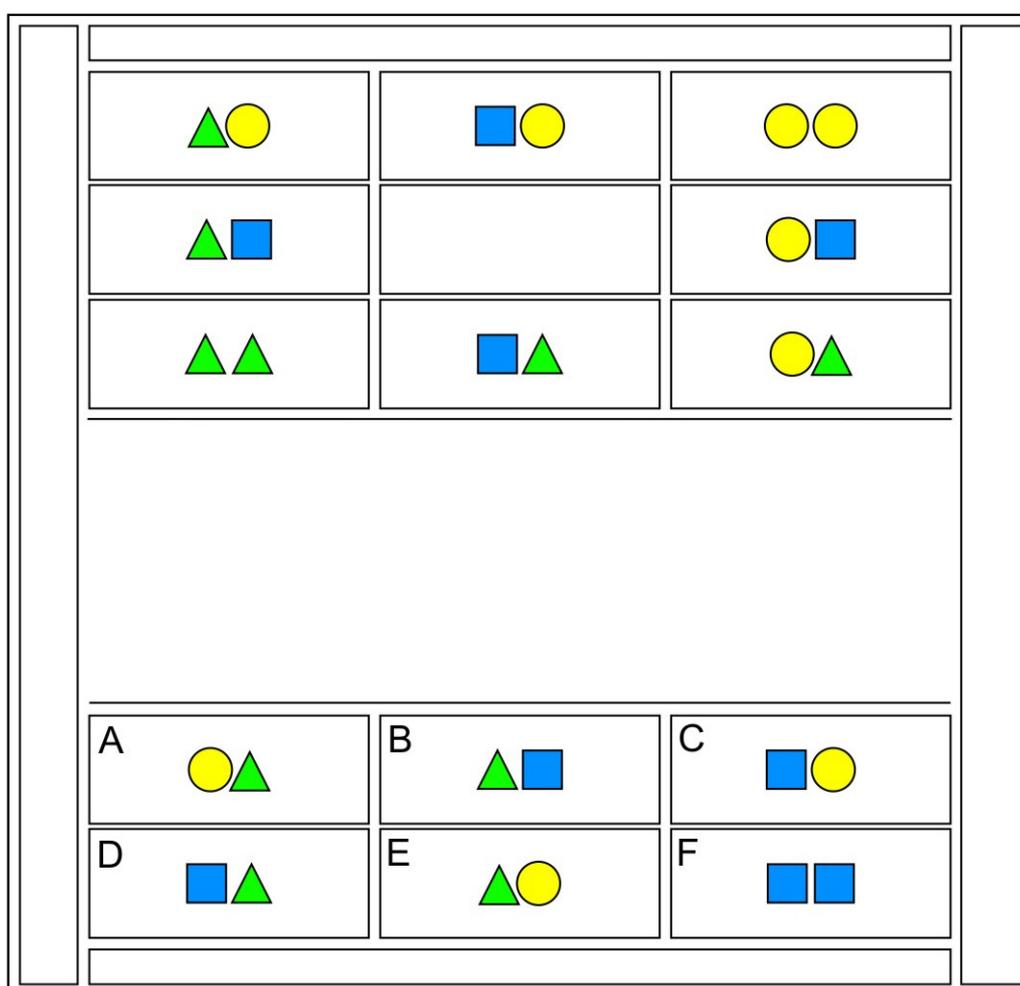


Figura 1 – exemplo de item

Neste caso, analisando o item por colunas (verticalmente) é possível perceber que o primeiro estímulo de cada célula sempre se mantém na célula seguinte. Logo, na primeira coluna o primeiro estímulo é sempre um triângulo, na segunda um quadrado e na terceira um círculo. Já o segundo estímulo segue uma sequência, iniciada por círculo, depois quadrado e por último triângulo. Dessa forma, o primeiro estímulo faltante seria um quadrado (estímulo que se repete como primeiro da coluna) e o segundo também seria um quadrado (estímulo que ocupa a segunda posição da sequência).

Analisando-o por linhas (horizontalmente), verifica-se que o primeiro estímulo de cada célula segue uma sequência, iniciada por triângulo, seguido de quadrado e círculo. Já o segundo estímulo se repete em cada linha, pois na primeira sempre é um círculo e na terceira um triângulo. Dessa forma, chega-se novamente à opção “F”, dois quadrados, como sendo a correta, pois o segundo estímulo evidente nas células da segunda linha também é um quadrado. Percebe-se que este item possui as regras de identidade e sequencia (Tabela 2), independentemente da forma como seja resolvido.

Para o alcance dos objetivos deste trabalho é preciso, portanto, utilizar-se de um modelo que contemple as regras existentes em cada item do teste. Uma vez identificadas tais regras, é preciso lançar mão da psicometria, não apenas pela necessidade estatística, mas também como subsídio para a interpretação dos resultados encontrados. Para Arias (1996) um dos objetivos mais importantes da psicometria é determinar a pontuação verdadeira de um teste, eliminando os erros de medida. Para tal, foram elaborados modelos formais em que aparecem três componentes básicos: as pontuações observadas, os erros de medida e as pontuações verdadeiras. Dentre estes modelos destacam-se a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI).

De uma forma geral, na Teoria Clássica dos Testes há uma preocupação com o resultado final total, ou seja, com o valor obtido a partir da soma das respostas atribuídas por uma pessoa ao teste. Por sua vez, a partir da Teoria de Resposta ao Item procura-se entender cada item e a probabilidade de cada um deles ser respondido de forma correta ou não – no caso de testes de aptidão – ou aceitos ou rejeitados, nos casos de testes de interesses e atitudes. Logo, enquanto na TCT se busca produzir testes de qualidade, na TRI procura-se produzir tarefas (itens) de qualidade (Pasquali, 2003).

Os postulados da TCT permitem várias derivações. Entretanto, é importante levantar algumas críticas que lhe são feitas, como a) o fato de ter sido desenvolvida exclusivamente a partir de dados de medida da inteligência, b) os psicólogos que a criaram terem preocupação predominantemente estatística e c) o materialismo presente na época ter sido a base de sua orientação. Entenda-se por materialismo a crença de que toda realidade se reduz a propriedades físicas (Pasquali, 2003).

Dentre os modelos que surgiram com o intuito de solucionar os problemas da TCT, destaca-se a TRI (Arias, 1996; Pasquali, 2003). Apesar dos pressupostos dessa teoria terem sido levantados por Thurstone nos anos 1930, só após os anos 1950 os psicometristas começaram a encontrar soluções para os problemas levantados, baseados nos trabalhos de Lazarsfeld (1950), Lord (1952) e Rasch (1960). Tais trabalhos tornaram-se a base da TRI, mas somente a partir de 1980 esta teoria passou a ser utilizada de forma mais ampla, uma vez que os softwares existentes até então não conseguiam lidar com a complexidade dos algoritmos matemáticos propostos (Pasquali, 2007).

A habilidade latente, denominada *teta*, não pode ser mensurada diretamente e equivale à dimensão psicológica que está sendo medida em um teste. A TRI é um modelo

matemático a partir do qual se busca especificar a relação entre a probabilidade de uma pessoa acertar determinado item e a habilidade ou traço latente requerido. Tal relação é descrita pela Curva Característica do Item (CCI) e se baseia na unidimensionalidade e na independência local (Conde & Laros, 2007). Por unidimensionalidade entende-se que um traço latente ou aptidão predominante é responsável pela realização de um grupo de itens. Já o postulado de independência local afirma que as respostas dos sujeitos a dois itens, quaisquer que sejam, são estatisticamente independentes. Dessa forma, a probabilidade de acerto ou erro de um item depende apenas do traço latente ou aptidão apresentado pelo sujeito (Arias, 1996; Conde & Laros, 2007; Pasquali, 2007).

Muitos modelos matemáticos poderiam expressar a CCI, mas três são predominantes, tendo por base o número de parâmetros que cada um considera. Esses parâmetros são a dificuldade, discriminação e resposta dada ao acaso (Pasquali, 2007). O modelo utilizado nas análises aqui realizadas corresponde ao modelo de um parâmetro ou de Rasch (1960), uma vez que trabalhou-se apenas com a dificuldade dos itens.

De acordo com os fundamentos teóricos trazidos até aqui, evidencia-se que as características do item, como seu grau de dificuldade, determinam a probabilidade de acertos ou erros, aliadas à grandeza do *teta* da pessoa que responde ao teste (Pasquali, 2007). Dessa maneira, as evidências de validade pelo processo de respostas serão encontradas a partir da análise dos itens. Além disso, pretende-se também contribuir com um novo estudo utilizando essa evidência.

OBJETIVO GERAL

Buscar evidências de validade pelo processo de resposta para o Teste de Raciocínio Inferencial – RIn (Sisto, 2006)

OBJETIVOS ESPECÍFICOS

- Identificar o índice de dificuldade de cada item do teste;
- Realizar uma análise qualitativa dos componentes e das regras presentes em cada item;
- Criar um modelo descritivo das regras identificadas que permita a classificação de todos os itens do teste;
- Correlacionar o nível de dificuldade com os componentes dos itens e com as regras identificadas.
- Identificar quais regras tem maior relação com a dificuldade dos itens.

MÉTODO

Fonte de dados

Para a realização da presente pesquisa foi utilizado um banco de dados contendo as respostas de 503 testes realizados por pessoas que haviam participado de um processo de avaliação psicológica para fins organizacionais durante os anos de 2006 a 2009. As características do grupo que compôs o banco de dados, no que tange ao sexo e escolaridade, estão dispostas na Tabela 1. Percebe-se que a amostra teve uma predominância de pessoas do sexo masculino (78,9%) e com Ensino Superior (63,6%).

Tabela 1. Distribuição da amostra quanto ao sexo e escolaridade

	Feminino		Masculino		Total	
	N	%	N	%	N	%
Ensino Médio	27	25,5	156	39,3	183	36,4
Ensino Superior	79	74,5	241	60,7	320	63,6
Total	106	21,1	397	78,9	503	100

Quanto à faixa etária, as idades variaram entre 16 e 70 anos. A idade média dos participantes foi de 29,42 anos (mediana de 28 e moda de 27 anos) e o desvio-padrão de 8,596. Essa frequência pode ser visualizada na Figura 2.

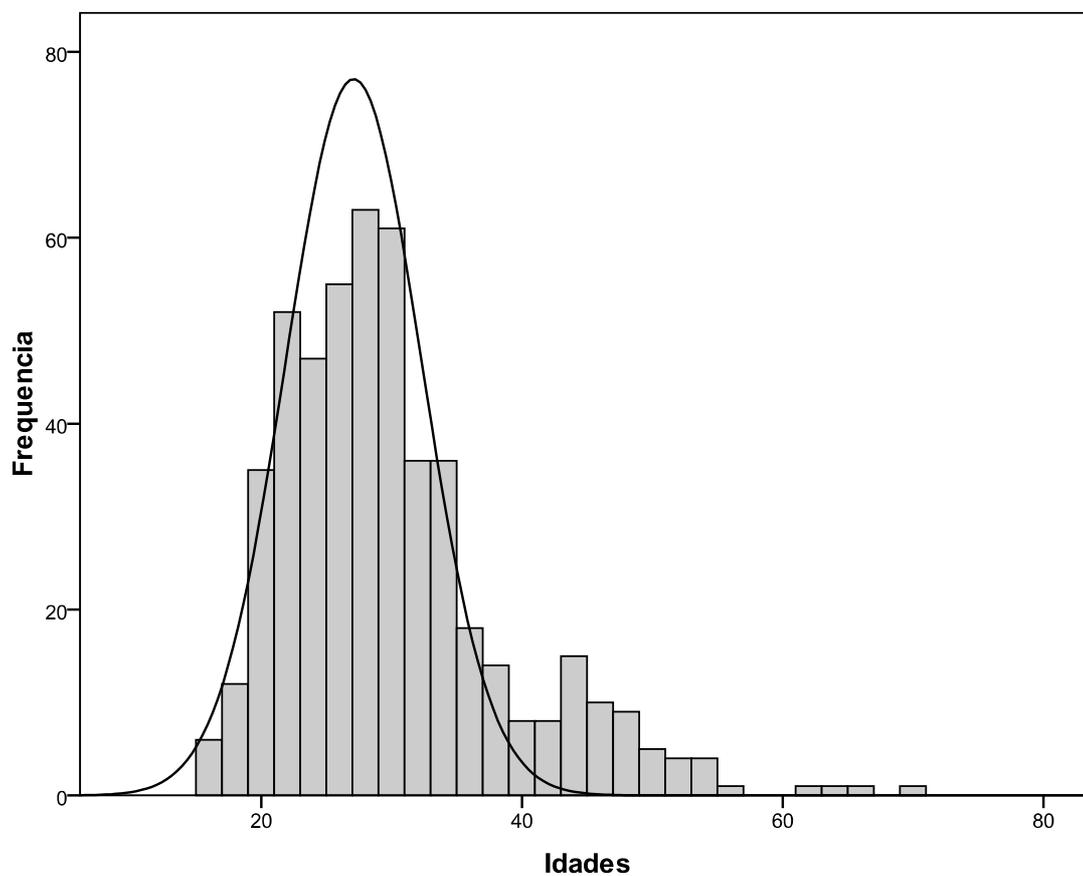


Figura 2. Frequência dos participantes por idade

Instrumento

O instrumento utilizado para análise dos dados é o Teste de Raciocínio Inferencial – RIn (Sisto, 2006), que avalia a inteligência ou fator *g* a partir do raciocínio inferencial. Pode ser aplicado em pessoas de 10 a 70 anos, individual ou coletivamente para um máximo de 30 pessoas e com um tempo pré-determinado de 25 minutos para a realização.

Sua aplicação deve obedecer aos critérios essenciais esperados em um ambiente de testagem, como iluminação e temperatura adequadas, ausência de ruídos distratores e os participantes devem estar a uma distância que permita a realização independente do instrumento (Sisto, 2006). Tais recomendações foram seguidas quando da coleta dos protocolos que serão utilizados neste estudo. A seguir serão apresentadas algumas informações referentes aos estudos psicométricos realizados pelo autor do teste e que fazem parte de seu manual de utilização. Os estudos serão descritos tais como apresentados por seu autor, o que não impede que algumas críticas sejam traçadas. Como exemplos, pode-se citar o fato da análise fatorial não ser descrita de forma clara, não ser apresentada a carga fatorial de cada item e ter como conclusão a realização de novos estudos para confirmar a existência de um único fator. O mesmo ocorre em relação a alguns procedimentos adotados, como no caso da dimensionalidade ter-se optado por utilizar apenas 10 dos 40 itens.

As quarenta questões do RIn (Sisto, 2006) estão divididas em quatro séries (A, B, C e D) compostas por doze, oito, seis e quatorze itens, respectivamente. Cada questão é apresentada em matrizes (linhas X colunas) de três por dois ou três por três, com exceção da série B, cujas linhas da matriz de dois por cinco são agrupadas. Em cada problema existem células onde estão de uma a quatro figuras geométricas (quadrado, círculo ou triângulo) agrupadas horizontal, vertical ou transversalmente, além de uma célula em branco. A pessoa que responde o teste deve escolher, dentre as opções de resposta fornecidas, qual completaria corretamente o raciocínio apresentado, considerando que essa resposta ocuparia o lugar da célula em branco.

Antes de iniciar a primeira série do teste é apresentado um exemplo que serve como treino para que o avaliado entenda o funcionamento do teste. No final da página de exemplo é apresentada a resposta correta do exercício. Cada uma das séries do teste está organizada por ordem de dificuldade, cujo item inicial permite a compreensão do raciocínio esperado para a realização dos exercícios seguintes. A correção é realizada atribuindo-se um ponto a cada resposta correta. Somando-se os acertos das quatro séries obtém-se a pontuação total, que deve ser comparada às Tabelas Normativas do manual, onde serão verificados o percentil e a classificação obtidos. O manual do teste contempla tabelas para os estados de São Paulo e Minas Gerais, tanto por níveis de ensino (de Fundamental à Especialização) quanto por faixa etária e os resultados podem ser interpretados pela pontuação bruta ou por medidas derivadas do modelo Rasch (Sisto, 2006).

A coleta de dados para a normatização do teste ocorreu no ano de 2005, contemplando 1102 pessoas de seis cidades do estado de São Paulo e duas de Minas Gerais. A idade variou de 11 a 73 anos ($M=18,99$ anos e $DP=6,06$,) e 60% da amostra pertencia ao sexo feminino. Em relação à escolaridade, 46,55% tinham Ensino Superior, 35,57% Ensino Médio e 17,88% Fundamental. Os estudantes de Ensino Superior provinham de cursos de áreas distintas, como Administração, Educação Física, Ciência da Computação e Medicina, dentre outros.

Quanto às análises psicométricas, para escalonar os itens em termos unidimensionais foi utilizada a medida Rasch, cuja média foi de 0,00 (desvio-padrão de 1,28 e erro-padrão de 0,08). Os índices de *infit* de todos os itens foram adequados, enquadrando-se nos intervalos de 0,70 a 1,30 e, em relação ao *outfit*, apenas dois itens (A12 e C4) não se ajustaram ao esperado. Entretanto, ambos os valores encontrados (1,31) estão

bem próximos ao esperado e por isso os itens foram mantidos. Houve também uma boa adequação entre a medida Rasch e a pontuação máxima obtida no teste (Sisto, 2006).

As análises a partir da Teoria Clássica dos Testes indicaram que a correlação item-total de todos os itens foi superior a 0,30. A dificuldade média do teste foi de 46,51, sendo 27,81 para a série A, 40,80 para a série B, 57,12 para a série C e 60,31 para a série D. De acordo com estes resultados o autor considerou que os itens estavam bem alocados, uma vez que as séries possuíam níveis de dificuldade crescentes e também havia correspondência entre as porcentagens mais altas e as alternativas corretas, com exceção de apenas três itens, cujos percentuais mais altos não foram nas respostas corretas. Também em três itens houve um percentual mais alto de respostas em branco do que corretas e quinze itens tiveram um percentual de acerto abaixo de 50%.

Em relação às evidências de validade, os estudos foram apresentados em dois tópicos no manual do teste. O primeiro refere-se às evidências relativas à estrutura interna e conta com três estudos, sendo eles o *Funcionamento Diferencial do Item (DIF)*, *validade de constructo através da análise fatorial* e verificação da possibilidade de aceitação da *unidimensionalidade* do teste pela análise de resíduos usando o modelo Rasch. No segundo tópico são apresentados dois estudos de validade de critério, sendo o primeiro relativo *ao desenvolvimento* e o segundo *ao desempenho escolar de estudantes de Ensino Médio*.

O Funcionamento Diferencial dos Itens substitui o termo viés. Tem por objetivo identificar os itens em que há diferentes probabilidades de acerto para subgrupos de uma determinada população, na qual as pessoas têm o mesmo nível de habilidade na variável avaliada. Utilizando-se a TRI, cuja vantagem é a invariância dos parâmetros para

subpopulações e os valores de p dependentes desta, verificou-se que apenas dois itens (A1 e B6) diferenciavam os gêneros e, como cada um favorecia um sexo distinto, as interferências se contrabalanceavam. De acordo com o autor do teste, esse tipo de estudo relaciona-se às implicações éticas, jurídicas e sociais ao permitir uma subavaliação sistemática das capacidades de alguns grupos em função de variáveis como sexo, cultura ou outras características diferenciadoras, tendo por justificativa a neutralização desse tipo de efeito (Sisto, 2006).

As séries do RIn (Sisto, 2006) foram alocadas por ordem geral de dificuldade e cada uma delas possui tanto itens fáceis quanto difíceis. Uma análise fatorial foi realizada considerando separadamente as quatro séries. A partir de medida de adequação de Kaiser-Meyer-Olkin ($KMO = 0,69$) e do teste de esfericidade de Bartlett ($\chi^2 = 796,893$, $df = 6$ e $p = 0,000$) pode-se aferir que a variância explicada não foi muito alta, embora a matriz fosse apropriada para a extração de fatores. Dessa forma, sugeriu-se cautela na interpretação das informações. Foi utilizado como método de extração a fatoração por eixos principais (PAF) com rotação *oblimin*. Apenas um fator alcançou *eigenvalue* igual ou superior a 1, explicando 52,41% da variância. Os resultados encontrados sugeriram a unidimensionalidade do teste e permitem inferir que o mesmo é composto por apenas um fator geral, fornecendo evidências de validade em relação à estrutura interna, embora seu autor tenha sugerido a realização de outras análises para confirmar a existência de um único fator.

Quanto à dimensionalidade, considerando os resíduos de uma análise de componentes principais, um fator explicou 3,5 de 40 unidades de variância residual, uma por item e a proporção de Yardstick foi de 12,6/1. Para verificar a possibilidade de ser

necessário mais de um fator, uma vez que o *eigenvalue* foi superior a 1,4, foi realizada uma nova análise contendo cinco itens (D1, D2, D3, D7 e D8) com cargas positivas e cinco (A2, A6, A7, A10 e B2) com cargas negativas. Para calcular as pontuações das pessoas, cada conjunto de itens foi considerado como um teste independente e submetido ao modelo de Rasch. Posteriormente, foram correlacionadas as pontuações de uma mesma pessoa em cada um dos conjuntos, encontrando-se um valor de $r=0,71$ (correlação de Pearson corrigida pela fórmula de Spearman-Brown). Tal coeficiente possibilita a interpretação de que os dois conjuntos com resíduos mais discrepantes concentram boa parte da variância explicada pela comunalidade entre eles, possibilitando a aceitação da unidimensionalidade do teste (Sisto, 2006).

As evidências de validade relativas ao desenvolvimento tiveram por objetivo verificar a necessidade de serem estabelecidas normas por idade e gênero, caso as diferenças encontradas não fossem atribuídas ao acaso. Apesar de muito baixas, as correlações entre os sexos em separado foram significativas, sendo que para o grupo masculino houve um aumento da pontuação no teste com o avanço da idade e para o grupo feminino ocorreu o oposto. Em um segundo estudo os grupos foram divididos em função da idade, sendo criado um grupo para pessoas com até 22 anos e outro para pessoas acima dessa idade. Novas correlações foram calculadas e constatou-se um aumento da pontuação até os 22 anos e uma diminuição para as demais faixas etárias, o que vai ao encontro ao que normalmente é encontrado na literatura. A diminuição precoce da pontuação, por volta dos 20 a 30 anos, permite inferir que trata-se da inteligência fluida ou fator *g*, uma vez que a inteligência cristalizada tem uma diminuição mais tardia (Sisto, 2006).

Para verificar a correlação entre a pontuação no teste e o desenvolvimento escolar, 228 estudantes de nível socioeconômico médio ou superior, com idade entre 14 e 19 anos ($M=16,60$ e $DP=1,19$) e predominantemente do sexo feminino (62,72%) tiveram as médias de suas notas bimestrais correlacionadas às pontuações obtidas no teste. Apesar de algumas diferenças nas médias das notas entre os gêneros, a divisão da amostra em função dessa variável não provocou grandes alterações nos resultados, que indicaram correlações mais altas e significativas entre os alunos do terceiro ano do Ensino Médio, seguidos pelos alunos do segundo e primeiro ano, respectivamente. Tal resultado pode ter sido gerado pelo fato dos alunos da última série se dedicarem mais aos estudos em função de sua preparação para o vestibular e são mais um indicativo de validade de critério para o RIn (Sisto, 2006).

Os estudos de precisão em relação às faixas etárias e ao sexo foram realizados a partir do coeficiente alfa de Cronbach e do método das duas metades de Spearman-Brown, cujos coeficientes encontrados foram de 0,71 a 0,85 e de 0,73 a 0,93, respectivamente. Também foi calculada a precisão do teste como um todo pelo modelo Rasch e todos os indicadores foram satisfatórios. Além disso, verificou-se que a exclusão de qualquer item do teste não aumentaria sua precisão.

Além dos estudos citados que, conforme dito no início deste capítulo, fazem parte do manual do RIn (Sisto, 2006), Rabelo (2008) buscou evidências de validade para as Matrizes Progressivas Avançadas de Raven utilizando o RIn (Sisto2006), considerando que ambos mediam o fator g tal como proposto por Spearman (1927). Sua amostra foi composta por 291 alunos do Ensino Superior, cuja faixa etária foi dos 17 aos 63 anos de idade ($M=28$ e $DP=9,42$). Todas as correlações encontradas entre os dois testes foram positivas e, com exceção da série D, cuja correlação foi baixa ($r = 0,30$), as correlação com as três demais

séries do RIn (Sisto, 2006) foram moderadas ($r = 0,52$ com a série A, $r = 0,47$ com a série B e $r = 0,49$ com a série C) e com a pontuação total foi alta ($r = 0,61$). O autor concluiu que os resultados foram ao encontro do que era esperado, indicando que tanto o Raven quanto o RIn (Sisto, 2006) mediam o mesmo construto. Não foram encontradas diferenças significativas no que se referia ao sexo e à idade, sendo sugerido que novos estudos fossem realizados. Embora não fosse o objetivo central do estudo, Rabelo (2008) também verificou a precisão do RIn (Sisto, 2006) por meio do coeficiente alfa de Cronbach, encontrando um coeficiente de 0,80.

Procedimentos

A presente pesquisa foi realizada a partir de dados coletados previamente. Entretanto, no momento em que realizaram o teste todos os participantes assinaram um Termo de Consentimento Livre e Esclarecido e foram informados quanto à possibilidade da utilização posterior de seus resultados, resguardando-se a preservação e sigilo de qualquer informação que pudesse identificá-los. Este projeto foi encaminhado e aprovado pelo Comitê de Ética e Pesquisa da Universidade São Francisco, estando registrado no CONEP sob o número 0062.0.142.000-11.

Como não foi necessária a aplicação do teste, uma vez que o estudo teve como fonte um banco de dados pré-existente, o primeiro passo para o alcance dos objetivos deste trabalho foi escalonar os itens quanto à dificuldade. Foram identificados os índices de dificuldade de cada um dos 40 itens do teste, a partir da Teoria de Resposta ao Item (TRI),

mais especificamente do modelo de Rasch (1960). Com base nessas informações, os itens foram classificados por ordem de dificuldade. Em seguida cada item foi decomposto considerando o tamanho da matriz (linhas e colunas); a quantidade de estímulos presentes no problema e nas opções de resposta; a célula (linha e coluna) onde se encontra a lacuna para resposta; o agrupamento de estímulos em uma mesma célula, fosse no problema ou nas opções de resposta; a direção em que os itens são apresentados (vertical, horizontal ou transversal) e o tipo ou quantidade de inferências possíveis (número de regras) para se chegar à resposta correta.

Para confirmar a confiabilidade das características atribuídas aos itens e das regras identificadas pelo autor deste estudo, foi solicitado a outros dois psicólogos com experiência na área de avaliação psicológica e usuários do RIn (Sisto, 2006) que realizassem a mesma análise, individualmente. Uma descrição dos resultados dessa análise será apresentada no tópico Análise de Juízes.

Com base nos modelos discutidos na introdução deste trabalho e a partir da identificação das regras e componentes de cada item do RIn (Sisto, 2006), foi criado um modelo que tinha por objetivo principal contemplar as regras presentes em todos os itens do teste, permitindo assim que fossem realizadas as correlações necessárias para o alcance dos objetivos propostos.

Procurou-se estabelecer para cada regra uma nomenclatura que fosse próxima à adotada por outros autores (quando correspondentes), assim como foram acatadas sugestões dadas pelos juízes que avaliaram o teste. Uma descrição desse modelo será apresentada a seguir (Tabela 2).

Tabela 2. Taxonomia para análise dos itens

Regra	Descrição
Identidade	Repetição de um estímulo ou conjunto de estímulos
Inversão	Alternância na ordem do estímulo ou de um conjunto de estímulos, mas não de posição (horizontal, vertical ou transversal)
Soma	Um ou mais estímulos são somados (agrupados ou unidos)
Sequencia	Repetição de estímulos em uma sequencia constante
Subtração	Um ou mais estímulos apresentados são suprimidos
Movimento	Um ou mais estímulos se movem (horizontal, vertical ou transversalmente)
Quantidade	Há uma determinada quantidade de cada estímulo por linhas ou colunas

Uma vez decomposto cada item, foram verificadas as correlações entre o nível de dificuldade do mesmo e cada um de seus componentes. Em outras palavras, se haviam correlações entre os distintos níveis de dificuldade e os atributos dos itens. O mesmo foi feito em relação às regras identificadas. A partir dessa análise, foi possível verificar quais elementos exerciam maior influência sobre o processo de resposta.

Considerando que um mesmo item poderia ter mais de uma regra, também foram verificadas as correlações entre estas e realizada uma análise de regressão para que fosse identificado quais regras contribuiriam de forma mais consistente para a dificuldade dos itens e, conseqüentemente, interfiram no processo de resposta.

Todas as análises foram realizadas utilizando os softwares estatísticos SPSS (versão 18.0) e Winsteps.

APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

De acordo com o modelo de Rasch (1960) a característica de um item que pode influenciar o desempenho de uma pessoa é a sua dificuldade, representada na Teoria de Resposta ao Item (TRI) pela letra “*b*”. Para verificar a equivalência da dificuldade dos itens da amostra utilizada neste estudo em relação aos dados da amostra de normatização do teste, foram comparadas as dificuldades dos itens de cada série de ambas as amostras, cujos resultados serão dispostos nas Tabelas 3, 4, 5 e 6. Os parâmetros utilizados para considerar a magnitude das correlações são os propostos por Sisto (2007), que classifica as correlações como nulas (até 0,20), baixas (de 0,20 a 0,40), moderadas (de 0,40 a 0,60), altas (de 0,60 a 0,80) e muito altas (acima de 0,80).

Tabela 3. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série A

Item	Dificuldade		Correlação	Significância
	Amostra normativa	Amostra atual		
A1	-2,77	-3,11		
A2	-1,08	-0,65		
A3	-1,2	-0,67		
A4	-2,63	-2,34		
A5	-2,32	-2,11		
A6	-0,39	0,14	0,985**	0,000
A7	-1,59	-1,29		
A8	-0,02	0,04		
A9	-1,46	-1,44		
A10	-1,19	-1,21		
A11	0,02	0,18		
A12	1,55	2,10		
Média	-1,09	-0,86		

**significante a 0,01

A Tabela 3 corresponde às médias das dificuldades dos itens da Série A. Foi encontrada uma correlação muito alta ($r = 0,985$) entre as médias de ambas as amostras. Segundo Sisto (2006) essa é a série mais fácil do teste, o que foi confirmado pelos dados da amostra atual, uma vez que foi a única das quatro séries cuja média de dificuldade dos itens foi negativa ($M = -0,86$). Deve-se considerar, porém, que embora a correlação total entre os itens dessa série tenha sido alta, se os itens fossem classificados por índice de dificuldade, apenas seis (50%) ocupariam a mesma ordem.

Na Tabela 4 estão dispostos os resultados da correlação entre os itens da Série B. Apesar da média para a amostra atual ser superior à da amostra de padronização, a correlação encontrada ($r = 0,967$) também foi muito alta (Sisto, 2007), embora três dos oito itens não ocupem a mesma posição hierárquica em ambas as amostras no que tange à dificuldade.

Tabela 4. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série B

Item	Dificuldade		Correlação	Significância
	Amostra normativa	Amostra atual		
B1	-0,71	0,04	0,967**	0,000
B2	-0,42	-0,07		
B3	-0,70	-0,53		
B4	-0,54	-0,03		
B5	-0,81	-0,78		
B6	0,38	0,85		
B7	1,12	1,90		
B8	0,02	0,42		
Média	-0,20	0,22		

**significante a 0,01

A correlação entre a dificuldade dos itens da Série C foi de $r = 0,997$, sendo esse o maior valor encontrado entre as quatro séries do teste (Tabela 5). Embora Sisto (2006) tenha mencionado que as séries do teste possuíam uma dificuldade crescente, o que tornaria a série D a mais difícil, para a amostra atual esse resultado não se confirmou, uma vez que a Série C apresentou a maior média de dificuldade dentre as demais. Tal resultado pode ser em consequência de que alguns dos itens com menor dificuldade também pertencerem a série D, o que faz com que a média (Tabela 6) seja influenciada por esses valores. Outro dado que pode ser verificado na série C (Tabela 5) é que essa foi a única dentre as quatro séries onde todos os itens, se escalonados por ordem de dificuldade, apresentariam a mesma posição hierárquica em ambas as amostras.

Tabela 5. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série C

Item	Dificuldade		Correlação	Significância
	Amostra normativa	Amostra atual		
C1	-0,06	0,01	0,997**	0,000
C2	-0,24	-0,14		
C3	0,51	0,56		
C4	1,66	1,93		
C5	0,83	0,90		
C6	0,97	1,09		
Média	0,61	0,73		

**significante a 0,01

Os itens da Série D também tiveram um correlação muito alta ($r = 0,964$), conforme apresentado na Tabela 6. Dos quatorze itens que compõe essa série, sete ocupam a mesma

posição no que tange à dificuldade em ambas as amostras e sete apresentam posições distintas.

Tabela 6. Correlação entre a dificuldade dos itens da amostra de padronização do teste e da amostra atual para a Série D

Item	Dificuldade		Correlação	Significância
	Amostra normativa	Amostra atual		
D1	-1,25	-2,34		
D2	-1,35	-2,71		
D3	-0,35	-0,51		
D4	2,08	2,38		
D5	2,04	2,30		
D6	1,27	0,99		
D7	0,09	-0,66	0,964**	0,000
D8	0,17	-0,44		
D9	1,13	1,00		
D10	1,60	1,48		
D11	1,74	1,72		
D12	0,82	-0,27		
D13	1,34	0,65		
D14	1,72	0,62		
Média	0,78	0,30		

**significante a 0,01

Os dados apresentados indicam que todas as correlações encontradas foram muito altas e positivas, além de significativas. Considerando os quarenta itens do teste, a correlação entre a dificuldade das duas amostras também foi muito alta ($r = 0,926$; $p < 0,01$). Dessa forma, pode-se inferir que, independentemente da amostra utilizada, o nível de dificuldade dos itens é equivalente. Conclui-se, portanto, que os resultados obtidos a partir das análises que serão feitas com o banco de dados atual podem ser considerados válidos para amostras distintas do teste. Entretanto, ao levar em consideração a dificuldade de cada

item individualmente, foi possível perceber que apenas na série C, a mais difícil nessa amostra, os itens ocupam a mesma ordem. Também não se confirmou a informação de Sisto (2006) quanto ao escalonamento das séries por ordem de dificuldade, uma vez que a série D apresentou uma dificuldade inferior à série C. A seguir serão apresentados os índices de *infit* e *outfit* de cada item da amostra atual (Tabela 7).

Tabela 7. Índices de *infit* e *outfit* dos itens do teste

Item	<i>Infit Mnsq</i>	<i>Outfit Mnsq</i>	Item	<i>Infit Mnsq</i>	<i>Outfit Mnsq</i>
A1	1,06	1,17	C1	0,85	0,78
A2	0,89	0,78	C2	1,03	1,03
A3	0,91	0,81	C3	0,98	0,99
A4	0,88	0,63	C4	1,13	1,33
A5	0,89	0,65	C5	0,97	0,96
A6	0,84	0,80	C6	0,91	0,91
A7	0,91	0,68	D1	1,01	1,19
A8	0,88	0,84	D2	0,92	0,84
A9	0,87	0,62	D3	0,91	0,85
A10	0,86	0,68	D4	1,00	1,14
A11	1,01	1,02	D5	1,18	1,50
A12	1,40	2,47	D6	1,18	1,28
B1	0,91	0,83	D7	1,05	1,04
B2	0,98	0,96	D8	0,86	0,77
B3	1,12	1,28	D9	0,99	1,02
B4	0,88	0,81	D10	1,19	1,32
B5	0,92	0,82	D11	1,06	1,33
B6	1,08	1,09	D12	1,09	1,11
B7	1,02	1,13	D13	1,14	1,21
B8	1,11	1,12	D14	1,08	1,12

Em relação aos índices de *infit* e *outfit*, no manual do teste foi mencionado que todos os itens estavam adequados quanto ao primeiro e apenas os itens A12 e C4 não se

ajustavam ao intervalo esperado quanto ao segundo. Considerando os parâmetros de Linacre (2002) segundo o qual o intervalo de *infit* e *outfit* deve ser entre 0,70 e 1,50, na amostra atual apenas o item A12 apresentou um *outfit* bastante acima do desejado, além de ter sido o único com correlação ponto bi-serial negativa (-0,13). Ressalta-se que os juízes que analisaram o teste discordaram quanto à alternativa dada como correta pelo autor do teste para esse item e concordaram quanto à outra opção de resposta ser a correta. Por sinal, a resposta considerada correta pelos juízes é a mesma escolhida pelas pessoas de maior habilidade que responderam ao teste. Todas essas observações podem indicar que há um problema em relação a esse item, já que pessoas com maior habilidade o erram os parâmetros psicométricos não indicam resultados satisfatórios.

Análise de Juízes

De acordo com Urbina (2007) quando existe a possibilidade de um elemento subjetivo na avaliação de um teste podem haver erros que interfiram em sua fidedignidade. Dessa forma, se um item fosse decomposto de forma errônea, ou a regra inerente a esse item fosse identificada incorretamente, as correlações encontradas não seriam reais e os resultados, conseqüentemente, prejudicados. Como descrito na introdução, nenhum dos modelos teóricos encontrados contemplava totalmente as regras dos itens do RIn (Sisto, 2006), tendo sido necessária a criação de um modelo adaptado (disponível na Tabela 1) para tal propósito.

Partindo do pressuposto de que o modelo teórico adaptado pelo autor poderia conter erros, foi solicitado a dois psicólogos com conhecimento do teste e experiência prática em avaliação psicológica, que analisassem cada item e descrevessem as regras encontradas, as quais foram posteriormente correlacionadas (Tabela 8). Tal procedimento, similar ao que é nomeado por Urbina (2007) por *fidedignidade do avaliador*, teve por objetivo confirmar a confiabilidade das regras criadas e das interpretações feitas a partir de suas análises. Além disso, vão ao encontro de uma das preocupações explicitadas nos Standards (APA, AERA & NCME, 1999) quanto à necessidade de que sejam verificados se fatores irrelevantes não poderiam influenciar as avaliações feitas por determinados juízes.

Tabela 8. Correlação entre a análise de juízes quanto às regras identificadas em cada item

	Regras Horizontais		Regras Verticais	
	Autor	Juiz 1	Autor	Juiz 1
Juiz 1	0,816**	-	0,994**	-
Juiz 2	0,640**	0,793**	0,943**	0,936**

**significativa a 0,01

Todas as correlações encontradas foram positivas, de moderadas a muito altas e significativas. Em relação às inferências horizontais, a correlação entre as regras identificadas pelo autor e o primeiro juiz foi alta (0,816) e com o segundo juiz moderada (0,640). Já entre estes dois últimos, a correlação foi alta (0,793). Quanto às regras verticais, todas as correlações foram muito altas, entre 0,936 e 0,994 (Sisto, 2007). Ressalta-se que o

juiz 2 não identificou regras para alguns itens no sentido horizontal, o que pode explicar o fato das correlações com este juiz e os demais terem sido um pouco menores nesse quesito, embora significativas. Constatou-se, frente a esses resultados, a confiabilidade do modelo criado.

Correlações entre os componentes dos itens do teste e a dificuldade dos itens

Cada item do teste foi analisado individualmente e ao todo foram extraídos quatorze componentes que contemplam todos os itens, sendo eles 1) localização da célula para resposta à esquerda, 2) centro ou 3) direita; 4) quantidade de opções de resposta e 5) localização da célula contendo a resposta correta; 6) quantidade de linhas, 7) colunas e 8) tamanho da matriz; 9) tipos de estímulos diferentes no problema e 10) na resposta; 11) quantidade total de estímulos no problema e 12) na resposta e 13) agrupamento de estímulos no problema e 14) na resposta. Paralelamente à apresentação dos resultados das correlações, será realizada uma descrição e fornecidos alguns exemplos de cada um dos componentes identificados. Para isso, será utilizado como modelo um item do teste (Figura 3).

Todos os itens do teste possuem uma célula em branco, a qual se espera que seja preenchida com uma das opções de resposta fornecidas, de forma que a opção escolhida complete corretamente o raciocínio desejado. A localização dessa célula não é fixa. No caso da Figura 3, essa célula está na terceira coluna da segunda linha, mas em outros itens

pode ocupar uma posição diferente. Dessa forma, os problemas foram divididos como tendo a lacuna para resposta à esquerda, no centro ou à direita (como no caso da Figura 3).

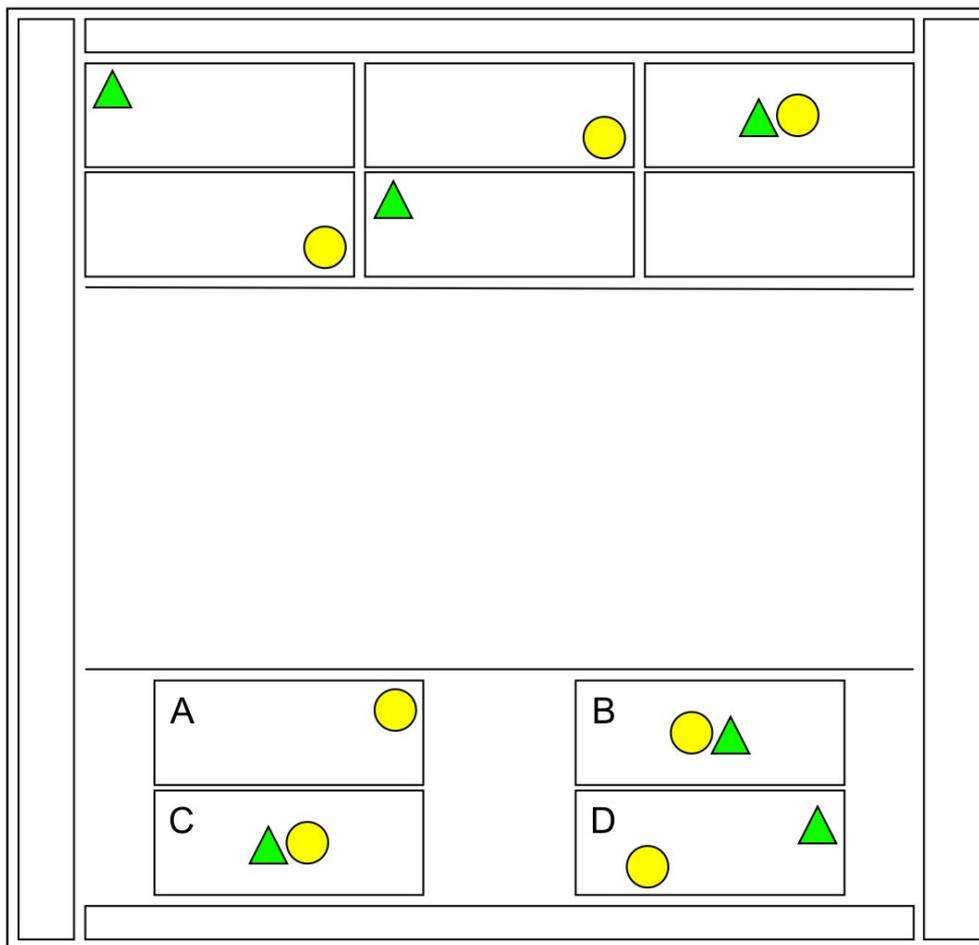


Figura 3. Item utilizado como modelo para descrição dos componentes

Os resultados (Tabela 9) indicaram que esse componente não interfere na dificuldade dos itens, uma vez que não há nenhuma correlação significativa entre esta última e a localização da célula em branco no problema. Dessa maneira, pode-se considerar

que, independentemente de onde esteja localizada a célula em branco no problema, a dificuldade do item não sofre alteração.

Tabela 9. Correlação entre a dificuldade dos itens e a localização da célula para resposta no problema

Componente	Correlação	Significância
Localização da célula para resposta à esquerda	-0,221	0,170
Localização da célula para resposta no centro	0,174	0,282
Localização da célula para resposta à direita	0,007	0,964

Em relação às opções de resposta, estas podem variar em número (4 ou 6 opções diferentes) e na posição onde a célula com a resposta correta está localizada.

Tabela 10. Correlação entre a dificuldade dos itens com a quantidade de opções de resposta e com a localização da célula com a resposta correta

Componente	Correlação	Significância
Quantidade de opções de resposta	0,314*	0,048
Resposta correta "A"	-0,004	0,980
Resposta correta "B"	-0,072	0,659
Resposta correta "C"	0,141	0,387
Resposta correta "D"	-0,184	0,255
Resposta correta "E"	0,015	0,929
Resposta correta "F"	0,199	0,218

*significativa a 0,05

Na Tabela 10 são apresentados os resultados obtidos quanto a estes dois componentes, ou seja, à quantidade de opções de respostas e à posição da resposta correta. É possível perceber que há uma correlação significativa, embora baixa, entre a quantidade de opções de resposta e a dificuldade do item. Embora essa relação seja baixa, evidencia-se que os itens mais difíceis possuem mais opções de resposta a serem escolhidas. Já em relação à localização da resposta correta, nenhuma correlação significativa foi encontrada.

Quanto ao tamanho da matriz, no caso da Figura 3 existem duas linhas e três colunas, ou seja, uma matriz de 2X3. Os demais itens do teste, além desse tipo de matriz, também apresentam matrizes de 2X5 e 3X3. Para avaliar a relação entre essa variável e a dificuldade dos itens, foram correlacionados a quantidade de linhas e de colunas da matriz, bem como seu tamanho total (Tabela 11).

Tabela 11. Correlação entre a dificuldade dos itens e a quantidade de linhas, colunas ou tamanho da matriz

Componente	Correlação	Significância
Quantidade de linhas da matriz	0,314*	0,048
Quantidade de colunas da matriz	0,083	0,613
Tamanho da matriz	0,389*	0,013

*significativa a 0,05

Os resultados, apresentados na Tabela 11, evidenciaram uma correlação baixa, mas significativa, para os componentes quantidade de linhas da matriz e tamanho total, sendo a primeira correlação de 0,314 e a segunda de 0,389, ambas significativas a 0,05. Apenas com a quantidade de colunas da matriz não houve correlação. Uma hipótese traçada para

esse resultado é de que quanto maior a quantidade de linhas ou o tamanho da matriz, mais análises são necessárias para que se obtenha a resposta correta, o que pode dificultar o item.

Verificada a influência do tamanho da matriz, buscou-se analisar se o número de estímulos diferentes presente nos itens, seja na pergunta ou na resposta, estava relacionado ao nível de dificuldade do mesmo. Além dos dois estímulos (círculo e triângulo) presentes na Figura 3, alguns itens do teste têm um terceiro estímulo (quadrado). É possível perceber, conforme os resultados da Tabela 12, que existe uma correlação significativa, embora baixa ($r = 0,382$), entre estímulos diferentes presentes nos problemas, o que não ocorre em relação às opções de resposta. Esses resultados indicam que a presença de mais de um tipo de estímulo pode estar associada à dificuldade do item apenas em relação ao problema e não quanto às opções de resposta.

Tabela 12. Correlação entre a dificuldade dos itens e o tipo de estímulos presente nos problemas e nas opções de resposta

Componente	Correlação	Significância
Tipos de estímulos diferentes no problema	0,382*	0,015
Tipos de estímulos diferentes na resposta	0,066	0,684

*significativa a 0,05

Na análise anterior (Tabela 12) foram considerados quais estímulos diferentes compunham o item (quadrado, triângulo ou círculo), mas não quantos no total. Além disso, os estímulos do teste podem aparecer tanto individualmente (apenas um estímulo por célula, conforme opção de resposta “A” da Figura 3) quanto combinados (mais de um estímulo por célula, conforme opções “B”, “C” e “D” da Figura 3). A essa junção de itens

em uma mesma célula denominou-se *agrupamento*. Foram calculadas, portanto, as correlações com a quantidade total de estímulos presentes nos problemas ou nas respostas, bem como com os agrupamentos (Tabela 13).

Tabela 13. Correlação entre a dificuldade dos itens e a quantidade total de estímulos existentes nos problemas e nas opções de resposta

Componente	Correlação	Significância
Quantidade total de estímulos no problema	0,491**	0,001
Quantidade total de estímulos na resposta	0,529**	0,000
Agrupamento de estímulos no problema	0,542**	0,000
Agrupamento de estímulos na resposta	0,440**	0,005

**significativa a 0,01

De acordo com os resultados da Tabela 13, verifica-se que tanto quando se considera a quantidade total de estímulos do problema ($r = 0,491$) quanto nas opções de resposta ($r = 0,529$), as correlações encontradas foram significativas, positivas e moderadas. O mesmo acontece em relação ao agrupamento de estímulos. Dentre todos os componentes identificados, esses foram os que apresentaram maior relação com a dificuldade dos itens. Esse resultado sugere que os itens com maior número de estímulos, especialmente quando agrupados em uma única célula, são mais difíceis. Apesar de utilizar um teste com características diferentes das do RIn (Sisto, 2006), Primi (1998) encontrou um resultado semelhante no que tange à quantidade de estímulos.

Correlações entre as regras e a dificuldade dos itens

As análises apresentadas a seguir foram realizadas após a classificação das regras (Tabela 2) presentes em cada item do teste. Os itens podem ser resolvidos horizontalmente, verticalmente ou de ambas as formas. Foi correlacionada então a dificuldade com as possibilidades de resolução, conforme disposto na Tabela 14. Os resultados dispostos na Tabela 14 indicam que houve uma correlação moderada ($r = 0,570$) e significativa com as regras horizontais, ou seja, para as linhas. Já as regras verticais, que estariam relacionadas às inferências que podem ser realizadas por colunas, indicaram uma correlação positiva, mas baixa e não significativa.

Tabela 14. Correlação entre a dificuldade dos itens e as regras horizontais e verticais

	Regras Horizontais		Regras Verticais	
	Correlação	Significância	Correlação	Significância
Dificuldade do item	0,570**	0,000	0,207	0,199

**significativa a 0,01

Pode-se considerar, de acordo com esses resultados (Tabela 14), uma relação entre a dificuldade dos itens e as regras horizontais, o que não ocorre em relação às verticais. Como nem todos os itens permitem a realização de inferências tanto vertical quanto horizontalmente, foi verificado se a quantidade de inferências possíveis para os itens

interfere em sua dificuldade. Conforme pode ser visto na Tabela 15, os resultados indicam uma correlação negativa, porém moderada e significativa.

Tabela 15. Correlação entre a dificuldade dos itens e a quantidade de inferências

Componente	Correlação	Significância
Quantidade de inferências	-0,415**	0,008

*significante a 0,01

É possível supor, de acordo com os resultados encontrados (Tabela 15), que podem existir itens mais difíceis com menos inferências possíveis, assim como itens mais fáceis que permitam maior número de inferências. Uma hipótese traçada é a de que quando é possível utilizar mais de um tipo de inferência para encontrar a resposta correta, caso não se consiga chegar a essa resposta usando determinada inferência, pode-se recorrer à outra. Já no caso de itens que permitem apenas uma inferência, se essa não for satisfeita, não há outra forma de chegar à resposta desejada.

Assim como ocorre com o número de inferências, cada item do teste também pode conter uma ou mais regras. Foi realizada uma correlação entre cada regra e a dificuldade do item, para verificar quais regras teriam maior relação com essa característica. Os dados são apresentados na Tabela 16.

Tabela 16. Correlação entre o nível de dificuldade e as regras horizontais e verticais

Regras	Resolução Horizontal		Resolução Vertical	
	Correlação	Significância	Correlação	Significância
Identidade	-0,668**	0,000	-0,654**	0,000
Inversão	-0,362*	0,022	-0,344*	0,030
Soma	0,096	0,556	-	-
Sequencia	0,161	0,320	0,368*	0,020
Subtração	0,016	0,920	-	-
Movimento	-0,027	0,868	-0,121	0,459
Quantidade	0,324*	0,042	0,364*	0,021

**significativa a 0,01 *significativa a 0,05

Analisando a Tabela 16 é possível perceber que as regras Soma, Subtração e Movimento não tem uma relação significativa com a dificuldade dos itens. As duas primeiras não foram correlacionadas verticalmente por não existir nenhum item cuja resolução por colunas as contenha. Já a regra Identidade, por sua vez, tem as maiores correlações dentre as demais ($r = -0,668$; $p < 0,01$ horizontalmente e $r = -0,654$; $p < 0,01$ verticalmente), indicando ser a regra mais relacionada aos itens fáceis, uma vez que as correlações foram altas e significativas, porém negativas.

A regra Sequencia também se relaciona com a dificuldade dos itens ($r = 0,368$; $p < 0,05$), ainda que tal relação seja baixa e apenas quando se considera a resolução do item por colunas. O mesmo pode ser dito das regras Inversão e Quantidade, ainda que neste caso a relação, embora menor, exista tanto para a resolução por linhas quanto por colunas. Tal como a regra Identidade, a regra Inversão também está presente nos itens mais fáceis, uma vez que as correlações encontradas foram negativas, tanto horizontal quanto verticalmente.

As análises anteriores mostram que nem todas as regras identificadas no teste contribuem de forma significativa para a dificuldade dos itens. Soma, Subtração e Movimento, por exemplo, não apresentaram nenhuma correlação significativa. Uma vez que um mesmo item pode dispor de mais de uma regra, foi verificada a existência de correlações entre elas, conforme pode ser visualizado na Tabela 17.

Tabela 17. Correlação entre as regras horizontais e verticais

Regras	IdH	InvH	SeqH	MovH	QtdeH	SomaH	SubH	IdV	InvV	SeqV	MovV	QtdeV
IdH	1	0,587	-0,481	0,242	-0,231	-0,262	-0,111	0,544	0,187	-0,338	0,208	-0,053
InvH	0,587	1	-0,214	0,099	-0,167	-0,189	-0,080	0,320	0,531	-0,055	0,095	-0,167
SeqH	-0,481	-0,214	1	-0,320	-0,231	-0,262	-0,111	-0,481	-0,214	0,361	-0,198	-0,231
MovH	0,242	0,099	-0,320	1	-0,154	0,423	-0,074	0,102	0,099	-0,158	0,368	0,066
QtdeH	-0,231	-0,167	-0,231	-0,154	1	-0,126	-0,053	-0,231	-0,167	0,145	-0,095	0,444
SomaH	-0,262	-0,189	-0,262	0,423	-0,126	1	-0,061	0,061	0,189	-0,247	-0,108	-0,126
SubH	-0,111	-0,080	-0,111	-0,074	-0,053	-0,061	1	0,231	0,320	-0,105	-0,046	-0,053
IdV	0,544	0,320	-0,481	0,102	-0,231	0,061	0,231	1	0,587	-0,454	0,410	-0,231
InvV	0,187	0,531	-0,214	0,099	-0,167	0,189	0,320	0,587	1	-0,191	0,095	-0,167
SeqV	-0,338	-0,055	0,361	-0,158	0,145	-0,247	-0,105	-0,454	-0,191	1	-0,186	-0,036
MovV	0,208	0,095	-0,198	0,368	-0,095	-0,108	-0,046	0,410	0,095	-0,186	1	-0,095
QtdeV	-0,053	-0,167	-0,231	0,066	0,444	-0,126	-0,053	-0,231	-0,167	-0,036	-0,095	1

Legenda: Id = Identidade; Inv = Inversão; Seq = Sequencia; Mov = Movimento; Qtde = Quantidade; Soma = Soma; Sub = Subtração; H = Horizontal e V = Vertical

Essa correlação (Tabela 17) foi utilizada como base para a realização de uma análise de agrupamentos, cujos objetivos são “agregar objetos com base nas características que eles possuem” e formar grupos a partir das observações mais parecidas (Hair e cols., 2005, p.384). Tal procedimento também pode ser utilizado para a confirmação de hipóteses previamente estabelecidas (Hair e cols., 2005). Neste caso, a hipótese traçada seria de que algumas regras, quando combinadas, teriam maior relação com a dificuldade dos itens. Uma vez constatada a existência de algumas correlações moderadas entre as regras (em destaque na Tabela 17), realizou-se então uma análise de Cluster para verificação dos possíveis agrupamentos (Figura 4).

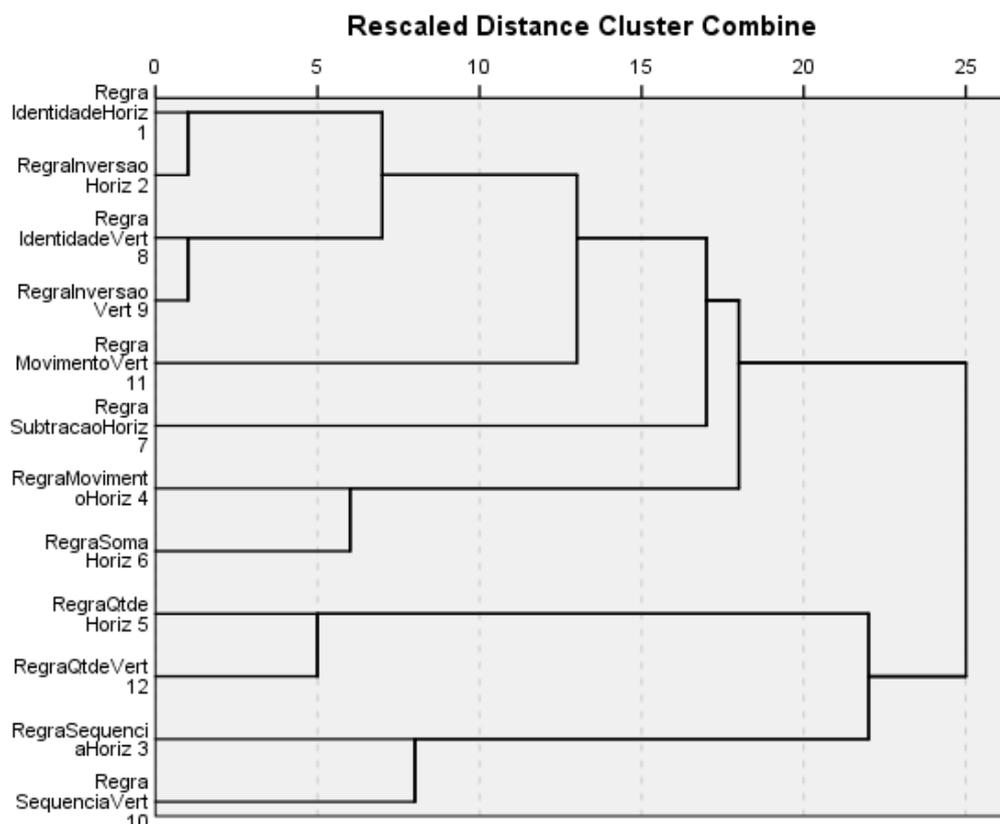


Figura 4 - Dendrograma

Um dendograma (Figura 4) é uma representação “dos resultados de um procedimento hierárquico no qual cada objeto é colocado em um eixo e o outro eixo representa os passos no procedimento hierárquico” até que todos os dados formem um único agrupamento (Hair e cols., 2005, p.382). É possível perceber que os primeiros agrupamentos formados foram entre a regra Identidade com a regra Inversão (tanto horizontal quanto verticalmente). Ambas, por sua vez, estão relacionadas aos itens mais fáceis do teste, conforme resultados anteriores. Já a regra Quantidade horizontal e vertical também formaram um grupo, assim como a regra Soma com a regra Movimento (horizontal), que não apresentaram nenhuma correlação com a dificuldade, sendo feitos outros agrupamentos até a finalização da análise.

Os achados anteriores (Tabela 13) evidenciaram que o número de elementos presentes nos problemas ou nas respostas e seu agrupamento contribuem significativamente para a dificuldade dos itens. Esse resultado corresponde ao que foi apontado por Primi (1998), segundo o qual, dentre os principais fatores de maior contribuição para a dificuldade dos itens em testes que envolvem analogias com figuras geométricas, está o número de figuras existentes no problema. Dessa forma, também foi incluída na análise a variável quantidade de estímulos agrupados no problema e na resposta (Tabela 18).

Na Tabela 18 são apresentados a média do teste e das regras considerando os 40 itens. Novamente as regras Soma e Subtração não foram consideradas verticalmente, por não existirem itens que as contemplem. Posteriormente, foi realizada uma análise de regressão para verificar qual seria o modelo mais indicado para o alcance dos objetivos propostos, ou seja, qual solução possibilitaria a melhor predição da dificuldade dos itens (Tabela 19).

Tabela 18. Estatística descritiva da análise de regressão

	Média	Desvio-padrão
Dificuldade do item	0,00	1,379
Quantidade estímulos agrupados no problema	1,93	1,439
Quantidade estímulos agrupados na resposta	2,13	1,285
Identidade horizontal	0,33	0,474
Inversão horizontal	0,20	0,405
Sequencia horizontal	0,33	0,474
Movimento horizontal	0,18	0,385
Quantidade horizontal	0,10	0,304
Soma horizontal	0,13	0,335
Subtração horizontal	0,03	0,158
Identidade vertical	0,33	0,474
Inversão vertical	0,20	0,405
Sequencia vertical	0,30	0,464
Movimento vertical	0,08	0,267
Quantidade vertical	0,10	0,304

A análise de regressão foi utilizada com o intuito de verificar a existência de uma combinação de regras que contribuíssem para dificuldade dos itens. Segundo Dancey e Reidy (2007) a análise de regressão é uma extensão da análise de correlação e pode ser utilizada para verificar o efeito de uma variável sobre outra. Esse tipo de análise possibilita uma estimativa de quanto uma variável pode mudar quando a outra muda. Aproximando

essa afirmação dos objetivos da presente pesquisa, seria possível, portanto, predizer a dificuldade do item a partir da utilização de determinada regra ou grupo de regras.

Tabela 19. Resumo dos modelos obtidos na análise de regressão

Modelo	R	R²	R² ajustado	Erro padrão
1	0,668	0,446	0,431	1,04052
2	0,752	0,566	0,542	0,93357
3	0,810	0,656	0,627	0,84288
4	0,836	0,699	0,665	0,79875

1 = Identidade horizontal; 2 = Identidade horizontal e vertical
 3 = Identidade horizontal, vertical e Sequencia Horizontal; 4 = Identidade horizontal, vertical, Sequencia Horizontal e Quantidade de estímulos agrupados na resposta

De acordo com o que pode ser visto na Tabela 19, o modelo 4 evidencia uma correlação muito alta ($r = 0,836$) e explica 69,9% da variância. Tal modelo é composto pelas regras Identidade Horizontal, Vertical, Sequencia Horizontal e Quantidade de estímulos agrupados na resposta. Uma ANOVA indicou que a diferença entre o modelo 4 e os demais é significativa ($F = 20,353$; $p < 0,001$). Por fim, foram verificados os coeficientes de regressão de cada uma das quatro regras que compõem o modelo 4 e realizado um teste t de Student para verificar a existência de diferenças significativas entre cada uma dessas regras e a dificuldade do item. Os resultados podem ser vistos na Tabela 20.

Tabela 20. Coeficientes de regressão, padronizados e teste *t* de Student

Regra	<i>b</i>	Erro padrão	β	<i>T</i>	<i>p</i>	Zero-order
Modelo 4	0,721	0,367		1,962	0,058	
Identidade horizontal	-1,509	0,340	-0,519	-4,439	0,000**	-0,668
Identidade vertical	-1,360	0,345	-0,467	-3,947	0,000**	-0,654
Sequencia horizontal	-0,895	0,329	-0,308	-2,722	0,010**	0,161
Qtde. est. agrup. na resposta	0,237	0,105	0,220	2,256	0,030*	0,440

**significativa a 0,01 *significativa a 0,05

Considerando os coeficientes padronizados (β) pode-se afirmar que as regras Identidade horizontal, vertical e Sequencia horizontal tornam os itens mais fáceis. Já o número de estímulos presentes nas opções de respostas torna os itens mais difíceis. Além disso, a diferença entre cada uma das regras e a dificuldade dos itens é significativa, demonstrando sua contribuição individual.

Os resultados encontrados, tanto nas correlações realizadas com os componentes dos itens do teste quanto com as realizadas com as regras, corroboram as duas hipóteses traçadas inicialmente, a saber, de que H1) os componentes dos itens do teste influenciam no processo de resposta e H2) regras diferentes implicam em distintos processos.

Quanto à primeira hipótese, os componentes que indicaram ter maior relação com a dificuldade dos itens foram o agrupamento de estímulos no problema, a quantidade de estímulos na resposta e no problema e o agrupamento de estímulos na resposta. Essa relação pode ser consequência do fato de que itens com mais estímulos, especialmente

quando estes estão agrupados, geram a necessidade de que sejam realizadas mais inferências para sua resolução. Pode-se supor, inclusive, uma maior necessidade de evocação de outros atributos cognitivos, como a memória, para a resolução desses itens.

No que tange à segunda hipótese, confirma-se a relação de diferentes regras com a dificuldade dos itens e, conseqüentemente com o processo de resposta necessário para resolvê-los. As regras Identidade e Sequencia, por exemplo, referem-se à repetição de estímulos, e se relacionaram aos itens mais fáceis. São regras comumente empregadas inclusive nos exemplos de como um teste deve ser realizado, como no caso do próprio RIn (Sisto, 2006), frente a facilidade de se identificar a resposta correta a partir delas. Já a regra Quantidade, por sua vez, exige que, para se resolver o item, seja verificado quantos estímulos de cada tipo estão presentes no problema. Dessa maneira, pode ter implícita a necessidade de que, após ser percebida, sejam realizadas inferências relacionadas a outras regras como, por exemplo, Identidade, Sequencia ou Inversão. Logo, é mais complexa do que as demais, o que pode explicar o fato de estar relacionada aos itens mais difíceis.

CONSIDERAÇÕES FINAIS

Assim como não existe um consenso quanto à definição da inteligência, o que pode inclusive ser um reflexo do interesse pelo tema e dos inúmeros estudos abordando este construto (Primi, 2002), é inegável a importância dessa característica e das decisões que são tomadas diariamente com base em resultados de testes que a avaliam (Sisto, 1975). Da mesma maneira, inúmeros autores ressaltam a importância da validade de um teste como um aspecto central e fundamental para subsidiar a confiabilidade das inferências que os mesmos permitem que sejam feitas (APA, AERA & NCME, 1999; Pasquali, 2003; Urbina, 2007). Partindo destas constatações, o objetivo central da presente pesquisa foi buscar evidências de validade pelo processo de resposta para o Teste de Raciocínio Inferencial – RIn (Sisto, 2006), contribuindo também para a atualização dos estudos com o referido instrumento, o que se configura como uma necessidade atual (Hutz, 2011).

Embora seja necessário que todos os testes psicológicos apresentem estudos de validade, independentemente de qual seja a evidência escolhida, optou-se pelo processo de resposta por ser uma das evidências menos utilizadas para tal objetivo, o que pode ser constatado pela escassez de referências bibliográficas encontradas a respeito dessa evidência. Além disso, o RIn (Sisto, 2006) não dispunha de um estudo dessa evidência de validade. Para validar o método utilizado, optou-se por inicialmente verificar a confiabilidade do modelo utilizado, uma vez que foi criado especificamente com base no teste fruto deste estudo, evitando assim erros conceituais ou a influência do pesquisador

(AERA, APA & NCME, 1999; Urbina, 2007). A partir desse modelo, foram realizadas as análises necessárias para o alcance dos objetivos.

Hair e cols. (2005) mencionam que muitos resultados de análises como a aqui proposta não podem ser generalizados e estão melhor relacionados à amostra utilizada. Essa afirmação não se aplica diretamente no caso dessa pesquisa, pelo fato de estar se analisando os itens e não as amostras. Ainda assim, a alta correlação ($r = 0,926$; $p < 0,01$) entre os dados atuais e os apresentados no manual do teste permite considerar a aplicação dos resultados encontrados a diferentes amostras, uma vez que a dificuldade dos itens foi a variável dependente utilizada para as análises realizadas.

Uma análise individualizada dos itens permitiu verificar a adequação dos mesmos ao teste no que tange aos índices de *infit* e *outfit*, tendo sido constatado que apenas o item A12 não apresentou um *outfit* satisfatório. Fato semelhante já havia ocorrido em relação a esse item nos estudos mencionados no manual do teste. Além disso, a concordância dos juízes quanto a um erro conceitual aliada ao fato de que os indivíduos de maior habilidade escolhem uma opção errônea, enquanto indivíduos menos hábeis o acertam, confirmam a necessidade de que o mesmo seja revisto.

Em relação às propriedades dos itens e às regras identificadas, foram feitas inúmeras análises para verificar quais destas realmente gerariam algum efeito sobre a dificuldade dos itens, tendo sido encontrados componentes e regras que se relacionam aos itens mais fáceis, aos mais difíceis ou que não tenham qualquer relação com a dificuldade dos mesmos.

Apesar de ter sido verificado uma maior relação entre as inferências horizontais e a dificuldade dos itens, deve se considerar que nem todos os itens permitem resolução tanto

horizontal quanto verticalmente. Em contrapartida, a correlação negativa com a quantidade de inferências possíveis sugere que itens para os quais é possível realizar mais de um tipo de inferência para se chegar à resposta correta são mais fáceis.

Em relação às regras encontradas pelos juízes e nas quais se pautaram as análises seguintes, verificou-se que a regra Identidade apresenta as maiores correlações com a dificuldade dos itens, seguida pela regra Sequencia vertical, Inversão e Quantidade, sendo que estas duas últimas contribuem tanto quando se considera a resolução por linhas quando por colunas.

A partir de uma análise de agrupamentos, foi possível perceber que existiam correspondências entre as regras Identidade com a regra Inversão (tanto horizontal quanto verticalmente), Quantidade horizontal e vertical, assim como entre a regra Soma e a regra Movimento (horizontal). Já a análise de regressão indicou que a maior influência sobre a dificuldade dos itens explicaria 69,9% da variância e seria consequência das regras Identidade Horizontal, Identidade Vertical, Sequencia Horizontal e Quantidade de estímulos agrupados na resposta. Com exceção desta última, as demais contribuiriam para tornar os itens mais fáceis, uma vez que as correlações foram negativas.

A quantidade de estímulos já foi anteriormente relacionada com outros achados científicos (Primi, 1998). Já as demais regras, possivelmente contribuem para tornar os itens mais fáceis dado o fato de serem mais simples e evidentes, uma vez que as regras de Identidade permitem a comparação de estímulos iguais e a regra Sequencia, como o próprio nome diz, corresponde a uma sequencia de estímulos que, quando encontrada, permite a resolução do item.

Os resultados aqui relatados confirmaram ambas as hipóteses traçadas para este trabalho. Embora nem todas as regras encontradas tenham se relacionado com a dificuldade do itens, pode-se afirmar que o fato do número de estímulos presentes ou agrupados no item, o tamanho da matriz e o número de inferências possíveis, assim como as regras já descritas, contribuem para a maior ou menor dificuldade dos itens. Dessa forma, pode-se considerar que a identificação destes elementos ou regras permite não apenas prever a dificuldade do item, como também pode servir de base para a criação de novos instrumentos similares. Em outras palavras, um autor de testes pode se basear nos componentes e regras descritos neste trabalho para criar itens possivelmente mais fáceis ou mais difíceis, uma vez que foram apresentadas relações para ambos.

Considera-se, portanto, que os objetivos traçados foram alcançados e os resultados fornecem evidências de validade pelo processo de resposta para o Teste de Raciocínio Inferencial – Rin (Sisto, 2006), uma vez que foram identificados os fatores que interferem no processo de resposta. Ainda assim, sugere-se a realização de estudos similares com testes que utilizem estímulos que não sejam figuras geométricas ou cujas características sejam diferentes das do instrumento analisado. Tais pesquisas permitiriam confirmar a relação entre as regras e componentes aqui encontrados com a dificuldade dos itens, assim como possibilitariam encontrar outras regras não contempladas pelo RIn (Sisto, 2006) que também exercem tal influência.

REFERÊNCIAS

- Almeida, L. S. (1988). *Teorias da inteligência*. Porto: Edições Jornal de Psicologia.
- Almeida, L. S. (1992). *Bateria de provas de raciocínio diferencial (BPRD)*. Braga.
- Almeida, L.S. & Primi, R. (2004). Perfis de capacidades cognitivas na bateria de provas de raciocínio (BPR-5). *Psicologia Escolar e Desenvolvimento*, 8 (2), 135-144.
- American Psychological Association, American Educational Research Association & National Council on Measurement in Education (1999). *Standards for educational and psychological tests*. Washington, DC: APA.
- Anastasi, A. (1961/1977). *Testes psicológicos*. São Paulo: EPU.
- Anastasi, A & Urbina, S. (1997) *Psychological Testing*. Upper Saddle River, New Jersey: Prentice Hall.
- Anastasi, A & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artes Médicas.
- Anzieu, D. (1981). *Os métodos projetivos*. Rio de Janeiro: Campus.
- Arias, R.M. (1996). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Editorial Síntesis.

Braga, M. (1938/1998). Validade e fidedignidade dos testes coletivos de inteligência. *Revista Brasileira Estudos Pedagógicos*, 79(193), 113-134.

Brody, N. (1992). *Intelligence*. San Diego: Academic Press.

Carpenter, P. A., Just, M. A. Shell, P. (1990) What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97(3), 404-431.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. New York: Cambridge University Press.

Catell, R. B. (1971). *Abilities: their structure, growth, and action*. Boston: Houghton Mifflin.

Colom, R. & Garcíapz, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences*, 32, 445-451.

Colom, R. (2008). *Nos limites da inteligência. É o ingrediente do êxito da vida?* São Paulo: Vetor.

Conde, F. N. & Laros, J. A. (2007). Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. *Avaliação Psicológica*, 6(2), 205-215.

Conselho Federal de Psicologia (2003). *Resolução 002/2003*. Recuperado em 25 junho 2010: <http://www.pol.org.br>.

Cornoldi, C. (2006). *European Journal of Cognitive Psychology*, 18 (1), 1-17.

Cronbach, L. J. (1996). *Fundamentos da testagem psicológica* (5ª ed.). Porto Alegre: Artes Médicas.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.

Cunha, N. B. & Santos, A. A. A. (2009). Validade por processo de resposta no teste de Cloze. *Revista de Psicologia*, 21(3), 549-562.

Dancey, C. P.; Reidy, J. (2007). *Estatística sem matemática para psicologia usando SPSS para Windows* (3ª ed.). Porto Alegre: Artes Médicas.

Evans, T. G. (1968). Program for the solution of a class of geometric-analogy intelligent-test questions. In: M. Minsky (org). *Semantic information processing*. Cambridge, MA: MIT Press.

Flanagan, D. P. & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: John Wiley & Sons, Inc.

Goodnow, J. J. (1976). The nature of intelligent behavior: questions raised by crosscultural studies. In: L. Resnick (org.). *The nature of intelligence*, 169-188. Hillsdale, NJ: Erlbaum.

Hair, J. F. & cols. (2005). *Análise multivariada de dados* (5ª ed.). Porto Alegre: Artes Médicas.

Horn, J. L. (1991). Measurement of Intellectual Capabilities: a review of theory. Em K.S. McGrew, J. K. Werder, & R. W. Woodcock. *Woodcock-Johnson technical manual*, 197-232. Chicago: Riverside.

Hutz, C. S. (2011). Manuais especificando seus contextos de aplicação e âmbitos de ação. In: *Ano da Avaliação Psicológica no Brasil – Textos geradores*. Brasília: Conselho Federal de Psicologia.

Irvine, J. T. (1978). Wolof magical thinking: culture and conservation revisited. *Journal Cross-Cultural Psychological*, 9, 300-310.

Jacobs, P. I. & Vandeventer, M. (1972) Evaluating the teaching of intelligence. *Educational and Psychological Measurement*, 32, 235-248.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In: S.A. Stauffer, L. Guttman, E.A. Suschman, P.F. Lazarsfeld, S.A. Star & J.A. Clausen (eds.). *Measurement and prediction*. Princeton, NJ: Princeton University Press.

Lemos, G. (2007) *Habilidades cognitivas e rendimento escolar entre o 5º e 12º ano de escolaridade*. Tese de Doutoramento, Departamento de Pós-Graduação em Psicologia, Universidade do Minho, Braga.

Linacre J. M. (2002). What do Infit and Outfit, Mean-Squared and Standardized mean? *Rasch Measurement Transactions*, 16 (2), 878. Recuperado em 19 março 2006: <http://209.238.26.90/rmt/rmt82a.htm>.

Lord, F. M. (1952). *A theory of tests scores*. Iowa City, IA: Psychometric Society.

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, 55(1), 137-150

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.

Messick, S. (1989). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.

Miranda, M. J. (2002). A inteligência humana: contornos de pesquisa. *Paidéia*, 12(23), 19-29.

Mulholland, T. M., Pellegrino, J. W. & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, 12, 252-284.

Muniz, J. (2004). La validación de los tests. *Metodología de las Ciencias Del Comportamiento*, 5 (2), 121-141.

Muniz, J. (Org.). (1996). *Psicometria*. Madrid: Universitas.

Nunes, C. H. S. & Primi, R. (2010). O Satepsi: desafios e proposta de aprimoramento. In: *Avaliação psicológica: Diretrizes na regulamentação da profissão*. Brasília: Conselho Federal de Psicologia.

Oliveira-Castro, J. M. & Oliveira-Castro, K. M. (2001). A função adverbial de “inteligência”: definições e uso em psicologia. *Psicologia: Teoria e Pesquisa*, 17(3), 257-264.

Pasquali, L. (1999). *Instrumentos psicológicos: manual prático de elaboração*. Brasília: LabPAM.

Pasquali, L. (2003). *Psicometria: teoria dos testes na psicologia e na educação*. Rio de Janeiro: Editora Vozes.

Pasquali, L. (2007). *TRI: Teoria, procedimentos e aplicações*. Brasília: LabPAM.

Primi, R. (1995) Inteligência, processamento de informação e teoria da gestalt: um estudo experimental. Campinas, 156p. Dissertação de Mestrado. *Instituto de Pós-Graduação em Psicologia*, PUCCAMP.

Primi, R. (1998). *Desenvolvimento de um instrumento informatizado para avaliação do raciocínio analítico*. Tese de Doutorado, Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Primi, R. (2002) Inteligência fluida: definição fatorial, cognitiva e neuropsicológica. *Paidéia*, 12(23), 57-75.

Primi, R. (2003). Inteligência: Avanços nos modelos teóricos e nos instrumentos de medida. *Avaliação psicológica*, 1, 67-77.

Primi, R., Muniz, M.; Nunes, C. H. S. (2009). Definições contemporâneas de validade de testes psicológicos. In: Hutz, C. S. (org.). *Avanços e polêmicas em avaliação psicológica*. São Paulo: Casa do Psicólogo.

Raven, J. C. (1962). *Matrizes progressivas - escala avançada série I e II*. Rio de Janeiro: CEPA.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: Mesa Press.

Roisenberg, M. (1999). *Inteligência artificial: abordagem simbólica*. Recuperado em 06 junho 2010: <http://www.inf.ufsc.br/~mauro/ine6102/slide/aula6-raciocinio/tsld001.htm>.

Rueda, F. J. M. & Sisto, F. F. (2008). Versão preliminar do teste pictórico de memória: estudo de validade. *Estudos de Psicologia*, 25(2), 223-231.

Santos, A. A. A. (2011). O possível e o necessário no processo de avaliação psicológica. In: *Ano da Avaliação Psicológica no Brasil – Textos geradores*. Brasília: Conselho Federal de Psicologia.

Serpell, R. (1982). Measures of perception, skills and intelligence. In: *W.W. Hartup (org.). Review of child development research*, 6, 392-440. Chicago: University Chicago Press.

Sisto, F. F. (1975). *El factor g y el pensamiento operatório formal (grupo INCR)*. Tesis Doctoral. Madrid: Universidad Complutense de Madrid.

Sisto, F. F. (2006). *Teste de raciocínio inferencial RIn: manual*. São Paulo: Vetor Editora.

Sisto, F. F. (2007). Delineamento correlacional. In M. N. Baptista & D.C. Campos, *Metodologia de Pesquisa em Ciências: análises quantitativa e qualitativa*. Rio de Janeiro: LTC.

Spearman, C. (1927). *Las habilidades del hombre: su naturaleza y medición*. Buenos Aires: Paidós.

Sternberg, R. J. (1977) A component process in analogical reasoning. *Psychological Review*, 84(4), 353-378.

Sternberg, R. J. (1979). The nature of mental abilities. *American Psychologist*, 34(3), 214-230.

Sternberg, R. J. (1981) The evolution of theories of intelligence. *Intelligence*, 5, 209-230.

Sternberg, R. J. (1983). Components of human intelligence. *Cognition*, 15, 1-48.

Sternberg, R. J. (1984) Toward a triarchic theory of human intelligence. *The Behaviour and Brain Sciences*, 7, 269-315.

Sternberg, R. J. (1992). *As capacidades intelectuais humanas: uma abordagem em processamento de informações*. Porto Alegre: Artmed.

Sternberg, R. J.; Detterman, D.K. (Orgs.) (1986). What is intelligence? *Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.

Tyler, L. E. (1973). *Testes e medidas*. (3ª ed.). Rio de Janeiro: Zahar.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Thurstone, L. L. & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.

Werlong, B. S. G., Villemor-Amaral, A. E., & Nascimento, R. S. G. F. (2010). Avaliação psicológica, testes e possibilidades de uso. In: *Avaliação psicológica: Diretrizes na regulamentação da profissão*. Brasília: Conselho Federal de Psicologia.