

UNIVERSIDADE DE BRASÍLIA - UNB
INSTITUTO DE PSICOLOGIA - IP
DEPARTAMENTO DE PSICOLOGIA SOCIAL E DO TRABALHO - PST
LABORATÓRIO DE PESQUISA EM AVALIAÇÃO E MEDIDA - PAM

A

CONSTRUÇÃO DE UM TESTE DE
RACIOCÍNIO VERBAL (RV)

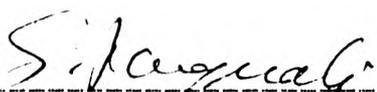
WAGNER BANDEIRA ANDRIOLA

Orientador: Prof. Dr. Luiz Pasquali

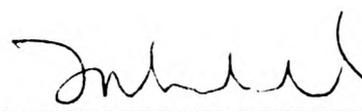
Brasília, junho de 1993

Dissertação apresentada ao Instituto de
Psicologia da Universidade de Brasília, como requisito
parcial à obtenção do grau de Mestre em Psicologia,
área de concentração Psicologia Social e do Trabalho.

Aprovada por:



LUIZ PASQUALI - ORIENTADOR



TIMOTHY MARTIN MULHOLLAND



EUNICE M. SORIANO DE ALENCAR

Brasília, 11 de junho de 1993

Pensamento Analógico:

Na guerra os méritos da vitória são sempre do GENERAL. A responsabilidade da derrota é sempre do SOLDADO (Carlos Drummond).

Na vida acadêmica os méritos da vitória são sempre do PhD. A responsabilidade da derrota é sempre do ORIENTANDO (Wagner Andriola).

Dedicatória

A Deus
... por ter me concedido a capacidade de fazer ciência.

A minha namorada Cristiany Gomes da Nóbrega
... pelos momentos felizes que juntos passamos, mesmo à
distância, e pela dedicação de esperar por mais um período
de férias.

Aos meus pais
Vicente Andriola e Maria Bandeira Andriola
... por terem me educado e dado o apoio para ser o que sou
hoje.

A minha família 'candanga'
Mardônio Rique Dias, Cristina M. S. B. Dias,
Rodrigo, Marília e Mirian
... pelos momentos alegres e felizes que me proporcionaram.

Aos meus Irmãos
Walder, Walter, Magda, Walker e Juliana
... pelo carinho de sempre.

Aos meus tios
João, Paulo, Francisco, Idalina, Zélia, Antônia, Helena,
Teresinha e Cotinha,
... pelos valiosos conselhos e momentos felizes que me
proporcionaram na infância.

Em especial
aos meus tios
Edmilson, Hilda, Francisca, Teresa, Tizinha, Eunice, José,
Valmir e Euclides
... por me mostrarem o valor da humildade e simplicidade.

Aos meus sobrinhos
Amanda e Alex
... pela inocência que apresentam ao mundo.

Em memória de

Meus avós
Francisco Andriola e Ana Otília Andriola,
Raimundo Pereira da Silva e Júlia Bandeira da Silva
... onde estiverem, saibam do meu amor, carinho e admiração;

Minha sobrinha
Maria Camila
... pela pureza e inocência.

Agradecimentos

Ao amigo e orientador, tanto acadêmico quanto mundano,
Prof. Mardônio Rique Dias. Soube, como ninguém, iniciar-me e
orientar-me como pesquisador e pessoa.

Ao Prof. Leandro da Silva Almeida,
pelas valiosas idéias.

Ao Prof. Luiz Pasquali,
pelos conhecimentos transmitidos e valiosas orientações.

À Prof^{ma}. Eunice M. S. de Alencar,
pelas importantes e detalhadas contribuições.
Sobretudo, pelo momento de desabafo que me proporcionou.

Ao Prof. Timothy M. Mulholland,
pela paciência e educação.

Ao Prof. Hartmut Günther,
por ter aceito fazer parte da banca examinadora.

Aos Prof^{mas}. Maria Ângela G. Feitosa e
Antônio de Freitas Ribeiro,
pelas sugestões certas feitas no momento adequado.

Aos Prof^{mas}. Balsem Pinelli Jr. e
Francisco Martins Catunda, pela amizade e consideração.

Ao Prof. Cláudio Vaz Torres, pela disponibilidade de sempre.

Ao companheiro de mestrado Valdiney Veloso Gouveia,
pelos desafios vencidos conjuntamente.

Aos colegas do alojamento estudantil, Domingos Sávio Coelho,
José William e Antônio Carlos Neves,
pelas vitórias, conseguidas em conjunto, sobre os momentos
mais difíceis e infelizes.

A todos os funcionários da secretaria do Instituto de
Psicologia,
em especial aos Sr^{es}. Elson Rodrigues de Souza e
Antônia Alves de Sousa, pela disponibilidade de sempre.

A todos os companheiros de mestrado,
em especial ao Luiz Avelino e Ivo Donner pela amizade e
conselhos.

Aos meus alunos de TEP (1.93),
pelo aprendizado que me proporcionaram.

Ao CNPq,
pelo investimento para a execução deste trabalho.

	Sumário	
Dedicatória		iv
Agradecimentos		v
Sumário		vi
Sumário das Tabelas		viii
Sumário das Figuras		ix
Sumário dos Anexos		x
Resumo		xi
Abstract		xii
I - Introdução		1
1 - Os Testes Psicológicos: a realidade brasileira		1
2 - Porque construir um Teste de Raciocínio Verbal		10
2.1 - Utilidade de um Teste de Raciocínio		10
2.2 - Críticas aos Testes de Conteúdo Verbal		17
3 - Fundamentação Teórica do Teste RV		21
4 - O Raciocínio segundo diferentes enfoques		24
5 - A Definição de Raciocínio segundo a Perspectiva Diferencialista		28
6 - Como mensurar o Raciocínio Verbal?		32
II - A Construção do Teste de Raciocínio Verbal (RV)		41
1 - Formato dos itens componentes do Teste RV		41
1.1 - Como aumentar a complexidade de uma analogia?		41
2 - Construção dos Itens		45
3 - Seleção dos Itens		47
3.1 - Estudo Piloto I		48
3.2 - Estudo Piloto II		50

III - Estudo Final do Teste RV	57
1 - Metodologia	57
1.1 - Amostra	57
1.2 - Instrumento	58
1.3 - Procedimento	59
2 - Análises estatísticas	60
3 - Resultados	60
3.1 - Validação	60
3.1.1 - Validade de Construto	60
3.1.2 - Validade Preditiva	64
3.2 - Precisão	68
4. - Análise dos Itens do RV	68
4.1 - Distribuição dos Itens no Contínuo de Dificuldade	69
4.2 - Poder de Discriminação dos Itens	71
5 - Construção das Formas Paralelas do Teste RV	74
6 - Elaboração das Normas Percentílicas	77
7 - Cálculo do tempo de Aplicação do Teste RV	83
IV - Conclusões	83
V - Referências Bibliográficas	87
VI - Anexos	96

Sumário das Tabelas

Tabela 01 - Correlações entre o Desempenho de soldados em tarefas de trabalho e em medidas de Capacidade Verbal	15
Tabela 02 - As cinco Categorias utilizadas para a Construção dos Itens	46
Tabela 03 - Características da Amostra de Estudantes do Estudo Piloto I	49
Tabela 04 - Características da Amostra de Estudantes do Estudo Piloto II	51
Tabela 05 - Distribuição dos 63 Itens segundo o índice de Dificuldade	52
Tabela 06 - Poder de Discriminação dos 63 Itens	54
Tabela 07 - Descrição da Amostra utilizada no Estudo Final do Teste RV	58
Tabela 08 - Cargas Fatoriais dos 40 Itens do Teste RV no Fator de Raciocínio Verbal	63
Tabela 09 - Distribuição dos 40 Itens do Teste RV segundo o índice de Dificuldade	69
Tabela 10 - Poder de Discriminação dos 40 Itens do Teste RV ..	73
Tabela 11 - Distribuição dos Itens nas Formas Paralelas	75
Tabela 12 - Normas para a Forma Geral por Sexo e Idade	80
Tabela 13 - Normas para a Forma 'A' segundo o Sexo e a Idade .	81
Tabela 14 - Normas para a Forma 'B' segundo o Sexo e a Idade .	82

Sumário das Figuras

Figura 1 - Processos Cognitivos propostos por Spearman	34
Figura 2 - Processos Cognitivos propostos por Sternberg	36
Figura 3 - Modelos Simplificados propostos por Pellegrino	39

Sumário dos Anexos

Anexo I - Testes Construídos no Brasil entre os anos 30 e 60	96
Anexo II - Versão de 51 itens do Teste RV	105
Anexo III - Forma Geral do Teste RV	109
Anexo IV - Formas Paralelas do Teste RV	114

Resumo

A presente pesquisa teve como objetivo construir um Teste de Raciocínio Verbal para a população estudantil brasileira. Foram realizados dois estudos pilotos com os objetivos de 1) estudo semântico das instruções e dos itens, seleção da população a qual se destinaria o teste e 2) estudo inicial dos parâmetros psicométricos dos itens.

O estudo final realizado com o Teste de Raciocínio Verbal, utilizou uma amostra de 730 estudantes oriunda da cidade de João Pessoa-PB, composta por homens (n=341) e mulheres (n=389) distribuídos segundo três séries escolares 1º (n=190), 2º (n=321) e 3º (n=219). A maior parte dos estudantes possuía idades entre a faixa de 16 a 19 anos (n=398).

As análises estatísticas feitas com o Teste de Raciocínio Verbal objetivaram verificar a validade de construto (através da análise fatorial), a validade preditiva sobre o desempenho escolar (através do coeficiente de correlação de Pearson) e a precisão (através do alfa de Cronbach). O estudo dos itens consistiu na análise do índice de dificuldades (através dos percentuais de acerto) e do poder de discriminação de grupos-critério (através do teste 't' de Student).

Para a validação de construto, a análise fatorial utilizando o método do "Principal Axis Factoring - PAF", demonstrou a existência de um fator - denominado de raciocínio verbal - explicando 21.3% da variância total e "eigenvalue" de 8.54.

A validade preditiva, obtida através da correlação entre escores no teste e últimas médias bimestrais na disciplina de português, resultou num $r=.14$. Uma correlação baixa como essa, ainda que significativa, revela a necessidade de estudos mais aprofundados sobre a validade preditiva do teste em questão.

A precisão, obtida através do alfa de Cronbach, foi de .90, altamente significativa.

O estudo dos itens mostrou uma distribuição adequada dos mesmos sobre o contínuo do raciocínio - segundo o índice de dificuldade - além de alto poder de discriminação de todos eles.

A forma final do Teste de Raciocínio Verbal (RV) ficou composta por 40 itens. Duas formas alternativas, 'A' e 'B', foram organizadas. Ambas as formas ficaram compostas por 22 itens com uma precisão de .84, para cada uma delas.

As normas percentílicas - provisórias - foram estabelecidas segundo o sexo e a idade.

O presente trabalho resultou num Teste de Raciocínio Verbal (RV), composto por 40 itens, com duas formas alternativas, com conteúdo atualizado, parâmetros psicométricos adequados e normas - ainda que provisórias - estabelecidas para a população estudantil brasileira.

Abstract

The present research was aimed at the elaboration of a Test for Verbal Reasoning for the Brazilian student population. Two pilot studies were done with the purpose to: 1) make a semantic analysis of the items and instructions, as well as to define the population for which the test was being constructed and 2) to perform an initial analysis of the psychometric parameters of the items.

The study for the final analysis of the test was based on a sample of 730 students from the city of João Pessoa-PB. The sample comprised 341 men and 389 women, belonging to the three high school grades. The majority ($n=398$) of the students fell within 16 and 19 years of age.

The statistic analysis of the data meant to define the construct validity (factor analysis), the predictive validity against school achievement (Pearson correlation), and the reliability (Cronbach's alpha). For the item analysis were studied the indexes of difficulty (percent of correct answers) and discrimination (Student's 't' for criteria groups).

For construct validation was used the Principal Axis Factoring (PAF) which showed the presence of one major factor - verbal reasoning - with eigenvalue of 8.54 and explaining 21.3% of the total variance.

The predictive validity was obtained through the correlation between the test scores and the final grades in Portuguese language. The obtained correlation was $r=.14$. Although significant, this low correlation suggests the need for further study in this particular issue.

The reliability coefficient was .90.

The item analysis showed that their distribution on the verbal reasoning continuum, as well their discrimination power were adequate.

The final form for the Reasoning Verbal Test (RV) was composed of 40 items. Two alternative forms, 'A' and 'B', were organized. Both were composed of 22 items with identical reliability coefficients of, that is, .84.

Preliminary percentile norms were elaborated by sex and age.

The present work resulted in the Reasoning Verbal Test (RV), composed of the 40 items, adequate psychometric parameters and preliminary norms for the Brazilian student population.

I - Introdução

1. Os Testes Psicológicos: a Realidade Brasileira

Nos Estados Unidos a prática do uso de testes psicológicos após a primeira guerra mundial foi ampla (Haney, 1981; Anastasi, 1988). Só no ano de 1954 foram utilizados nas escolas e universidades cerca de 81.000.000 de testes padronizados, para ascender para 150.000.000 no ano de 1964 (Baquero, 1974).

Segundo Reschly (1981), na área de Psicologia, aproximadamente 15.000 psicólogos americanos fazem uso dos testes. Sendo que "...os testes grupais são mais frequentemente usados, com milhões deles sendo ministrados a cada ano nas escolas, indústrias e organizações militares" (Tyler, 1981:71).

É bom salientar que na realidade norte-americana sempre se deu ênfase aos aspectos psicométricos dos testes psicológicos, quando de sua construção. Para se ter uma idéia, em 1938 Buros publicou o livro "Mental Measurement Yearbook" contendo uma lista de cerca de 4.000 testes validados para uso (Haney, 1981).

Mais recentemente Keyser e Sweetland (1984) organizaram seis volumes sobre os principais testes utilizados no mercado norte-americano, onde o objetivo foi fornecer uma descrição geral

(e crítica) sobre os aspectos psicométricos dos mesmos. Nesta obra são analisados 614 instrumentos psicológicos.

No Brasil, entre os anos 30 e 60, atingiu-se o auge na construção de testes psicológicos (Nick, 1988).

Apesar da maioria dos autores preocuparem-se com os parâmetros psicométricos de seus testes - tais como validade e precisão - realizando estudos para a determinação dos mesmos, alguns dos criadores relegaram tais estudos para segundo plano⁴.

Tem-se que salientar que, embora tenham existido tais estudos no país, o uso das técnicas estatísticas era bastante precário - e porque não dizer cheio de erros. Naquela época, o uso do computador como uma ferramenta auxiliar para a área ainda era utopia. Sendo assim, apesar de considerados válidos tais estudos, é necessário admitir a existência de inúmeros erros - sobretudo metodológicos, de análises estatísticas e de amostragem - sobre os resultados (Nick, 1988). Apesar de tudo, pode-se afirmar que os parâmetros psicométricos de tais instrumentos, resultantes destas pesquisas, podem ser considerados, em geral, razoáveis para aquela época.

Atualmente pode-se contar com o auxílio do computador para análises estatísticas mais apropriadas, precisas e rápidas. Tal fato permite novos estudos sobre tais instrumentos, com um primeiro objetivo de minimizar os erros - já mencionados - subjacentes às suas construções.

⁴ O Anexo I contém uma relação de testes construídos no Brasil entre os anos 30 e 60, com uma breve descrição dos estudos de seus parâmetros psicométricos. Fonte Van Kolck, 1981a e 1981b.

A construção de um Teste de Raciocínio Verbal (RV)

Embora tenha havido esta fase inicial de interesse na criação de testes, que resultou na introdução dos mesmos no mercado, não houve preocupação - nem pelos profissionais da área nem pelas empresas responsáveis pelos direitos de reprodução - pela atualização dos conteúdos de tais instrumento e pelo estudo de seus parâmetros psicométricos de forma sistemática.

Para Klausnitzer (1985:15-16) "de todos os chamados testes psicológicos atualmente existentes no país, somente um número pequeno corresponde às mínimas exigências científicas".

Assim, os testes atualmente colocados no mercado são, na maioria das vezes, tratados sobretudo como produto comercial, deixando-se a questão científica dos mesmos em segundo plano.

Segundo Kroeff (1988), no Brasil há necessidade de mudanças radicais na área de testes psicológicos. Para tanto, tem-se que "a) chegar a um amplo acordo sobre quais instrumentos deveriam sofrer adaptações, normatizações e por quanto tempo deveria durar a revisão de um teste; b) pesquisar sobre a utilidade, validade, fidedignidade, capacidade diagnóstica e preditiva dos testes mais usados no país; c) formar recursos humanos na área da medida psicológica através de cursos de especialização e d) desenvolver uma tecnologia nacional para criar e desenvolver medidas psicológicas realmente adaptadas as nossas necessidades e peculiaridades, e capaz de adaptar ao nosso país aqueles instrumentos estrangeiros que mereçam este esforço" (pág. 537).

Seguindo esta vertente de pensamento, algumas pesquisas têm sido feitas, ainda que de forma precária, sobre as qualidades psicométricas de testes que são utilizados no Brasil.

Na Universidade de Brasília, no Instituto de Psicologia e mais especificamente no Departamento de Psicologia Social e do Trabalho, existe o Laboratório de Pesquisa em Avaliação e Medida (PAM), cuja filosofia de trabalho coaduna-se com as idéias expostas por Kroeff (1988).

Entre os muitos estudos - de adaptação e construção de testes - realizados pelo PAM, pode-se citar os trabalhos já concluídos de Ramos (1990) sobre a Escala de Atração Intersexual; Pinelli Jr. (1990) sobre o Teste de Integração Viso-Motora (VMI), Pasquali e Gouveia (1990) sobre a Escala de Assertividade Rathus (RAS); Pasquali, Wechsler e Bensusan (1991) sobre as Matrizes Progressivas do Raven Infantil; Pasquali, Azevedo e Andriola sobre as Matrizes Progressivas do Raven Geral (1992); Pasquali, Gouveia, Andriola, Miranda e Ramos (1993) sobre o questionário de Saúde Geral de Goldberg (QSG). Ainda em fase de coleta e análise de dados, estão os trabalhos de Pasquali (1993) sobre a adaptação do "Edwards Personal Preference Schedule" (EPPS) e de Andriola, Dias e Pasquali (1993) sobre a Escala Multifatorial para Estudantes Pós-graduandos.

Os resultados obtidos com o estudo de instrumentos adaptados para o Brasil indicam algumas diferenças psicométricas entre tais adaptações e as formas originais.

O Teste do Desenvolvimento da Integração Viso-Motora (VMI), foi adaptado para o Brasil por Pinelli Jr. (1990), através da validação de construto e da validação concorrente. Para o estudo da validade de construto as análises fatoriais indicaram a existência de dois fatores no VMI: 1) o fator viso-motor e 2) o fator integração viso-motora. Para a validação concorrente foram utilizados o Teste Raven - Escala Especial ($r = .50$ a $.69$.) e o Bender-Koppitz ($r = .54$ a $.67$). A precisão do teste foi calculada através do alfa de Cronbach ($.84$). Como conclusão, o autor sugere mais pesquisas sobre o VMI utilizando uma amostra maior, visto que os resultados fazem supor a existência de um fator de ordem superior, mas que requer pesquisas mais abrangentes (Pinelli Jr. & Pasquali, 1992).

A Escala de Assertividade Rathus (RAS), em sua adaptação brasileira, mostrou que tal instrumento é composto por um grande fator, no qual 20 dos 30 itens originais da escala possuíam cargas maiores que $.30$. Tal fator foi definido como Inibição vs Desinibição, apresentando uma consistência interna (Alfa de Cronbach) de $.81$. Dois outros fatores apareceram, mas com uma estrutura bastante indefinida. A conclusão do estudo foi a de que a escala é útil para medir Inibição vs. Desinibição, porém apresenta-se modesta para medir o construto assertividade (Pasquali & Gouveia, 1990).

Com relação às Matrizes Progressivas do Raven Infantil, os resultados mostraram que não se trata de um teste unifatorial, mas um teste que cobre quatro conceitos distintos: percepção

espacial, raciocínio analógico concreto, raciocínio analógico abstrato e raciocínio dedutivo. Embora composto de quatro fatores, o Raven Infantil, como instrumento psicométrico, não é adequado para avaliar todos estes fatores. Apenas o fator 'raciocínio analógico abstrato' é avaliado adequadamente pelo instrumento. Embora este fator contenha apenas 31 itens, mostra-se mais consistente que o teste original de 36 itens (Pasquali & colaboradores, 1991).

O Teste das Matrizes Progressivas do Raven Geral avalia mais de um fator, mais especificamente quatro, embora apenas três (analógico-abstrato, analógico-concreto e dedutivo) apresentem alta consistência interna, alfa de Cronbach de .88, .86 e .84 respectivamente. Apesar disso "apenas o fator dedutivo apresenta uma distribuição de itens cobrindo toda a gama de intensidade da aptidão envolvida, isto é, os itens se distribuem sobre todo o contínuo de dificuldade. Ao contrário, os fatores analógico-abstrato e sobretudo o fator analógico-concreto só contêm itens fáceis e muito fáceis, tornando-se, por isso, incapazes de discriminar entre sujeitos de habilidades diferentes nestas duas dimensões" (Pasquali & colaboradores, 1992:18).

O Questionário de Saúde Geral de Goldberg (QSG) considerado como um instrumento unidimensional, mostrou ser composto por cinco fatores - capacidade de desempenho ou realização, desejo de morte, tensão ou stress psíquico, distúrbios psicossomáticos e distúrbios do sono, cuja consistência interna medida pelo alfa de Cronbach foi de .89 para

A construção de um Teste de Raciocínio Verbal (RV)

os três primeiros fatores, .80 e .84 para os dois últimos, e .95 para a forma total. Como conclusão, os autores destacam que "os fatores resultantes desta adaptação brasileira mostram-se melhor configurados do que os encontrados por Goldberg no estudo original. Esta ocorrência pode dever-se, sobretudo, ao tipo de análise utilizada no presente estudo (PAF), em contraste com a análise dos componentes principais feita pelo autor" (Pasquali & colaboradores, 1993: 16).

Tomando como exemplo estes estudos, pode-se observar que alguns testes bastante conhecidos e aplicados - como por exemplo o Raven Infantil e o Raven Geral - apresentam diferenças psicométricas com respeito às formas originais, que são ignoradas por quem os utilizam. Como resultado, é plenamente justificável o descrédito da população sobre tais testes, visto que os mesmos estão medindo apenas uma parte do que realmente se propõem a medir - problema da representação do traço latente.

Um outro ponto de vital importância para a área - que é também relegado a segundo plano - é que além da maioria dos testes que são utilizados no país serem adaptações de outras realidades, não sofrem revisões ou atualizações periódicas do seu conteúdo - principalmente o verbal.

Segundo o próprio Centro de Psicologia Aplicada - CEPA (sem data-a), no manual da Escala de Inteligência Wechsler para Crianças (WISC), "o fato de publicar o CEPA um teste estrangeiro, simplesmente 'traduzido' para nossa língua, traz em si alguns inconvenientes e algumas vantagens. O grande inconveniente é

colocar em uso, para todos os psicólogos brasileiros, material padronizado em outro país, de características muito diferentes das nossas. Tal fato, teoricamente, invalidaria todo o esforço da tradução e publicação do WISC em nosso meio, com o escopo de ser usado como medida de inteligência para crianças brasileiras. A vantagem é fornecer aos psicólogos brasileiros material traduzido de maneira unificada, apoiado nas adaptações, decorrentes do uso constante durante sete anos, na Clínica Psicológica do Instituto de Psicologia da Pontifícia Universidade Católica de São Paulo" (pág. 9). É bom frisar que o teste foi publicado no Brasil em 1949, sendo utilizado desde então sem atualização do seu conteúdo nem elaboração de normas brasileiras.

Um outro exemplo é o da Escala de Inteligência Wechsler para Adultos (WAIS), cuja adaptação é datada de 1950 (CEPA, sem data-b). A amostra utilizada para a sua padronização foi de 2.735 homens e 1.175 mulheres da zona urbana e rural da cidade de São Paulo, com idades variando de 11 a 49 anos. Desde a sua primeira adaptação, não houve preocupação do CEPA de atualização do conteúdo do teste. A seguir são apresentados dois exemplos de itens de conteúdo verbal da primeira adaptação:

- A Moeda Argentina chama-se...
1. escudo 2. peso 3. peseta 4. duro 5. franco
- Se uma pessoa comprar 7 selos de 20 centavos e pagar com uma nota de 5 cruzeiros, quanto receberá de troco?

Os exemplos acima mostram a necessidade de revisão e atualização do teste, na parte verbal, uma vez que tais conteúdos mostram-se em uso.

O descaso dos profissionais responsáveis pela aplicação de testes psicológicos no Brasil, pelo estudo dos parâmetros psicométricos e pela atualização dos conteúdos dos mesmos, resultou em consequências desastrosas para a área, entre as quais voltamos a destacar o descrédito da sociedade sobre a sua utilidade. A Folha de São Paulo, em artigo datado de 22 de março de 1992, cujo título era "Empresas aposentam os testes psicotécnicos", levanta a questão sobre a utilidade dos testes psicológicos, pondo em dúvida a sua eficácia em medir o que se propõem (Sanger, 1992).

Diante dos argumentos anteriormente apresentados, é que surge a necessidade do resgate da posição legítima dos instrumentos de avaliação psicológica no contexto da Psicologia no Brasil. Com o duplo caráter de criação de testes brasileiros e de preocupação pela qualidade científica dos mesmos - estudo dos parâmetros psicométricos. Objetiva-se de forma indireta, recuperar o prestígio da classe dos psicólogos, já que o mesmo depende em grande parte dos testes psicológicos, "motivo pelo qual a noção de valor do teste deverá ser cercada de muito cuidado: 'Teste psicológico' deve ser marca registrada, a fim de garantir uma absoluta e incontestável qualidade do método científico" (Klausnitzer 1985:15).

2 - Porque construir um Teste de Raciocínio Verbal

2.1 - Utilidade de um Teste de Raciocínio

A palavra teste é originária do latim 'testis' (testemunha) derivando para o inglês 'test' (prova). Em sentido etimológico, fazer um teste seria dar testemunho ou fazer uma prova (Erthal, 1987).

No campo científico, fazer um teste significa medir, ou seja, comparar os resultados com um critério determinado (Klausnitzer, 1985).

Restringindo o conceito para o campo psicológico, um teste é um procedimento sistemático para comparar comportamentos de duas ou mais pessoas (Cronbach, 1960).

Para Van Kolck (1981a) um teste é um procedimento que objetiva medir um fenômeno atendendo às exigências do método experimental.

Só que para medir, faz-se necessário cumprir dois critérios básicos: 1) deve existir um ponto ou estado de referência, isto é, um critério propriamente dito e 2) deve haver uniformidade do estímulo, ou seja, uma situação estimuladora padronizada (Anastasi, 1988).

Segundo ainda Anastasi (1988), um teste psicológico deve ser uma medida objetiva e padronizada de uma amostra do comportamento que se deseja mensurar, ressaltando que tal amostra

A construção de um Teste de Raciocínio Verbal (RV)

deve ser cuidadosamente escolhida, para que represente o comportamento que se quer medir.

Os testes psicológicos foram criados e são utilizados na maioria das atividades humanas, principalmente para a determinação e análise de diferenças individuais na inteligência geral, aptidões específicas, conhecimentos escolares, orientação vocacional e dimensões não intelectuais da personalidade (Freeman, 1962; Ausubel & Robinson, 1969; Wood, 1975; Schmidt & Hunter, 1981; Anastasi, 1988).

Com esses objetivos, os testes são utilizados nas áreas militar, de educação, comércio, indústria - principalmente para seleção e classificação de sujeitos em categorias, classes e setores diferenciados - e na clínica - para o diagnóstico individual (Freeman, 1962; Rapaport, Gill & Schafer, 1968).

No caso dos testes de raciocínio verbal, são utilizados especificamente para o diagnóstico e prognóstico escolar, além de orientação vocacional (Almeida, 1986b).

Nos Estados Unidos, um dos mais conhecidos testes sobre o raciocínio verbal é o "Miller Analogies Test (MAT)". Tal instrumento é utilizado sobretudo para a seleção de universitários, para mensurar o nível de erudição e para a seleção e avaliação na indústria (Sternberg, 1977). É ainda muito empregado no Canadá, Grã-Bretanha, Austrália e Filipinas, na seleção de estudantes para pós-graduação (Van Kolck, 1981a). É composto por 100 itens em forma de analogias verbais, cobrindo uma grande variedade de campos de aprendizagem e especialização.

As relações a serem descobertas em cada item não são uniformes, implicando que o sujeito deve receber orientação para analisá-las uma de cada vez. O tempo concedido para responder o teste é de 50 minutos.

Ainda nos Estados Unidos, mais especificamente no estado de Washington, o governo utiliza uma bateria (Bateria Washington) para avaliar a aptidão acadêmica de alunos do segundo grau que visam ingressar na universidade (Hunt, 1992). Tal bateria é composta de sete subtestes, a saber, 1) compreensão de leitura, 2) vocabulário, 3) gramática, 4) habilidades quantitativas, 5) raciocínio mecânico, 6) raciocínio espacial e 7) desempenho em matemática.

Frederick L. e Rose G. Anderson, especificamente diretor de pesquisas do Departamento de Instituições Públicas do Estado de Minnessota e diretor da Divisão de Crianças do Centro de Serviços Psicológicos da cidade de Nova Iorque, construíram o "Kuhlmann-Anderson Intelligence Test" publicado pela primeira vez em 1927. O objetivo do teste é o de medir os níveis do desenvolvimento mental necessários ao sucesso no trabalho escolar. As atividades requeridas no teste são várias: completar um quadro; localizar a parte incorreta ou supérflua em um desenho; classificar objetos relacionados entre si; copiar ou completar desenhos; emparelhar figuras; completar séries; seguir direções; encontrar peças que devem ser colocadas juntas para reproduzir determinada figura; contar; resolver problemas numéricos (Van Kolck, 1981a).

Na Inglaterra, País de Gales e Escócia, os testes de raciocínio verbal têm sido utilizados para a seleção de estudantes de 2º grau. Entre os mais conhecidos, estão o "Moray House", produzido pela "Godfrey Thomson Unit for Educational Research", e alguns outros produzidos pela "National Foundation for Educational Research in England and Wales - NFER". Tais testes são compostos por itens em forma de analogias, associação de palavras, código, sentido verbal e completamento de sentenças (Butcher, 1972).

Em Portugal, Almeida (1986a) construiu uma Bateria de Provas de Raciocínio Diferencial (BPRD), composta por cinco sub-testes (verbal, numérico, espacial, mecânico e abstrato), tendo investigado a relação entre o desempenho de alunos secundaristas na BPRD e o sucesso escolar dos mesmos. Tal estudo indicou que "os alunos com melhores resultados nas provas (BPRD) apresentavam classificações escolares mais elevadas, uma menor taxa de insucesso escolar no final do ano letivo em que realizaram as provas e um menor número de reprovações em anos escolares anteriores" (Almeida, 1988a: 233).

Para Hunt (1992), os testes de capacidade verbal estão entre os melhores previsores isolados de sucesso na escola. Para reforçar tal posição, citamos a afirmação de Matarazzo (apud Hunt 1992:46) de que "os estudos realizados nos anos 50 em diante encontraram, consistentemente, correlações que variam de .40 a .60 entre os testes de capacidade verbal e o desempenho acadêmico".

Ainda segundo Hunt (1992: 44), "as pessoas que possuem uma alta capacidade verbal parecem saber mais acerca do nosso mundo". Num estudo realizado pelo próprio autor, estudantes universitários foram submetidos a três testes de conhecimentos fatuais: 1) acerca do humanismo, 2) sobre ciências sociais e 3) sobre ciências naturais. Os mesmos estudantes receberam um escore de "capacidade verbal", baseado em um composto de testes sobre compreensão de leituras, vocabulário e gramática. As correlações entre os testes de capacidade verbal e os testes de conhecimento foram de .42 para humanismo, .57 para ciências sociais, e .28 para as ciências naturais. O autor conclui então que "quanto mais altas as pontuações dos estudantes nos testes de capacidade verbal, maiores tendiam a ser seus pontos nos testes de conhecimento. A boa capacidade verbal está positivamente correlacionada com a aquisição de conhecimentos" (pág. 46).

Mas não é só na área acadêmica que os testes predizem o sucesso, fora dela também. Sticht (apud Hunt 1992) apresenta alguns dados sobre o desempenho de soldados em testes de conteúdo escrito e verbal e a capacidade para realizar algumas tarefas. As capacidades dos militares foram avaliadas pelo desempenho tanto de campo (atividades específicas) quanto escrito. A tabela 01 fornece tais informações:

A construção de um Teste de Raciocínio Verbal (RV)

Tabela 01 - Correlações entre o desempenho dos soldados em tarefas de trabalho e em medidas da capacidade verbal

	Capacidade	
	Leitura	Compreensão auditiva
Equipes de veículos blindados	.32	.29
Mecânicos de veículos	.26	.38
Oficiais intendentes	.40	.42
Cozinheiros	.34	.28

Os resultados indicam que, apesar das habilidades envolvidas para tripulações de carros blindados, mecânicos de veículos, cozinheiros e oficiais de intendência não serem aquelas normalmente consideradas como estando envolvidas nas "tarefas acadêmicas", as correlações entre tarefas verbais e as relacionadas ao trabalho eram apenas ligeiramente mais baixas do que as encontradas entre a capacidade verbal e a avaliação acadêmica. Sticht (apud Hunt 1992:46) afirma ainda que "... o resultado não é surpreendente. Em nossa sociedade, todos aprendem muito tanto pela leitura quanto pela instrução verbal de como realizar as coisas".

Almeida (1986b) ressalta ainda a importância dos testes de raciocínio para a orientação vocacional. Segundo o mesmo, parece evidente que quanto mais o aluno avança no nível de escolaridade mais os programas das disciplinas utilizam novas dimensões do raciocínio (resolução de problemas, abstração, inferência, sentido crítico, generalização de princípios).

Acrescenta, ainda, que os currículos e programas progressivamente tendem a ser mais específicos em seus conteúdos, obrigando o aluno "a responder através da realização de escolhas cada vez mais pormenorizadas e decisivas para os seus projetos escolares e profissionais futuros" (pág. 203). Finalmente, ressalta que a relevância dos testes de raciocínio com este propósito dá-se principalmente pela aproximação entre os conteúdos curriculares e os conteúdos de tais testes (verbal, numérico, espacial, abstrato e mecânico).

Um outro campo de aplicação de um teste de raciocínio verbal é o de diagnosticar problemas de aprendizagem em disciplinas como português. Em tal disciplina, principalmente para alunos de 2º grau, há um tópico específico para o trabalho com conteúdos verbais: a elaboração de redação.

Para Bueno (1986), 'redação' é o ato ou efeito de redigir e 'redigir' seria o ato de escrever com ordem e método. Pode-se dizer, então, que uma redação consiste numa exposição lógica, organizada e concatenada de idéias, utilizando-se símbolos verbais. Num teste de raciocínio verbal, cada item tem uma lógica e organização própria. Dessa forma, provavelmente, baixos escores no teste poderão indicar dificuldades de aprendizagem relativos aos conteúdos verbais.

Não existe no Brasil um instrumento adequado para a predição do desempenho acadêmico, diagnóstico de problemas de aprendizagem ou para a orientação vocacional que tenha sido desenvolvido para a nossa população estudantil. Apoiado em tal

argumento e nos fatos anteriormente explicitados sobre a qualidade dos testes adaptados para o país, pode-se sustentar que o objetivo primordial para a construção de um teste brasileiro de raciocínio verbal é o de proporcionar um instrumento que possa ser utilizado para tais fins, com conteúdo atualizado e cujas características psicométricas tenham sido estudadas cientificamente.

2.2 - Críticas aos Testes de Conteúdo Verbal

Hunt (1978) argumenta que as diferenças no raciocínio verbal podem ser compreendidas em termos de diferenças na velocidade individual de acesso à informação léxica na memória de longa duração. O termo 'léxico' refere-se ao conjunto de vocábulos que fazem parte de determinada língua. Dessa forma, um indivíduo que conheça um maior número de vocábulos poderá ter uma maior velocidade de acesso ao léxico na memória de longa duração, resultando em um desempenho melhor em tarefas de conteúdo verbal. Em outras palavras, um sujeito que possua um vasto vocabulário terá uma maior probabilidade de sucesso num teste de conteúdo verbal que um indivíduo que possua um vocabulário pobre.

Continuando na linha da observação feita anteriormente, afirma-se que a aquisição do vocabulário é reflexo de experiências e conhecimentos mais frequentes da classe média, favorecendo assim um melhor desempenho de indivíduos advindos de

tal classe em situação de teste (Gordon & Terrell, 1981; Bersoff, 1981; Alencar, 1986).

As opiniões de Garcia (1981) e Olmedo (1981) a respeito dos testes verbais confluem para o mesmo ponto: que na realidade o que medem são formas específicas de competências linguísticas e lógicas extremamente valorizadas pelas classes média e alta das sociedades ocidentais.

Já para Sawrey e Telford (1978) a problemática da influência cultural existe; porém o que tem que ser levado em conta, no caso específico dos testes de aptidão escolar, é a compatibilidade entre os currículos escolares e o conteúdo de tais testes.

Para Roazzi, Spinillo e Almeida (1991) há que se levar em conta que as habilidades intelectuais que são relevantes para uma determinada cultura podem não o ser para outras, implicando que processos mentais importantes para uma cultura podem não o ser para outra. Neste sentido, um aspecto que deve ser analisado com muito cuidado refere-se à forma de correção dos itens, ou seja, aos parâmetros adotados para se considerar uma resposta certa ou errada.

Alderton e Larson (1990) afirmam que, considerar uma única resposta como sendo a certa, significa ignorar a existência de um contínuo de conhecimento entre a resposta errada e a certa, visto que a dicotomização entre respostas certas e erradas é algo artificial. Strommen, McKinney e Fitzgerald (apud Roazzi & colaboradores, 1991) argumentam ainda que a resposta não

A construção de um Teste de Raciocínio Verbal (RV)

convencionada pela maioria como sendo a correta é prejudicada, manifestando-se desta maneira a tendenciosidade da classe social embutida nos testes.

Há críticas até quanto ao procedimento de avaliação (material e situação de teste) que podem favorecer indivíduos mais familiarizados com situações acadêmicas (Roazzi & colaboradores, 1991).

Shimberg (1981) tece críticas quanto ao índice de dificuldade dos itens, que na maioria dos testes, não seguem uma sequência linear crescente.

Um outro alvo das críticas aos testes de conteúdo verbal, refere-se às repercussões provocadas pelos escores obtidos pelo indivíduo (Brody & Brody, 1976; Hargadon, 1981; Reschly, 1981; Linn, 1982; Alencar, 1986), visto que há decisões a serem tomadas a partir de tais escores.

Neste sentido a posição de Freeman (1962) é a seguinte: basicamente existem dois tipos de decisão a partir dos resultados de um teste: 1) decisão pessoal - em orientação vocacional por exemplo, o sujeito tomará uma decisão pessoal sobre a área de conhecimento que seguirá na universidade - e 2) decisão institucional - por exemplo em seleção profissional, o encarregado da seleção escolherá o candidato mais apto para o cargo oferecido pela empresa. Shimberg (1981) acrescenta a esse argumento o seguinte: supondo que os parâmetros psicométricos do teste sejam confiáveis e que o mesmo foi escolhido acertadamente para o objetivo de sua aplicação, haverá uma maior probabilidade

de decisões acertadas. Ocorrendo o inverso, os resultados poderão ser catastróficos tanto para o indivíduo quanto para a empresa.

Klausnitzer (1985:21) ainda acrescenta que "é verdadeiramente surpreendente que uma sociedade ainda subsista razoavelmente sabendo-se que seus membros deixam inteiramente ao acaso a escolha de suas profissões, dos rumos de suas carreiras, da escolha dos cônjuges etc". Segue afirmando que os testes de aptidão, por possuírem alto grau de credibilidade, deveriam ser obrigatórios antes de cada decisão que se relacione com a capacidade de desempenho. Completa seu pensamento com um posicionamento favorável à utilização dos testes, visto que atualmente precisa-se tomar decisões rápidas em vários setores do trabalho ou da vida cotidiana e tais decisões devem ser bem fundamentadas e com resultados fidedignos a longo prazo.

Apesar da quantidade de críticas, a utilidade de um teste de raciocínio verbal para predição do desempenho acadêmico, diagnóstico de problemas de aprendizagem ou para a orientação vocacional, parece ser bastante evidente. Parece ser claro, também, que a população à qual se destinará tal teste deve possuir um vocabulário mínimo para a compreensão dos termos utilizados nos itens. No caso específico do teste de raciocínio verbal, desenvolvido no presente estudo, pelo fato de ter sido destinado à uma população estudantil, essa fonte de viés, provavelmente, foi minimizada.

3 - Fundamentação Teórica do Teste de Raciocínio Verbal (RV)

Para Nick (1988), torna-se imprescindível a ligação da teoria psicológica subjacente ao teste com a práxis da avaliação. Neste sentido apresenta-se a seguir o modelo teórico que embasou a construção do Teste RV.

O Teste de Raciocínio Verbal (RV) está apoiado na Teoria dos Fatores Primários, postulada por L. L. Thurstone em 1938.

Thurstone construiu sua teoria sobre os fatores primários a partir do estudo de um conjunto de dados resultante da aplicação de 56 diferentes testes a uma amostra de 218 estudantes da Universidade de Chicago (Cronbach, 1960; Brody & Brody, 1976; Alencar, 1986; Almeida, 1988b).

Utilizando a técnica da Análise Fatorial para o tratamento estatístico dos dados, concluiu, a partir da matriz de intercorrelações obtidas entre os testes, que não havia qualquer indicação para a existência de um fator geral - fator 'g' - encontrando pelo contrário nove fatores independentes: 1) espacial - S; 2) perceptivo - P; 3) numérico - N; 4) compreensão verbal - V; 5) fluência verbal - W; 6) memória - M; 7) indução - I; 8) raciocínio aritmético - R e 9) dedução - D.

Para identificar tais fatores, Thurstone postulou a lei da estrutura simples, que se baseia na total independência ou oposição entre os fatores. Em outras palavras, um item que esteja

representando a fluência verbal, terá correlações próximas de 1 naquele fator e correlações próximas de zero nos outros fatores.

Thurstone encontrou dificuldades empíricas para a comprovação do postulado da estrutura simples, ou seja, da independência total entre os fatores. Optou então pela aceitação da existência de alguma intercorrelação entre os fatores. Em outras palavras, duas aptidões ou fatores podem ser distintos mas podem não ser totalmente independentes entre si (Almeida, 1988b). Por exemplo, itens para mensurar a fluência verbal poderiam também estar relacionados com a compreensão verbal.

Esta mudança de posição de Thurstone resultou no agrupamento de três fatores propostos inicialmente como independentes - os fatores de indução, dedução e raciocínio aritmético - com o novo nome de fator de raciocínio.

Dessa forma, os fatores primários propostos por Thurstone (1948), em número de sete, são descritos a seguir:

1 - Espacial (S): capacidade de visualização de objetos num espaço bi ou tridimensional. Os itens são normalmente desenhos ou figuras geométricas simples que se movimentam em várias direções, ou assumem diferentes formas, de acordo com a perspectiva em que são observados.

2 - Velocidade perceptiva (P): capacidade de responder, com acuidade e rapidez, pequenas diferenças ou semelhanças entre duas ou mais figuras. Os itens típicos deste tipo de teste são representados por figuras.

A construção de um Teste de Raciocínio Verbal (RV)

3 - Numérico (N): capacidade de lidar com números e efetuar operações aritméticas simples, de forma rápida. Itens característicos de tal capacidade, são operações - de adição, subtração, multiplicação e divisão - em que o sujeito assinala a resposta correta ou itens em que o próprio respondente fornece o resultado das operações envolvidas.

4 - Compreensão verbal (V): capacidade de compreensão de idéias expressas através de palavras. São itens típicos deste fator a geração de palavras que expressem antônimos e sinônimos, bem como a compreensão de sentenças - interpretação sobre uma afirmativa.

5 - Fluência verbal (W): capacidade de produzir palavras de forma rápida e correta. Itens característicos para a realização desta tarefa, são aqueles em que apresenta-se uma letra ao sujeito e pede-se para que faça o maior número de palavras que comecem com tal letra.

6 - Memória (M): capacidade de evocar estímulos de forma correta. São característicos itens formados por pares de palavras, frases ou números. A tarefa do sujeito consiste em completar o par de elementos quando lhe for apresentado o primeiro componente de tal par.

7 - Raciocínio (R): capacidade de resolver problemas lógicos. Cada item pressupõe a descoberta e a aplicação de uma lei geral de sucessão de dígitos (letras ou números), conjunto de dígitos (palavras), ou de transformação de figuras (espaciais, mecânicas, concretas ou abstratas).

Como o objetivo do presente trabalho foi o de desenvolver um teste sobre raciocínio verbal, em seguida será dada maior atenção ao fator raciocínio.

4. O Raciocínio segundo diferentes Enfoques

Existem três grandes vertentes de estudiosos do raciocínio. Comumente são taxados como "diferencialistas", "desenvolvimentistas" e "cognitivistas" (Almeida, 1988b).

Na vertente "diferencialista", os grandes expoentes são L. L. Thurstone, P. E. Vernon e J. P. Guilford.

Os trabalhos destes estudiosos enfatizam a realização cognitiva do indivíduo como sendo resultante do uso diferenciado das aptidões mentais para a realização de tarefas (diferenças individuais), bem como do conteúdo subjacente às tarefas (especificidade do conteúdo). Acrescentam ainda que a realização cognitiva poderá exigir maior ou menor grau da mesma aptidão, dependendo da especificidade do conteúdo. Em outras palavras, o conteúdo da tarefa pode requerer maior ou menor grau cognitivo daquela aptidão, ou até de outras aptidões, para a sua realização. Por exemplo, um item para mensurar o raciocínio verbal cuja palavras componentes estejam em desuso, provavelmente este item irá requerer o uso de outras aptidões - como a fluência verbal ou o conhecimento vocabular -, além daquela a que se destina a tarefa para a sua realização.

A construção de um Teste de Raciocínio Verbal (RV)

Para estes estudiosos, há diferentes aptidões, todas exigindo diferentes processos cognitivos. No caso do raciocínio, seria característico o uso de processos cognitivos associados à capacidade de apreensão de relações e à sua aplicação.

Os trabalhos desenvolvidos pelos estudiosos desta vertente são caracterizados, sobretudo, pelo uso da análise fatorial como método estatístico para o estudo dos seus modelos teóricos.

No grupo dos "desenvolvimentistas" destaca-se J. Piaget. Basicamente enfatiza as diferenças entre adultos e crianças quanto ao uso dos processos cognitivos para a resolução de tarefas cotidianas (Stenberg, 1980).

Centra-se no estudo da influência de determinantes hereditários e ambientais para o desenvolvimento da capacidade intelectual do sujeito. A evolução intelectual seria resultado de etapas ou estádios sequenciais e invariantes. Estes estádios sofreriam influências do meio e também influenciariam o meio, em outras palavras, o sujeito incorpora aspectos do mundo exterior às suas estruturas internas já construídas - processo de assimilação - e também adequa as suas estruturas à nova realidade ou aos novos conhecimentos adquiridos - processo de acomodação. Para tal desenvolvimento, Piaget considerou a influência da maturação do sistema nervoso central e periférico, da ação sobre os objetos, dos fatores sociais e do fator de equilíbrio ou auto-regulação - representando este último fator a adaptação do indivíduo com o seu meio.

Ainda segundo Piaget, o desenvolvimento intelectual se iniciaria com o nascimento atingido o apogeu no final da adolescência. Na adolescência a capacidade intelectual estaria apta à lidar com conceitos que envolvem idéias de abstração, dedução, indução, lógica etc. Tal capacidade se aproxima do que foi denominado de 'raciocínio' pelos "diferencialistas". Como método de trabalho, Piaget empregava a metodologia experimental para o estudo do seu modelo teórico.

Sendo a vertente de estudos mais recente, a linha dos "cognitivistas" caracteriza-se por estudiosos como M. I. Posner, R. Glaser, J. W. Pellegrino e R. J. Sternberg. Os estudos deste grupo enfatizam os processos cognitivos básicos usados na resolução de uma determinada tarefa.

Para estes pesquisadores, existem componentes cognitivos universais subjacentes a toda realização intelectual dos sujeitos. Em outras palavras, independentemente do conteúdo da tarefa, os processos cognitivos requeridos são sempre os mesmos. A ênfase está na importância relativa de cada processo cognitivo para a resolução de tarefas diferentes. Por exemplo, em tarefas de diferenciação de sinais visuais, os processos de codificação são muito exigidos, enquanto para a resolução de uma analogia os processos de escolha e decisão seriam mais importantes, embora nas duas tarefas sejam imprescindíveis a utilização de todos eles (Sternberg, 1977b).

Para Sternberg, o importante seria enfocar as análises dos processos cognitivos exigidos para a resolução das tarefas e

não tanto das diferenças entre os indivíduos. Sua posição está embasada na visão de que os indivíduos, independentemente do seu sexo ou idade, realizam as mesmas tarefas tendo que utilizar os mesmos processos cognitivos. Em outras palavras, os processos cognitivos são mais específicos das tarefas que dos sujeitos, sendo as diferenças entre os indivíduos devidas à maior ou menor eficiência no uso de tais processos.

Os processos cognitivos enfocados pelos "cognitivistas", são próximos do conceito de 'raciocínio' dos "diferencialistas" e do 'fator de auto-regulação' proposto pelos "desenvolvimentistas".

Os estudiosos da vertente "cognitivista" também empregam a metodologia experimental. Recorrem a situações laboratoriais em que tarefas e situações de realização são manipuladas para verificação de sua influência sobre o uso dos processos cognitivos.

Apesar das diferenças subjacentes às várias abordagens, pode-se afirmar que as três vertentes confluem para pelo menos dois pontos em comum: 1) o de que a aptidão em raciocinar (como querem os 'diferencialistas'), o fator de auto-regulação (como querem os 'desenvolvimentistas') e os componentes dos processos cognitivos (como querem os 'cognitivistas') poderiam ser referidos por um fator único, o raciocínio, (responsável pela utilização de processos cognitivos característicos dele) utilizado para a resolução de diferentes tarefas; 2) as três vertentes poderiam ser rotuladas conjuntamente de

"cognitivistas", já que o 'raciocínio' é tido como um processo cognitivo (processo latente) em todas elas.

Aceitando-se que este fator é utilizado, sobretudo, na solução de problemas, que a partir da infância os indivíduos estão sempre se deparando com problemas e que a solução dos mesmos é influenciada diretamente pela natureza da tarefa e o tipo de conhecimento trazido pelo respondente, pode-se concluir que o uso de tal fator é algo rotineiro (Chi & Glaser, 1992). Para estes autores, um problema é uma situação na qual o indivíduo tenta alcançar um objetivo (resolver o problema) através de meios apropriados (utilização adequada de processos cognitivos).

Dessa forma, a maneira mais adequada para que tal fator - denominado aqui de raciocínio - seja estudado é através da apresentação de problemas ao indivíduo, solicitando-lhe que os resolva.

A seguir será dada uma maior ênfase ao raciocínio segundo a ótica diferencialista.

5 - A Definição de Raciocínio segundo a Perspectiva Diferencialista

Almeida (1986b) caracteriza a investigação e a prática psicológica da perspectiva diferencial, com referência ao fator raciocínio, segundo três aspectos fundamentais. O primeiro refere-se à possibilidade de estabelecimento de uma relação entre

o desempenho dos sujeitos no teste e a sua realização escolar, permitindo assim prever a margem de sucesso esperado para cada indivíduo, bem como a decisão individual sobre seu futuro escolar e profissional.

O segundo aspecto seria resultante do postulado de que "as características intelectuais dos indivíduos permaneceriam estáveis ao longo da sua existência, designadamente após os primeiros anos de maior desenvolvimento psicofisiológico" (Almeida, 1986b: 202). Tal postulado fundamenta a existência de diferenças individuais permanentes quanto às capacidades intelectuais, enfocando a influência hereditária. Apesar disso, não é negada a influência das aprendizagens ou experiências sociais sobre a utilização do raciocínio.

O terceiro aspecto refere-se à distribuição dos escores obtidos pelos sujeitos seguirem as leis da curva normal, ou seja, há maior número de indivíduos cujos resultados situam-se em torno da média da população, e indivíduos cujos escores situam-se simetricamente acima e abaixo de tal valor médio. Tal lei permite a generalização dos resultados obtidos com a amostra, para a população da qual é originária.

Como o raciocínio pode ser usado em problemas com conteúdos diferentes, os teóricos psicometristas acreditam na existência de um fator de raciocínio - expresso pela capacidade de estabelecer e aplicar relações entre elementos, ou em outras palavras, expresso pela capacidade de induzir e deduzir relações

- e em aptidões primárias ou específicas - expressas pelas diferenças de contexto ou conteúdo dos elementos.

Dessa maneira os testes de raciocínio têm como objetivo mensurar o construto raciocínio, nos seus componentes indutivo e dedutivo, em diferentes contextos ou conteúdos (verbal, numérico, espacial, abstrato e mecânico).

Dependendo do conteúdo que compõe o problema - especificidade do conteúdo - haverá a contaminação do item por tal conteúdo. O argumento subjacente à tal afirmativa é que um item "puro", ou seja, que represente e mensure um único fator, nunca é conseguido na prática. Por exemplo, um item de conteúdo numérico para mensurar a aptidão em raciocinar com números, provavelmente terá uma alta comunalidade com o fator de aptidão numérica. Em outras palavras, um item, na realidade, não mensura de forma isolada um único atributo ou fator. É contaminado pela comunalidade do item com fatores "distintos".

Assim, para Bueno (1986), de forma bastante genérica, o termo 'raciocínio' refere-se à operação mental que se resume em equacionar um problema e resolvê-lo.

Para Almeida (1988a), duas características básicas poderiam descrever o raciocínio: 1) a generalidade de ocorrência e 2) o componente relacional. A primeira refere-se aos níveis em que o raciocínio é utilizado para a resolução de tarefas. É utilizado desde as mais simples tarefas até as mais complexas, desde tarefas tipicamente intelectuais até situações quotidianas de aprendizagem e de resolução de problemas.

A construção de um Teste de Raciocínio Verbal (RV)

A segunda refere-se a relação do construto "raciocínio" ante as abordagens tidas como "cognitivistas", "desenvolvimentistas" e "diferencialistas". Em todas estas abordagens o raciocínio é tido como uma função mental que no tratamento da informação vai além do trabalho de codificação e armazenamento, embora mesmo em tais fases seja requerido ou tido como importante.

Pode-se então depreender que o raciocínio é um mecanismo cognitivo que é utilizado para a solução de problemas (simples ou complexos), em suas mais diferentes formas de expressão (verbal, numérico, abstrato, espacial e mecânico) através de seus componentes indutivo e dedutivo.

Sendo assim, Almeida (1988a) propõe que o raciocínio poderia ser caracterizado pela aptidão - compreendida aqui como sendo uma menor ou maior facilidade do indivíduo em realizar tarefas específicas - do sujeito em:

- identificar os elementos componentes de um problema;
- conceitualizar um problema ou a sua formulação;
- conceber formas alternativas de resolver um problema;
- avaliar as diferentes formas alternativas elaboradas para resolver um problema;
- retirar conclusões lógicas da informação fornecida e processada, quer para a formulação do problema quer para a supervisão ou monitorização nos vários momentos da sua resolução;
- utilizar os componentes indutivos (ou de inferências) e dedutivos (ou de aplicação) nos procedimentos anteriores;

- utilizar os procedimentos anteriores independentemente do conteúdo e da forma das situações;

- avaliar a adequação da resposta elaborada tomando em consideração mais a especificidade da situação que a "opinião pessoal" sobre a mesma.

Aceitando-se tais componentes cognitivos como caracterizadores do raciocínio, a pergunta que se apresenta agora é: como operacionalizar situações em que o raciocínio seja utilizado? É o que se tenta responder a seguir.

6 - Como mensurar o Raciocínio Verbal?

Apesar da diversidade de Testes de Inteligência e das diferenças teóricas subjacentes aos mesmos, há um elemento em comum entre eles: o fato da maioria deles possuírem itens (verbais ou não) em forma de analogia.

Segundo Bueno (1986), analogia seria um ponto de semelhança entre coisas diferentes.

Para Aristóteles, analogia seria um termo que designaria uma igualdade de proporções envolvendo pelo menos quatro elementos. As analogias foram usadas extensivamente por Platão, Aristóteles e Tomás de Aquino para ajudar no desenvolvimento e explicação dos significados dos conceitos matemáticos, filosóficos e metafísicos (Pellegrino 1992).

A primeira aparição de itens em forma de analogia nos testes de inteligência ocorreu simultaneamente no Teste de

Relações Mistas de Woodworth e Well (1911), nos Estados Unidos, e no Teste de Completação de Analogias de Burt (1911), na Inglaterra (Pellegrino, 1992).

Para Oppenheimer (1956), o uso de analogias é algo inevitável no pensamento humano.

As analogias são utilizadas nos Testes de Inteligência por permitirem o uso de processos cognitivos, quando de sua resolução, característicos do que no presente trabalho definiu-se como 'raciocínio'.

Segundo Spearman (apud Sternberg 1977a) a solução para uma questão em forma de analogia, requer do indivíduo o uso de três processos cognitivos:

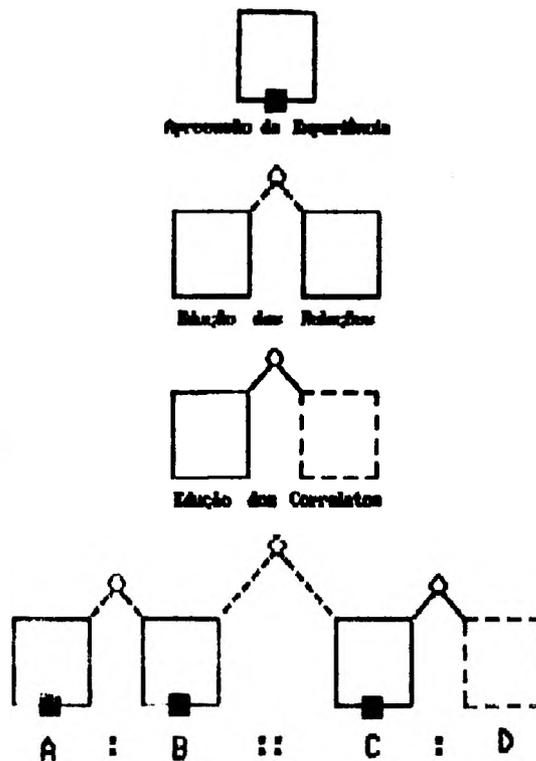
1 - apreensão da experiência: relaciona-se com a maior ou menor capacidade do indivíduo em apreender ou codificar a informação que lhe chega aos sentidos;

2 - educação de relações: tem a ver com a capacidade do indivíduo em inferir relações entre os elementos individuais da informação apresentada (às vezes esta tarefa poderá requerer a evocação de informações armazenadas na memória de longa duração);

3 - educação dos correlatos: é a aplicação da relação inferida a um novo domínio (fundamento).

Os três processos são ilustrados na figura 01.

Figura 01 - Processos cognitivos propostos por Spearman, 1923



Fonte: Sternberg, 1977a.

Tenta-se, em seguida, explicar como tais processos são utilizados para responder um item em forma de analogia, através do exemplo a seguir:

Sapato:Pé :: Luva: ? a.Braço b.Mão c.Dedos
A : B C : D

Considerando, então, a analogia acima ilustrada, A:B :: C:D, o primeiro princípio será posto em prática a partir do momento em que o indivíduo lê as palavras, isto é, ele precisa

A construção de um Teste de Raciocínio Verbal (RV)

apreender o significado dos conceitos que são os componentes da analogia. O segundo princípio consiste em descobrir uma relação entre os conceitos do primeiro par. No terceiro princípio, o indivíduo aplica a relação descoberta anteriormente a um novo par de conceitos, no presente caso, entre a palavra Luva e as possíveis respostas apresentadas. A tarefa do indivíduo consistiria em identificar o correlato (no caso 'Mão') específico para o termo 'Luva' a partir de uma relação apresentada.

Já para Sternberg (1977a) a resolução de uma questão em forma de analogia envolve basicamente seis princípios:

1. codificação ("encoding"): refere-se a percepção, codificação e registro na memória operativa (de trabalho ou a curto prazo) das características relevantes da informação (problema) para a sua posterior solução;

2. inferência ("inference"): é o estabelecimento de uma relação entre os dois primeiros elementos da analogia e sua retenção na memória de trabalho;

3. mapeamento ("mapping"): é a tentativa de descobrir uma relação (correspondência) entre o primeiro e o terceiro elemento da analogia, retendo-a na memória;

4. aplicação ("application"): é a comparação do terceiro elemento da analogia com cada uma das alternativas propostas como respostas, para verificar se e com qual delas a relação inferida anteriormente se realiza;

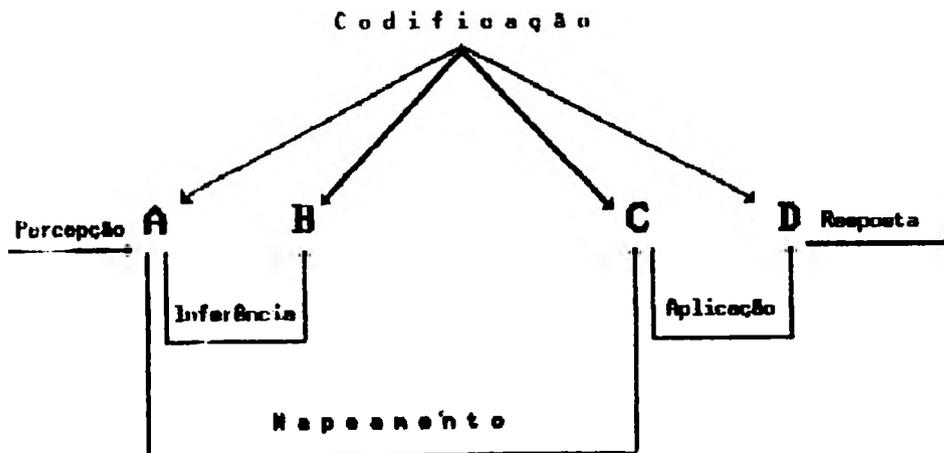
5. justificação ("justification"): há analogias em que nenhuma opção aparece como sendo a resposta correta (por possuir

regras ambíguas ou permitir várias respostas aparentemente corretas); neste caso, o indivíduo justifica a sua resposta como sendo superior às demais possíveis;

6. resposta ("response"): é a complementação da analogia, através da indicação da alternativa considerada como correta.

A figura 02 mostra a representação esquemática dos princípios propostos por Sternberg (1977a).

Figura 02 - Modelo proposto por Sternberg, 1977



Fonte: Sternberg, 1977a.

A construção de um Teste de Raciocínio Verbal (RV)

Tomando o exemplo anterior, tem-se:

Sapato:Pé : : Luva: ? a.Braço b.Mão c.Dedos
A : B C : D

O primeiro princípio será colocado em prática a partir do momento em que o indivíduo, lendo as palavras, apreende o seu significado e registra as informações em sua memória operativa. O segundo princípio consiste em tentar estabelecer uma relação entre o primeiro par de conceitos. No terceiro, há tentativas para a descoberta de uma relação (correspondência) entre o primeiro e o terceiro elemento da analogia, com posterior retenção de tal informação na memória. No quarto princípio, há uma procura da relação inferida entre o terceiro elemento e cada uma das alternativas propostas como respostas. No caso do presente exemplo o indivíduo utilizaria o quinto princípio, visto existirem duas respostas ('Mão' e 'Dedos') que podem, a primeira vista, ser consideradas corretas. O sexto princípio seria a indicação da resposta considerada mais correta.

Pellegrino (1992) apresenta uma descrição geral dos processos necessários para a solução de analogias. Segundo o mesmo, os processos cognitivos se ajustariam a três classes gerais. A primeira consistiria na descoberta de atributos ou processos de codificação, ou seja, os atributos importantes de cada termo individual na analogia devem ser representados na memória. Para os itens verbais, isso envolveria a ativação de um conjunto de aspectos semânticos associados a um conceito.

A segunda classe consistiria em processos de comparação de atributos usados para pares específicos de itens. Um dos processos de comparação seria a inferência, e lidaria com o relacionamento entre os dois primeiros termos da analogia. No caso de estímulos verbais, o processo de inferência teria o papel de descobrir aspectos semânticos que poderiam unir os dois conceitos. Outro processo utilizado é o mapeamento, cujo objetivo seria encontrar correspondências entre o primeiro termo e o terceiro ('A' e 'C') de uma analogia. A aplicação seria um outro processo de comparação de atributos e, consistiria na aplicação da regra específica inferida do par 'A-B' para os atributos do termo 'C', a fim de produzir uma solução ('D') adequada para a analogia. "A inferência, mapeamento e aplicação são processos componentes associados com os termos-raízes de uma analogia, isto é, os termos 'A', 'B', e 'C'" (Pellegrino, 1992: 222).

A terceira classe consistiria de processos de avaliação, que determinariam a propriedade de qualquer completção ou termo 'D'. Em analogias simples, onde o termo 'D' correto pode ser facilmente identificado, a avaliação torna-se também um processo de confirmação, visto que os aspectos do termo final são combinados e comparados à resposta ideal. Este processo leva à rejeição de completações inapropriadas e à seleção da resposta correta.

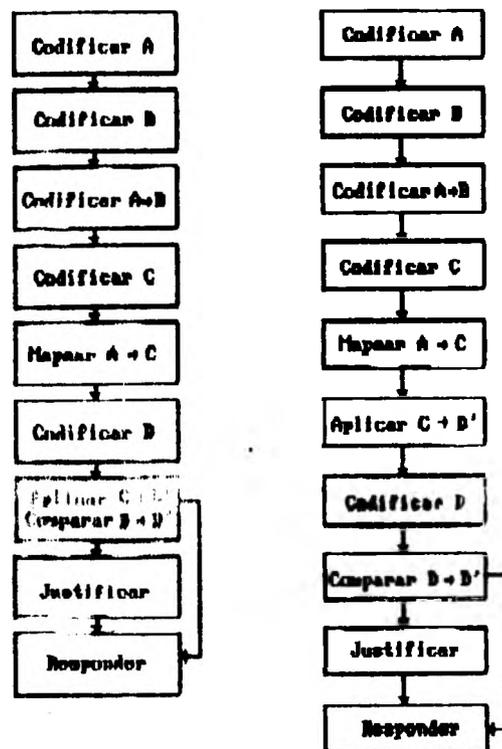
Entretanto, se a analogia envolver regras ambíguas ou proporcionar várias respostas possíveis, haverá o adcionamento de um outro processo na escolha da resposta correta: a

A construção de um Teste de Raciocínio Verbal (RV)

justificação. Neste caso a avaliação das alternativas proporciona dois ou mais termos de completção que parcialmente irão se ajustando aos aspectos de uma resposta aceitável, através de comparações e discriminações. O processo final na solução de analogias seria a execução de uma resposta manifesta.

A figura 03 apresenta os modelos de processos simplificados utilizados para a solução de analogias, propostos por Pellegrino (1992).

Figura 03 - Modelos simplificados propostos por Pellegrino, 1992



Fonte: Pellegrino, 1992

Tomando a analogia anteriormente citada como exemplo, observa-se como os processos são utilizados:

Sapato:Pé : : Luva: ? a.Braço b.Mão c.Dedos
A : B C : D

No modelo da esquerda a sequência será a seguinte. O sujeito deverá codificar o termo 'A' ('Sapato') e o termo 'B' ('Pé'). Em seguida infere uma relação entre 'A' e 'B'. O termo 'C' ('Luva') é então codificado, e o processo de mapeamento de 'A-C' é executado. É então feita a codificação de 'D', e um processo de aplicação-confirmação combinado é realizado, resultando na aceitação ou rejeição do termo 'D'. Como o item está em um formato de escolha forçada, o ciclo continua até que todas as opções (D') tenham sido avaliadas e a resposta correta identificada.

No modelo da direita, a sequência dos processos permanece a mesma, exceto que a aplicação (C → D') ocorre antes da codificação do termo 'D', que é acompanhada por um processo de comparação (D → D'). Caso a resposta não seja confirmada como a correta o ciclo continua.

As posições citadas, apesar das diferenças subjacentes, convergem para a posição de que itens em forma de analogias são os que mais fidedignamente caracterizam a utilização do raciocínio. Baseado nesta argumentação, os itens do teste do presente trabalho, se apresentam em formato de analogias de conteúdos verbais.

II - A Construção do Teste de Raciocínio Verbal (RV)

1 - Formato dos itens componentes do Teste RV

Basicamente os itens do presente teste foram construídos seguindo duas formas de analogias: 1) do tipo A:B::C:D e 2) do tipo A:C::B:D.

No primeiro tipo, tem-se uma relação entre os dois primeiros elementos da analogia (A:B) que é repetida ao outro par de elementos (C:D). Um exemplo típico deste tipo de analogia é a que se segue:

Alto : Baixo :: Estreito : largo.

No segundo tipo, a relação existente se dá entre o primeiro e o terceiro elementos da analogia (A:C), que é também aplicada ao segundo e quarto (B:D). Pode ser exemplificado pela analogia:

Soldado : Médico :: Fuzil : Bisturi.

1.1 - Como aumentar a complexidade de uma analogia?

Muitos trabalhos têm sido realizados sobre a resolução de analogias. Na maior parte o objetivo é estudar a influência da manipulação de algumas variáveis, consideradas relevantes ao problema, sobre a qualidade da resposta (número de acertos) e o tempo de resposta. Os resultados fornecem pistas a respeito dos

processos cognitivos utilizadas para a resolução de analogias. Os estudos sobre tais processos são muitos, porém são destacados aqui os de Carpenter e Just (1975), Sternberg (1977ab), Hunt (1978) e Sternberg e Nigro (1980). Este grupo de estudiosos defende a universalidade dos processos cognitivos utilizados para a resolução de analogias, independente do conteúdo específico (Marmèche, Mathieu & Barais, 1985).

Entre as muitas variáveis pesquisadas, citam-se entre outras, a diferença do conteúdo específico da analogia (Sternberg, 1977b; Hunt, 1978), o número de transformações sofridas pela analogia e o número de elementos componentes da mesma (Mulholland, Pellegrino & Glaser, 1980; Sternberg & Nigro, 1980), o número de alternativas apresentadas e o tipo de relação entre as palavras (Sternberg & Nigro, 1980), o conhecimento anterior do sujeito com o conteúdo da analogia (Gick & Holyoak, 1980), o formato da analogia (Sternberg, 1977b), além de itens em formato de sentenças afirmativas e negativas (Carpenter & Just, 1975).

Os resultados revelam que, quanto a diferença de conteúdo, as analogias verbais exigem maior tempo para serem respondidas que analogias geométricas ou do tipo pessoa-parte ('people-piece') (Sternberg, 1977b; Hunt, 1978).

Quanto ao número de transformações sofridas pela analogia e o número de elementos componentes da mesma, são variáveis utilizadas para o aumento da dificuldade da analogia. Quanto mais elementos estiverem presentes na analogia e quanto

maior o número de transformações tiver sofrido a mesma, mais difícil se tornará a mesma (Mulholland & colaboradores, 1980; Sternberg & Nigro, 1980).

No caso específico de analogias de conteúdo verbal, os resultados relativos ao número de alternativas apresentadas e ao tipo de relação entre as palavras, revelam que quanto menos alternativas maior número de acertos, visto que, a probabilidade de acerto ao acaso aumenta consideravelmente. Quanto ao tipo de relação entre as palavras, a mais difícil envolve a idéia de sinonímia e a mais fácil de funcionalidade ou uso (Sternberg & Nigro, 1980).

Sobre o conhecimento anterior, há um consenso de que, quanto mais familiar o conteúdo da analogia for para o sujeito mais rapidamente será respondida e maior número de acertos serão conseguidos (Gick & Holyoak, 1980; Perfetti, 1992). Para Hunt (1978), a capacidade cognitiva do indivíduo depende do conhecimento anterior e do uso adequado dos processos cognitivos para o processamento da informação.

Com relação a itens em formato de sentenças afirmativas e negativas, os resultados mostram que sujeitos mais novos têm mais dificuldade para a compreensão de sentenças negativas que afirmativas (Carpenter & Just, 1975).

Pode-se concluir a partir destas pesquisas, que no caso de analogias verbais, variáveis como o tempo de resposta, o número de alternativas apresentadas, o tipo de relação entre as palavras, o conhecimento anterior do sujeito (familiaridade com

as palavras) e o formato da analogia, são parâmetros relevantes para o aumento da dificuldade do item.

No caso específico do presente estudo, para a construção dos itens do teste, controlaram-se três variáveis do conjunto mencionado anteriormente, que poderiam influenciar o nível de dificuldade: 1) o formato da analogia; 2) o tipo de relação entre as palavras e 3) o nível de conhecimento das palavras componentes da analogia (a saber: um conceito concreto é considerado mais familiar que um abstrato).

Uma outra variável considerada, e não estudada de forma sistemática, foi a proximidade semântica das alternativas propostas com a resposta considerada como correta. Hipotetizou-se que, quanto mais próximas (conceitualmente) as alternativas apresentadas da resposta considerada como a correta, mais difícil seria o item.

Dessa maneira, hipotetizou-se que uma analogia do tipo $A:B :: C:D$ pode ser considerada mais fácil que uma analogia do tipo $A:C :: B:D$. Outra hipótese é a de que o tipo de relação entre as palavras pode dificultar a resposta a um item. A terceira hipótese é de que o nível de abstração da palavra pode tornar mais difícil uma analogia. Uma última hipótese é a de que uma analogia cujas alternativas apresentem uma resposta sem associação imediata com as outras, torna-se mais fácil que uma analogia cujas alternativas apresentem certo nível de proximidade semântica entre si.

2 - Construção dos itens

Por serem os itens em formato de analogias verbais e, segundo Gruber (1975), ser possível a construção de cerca de 40 tipos de relações analógicas diferentes entre as palavras, obrigatoriamente o instrumento teria que ser composto pelo maior número possível de itens cobrindo tais relações com o objetivo de garantir a representatividade do traço latente - no caso o raciocínio verbal - a ser mensurado.

Como o objetivo do teste é mensurar, especificamente, a aptidão do indivíduo em raciocinar com palavras, espera-se que, nas analogias, as palavras componentes sejam compreensíveis e utilizadas no dia-a-dia do respondente, caso contrário o problema a resolver seria primariamente o de compreensão de palavras e/ou fluência verbal.

Para Sternberg e Nigro (1980) e Almeida (1986a), as relações ideais para este tipo de prova são as que envolvem noções de funcionalidade, continuidade, antonímia, sinonímia e espécie, visto serem muito utilizadas no cotidiano pelos estudantes.

Dessa forma, das 40 relações analógicas propostas por Gruber (1975), cinco delas, expressavam algumas das noções sugeridas por Sternberg e Nigro (1980) e Almeida (1986a), e o restante referia-se a conhecimentos bastante específicos de

história geral, de geografia, de política americana, de gramática, de mitologia grega e romana etc. As cinco categorias a serem utilizadas no presente instrumento são caracterizadas e exemplificadas na tabela 02.

Tabela 02 - As cinco categorias utilizadas para a construção dos itens

Categorias	Especificações	Exemplos
1.Causalidade	<ul style="list-style-type: none"> - Causa-efeito - Implicação - Decorrência de fatos, eventos, atitudes 	<p>Inverno : Frio :: Verão : Calor</p> <p>Dormir : Escurecer :: Acordar : Clarear</p> <p>Esperança : Futuro :: Remorso : Passado</p>
2.Significação	<ul style="list-style-type: none"> - Sinônimos - Antônimos 	<p>Inerte : Estático :: Dinâmico : Ativo</p> <p>Minoria : Maioria :: Pouco : Muito</p>
3.Intensidade/ Qualidade	<ul style="list-style-type: none"> - Mudança de quantidade, de intensidade - Mudança de grau, de posição, de estado 	<p>Grande : Gordo :: Magro : Pequeno</p> <p>Leite : Iogurte :: Trigo : Pão</p>
4.Conjunto	<ul style="list-style-type: none"> - Parte-todo - Noção de categoria, espécie 	<p>Jogador : Time :: Aluno : Classe</p> <p>Girafa : Herbívoro :: Urubu : Carnívoro</p>
5.Funcionalidade	<ul style="list-style-type: none"> - Relação de função ou de uso - Profissão-atividade - Contexto-Profissão 	<p>Bússola : Navegador :: Semáforo : Motorista</p> <p>Professor : Ensinar :: Juiz : Julgar</p> <p>Quartel : Soldado :: Fábrica : Operário</p>

Seguindo as especificações caracterizadoras de tais categorias, um conjunto inicial de 101 analogias foi construído.

3 - Seleção dos Itens

O passo inicial na seleção dos itens, consistiu numa análise semântica dos 101 itens, feita por 14 juizes, todos psicólogos - dos quais oito eram professores universitários e seis alunos de pós-graduação em psicologia - com os objetivos de 1) verificar a existência de itens que fornecessem mais de uma resposta correta e 2) verificar a clareza das regras envolvidas nas analogias.

Um exemplo de observação feita a respeito de itens que forneciam mais de uma resposta correta, segue abaixo:

Beber está para Água, assim como Comer está para:

a. Alimento b. Comida c. Engolir d. Conduzir e. Mastigar

As letras 'a' e 'b' podem ser consideradas como corretas.

Quanto às regras das analogias, uma observação típica é a do exemplo abaixo:

Pássaro está para Gaiola, assim como Homem está para:

a. Prisão b. Casa c. Comida d. Carro e. Algema

Se o termo "gaiola" for definido como sendo o lugar onde vive o pássaro, então a letra 'b' seria a correta. Por outro lado se "gaiola" for definido como lugar em que é aprisionado o pássaro, então a letra 'a' seria a correta.

Após tais observações, os itens que apresentavam mais de uma resposta correta ou que possuíam regras ambíguas foram alterados ou eliminados.

Como resultado dessa etapa de trabalho, 81 itens foram considerados satisfatórios aos objetivos do teste, compondo assim a primeira versão do teste.

Em seguida, um primeiro estudo piloto foi realizado com o conjunto de 81 itens.

2.1 - Estudo Piloto I

Os objetivos deste estudo foram especificamente 1) detectar palavras desconhecidas ou não utilizadas no cotidiano pelos alunos, visto que, se tal acontecesse, o teste estaria a mensurar o vocabulário ou erudição e não o mecanismo de raciocínio verbal; 2) decidir sobre a população a qual se destinaria o teste; 3) verificar a frequência com que as alternativas de resposta aos itens eram marcadas e 4) testar a clareza das instruções.

Para se atingirem os objetivos acima, estudou-se uma amostra oriunda da cidade de João Pessoa-PB, composta por 95 alunos do I e II graus. A tabela 03 fornece informações adicionais desta amostra.

A construção de um Teste de Raciocínio Verbal (RV)

Tabela 03 - Número de alunos por série escolar

Graus	Séries	N
I	7 ^a	20
	8 ^a	21
II	1 ^a	19
	2 ^a	17
	3 ^a	18
Total		95

Os resultados, conseguidos a partir da aplicação do conjunto de 81 itens, revelaram a necessidade da eliminação ou reformulação de itens com os seguintes problemas:

- analogias em que as palavras componentes eram rebuscadas ou pouco usadas pelos alunos. Ex: Iogurte, Nau, Pônei, Carruagem etc.

- analogias cujas alternativas obtiveram frequência de resposta inferior a 10%. Visto que, cada item possuía cinco alternativas e que mesmo sendo respondidas ao acaso - no 'chute' - uma determinada alternativa poderia obter 20% de resposta, considerou-se que 10% era uma frequência muito baixa. Provavelmente, tal alternativa possuía uma distância semântica muito grande em relação à resposta correta, acarretando uma frequência baixa de escolha por parte dos respondentes. Estas alternativas foram então modificadas.

Decidiu-se também, que o teste seria direcionado a secundaristas da 1^a à 3^a série, visto que os alunos de I grau - respectivamente 7^a e 8^a série - mostraram pouco conhecimento com

grande parte das palavras utilizadas no teste, resultando em dificuldades para a compreensão da tarefa a ser executada.

Quanto às instruções, procedeu-se a pequenas alterações, mas de forma geral foram compreendidas claramente pelos alunos.

Após essa etapa de estudos, uma nova versão do teste foi organizada com 63 itens e submetida a um segundo estudo piloto.

2.2 - Estudo Piloto II

O objetivo desta etapa de trabalho foi um estudo inicial dos parâmetros psicométricos do grupo de 63 itens, tais como o índice de dificuldade e o poder de discriminação dos mesmos.

Segundo Lord (1980), a técnica mais moderna e pertinente para o estudo dos parâmetros dos itens - tais como, dificuldade, discriminabilidade e controle de 'chute' - é a ICC ("item characteristic curve") dentro dos modelos da IRT ("item response theory").

Há vários modelos de IRT. Entre os mais conhecidos e usados estão o modelo logístico de um parâmetro (Rasch, 1966), que avalia somente a dificuldade dos itens, e o modelo logístico de três parâmetros (Lord, 1980) para avaliar a dificuldade, a discriminabilidade e a resposta ao acaso.

Tais tipos de análises exigem modernos 'softwares', que são de difícil acesso no Brasil. Dessa maneira, as análises

A construção de um Teste de Raciocínio Verbal (RV)

feitas no presente trabalho resultam de técnicas tradicionais de análise de itens, que se mostram precárias para esta tarefa.

Para se atingirem os objetivos desta fase de trabalho, foi utilizada uma amostra de 250 estudantes secundaristas de escolas privadas da cidade de João Pessoa. Os estudantes de escolas públicas não foram utilizados em tal estudo, visto que tais instituições encontravam-se em greve no período da coleta de dados.

A amostra de estudantes secundaristas se distribuiu mais ou menos equitativamente entre as três séries, havendo a predominância do sexo feminino (59,6%) e de idades na faixa de 14 a 17 anos (78,8%).

A tabela 04 fornece mais informações a respeito da amostra utilizada.

Tabela 04 - Características da amostra de estudantes (n=250)

Variável	Nível	%*	Média	DP
Sexo	1. Masc	40,4	-	-
	2. Fem	59,6	-	-
Idade	14 - 17 anos	78,8	15,48	1,74
	18 - 21 anos	21,2	-	-
Série	1. 1ª série	34,0	-	-
	2. 2ª série	28,8	-	-
	3. 3ª série	37,2	-	-

* As percentagens foram feitas dentro de cada variável.

De posse dos dados, a primeira análise feita foi com relação à distribuição dos itens segundo o índice de dificuldade. Tal análise revelou que, do total de 63 itens, sete eram muito

fáceis (entre 80 e 100% de acerto), 23 eram fáceis (entre 60 e 80% de acerto), 21 localizados na faixa intermédia de dificuldade (entre 40 e 60% de acertos), 10 eram difíceis (entre 20 e 40% de acertos) e apenas dois eram muito difíceis (entre 0 e 20% de acerto). A tabela 05 mostra a distribuição dos itens em cinco faixas de dificuldade (percentual de acerto).

Tabela 05 - Distribuição dos itens segundo o índice de dificuldade

% de acerto	ITENS.	N
Entre 0 e 20%	58 e 59.	02
Entre 20 e 40%	34, 35, 36, 38, 41, 51, 52, 55, 57 e 62.	10
Entre 40 e 60%	04, 09, 16, 18, 24, 27, 30, 32, 37, 39, 40, 43, 44, 46, 49, 50, 53, 56, 60, 61 e 63.	21
Entre 60 e 80%	06, 10, 11, 12, 14, 15, 17, 19, 20, 21, 22, 23, 25, 26, 28, 29, 31, 33, 42, 45, 47, 48 e 54.	23
Entre 80 e 100%	01, 02, 03, 05, 07, 08 e 13.	07
Total		63

Para Baquero (1974), uma distribuição ideal dos itens no contínuo de dificuldades deveria ser a seguinte:

- 10% de itens muito difíceis;
- 20% de itens difíceis;
- 40% de itens medianos;

A construção de um Teste de Raciocínio Verbal (RV)

- 20% de itens fáceis;
- 10% de itens muito fáceis.

No presente caso, tem-se a seguinte distribuição:

- 3% de itens muito difíceis;
- 16% de itens difíceis;
- 33% de itens medianos;
- 37% de itens fáceis;
- 11% de itens muito fáceis.

A faixa mais comprometida é a de itens "muito difíceis", com apenas dois itens (3% do total) representando-a. Na faixa de itens fáceis, alguns itens poderiam ser eliminados, visto existir uma concentração grande dos mesmos (37% do total). Nas demais faixas a distribuição é aceitável.

A segunda análise do grupo de itens foi com relação ao poder discriminativo dos mesmos. Para isso foi adotada a técnica de dividir a amostra em grupos-critério, segundo os escores obtidos no teste. O grupo inferior, formado pelos 30% dos escores mais baixos, e o grupo superior, constituído pelos 30% de escores mais altos. O valor de 30% para a formação dos grupos critério, é considerado ideal para este tipo de estudo (Anastasi, 1988).

A partir de tal divisão, submeteu-se o conjunto de itens individuais do teste à análise do teste 't' de Student. A lógica subjacente a tal tipo de análise é que, se existir diferença significativa entre a média dos escores dos dois grupos em um item específico, este item é considerado discriminativo,

acrescentando-se que a média dos escores em cada item deve ser sempre maior no grupo superior (Anastasi, 1988).

Dos 63 itens, 51 mostraram discriminar ou diferenciar os grupos-critério, adotando-se um $p \leq .01$; apenas no item 58 a média do grupo inferior foi maior que a do grupo superior. A tabela 06 mostra o resultado de tais análises.

Tabela 06 - Poder de discriminação dos itens

Itens	Gpo Inf (n=77)		Gpo Sup (n=81)		t	p
	Média	DP	Média	DP		
01	.74	.44	.91	.28	2.92	0.00
02	.76	.43	.99	.11	4.42	0.00
03	.91	.29	.98	.16	1.78	0.08*
04	.44	.50	.80	.40	4.99	0.00
05	.79	.41	.95	.22	3.02	0.00
06	.65	.48	.89	.32	3.68	0.00
07	.84	.37	.95	.22	2.21	0.03*
08	.83	.38	.94	.24	2.11	0.04*
09	.38	.49	.83	.38	6.45	0.00
10	.47	.50	.94	.24	7.44	0.00
11	.55	.50	.98	.16	7.20	0.00
12	.44	.50	.93	.26	7.56	0.00
13	.65	.48	.95	.22	5.03	0.00
14	.61	.49	.93	.26	5.00	0.00
15	.61	.49	.90	.30	4.47	0.00
16	.42	.50	.73	.45	4.15	0.00
17	.55	.50	.86	.35	4.63	0.00
18	.23	.43	.93	.26	12.21	0.00
19	.38	.49	.93	.40	8.74	0.00
20	.73	.45	.80	.40	1.11	0.27*
21	.52	.50	.91	.28	6.03	0.00
22	.49	.50	.94	.24	7.02	0.00
23	.38	1.24	.41	.67	2.00	0.05*
24	.31	.47	.80	.40	7.08	0.00
25	.74	.44	.85	.36	1.74	0.08*
26	.55	.51	.90	.30	5.38	0.00
27	.68	.47	.74	.44	0.90	0.37*
28	.68	.47	.89	.32	3.33	0.00
29	.47	.50	.81	.39	4.83	0.00
30	.38	.49	.67	.47	3.79	0.00

A construção de um Teste de Raciocínio Verbal (RV)

Continuação da Tabela 06

Itens	Gpo Inf (n=77)		Gpo Sup (n=81)		t	p
	Média	DP	Média	DP		
31	.48	.50	.89	.32	6.07	0.00
32	.39	.49	.77	.43	5.13	0.00
33	.44	.50	.91	.28	7.26	0.00
34	.35	.48	.54	.50	2.47	0.01
35	.17	.38	.77	.43	9.33	0.00
36	.22	.42	.48	.50	3.55	0.00
37	.27	.45	.78	.42	7.31	0.00
38	.23	.43	.41	.49	2.37	0.01
39	.21	.41	.86	.35	10.89	0.00
40	.21	.41	.86	.35	10.89	0.00
41	.24	.43	.65	.48	5.84	0.00
42	.36	.48	.80	.40	6.19	0.00
43	.43	.50	.63	.49	2.57	0.01
44	.10	.31	.63	.49	8.17	0.00
45	.62	.49	.77	.43	1.95	0.05*
46	.27	.45	.60	.49	4.44	0.00
47	.55	.50	.79	.41	3.18	0.00
48	.62	.49	.94	.24	5.10	0.00
49	.22	.42	.65	.48	6.08	0.00
50	.13	.34	.62	.49	7.31	0.00
51	.17	.38	.70	.46	8.02	0.00
52	.17	.38	.48	.50	4.44	0.00
53	.42	.50	.59	.49	2.25	0.03*
54	.51	.50	.90	.30	5.95	0.00
55	.11	.32	.43	.50	4.74	0.00
56	.17	.38	.67	.47	7.32	0.00
57	.14	.35	.54	.50	5.83	0.00
58	.10	.31	.09	.30	-0.11	0.91*
59	.5	.48	.75	.43	1.25	0.21*
60	.37	.48	.75	.43	5.31	0.00
61	.43	.50	.71	.45	3.79	0.00
62	.21	.41	.44	.50	3.27	0.00
63	.39	.49	.57	.50	2.26	0.03*

Os itens com * foram eliminados da nova versão do teste.

Os itens 03, 07, 08, 20, 23, 25, 27, 45, 53, 58, 59 e 63 mostraram-se pouco discriminativos, podendo por este critério serem eliminados do teste. Porém, há que se levar em conta também, a localização dos mesmos no contínuo de dificuldade.

Dessa maneira, os itens 03, 07 e 08 que são itens 'muito fáceis' (todos encontram-se na faixa entre 80 e 100% de acerto) podem ser eliminados do teste. O mesmo ocorre com os itens 20, 23, 25, 45 e 53, que localizam-se na faixa entre 60 a 80% de acertos (itens fáceis), os itens 27 e 63, que estão na faixa intermédia de acertos (entre 40 e 60%). No caso dos itens 58 e 59 que são itens 'muito difíceis' (localizam-se na faixa entre 0 e 20% de acertos), apesar dos mesmos serem os únicos representantes de itens 'muito difíceis', também foram eliminados do teste².

Através desta tomada de decisão, a nova forma do teste ficou composta por 51 itens. Tais itens ficaram assim distribuídos no contínuo de dificuldade:

- 00% de itens muito difíceis;
- 20% de itens difíceis;
- 37% de itens medianos;
- 35% de itens fáceis;
- 08% de itens muito fáceis.

Apesar de não seguir a distribuição idealizada por Baquero (1974) no que tange ao contínuo de dificuldade, o conjunto dos 51 itens apresenta um aceitável poder de discriminação e de representação da amplitude do traço envolvido no teste.

² Em futuras pesquisas estes dois itens deverão ser novamente estudados.

A construção de um Teste de Raciocínio Verbal (RV)

Nesta nova versão os itens foram organizados, em série crescente de dificuldade, segundo sua posição no contínuo de dificuldade.

III - Estudo Final do Teste RV

1 - Metodologia

1.1 - Amostra

Como o teste é direcionado para estudantes do II grau, nas três séries que o compõem, a amostra selecionada advém de tal população.

Foram 730 estudantes do II grau, da rede particular e pública de ensino da cidade de João Pessoa-PB, com pequena preponderância de estudantes pertencentes à escola pública (58,5%), cursando a 2ª série (43,8%) e do sexo feminino (53,3%). A idade média foi de 17,7 anos, com amplitude de 12 a 41. Uma informação adicional sobre as escolas utilizadas no estudo, é que de um total de oito escolas, quatro eram da rede pública e quatro da rede privada de ensino.

A tabela 07 descreve mais detalhadamente as características da amostra estudada.

Tabela 07 - Descrição da amostra utilizada no estudo (n=730)

Variável	Nível	%	Média	DP
Tipo de Escola	1. Pública	58.5		
	2. Particular	41.5		
Série	1. 1ª série	26.2		
	2. 2ª série	43.8		
	3. 3ª série	30.0		
Sexo	1. Masculino	46.7		
	2. Feminino	53.3		
Idade	1. Entre 12 e 15 anos	23.2		
	2. Entre 16 e 19 anos	54.5	17.7	3.12
	3. Mais de 20 anos	22.3		

1.2 - Instrumento

Foi utilizada a versão do Teste RV composta por 51 itens em formato de analogias verbais, seguindo uma ordem crescente de dificuldade (Anexo II).

Trata-se de um teste de conteúdo verbal, de lápis-papel e de apuração objetiva. É composto por um caderno que contém as instruções e as questões, além de uma folha de respostas. Na folha de respostas há uma parte destinada à caracterização do respondente (dados sócio-demográficos).

1.3 - Procedimento

O próprio autor atuou como aplicador em conjunto com uma auxiliar de pesquisa.

A coleta de dados ocorreu entre os dias 01 e 19 de março de 1993, na cidade de João Pessoa-PB.

A direção das escolas determinava as séries/turmas a serem avaliadas em cada sessão, em função de sua disponibilidade de horários.

Na etapa de aplicação, era explicado aos respondentes o objetivo da pesquisa, após o qual distribuía-se a folha de respostas. Após tal distribuição, era solicitado aos sujeitos preencherem o espaço destinado à caracterização individual.

Em seguida, distribuía-se o caderno de questões e passava-se à leitura das instruções pelo aplicador. Caso não houvessem dúvidas quanto à tarefa a ser executada, iniciava-se o teste, caso contrário explicavam-se novamente as instruções.

Após essa fase, os respondente teriam tempo livre para responder a todas as questões. Era solicitado aos mesmos que não deixassem questões sem respostas.

2 - Análises Estatísticas

Com o objetivo de estudar a validade de construto, procurou-se verificar a estrutura fatorial e a consistência interna do teste RV.

Para estabelecer a validade preditiva, correlacionaram-se os escores obtidos no teste RV com as médias dos respondentes na disciplina de português obtidas no último bimestre letivo.

No estudo dos itens, verificou-se a distribuição dos mesmos no contínuo de dificuldade e o poder de discriminação em função de grupos critério.

Para a avaliação da precisão do teste RV foi utilizado o índice alfa de Cronbach.

3 - Resultados

3.1 - Validação

3.1.1 - Validade de Construto

A validade de construto de um instrumento refere-se ao mesmo estar realmente medindo aquilo (isto é, o traço latente) que se propõe (Kerlinger, 1986; Richardson, Peres, Correia, Peres & Wanderley, 1985).

É consenso entre psicometristas, tais como Thurstone (1947), Guilford (1954), Fruchter (1954), Cronbach (1960),

A construção de um Teste de Raciocínio Verbal (RV)

Nunnally (1967), Harman (1967) e Williams (1968), ser a técnica da análise fatorial - por excelência - uma ferramenta estatística imprescindível ao estudo da validade de construto - ou seja, verificar a adequação da representação comportamental do traço latente a ser mensurado. Desta maneira, utilizou-se a mesma nesta etapa de análises.

Procedeu-se inicialmente a uma análise exploratória dos componentes principais (PC) com rotação ortogonal (varimax) (SPSS PC+ Version 1.0, 1990). Como resultado, surgiram 11 fatores com "eigenvalue" igual ou maior que 1,0 que juntos explicavam 40,8% da variância total dos escores. Apareceu um grande fator, com "eigenvalue" de 8,51, que sozinho explicava 20,3% de tal variância, além de dois outros fatores menores, com "eigenvalues" de 1,8 e 1,4, explicando juntos 7,7% da variância total.

Numa segunda etapa, procedeu-se a uma análise fatorial propriamente dita, utilizando-se o método do "Principal Axis Factoring" (PAF). Nesta análise, partiu-se da suposição da existência de um fator único de raciocínio verbal, conforme propõe a teoria de Thurstone (1948), razão pela qual foi realizada essa análise confirmatória de um fator.

O fator resultante, denominado de raciocínio verbal, foi responsável pela explicação de 17,5% da variância total, com "eigenvalue" de 8,54.

Esta análise levou em conta os 51 itens do teste RV; contudo, 11 itens mostraram ter cargas fatoriais menores que .30, sendo eliminados do teste. Nova análise fatorial confirmatória de

um fator foi realizada, desta vez com o conjunto dos 40 itens resultantes da análise anterior. O fator resultante explicava, desta vez, 21,3% da variância total, cujo valor do "eigenvalue" foi de 8,54.

Tal fator - raciocínio verbal - ficou representado pelo conjunto de 40 itens. A tabela 08 fornece as cargas fatoriais dos itens, as correlações item-total e a comunalidade dos mesmos no fator.

A construção de um Teste de Raciocínio Verbal (RV)

Tabela 08 Cargas fatoriais dos itens no Fator de Raciocínio Verbal

F A T O R			
ITEM	Cargas fatoriais	h^2	$r_{(t)}$
27	.61	.37	.57
06	.58	.34	.55
12	.58	.34	.54
07	.56	.31	.53
38	.55	.30	.51
28	.54	.29	.50
08	.53	.28	.50
43	.52	.27	.49
19	.51	.26	.49
13	.50	.25	.47
31	.49	.24	.47
32	.49	.24	.47
25	.49	.24	.46
24	.47	.22	.44
15	.46	.21	.44
23	.46	.21	.44
17	.45	.20	.43
47	.44	.20	.42
04	.44	.20	.42
26	.44	.20	.41
39	.43	.19	.41
18	.42	.18	.41
49	.40	.16	.39
09	.40	.16	.38
05	.37	.15	.37
50	.38	.14	.36
14	.37	.14	.35
29	.37	.14	.35
03	.37	.14	.35
33	.36	.13	.34
30	.36	.13	.34
10	.35	.12	.33
21	.35	.12	.33
35	.34	.12	.33
02	.34	.12	.33
20	.33	.11	.32
41	.33	.11	.32
45	.32	.10	.31
40	.30	.09	.29
22	.30	.09	.29

* h^2 significa a comunalidade dos itens no fator.

* $r_{(t)}$ significa a correlação de cada item com o total de itens.

Dessa maneira, a forma final do teste RV ficou representada pelo conjunto de 40 itens

3.1.2 - Validade Preditiva

Segundo Gronlund (1974), a validade preditiva diz respeito à predição de um comportamento futuro de um indivíduo ou a estimativa de sua situação atual com respeito a alguma variável de particular importância, mas diferente do instrumento.

Linn (1982), afirma que, a justificativa maior para o uso de testes como preditores à critérios externos, depende do quão alta é a validade preditiva do teste.

Para este tipo de estudo, geralmente procura-se correlacionar um critério externo - a maioria dos estudos utilizam as notas escolares - com os escores conseguidos no teste.

Porém, há problemas quando se adotam, como critério externo, as notas escolares. O principal deles diz respeito às notas serem resultantes de avaliações muito subjetivas e, conseqüentemente, sujeitas a uma grande variabilidade de erros de avaliação - vieses de julgamento, erros na medida para avaliação individual da aprendizagem etc. Dessa forma, as notas escolares não são o parâmetro ideal para esse tipo de estudo, mesmo assim são utilizadas como tal. A literatura sobre este problema é, em geral, bastante controversa, mostrando em alguns estudos

A construção de um Teste de Raciocínio Verbal (RV)

correlações significativas entre notas escolares e escores no teste enquanto outros não conseguem tais resultados (Linn, 1982).

Platz, McClintock e Katz (1959), realizaram estudo sobre a validade preditiva do "Miller Analogies Test (MAT)" sobre o 'sucesso' escolar de estudantes de pós-graduação em Psicologia, na Universidade de Michigan.

Concluíram que o MAT não é um bom instrumento para a seleção de estudantes para a pós-graduação em Psicologia, porém pode ser utilizado como o melhor preditor para avaliar o desenvolvimento científico de tais alunos.

Rawls, Rawls e Harrison (1969) encontraram resultados semelhantes sobre a validade preditiva do MAT com relação ao desempenho acadêmico de estudantes de psicologia, tanto na graduação como na pós. Enfatizam a importância de variáveis que podem ser adicionadas como fontes de informação, além do critério externo de desempenho acadêmico, como a experiência do indivíduo na área, idade e estado civil.

Hyman (1957), correlacionou os resultados de 38 alunos da pós-graduação em psicologia, da Universidade de Pittsburgh, em dois testes: o MAT e um Teste para Compreensão Verbal. O resultado foi um coeficiente de correlação (de Pearson) de .58. Em outras palavras, que o MAT também pode ser utilizado como preditor para a compreensão verbal.

Outros estudos, sobre outros testes, não têm conseguido tais resultados.

Por exemplo, no estudo realizado por Savastano e Moraes (1963), objetivou-se estudar a validade preditiva do Teste das Matrizes Progressivas de Raven sobre o aproveitamento de alunos dos cursos de pós-graduação em saúde pública. A amostra foi constituída por 186 alunos dos cursos de odontologia (n=52), engenharia da saúde (n=42) e medicina (n=92). Os coeficientes de correlações (r de Pearson) obtidos entre os escores do indivíduo no teste e o seu coeficiente de rendimento escolar, resultou em $r=.35$ (engenheiros), $r=.01$ (dentistas) e $r=.38$ (médicos).

Concluíram pela não existência de correlação entre os escores do Teste Raven e o coeficiente de rendimento escolar, porém acrescentam "que não se deve fazer essa correlação quando se utiliza para a avaliação do rendimento do aluno, as notas obtidas pelos indivíduos através do critério tradicional de notas finais. Sugerimos que um outro critério seja tomado para se correlacionar o aproveitamento do aluno através de suas realizações no decorrer do curso" (pág. 117).

A diferença entre os estudos que mostram ter um determinado teste validade preditiva e outros que não conseguem demonstrar empiricamente tal resultado, pode estar em um ou mais dos seguintes aspectos:

- o teste em questão pode, realmente, não ter validade preditiva sobre o critério externo adotado;
- o critério externo adotado pode não ser o ideal para ser predito pelo teste;

A construção de um Teste de Raciocínio Verbal (RV)

- o critério externo foi acertadamente adotado porém possui inúmeras variáveis incidindo sobre o mesmo (por exemplo diferentes avaliadores).

Como foi dito anteriormente, o critério externo de notas finais não é um parâmetro fidedigno, visto ser o coeficiente de rendimento escolar a média aritmética das notas de todas as disciplinas. Como cada disciplina possui um avaliador diferente, há inúmeras fontes de erro agindo sobre tais avaliadores.

Para o presente trabalho, como não se dispunha de outros critérios externos mais confiáveis que as notas escolares, resolveu-se assim mesmo, adotar tal critério e, assim, poder dispor de algum indicativo da validade preditiva do RV.

Para verificar a validade preditiva, procedeu-se a correlação (r de Pearson) entre os escores totais e as últimas médias obtidas na disciplina português. Esta correlação foi de .14 numa amostra de 404 estudantes (representando 55,34% da amostra total). O restante dos sujeitos ($n=326$) não forneceu as suas últimas médias bimestrais na disciplina de português.

Tal correlação evidentemente é fraca. Contudo, segundo Siegel (1956) e Levin (1978), se a amostra de uma determinada população tiver sido extraída de forma aleatória, pode-se testar a hipótese de que a associação obtida entre escore total no teste RV e médias em português não se deve meramente a um erro de amostragem. De fato, calculando-se a significância desta correlação, obteve-se o seguinte resultado: $t(402)=2,83$ que é

significativo a $p < .01$, indicando que a relação entre escores no teste RV e médias em português existe na população, ou seja, que os escores no teste RV apresentam-se como algo preditores para o desempenho acadêmico de estudantes do II grau na disciplina de português.

3.2 - Precisão

A precisão, propriamente dita, refere-se à mensuração do traço latente em toda a sua extensão cometendo a menor quantidade de erro possível (Souza, 1988). Já a consistência interna ou homogeneidade, refere-se ao aspecto do conjunto de itens realmente mensurarem um mesmo traço latente ou aspecto (Erthal, 1987).

O estudo da consistência interna dos 40 itens, foi feito através do alfa de Cronbach, que resultou num coeficiente de .90, altamente significativo e relevante.

4 - Análise dos Itens do RV

4.1 - Distribuição dos itens no contínuo de dificuldade

Procedeu-se a um estudo do percentual de acertos que cada item obteve, levando-se em consideração a amostra dos 730 estudantes.

Tem-se que, do total final de 40 itens, apenas o item 02 é muito fácil (entre 80 e 100% de acertos), 12 são fáceis (entre 60 e 80% de acertos), 12 de dificuldade média (entre 40 e 60% de acertos), 12 difíceis (entre 20 e 40% de acertos) e apenas o item 49 como um item muito difícil (entre 0 e 20% de acerto).

A tabela 09 descreve a distribuição dos itens em cinco faixas de dificuldades.

Tabela 09 - Distribuição dos itens segundo o índice de dificuldade

Faixas de Acerto	Itens	N	%
00 a 20%	49.	1	2,5%
20 a 40%	27, 28, 30, 31, 33, 35, 38, 39, 40, 43, 45, 47, 50.	13	32,5%
40 a 60%	8, 12, 13, 14, 17, 18, 19, 20, 24, 25, 26, 29, 32, 41.	14	35%
60 a 80%	3, 4, 5, 6, 7, 9, 10, 15, 21, 22, 23.	11	27,5%
80 a 100%	02.	1	2,5%
Total		40	100,0%

O conjunto de itens (n=40) foi então numerado novamente, respeitando-se a distribuição dos mesmos quanto ao índice de dificuldade. Por exemplo, o item 02 é o mais fácil, desta forma recebeu a numeração 01, enquanto o item 49 (mais difícil) recebeu a numeração 40. Tal operação foi realizada para todos os itens, resultando na forma final do RV.

Comparando distribuição dos 40 itens com a idealizada por Baque, J (1974), deveria-se aumentar o número de itens muito fáceis e muito difíceis, até que representassem juntos, 20% do total. As faixas de itens fáceis e difíceis, deveriam reduzir seu número, para que juntas possuíssem no máximo 40% do total. E a faixa de itens intermédios deveria ser representada por no máximo 40% do conjunto total de itens.

Alguns comentários tornam-se necessários a respeito dos itens.

Hipotetizou-se que uma variável a ser considerada em tal análise seria a familiaridade dos conceitos envolvidos na analogia. Os itens 02, 03, 04, 05 e 06 são compostos por palavras bastante usadas no dia-a-dia dos estudantes - conceitos concretos. Por exemplo, palavras tais como abelha, girafa, hidrogênio, ouro, calça, sapato, soldado etc. Já para os cinco itens mais difíceis, os conceitos empregados são mais abstratos. Por exemplo municipal, federal, exceder, irado etc. Tal resultado se aproxima do encontrado por Gick e Holyoak (1980).

Outra variável a ser considerada, refere-se ao tipo de relação entre as palavras. Os cinco itens mais fáceis do teste expressam relações de conjunto (parte-todo e espécie), intensidade (mudança de estado) e funcionalidade (função ou utilidade). Para os cinco itens mais difíceis as relações usadas são de significação (sinônimos), intensidade/qualidade (mudança de estado), funcionalidade (função) e conjunto (espécie), corroborando os resultados do estudo de Sternberg e Nigro (1980).

Uma terceira variável a considerar é o tipo de analogia. Esta variável não forneceu pistas relevantes, no presente estudo, sobre sua importância como possível fonte de dificuldade para um item, visto que o item mais fácil é do tipo A:C :: B:D e o item mais difícil também é do mesmo tipo.

Uma última variável estudada foi com respeito a proximidade semântica das alternativas e a resposta correta. Deve ser mais cuidadosamente estudada, visto que as alternativas do item mais difícil possuíam certa proximidade semântica da resposta correta, igualmente como os dois itens mais fáceis.

4.2 - Poder de discriminação dos itens

Com o conjunto dos 40 itens finais, estudou-se o poder de discriminação dos mesmos.

Para isso, procedeu-se a um teste 't' para diferença entre as médias dos itens, de dois grupos-critério. O grupo inferior ficou representado pelos 30% de estudantes com escores

mais baixos - escores menores que 13. O grupo superior ficou formado pelos 30% de estudantes com escores mais altos - escores maiores que 24.

Os resultados encontrados mostraram que todos os 40 itens discriminam o grupo superior-critério. A tabela 10 detalha os resultados.

A construção de um Teste de Raciocínio Verbal (RV)

Tabela 10 - Poder de discriminação dos itens

Item	Gpo Inf (n=187)		Gpo Sup (n=213)		t	p
	Média	DP	Média	DP		
02	.61	.49	.99	.12	10.11	.000
03	.48	.50	.94	.24	11.36	.000
04	.34	.47	.90	.30	14.12	.000
05	.48	.50	.93	.25	11.15	.000
06	.21	.41	.95	.25	20.76	.000
07	.23	.42	.94	.24	20.43	.000
08	.18	.39	.86	.34	18.40	.000
09	.33	.47	.86	.34	12.89	.000
10	.42	.50	.88	.32	10.76	.000
12	.09	.29	.83	.38	22.32	.000
13	.16	.37	.81	.40	16.77	.000
15	.33	.47	.91	.29	14.49	.000
16	.54	.50	.80	.40	5.50	.000
17	.22	.41	.82	.38	15.15	.000
18	.21	.41	.81	.39	15.24	.000
19	.18	.39	.82	.38	16.78	.000
20	.31	.46	.77	.42	10.30	.000
21	.40	.49	.86	.34	10.97	.000
22	.43	.50	.85	.36	9.60	.000
23	.19	.39	.80	.40	15.38	.000
24	.23	.42	.84	.37	15.44	.000
25	.16	.37	.76	.42	15.33	.000
26	.21	.41	.76	.43	13.24	.000
27	.08	.28	.81	.39	21.65	.000
28	.05	.23	.65	.48	16.23	.000
29	.27	.44	.77	.42	11.57	.000
30	.18	.39	.65	.48	10.96	.000
31	.06	.25	.61	.49	14.58	.000
32	.12	.33	.73	.44	15.96	.000
33	.20	.40	.60	.49	9.03	.000
35	.14	.34	.62	.49	11.63	.000
38	.09	.28	.68	.47	15.76	.000
39	.11	.32	.59	.49	11.59	.000
40	.24	.43	.64	.48	8.74	.000
41	.19	.39	.63	.48	9.92	.000
43	.05	.22	.59	.59	14.71	.000
45	.12	.33	.51	.50	9.16	.000
47	.05	.24	.50	.50	11.82	.000
49	.05	.23	.33	.47	7.61	.000
50	.09	.28	.52	.50	10.99	.000

5 - Construção de Formas Paralelas do Teste RV

Decidiu-se, a partir do conjunto de 40 itens, construir duas formas paralelas do teste RV. Alguns argumentos de ordem teórica e rática justificam tal decisão.

Como todos os 40 itens que compõem a forma geral possuem cargas fatoriais superiores a .30 (com média de .44), pode-se concluir que o atributo - raciocínio verbal - está muito bem representado pelos itens. Dessa forma, mesmo com a divisão do teste em duas formas paralelas, ambas continuariam a representar o raciocínio verbal de forma adequada.

Os argumentos de ordem prática são relativos a uma maior economia de tempo e dinheiro quando se utilizam formas menores, porém fidedignas, do mesmo teste. Além disso, as formas alternativas, proporcionam o controle sobre a aprendizagem do sujeito com respeito ao conteúdo dos itens. Com duas formas alternativas, apesar da similaridade do conteúdo de ambas, os itens são diferentes.

Assim, a forma geral do teste RV originou duas formas paralelas. Para tanto, dividiu-se a forma geral de 40 itens em duas partes. Para essa divisão, tomou-se a distribuição do itens no contínuo de dificuldade e sua respectiva carga fatorial. Dessa maneira, procurou-se equilibrar os itens componentes das duas versões, quanto a estes dois aspectos. A tabela 11 mostra tais informações

Tabela 11 - Distribuição dos itens nas Formas Paralelas

Item	% de acerto	Carga Fatorial	Itens	
			Forma 'A'	Forma 'B'
02	82	.34	02	02
03	73	.37	03	05
05	72	.39	04	15
04	69	.44	10	22
22	69	.30	15	07
10	68	.35	21	09
15	66	.46	06	06
07	65	.56		
21	64	.35		
06	63	.58		
09	63	.40		
14	59	.37	14	18
18	59	.42	17	20
17	57	.45	08	23
20	56	.33	29	24
08	51	.53	13	12
23	50	.46	26	40
29	49	.37	25	19
24	49	.47	27	32
13	48	.50	41	30
12	46	.58		
26	46	.44		
40	46	.30		
25	44	.49		
19	43	.51		
32	43	.49		
27	41	.61		
41	41	.33		
30	40	.36		
35	34	.34	35	33
38	32	.54	38	28
33	31	.36	39	45
45	31	.32	31	43
28	30	.54	50	47
39	29	.43		
31	28	.49		
43	28	.52		
50	28	.38		
47	25	.44		
49	13	.40	49	49
Total			22	22

Levando em conta essa distribuição, os itens foram, novamente, enumerados para cada forma. Por exemplo, o item 02 recebeu nova numeração e ficou como o número 01 para ambas as formas. O item 03 ficou com a numeração 02 para a forma 'A', enquanto o item 05 ficou com o número 02 para a forma 'B'. O restante dos itens também sofreu idêntica enumeração. Como resultado, as formas 'A' e 'B', ficaram compostas por 22 itens cada uma delas.

As formas paralelas foram estudadas quanto aos aspectos de dificuldade média dos itens e de consistência interna.

O cálculo da dificuldade média, para as formas 'A' e 'B', foi feito através da média de acertos dos itens dentro de cada forma. Dessa maneira, a dificuldade média dos itens da forma 'A' é de 47% e na forma 'B' é de 45,12%. Tal resultado mostra o quanto equilibradas são as formas, no tocante à dificuldade média dos itens.

A consistência interna das formas, calculado através do alfa de Cronbach, resultou em coeficientes de .84 para ambas. Pode-se afirmar que as duas formas, mesmo sendo compostas por apenas 22 itens, têm alta consistência interna.

6 - Elaboração das Normas Percentílicas

A elaboração de normas para a interpretação dos resultados nos testes tem merecido destaque na maioria dos trabalhos de construção e adaptação de instrumentos de avaliação (Anastasi, 1988).

As normas elaboradas permitem ultrapassar a pouca informação que os escores brutos obtidos no teste, geralmente expressos pelo número de respostas certas - principalmente em testes de aptidão -, possam a princípio fornecer (Freeman, 1962; Richardson e colaboradores, 1985; Anastasi, 1988).

A perspectiva clássica de avaliação psicológica valoriza os níveis de realização atingidos pelo conjunto da população ou por grupos populacionais específicos. Os resultados de cada indivíduo podem ser comparados com o restante dos membros do seu grupo de origem ou também com relação a sua realização em diferentes testes (Erthal, 1987).

Tais comparações consideram a proximidade ou o desvio existente entre o resultado obtido pelo sujeito e a realização de seu grupo de origem. Numa distribuição dos resultados segundo as leis da curva normal, a cada unidade de desvio-padrão corresponde uma determinada percentagem de sujeitos, constituindo essa percentagem uma base para a localização da posição do indivíduo comparada com os demais componentes do grupo (Roscoe, 1975; Levin, 1978; Kvanli, 1988).

No presente estudo foram elaboradas normas percentílicas, ou seja, o escore bruto obtido pelo indivíduo é

convertido num posto percentílico que indica a percentagem de sujeitos que obtiveram escores inferiores e superiores a tal (Magnusson, 1975).

Pelos resultados obtidos para as três formas (geral, 'A' e 'B') os homens obtiveram médias mais elevadas que as mulheres.

Tal resultado vai de encontro aos estudos de Waber (1976 e 1977) e Sanders e Soares (1986). Estes autores encontraram diferenças significativas entre homens e mulheres, quanto ao desempenho em provas de aptidão verbal, porém nestes estudos o desempenho das mulheres foi superior ao dos homens.

Segundo os estudos citados, a maior precocidade das mulheres na fase de transformação puberal proporciona um maior desenvolvimento das aptidões verbais.

No presente caso, a metodologia utilizada não permite de forma efetiva uma resposta clara e decisiva à questão das variáveis explicativas de tais diferenças, visto que as análises realizadas estão centradas na mera comparação dos resultados entre os sexos.

Quanto a variável idade, os resultados aproximam-se aos encontrados por Almeida (1988a). Na elaboração das normas da BPRD Almeida verificou uma certa progressão, quanto ao desempenho em provas de raciocínio, até as idades de 17-18 anos. A partir de tal idade ocorre uma certa estabilidade e mesmo decréscimo no desempenho.

A construção de um Teste de Raciocínio Verbal (RV)

Por outro lado, pode-se associar o decréscimo dos escores nas faixas etárias mais altas ao menor empenho por parte dos alunos mais adiantados na realização de tais tarefas. Ou ainda pelo fato do conteúdo da prova não se aproximar das opções curriculares já efetuadas pelo indivíduo (Almeida, 1988b).

Pelos resultados do presente estudo, há a tendência de decréscimo das médias quanto maior é a faixa etária, para ambos os sexos.

Enfatiza-se aqui, outra vez, a limitação metodológica para a explicação de variáveis relevantes para a compreensão de tal decréscimo. Mesmo assim, algumas hipóteses podem ser estudadas. Por exemplo, nas séries mais avançadas, ou nas faixas etárias maiores, o número de alunos com reprovações tende a ser mais elevado, podendo ser associado a isso níveis de realização intelectual mais baixo.

Apesar das limitações, as normas percentílicas foram construídas para três faixas etárias para ambos os sexos. As três faixas etárias foram construídas por representarem cada uma, aproximadamente, 30% dos sujeitos da amostra total.

As tabelas 12, 13 e 14 são relativas as normas percentílicas para homens e mulheres segundo a distribuição em três faixas etárias, respectivamente, para a forma geral, 'A' e 'B'.

Tabela 12 - Normas para a Forma Geral segundo o Sexo e a Idade

Percentil	Homens (n=341)			Escores para Mulheres (n=389)		
	12-15 anos	16-19 anos	+ de 20 anos	12-15 anos	16-19 anos	+ de 20 anos
100	39-40	38-40	33-40	39-40	37-40	36-40
95	38	35-37	29-32	34-37	32-36	25-35
90	37	34	27-31	32-33	28-31	22-24
85	36	30-33	24-26	30-31	25-27	20-21
80	35	28-29	22-23		23-24	18-19
75	34	27		28-29	21-22	16-17
70	33	25-26	21	27	19-20	15
65	31-32	23-24	20	25-26	18	
60	30	22	19	24	17	14
55	28-29	21	17-18	22		
50	27	20	16	21	16	13
45	25-26	19	14-15	20	14-15	12
40	24	18		19	13	11
35		17	13	17-18	12	10
30	23	15-16	12	15-16	11	09
25	22	14	11	14		08
20	18-21	13	10	13	10	
15	17	11-12	08-09	11-12	09	07
10	14-16	09-10		10	08	06
05	13	08	06-07	08-09	06-07	05
00	00-07	00-07	00-05	00-07	00-05	00-04
Média	26.50	20.25	16.30	21.02	16.22	12.35
DP	7.75	8.28	6.81	8.44	7.63	6.23

A construção de um Teste de Raciocínio Verbal (RV)

Tabela 13 - Normas para a Forma 'A' segundo o Sexo e a Idade

Percentil	Escores para Homens (n=341)			Escores para Mulheres (n=389)		
	12-15 anos	16-19 anos	+ de 20 anos	12-15 anos	16-19 anos	+ de 20 anos
100	22	19-22	18-22	19-22	18-22	18-22
95	19	17-18	16-17	18	17	12-17
90	18	16	15	16-17	14-16	11
85		15	14	15	13	10
80	17		13		12	09
75		14	12	14	11	08
70	16	13	11		10	
65	15			13	09	07
60		12	10	12		
55	14	11				
50	13	10	09	11	08	06
45			08	10	07	
40		09		09		
35	12		07		06	05
30		08	06	08		
25	11	07		07	05	04
20		06	05		04	
15	10			06		03
10	08-09	05	04	04-05	03	
05	06-07	04	03	03		02
00	00-05	00-03	00-04	00-02	00-02	00-01
Média	13.64	10.70	9.00	10.67	8.21	6.59
DP	3.97	4.40	3.86	4.34	4.02	3.20

Tabela 14 - Normas para a Forma 'B' segundo o Sexo e a Idade

Percentil	Escores para Homens (n=341)			Escores para Mulheres (n=389)		
	12-15 anos	16-19 anos	+ de 20 anos	12-15 anos	16-19 anos	+ de 20 anos
100	22	19-22	18-22	19-22	19-22	18-22
95	19	17-18	13-17	18	18	17
90	17-18	15-16	11-12	16-17	14-17	11-16
85		14		15	13	10
80		13	10	14	12	09
75	16		09	13	11	08
70		12			10	07
65	15	11	08	12	09	06
60	14	10		11		
55						
50	13	09	07	10	08	05
45	12	08		09	06-07	
40	11					
35		07	06	08	05	04
30	10		05	07		
25	09	06	04	06	04	
20	08	05		05		03
15	07	04	03	03-04	03	
10	06	03	02			02
05	05	02	01	01-02	01-02	01
00	00-04	00-01	00	00	00	00
Média	12.70	9.58	7.39	10.25	7.98	6.23
DP	4.08	4.40	3.40	4.60	4.10	3.64

A construção de um Teste de Raciocínio Verbal (RV)

7 - Cálculo do Tempo de Aplicação do Teste RV

Para este cálculo, foi cronometrado o tempo que cada sujeito levava para responder a forma de 51 itens. O tempo médio para responder ao teste foi de 35 minutos. Como este parâmetro estatístico é influenciado por valores extremos (Levin, 1978; Kvanli, 1988) - tanto altos quanto baixos - resolveu-se trabalhar com um parâmetro que representasse o tempo conseguido pela maioria da amostra para responder o teste RV. No presente caso, o valor relativo ao tempo que 80% da amostra levou para responder o teste RV (Forma Geral), foi de 28 minutos.

Dessa forma, adotou-se este valor como sendo o ideal para os sujeitos responderem o teste RV.

Adotou-se para as formas 'A' e 'B' a metade do tempo necessário para responder a forma geral, resultando em 14 minutos.

IV - Conclusões

Pode-se afirmar que o objetivo primordial do presente trabalho, o de construir um teste para a população estudantil brasileira, foi atingido.

Construiu-se um instrumento de medida, no caso para o raciocínio verbal, cujos parâmetros psicométricos foram

estabelecidos com uma amostra brasileira - apesar de ser bastante específica.

Tais parâmetros - validade e precisão - mostraram-se adequados, para os padrões científicos, podendo ser o teste utilizado de forma confiável para os objetivos que se propõe - de diagnosticar, prognosticar e, eventualmente, para a orientação vocacional.

As normas, provisórias, foram estabelecidas para a população a qual se destinou o teste - estudantes secundaristas.

As tarefas contidas no conteúdo do teste advêm das experiências e conhecimentos cotidianos dos alunos secundaristas. É portanto, um conteúdo atualizado.

Duas formas alternativas ao Rv, 'A' e 'B', foram organizadas com o mesmo nível de dificuldade. Com os objetivos de serem utilizadas quando, eventualmente, se dispuser de um tempo pequeno para a aplicação ou quiser se controlar a aprendizagem sobre a forma geral.

Apesar disso, há necessidade de muitos estudos sobre o teste resultante deste trabalho.

Por exemplo, é interessante estudos mais aprofundados sobre as variáveis que podem aumentar a dificuldade de um item em forma de analogia verbal.

Os parâmetros psicométricos devem ser analisados por técnicas mais modernas, tais como a "Item Response Theory - IRT", só conseguidas através de 'softs' importados.

A construção de um Teste de Raciocínio Verbal (RV)

A validação a nível nacional faz-se necessária, para isso uma amostra maior e representativa do país torna-se mister. Tal amostragem proporcionaria o estabelecimento e estudo mais aprofundado das normas.

A validade preditiva do RV também precisa de estudos mais aprofundados, adotando-se um critério externo mais preciso, e levando-se em conta outras variáveis tais como idade e estado civil.

Em síntese, o resultado do estudo é o surgimento de um Teste de Raciocínio Verbal (RV), com formas alternativas, construído para a população de estudantes secundaristas, com conteúdo recente e cujos parâmetros psicométricos mostraram-se bastante adequados, ou seja, pode-se usar o teste de forma confiável em tal população.

Para finalizar, quero comungar com o posicionamento de R. J. Sternberg (1992), através da citação de um trecho - de sua autoria - do artigo intitulado "A capacidade intelectual geral", sobre as teorias de cunho "psicométrico" e "cognitivo":

"É tentador considerarem-se as teorias sobre o processamento de informação como substitutas das teorias psicométricas. Na verdade, elas são substituições se nosso objetivo é compreender os processos de inteligência subjacentes aos testes e, até certo ponto, o desempenho no mundo real. Mas às vezes é errôneo descartar-se os resultados de todas as abordagens que casualmente não coincidem com a nossa própria e creio que isto é verdadeiro em relação à atual situação. Em grande parte,

Os mapas das teorias psicométricas e de processamento de informação podem ser sobrepostos. As teorias psicométricas enfocam, principalmente, a questão relativa de que constelações estruturais de processo, conteúdos, representações e similares dão surgimento a padrões estáveis de diferenças individuais. Portanto, esta última abordagem é útil para uma análise molecular de como as tarefas são solucionadas. A primeira abordagem, por sua vez, é útil para descobrir-se como estes processos, conteúdos e representações agrupam-se para formar fontes relativamente estáveis de diferenças individuais. Thurstone propôs que as análises fatoriais das capacidades humanas fossem apenas um início, precisando ser seguidas por análises detalhadas do processo experimental de desempenho em tarefas intelectuais. A partir desta perspectiva, com a qual concordo, a pesquisa sobre processamento de informação representa uma evolução natural da pesquisa psicométrica. Nem tanto a supera, quanto elabora sobre ela. Em particular, as análises da teoria em processamento de informação proporcionam o tipo de análise detalhada das capacidades humanas que deveria ser consequência dos tipos de análise ampla 'generalista' oferecidas pelas teorias psicométricas" (pp. 33 e 34).

V - Referências Bibliográficas

- Alderton, D. L. & Larson, G. E. (1990). Dimensionality of Raven's advanced progressive matrices items. *Educational and Psychological Measurement*, 50, 4, 887-900.
- Alencar, E. M. L. S. (1986). *Psicologia e Educação do Superdotado*. São Paulo: EPU.
- Almeida, L. S. (1986a). *Bateria de Provas de Raciocínio Diferencial: Manual*. Porto: Faculdade de Psicologia e de Ciências da Educação, Serviço de Consulta Psicológica e Orientação Vocacional.
- Almeida, L. S. (1986b). Aferição de uma Bateria de Provas de Raciocínio Diferencial para utilização nas atividades de orientação vocacional dos alunos do ensino secundário. *Psicologia Teoria e Pesquisa*, 2, 4, 201-212.
- Almeida, L. S. (1988a). *O Raciocínio Diferencial dos Jovens*. Porto: Instituto Nacional de Investigação Científica.
- Almeida, L. S. (1988b). *Teorias da Inteligência*. Porto: Edições Jornal de Psicologia.
- Anastasi, A. (1988). *Psychological Testing*. New York: MacMillan.
- Andriola, W. B.; Dias, M. R. & Pasquali, L. (1993). *Escala Multifatorial para estudantes Pós-graduandos*. Pesquisa em fase de coleta e análise dos dados. Brasília: PAM.

- Ausubel, D. P. & Robinson, F. G. (1969). **School Learning. An Introduction to Educational Psychology.** Nova Iorque: Holt, Rinehart and Winston.
- Baquero, G. (1974) **Testes Psicométricos e Projetivos. Medidas Psicoeducacionais.** São Paulo: Loyola.
- Bersoff, D. N. (1981). **Testing and the Law.** *American Psychologist*, 36, 10, 1047-1056.
- Brody, E. B. & Brody, N. (1976). **Intelligence: Nature, Determinants, and Consequences.** New York: Academic Press.
- Bueno, F. S. (1986). **Dicionário Escolar da Língua Portuguesa.** Rio de Janeiro: Fundação de Assistência ao Estudante.
- Butcher, H. J. (1972). **A Inteligência Humana.** São Paulo: Perspectiva.
- Carpenter, P. A. & Just, M. A. (1975). **Sentence Comprehension: A Psycholinguist Processing Model of Verification.** *Psychological Review*, 82, 1, 45-73.
- Cepa (Sem data-a). **Escala de Inteligência Wechsler para Crianças (WISC) - Manual.** Rio de Janeiro: Centro Editor de Psicologia Aplicada.
- Cepa (Sem data-b). **Teste Côletivo de Inteligência para Adultos (WAIS) - Manual.** Rio de Janeiro: Centro Editor de Psicologia Aplicada.
- Chi, M. T. H. & Glaser, R. (1992). **A capacidade para a solução de problemas.** In In R. J. Sternberg (org.). **As capacidades intelectuais humanas. Uma abordagem em processamento de informações.** Porto Alegre: Artes Médicas.

A construção de um Teste de Raciocínio Verbal (RV)

- Cronbach, L. J. (1960). *Essentials of Psychological Testing*. New York: Harper & Row.
- Erthal, T. C. (1987). *Manual de Psicometria*. Rio de Janeiro: Jorge Zahar.
- Freeman, F. S. (1962). *Teoria e Prática dos Testes Psicológicos*. Lisboa: Fundação Calouste Gulbenkian.
- Fruchter, B. (1954). *Introduction to Factor Analysis*. New Jersey: D. Van Nostrand Company.
- Garcia, J. (1981). The Logic and Limits of Mental Aptitude Testing. *American Psychologist*, 36, 10, 1172-1180.
- Gick, M. L. & Holyoak, K. J. (1980). Analogical Problem Solving. *Cognitive Psychology*, 12, 306-355.
- Gordon, E. W. & Terrell, M. D. (1981). The Changed Social Context of Testing. *American Psychologist*, 36, 10, 1167-1171.
- Gronlund, N. E. (1974). *Elaboración de Tests de Aprovechamiento*. México: Editorial Trillas.
- Gruber, E. C. (1975). *Preparation for Miller Analogies Test*. New York: Simon and Schuster.
- Guilford, J. P. (1954). *Psychometric Methods*. New York: McGraw-Hill.
- Haney, W. (1981). Validity, Vaudeville, and Values. *American Psychologist*, 36, 10, 1021-1034.
- Hargadon, F. (1981). Tests and College Admissions. *American Psychologist*, 36, 10, 1112-1119.
- Harman, H. H. (1967). *Modern Factor Analysis*. Chicago: The University of Chicago Press.

- Hunt, E. (1978). Mechanics of Verbal Ability. *Psychological Review*, 85, 2, 109-130.
- Hunt, E. (1992). A capacidade verbal. In R. J. Sternberg (org.). *As capacidades intelectuais humanas. Uma abordagem em processamento de informações*. Porto Alegre: Artes Médicas.
- Kerlinger, F. (1986). *Fundamentos de Pesquisa em Ciências Sociais*. São Paulo: EPU.
- Keyser, D. J & Sweetland, R. C. (ed. 1984). *Test Critiques*. Missouri: Test Corporation of America, Vol. 1 a 6.
- Klausnitzer, J. E. (1985). *Escola de Inteligência. Testes de Inteligência: Como devemos nos preparar a fim de realizá-los com sucesso e sua importância para treinar o raciocínio*. Rio de Janeiro: Tecnoprint.
- Kroeff, P. (1988). Síntese de posicionamentos a serem feitos quanto ao uso de testes psicológicos em avaliação psicológica. *Anais da 18ª Reunião Anual de Psicologia de Ribeirão Preto*. Ribeirão Preto: SBPRP.
- Kvanli, A. H. (1988). *Statistics. A Computer Integrated Approach*. Saint Paul: Harcourt Publishing Company.
- Levin, J. (1986). *Estatística Aplicada a Ciências Humanas*. São Paulo: Harbra.
- Linn, R. (1982). Admissions testing on trial. *American Psychologist*, 37, 3, 279-291.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.

A construção de um Teste de Raciocínio Verbal (RV)

- Magnusson, D. (1975). *Teoria de los Tests*. México: Editorial Trillas.
- Marmèche, E. C., Mathieu, J. & Barais, A. W. (1985). *Raisonnement Analogique et Résolution de Problèmes*. *L'Année Psychologique*, 85, 49-72.
- Mulholland, T. M., Pellegrino, J. W. & Glaser, R. (1980). *Components of Geometric Analogy Solution*. *Cognitive Psychology*, 12, 252-284.
- Nick, E. (1988). *Vivências relativas ao trabalho em avaliação psicológica. Dificuldades, limites e perspectivas para o Brasil*. *Anais da 16ª Reunião Anual de Psicologia de Ribeirão Preto*. Ribeirão Preto: SBPRP.
- Nunnally, J. C. (1967). *Psychometric Theory*. New York: McGraw-Hill.
- Olmedo, E. L. (1981). *Testing Linguistic Minorities*. *American Psychologist*, 36, 10, 1078-1085.
- Oppenheimer, J. R. (1956). *Analogy in Science*. *American Psychologist*, 11, 127-135.
- Pasquali, L. & Gouveia, V. V. (1990). *Escala de Assertividade Rathus - RAS: Adaptação Brasileira*. *Psicologia Teoria e Pesquisa*, 6, 3, 233-249.
- Pasquali, L.; Wechsler, S. & Bensusan, E. (1991). *Matrizes Progressivas do Raven Infantil - Ravin. Avaliação dos Parâmetros Psicométricos (Manual Técnico)*. Brasília: Laboratório de Pesquisa em Avaliação e Medida - PAM. (Não publicado).

- Pasquali, L.; Azevedo, M. M. & Andriola, W. B. (1992). **Matrizes Progressivas de Raven. Avaliação dos Parâmetros Psicométricos (Manual Técnico)**. Brasília: Laboratório de Pesquisa em Avaliação e Medida - PAM. (Não publicado).
- Pasquali, L.; Gouveia, V. V.; Andriola, W. B.; Miranda, F. J. & Ramos, A. L. M. (1993). **Questionário de Saúde Geral de Goldberg (QSG): Adaptação Brasileira. Psicologia Teoria e Pesquisa**. (Trabalho aceito para publicação).
- Pasquali, L. (1993). **Adaptação para o Brasil do "Edwards Personal Preference Schedule" - EPPS**. Pesquisa em fase de coleta e análise dos dados. Brasília: PAM.
- Pellegrino, J. W. (1992). **A capacidade para o raciocínio indutivo**. In R. J. Sternberg (org.). **As capacidades intelectuais humanas. Uma abordagem em processamento de informações**. Porto Alegre: Artes Médicas.
- Perfetti, C. A. (1992). **A capacidade para a leitura**. In R. J. Sternberg (org.). **As capacidades intelectuais humanas. Uma abordagem em processamento de informações**. Porto Alegre: Artes Médicas.
- Pinelli Jr., B. (1990). **Adaptação do VMI para uso no Brasil**. Dissertação de Mestrado. Brasília: UnB.
- Pinelli Jr., B. & Pasquali, L. (1992). **Validação do teste de desenvolvimento da integração viso-motora (VMI), para uso no Brasil**. *Psicologia Teoria e Pesquisa*, 8, 2, 187-205.

- Ramos, A. L. M. (1990). *Atração Intersexual por sexo, idade e prática religiosa*. Dissertação de Mestrado. Brasília: UnB. (Não publicada).
- Rapaport, D.; Gill, M. M. & Schafer, R. (1968). *Diagnostic Psychological Testing*. New York: International Universities Press.
- Rasch, G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (eds.). *Readings in Mathematical Social Sciences*. Cambridge: MIT Press, 89-107.
- Reschly D. J. (1981). Psychological Testing in Educational Classification and Placement. *American Psychologist*, 36, 10, 1094-1102
- Richardson, R. J.; Peres, J. A. S.; Correia, L. M.; Peres, M. H. M. & Wanderley, J. C. V. (1985). *Pesquisa Social. Métodos e Técnicas*. São Paulo: Atlas.
- Roazzi, A.; Spinillo, A. G. & Almeida, L. S. (1991). Definição e Avaliação da Inteligência: Limites e Perspectivas. In L. S. Almeida (org.). *Cognição e Aprendizagem Escolar*. Porto: Associação dos Psicólogos Portugueses.
- Roscoe, J. T. (1975). *Fundamental Research Statistics for the Behavioral Sciences*. New York: Holt, Rinehart and Winston.
- Sanders, B. & Soares, M. P. (1986). Sexual maturation and spatial ability in college students. *Developmental Psychology*, 22, 199-203.
- Sanger, M. (1992, março 22). Empresas aposentam os testes psicotécnicos. *Folha de São Paulo, Empregos*, 5.

- Navastano, E. (1963). Aplicação do Teste das Matrizes Progressivas de Raven com notas de aproveitamento de alunos dos cursos de pós-graduação em saúde pública. *Revista de Psicologia Normal e Patológica*, IX, 1/2, 105-119.
- Sawrey, J. M. & Telford, C. W. (1978). *Medidas Educacionais*. Rio de Janeiro: Livros Técnicos e Científicos.
- Schmidt, F. L. & Hunter, J. E. (1981). Employment Testing. Old Theories and new Research Findings. *American Psychologist*, 36, 10, 1128-1137.
- Shimberg, B. (1981). Testing for Licensure and Certification. *American Psychologist*, 36, 10, 1138-1146.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill Book Company.
- Souza, J. (1988). *Teoria Estatística Clássica dos Testes Subjetivos*. Brasília: Thesaurus.
- SPSS/PC+ Version 4.0 - Base System User's Guide (1990). SPSS: Statistical Data Analysis. Chicago: Marija J. Norusis/SPSS Inc.
- Sternberg, R. J. (1977a). *Intelligence, Information Processing, and Analogical Reasoning: The Componential Analysis of Human Abilities*. New Jersey: Lawrence Erlbaum Associates.
- Sternberg, R. J. (1977b). Component Processes in Analogical Reasoning. *Psychological Review*, 84, 4, 353-378.
- Sternberg, R. J. & Nigro, G. (1980). Development Patterns in the Solution of Verbal Analogies. *Child Development*, 51, 27-38

A construção de um Teste de Raciocínio Verbal (RV)

- Sternberg, R. J. (1992). A capacidade intelectual geral. In R. J. Sternberg (org.). **As capacidades intelectuais humanas. Uma abordagem em processamento de informações.** Porto Alegre: Artes Médicas.
- Thurstone, L. L. (1947). **Multiple-Factor Analysis. A Development and Expansion of Vectors of Mind.** Chicago: The University of Chicago Press.
- Thurstone, L. L. (1948). Psychological Implications of Factor Analysis. *The American Psychologist*, 3, 402-408.
- Tyler, L. E. (1981). **Testes e Medidas.** Rio de Janeiro: Zahar Editores.
- Van Kolck, O. L. (1981a). **Técnicas de Exame Psicológico e suas aplicações no Brasil.** Petrópolis: Vozes, Vol. 1.
- Van Kolck, O. L. (1981b). **Técnicas de Exame Psicológico e suas aplicações no Brasil.** Petrópolis: Vozes, Vol. 2.
- Waber, D. P. (1976). Sex differences in cognition: a function of maturation rate? *Science*, 192, 572-573.
- Waber, D. P. (1977). Sex differences in mental abilities, hemispheric lateralization, and rate of physical growth at adolescence. *Developmental Psychology*, 13, 29-38.
- Williams, F. (1968). **Reasoning with Statistics. Simplified Examples in Communications Research.** New York: Holt, Rinehart & Winston.
- Wood, D. A. (1975). **Elaboración de Tests: Desarrollo e Interpretación de los Tests de Aprovechamiento.** México: Editorial Trillas.

**Anexo I - Testes criados no Brasil entre os anos 30 e
60.**

Nome: Teste ABC

Autor: M. B. Lourenço Filho

Ano de criação: 1931

Parâmetros Psicométricos: nada consta sobre a validade e precisão.

Nome: Psicodiagnóstico Miocinético (PMK)

Autor: E. Mira y Lopez

Ano de criação: 1940

Parâmetros Psicométricos: validade concorrente entre "resultados no teste e julgamento por técnicos de um Serviço de Orientação e Seleção Profissional (r de .37 a .79)"; sobre a precisão, estudos "permitem considerá-lo satisfatório, no conjunto das medidas" (Van Kolck, 1981b:379).

A construção de um Teste de Raciocínio Verbal (RV)

Nome: Teste Jota

Autor: O. A. L. Martins

Ano de criação: 1944

Parâmetros Psicométricos: segundo Van Kolck (1981a:262), "o autor, afirmou ser razoável a expectativa de substancial validade para o teste em estudo"; a precisão foi calculada pelo método das metades ($r = .98$).

Nome: Teste Minhas Mãos (MM)

Autor: H. Antipoff

Ano de criação: 1947

Parâmetros Psicométricos: nada consta sobre validade e precisão.

Nome: Teste de Classificação de Objetos do ISOP

Autor: E. Schneider

Ano de criação: 1950

Parâmetros Psicométricos: validade concorrente com Testes de Inteligência Verbal e Abstrata (r de .83 a .84) e "com os estudos de caso em que se verifica a fidelidade das indicações diagnósticas de personalidades desajustadas" (Van Kolck, 1981b:39); sobre a precisão nada consta.

Nome: Teste Maia

Autor: Jacyo Maia

Ano de criação: 1950

Parâmetros Psicométricos: validade concorrente com o Teste INV ($r = .48$) e com testes de Português e Matemática (r de $.40$ a $.80$); nada consta sobre a precisão.

Nome: Questionário de Personalidade Dadahie

Autores: E. M. Andrade, R. Moraes & Z. S. Wendel

Ano de criação: 1950

Parâmetros Psicométricos: nada consta sobre a validade, precisão foi calculada pelo método das metades (r de $.69$ a $.84$).

Nome: Teste de Coordenação Bimanual

Autor: Daniel Antipoff

Ano de criação: 1951

Parâmetros Psicométricos: nada consta sobre validade e precisão.

Nome: Teste DEP

Autor: O. de B. Santos e colaboradores.

Ano de criação: 1952

Parâmetros Psicométricos: validade concorrente com o rendimento escolar (r de $.42$ a $.49$), precisão calculada pelo método das metades (r de $.87$ a $.88$).

A construção de um Teste de Raciocínio Verbal (RV)

Nome: Teste de Inteligência Não-Verbal (INV)

Autor: Pierre G. Weil

Ano de criação: 1955

Parâmetros Psicométricos: validade concorrente com o Stanford-Binet ($r = .73$), com o Goodenough ($r = .52$), com o Raven ($r = .84$), com os Cubos de Kohs ($r = .72$), com o Meili ($r = .52$) e com o teste de Jacyr Maia ($r = .48$); a precisão foi calculada pelo método das metades ($r = .93$).

Nome: Teste Sinal

Autor: J. Carvalhes

Ano de criação: 1955

Parâmetros Psicométricos: validade concorrente com o Teste de Pontilhagem ($r = .66$) e com Teste de Batida ($r = .45$); nada consta sobre a precisão.

Nome: Prova Perceptiva

Autor: C. C. de Souza

Ano de criação: 1955

Parâmetros Psicométricos: nada consta sobre a validade e precisão.

Nome: Método Projetivo para Avaliação da Motivação (MPAM)

Autor: A. L. Angelini

Ano de criação: 1955

Parâmetros Psicométricos: a validade foi feita com base em dois estudos: 1) discriminação em nível estatisticamente significativo das realizações dos grupos masculino e feminino e 2) correlações elevadas entre resultados altos no MPAM e nas provas de aprendizagem, para os grupos de ambos os sexos." (Van Kolck, 1981b:310); a precisão foi calculada pelo método teste-reteste ($r = .96$).

Nome: Teste Sena1 AG-3

Autor: Oswaldo B. Santos e colaboradores.

Ano de criação: 1959

Parâmetros Psicométricos: validade preditiva com a eficiência nos cursos em que os indivíduos haviam sido selecionados (r de .31 a .63) e concorrente com o Teste Cubos de Kohs ($r = .81$), com o DEP (r de .74 a .82), com o Teste de Inteligência Técnica de Stenquist-Weinberg ($r = .44$), com provas de matemática e português (r de .41 a .74); precisão calculada pelo método das metades ($r = .97$).

A construção de um Teste de Raciocínio Verbal (RV)

Nome: Bécasse: Teste Coletivo de Maturidade Escolar

Autor: B. K. Schoenfeldt

Ano de criação: 1961

Parâmetros Psicométricos: sobre a validade, afirma Van Kolck (1981b:75) que "o manual traz elementos de validação clínica, ao apresentar análises de casos em que o teste foi utilizado isoladamente ou no contexto de um exame psicoeducacional"; sobre a precisão nada consta.

Nome: Bateria CEPA

Autor: CEPA

Ano de criação: 1963

Parâmetros Psicométricos: validade concorrente com o Teste de Inteligência Geral de Meili ($r = .42$), com o Teste de Sinônimos (r de $.92$ a $.96$); a precisão foi calculada pelo método das metades ($r = .81$).

Nome: Inventário de Interesses

Autor: A. I. Anselmi

Ano de criação: 1963

Parâmetros Psicométricos: nada consta sobre validade e precisão.

Nome: CIA

Autores: R. de Moraes, E. M. de Andrade & D. de Godoy Alves

Ano de criação: 1964

Parâmetros Psicométricos: validade concorrente (não consta dados), precisão calculada pelo método das metades e também pelo método K-F (r de .45 a .95)

Nome: Bateria Fatorial do SENAC

Autor: SOEP

Ano de criação: 1964

Parâmetros Psicométricos: nada consta sobre a validade; a precisão foi calculada pelo método das metades (r de .86 a .96).

Nome: Bateria Intermediária do SENAC

Autores: E. Nick, M. A. Pinheiro, M. L. E. Silva & C. N. Cohen

Ano de criação: 1964

Parâmetros Psicométricos: nada consta sobre a validade; a precisão foi calculada pelo método das metades (r de .75 a .89) e teste-reteste (r de .54 a .77).

A construção de um Teste de Raciocínio Verbal (RV)

Nome: G-36 Teste Não-Verbal de Inteligência

Autor: E. R. Bocalandro

Ano de criação: 1966

Parâmetros Psicométricos: validade concorrente com o Teste Dominós de Anstey ($r = .84$), precisão calculada pelo método das metades ($r = .90$).

Nome: Teste dos Relógios

Autor: F. de V. Souza & S. V. Cambraia

Ano de criação: 1967

Parâmetros Psicométricos: validade concorrente com o Raven ($r = .57$) e com o INV ($r = .77$); a precisão foi calculada pelo método das metades ($r = .81$).

Nome: Definição de Palavras

Autor: L. C. Bonilha

Ano de criação: 1968

Parâmetros Psicométricos: nada consta sobre a validade; a precisão calculada pelo método das metades ($r = .93$).

Nome: Estudo da Motivação Profissional do Adolescente (EMPA)

Autor: M. A. Pinheiro & C. N. Cohen

Ano de criação: 1968

Parâmetros Psicométricos: nada consta sobre validade e precisão.

Nome: Teste de Sondagem Intelectual

Autor: J. A. Cunha & N. S. da Rocha

Ano de criação: 1969

Parâmetros Psicométricos: validade concorrente com o Teste INV
($r = .87$ a $.91$); precisão calculada pelo método das metades (r
 $= .87$ a $.91$).