



**ISPA** | Instituto Superior de Psicologia Aplicada

**VALIDAÇÃO DOS MÉTODOS DE  
SELECÇÃO DA ACADEMIA DE TÉCNICOS  
DE TELECOMUNICAÇÕES 2008**

**INÊS SANTOS PEREIRA**

**Orientador de Dissertação:  
RUI BÁRTOLO RIBEIRO**

**Coordenador de Seminário de Dissertação:  
RUI BÁRTOLO RIBEIRO**

**Tese submetida como requisito parcial para a obtenção do grau de:  
MESTRE EM PSICOLOGIA  
Especialidade em Psicologia Social e das Organizações**

2009

Dissertação de Mestrado realizada sob a orientação do Mestre Rui Bártolo Ribeiro, apresentada ao Instituto Superior de Psicologia Aplicada para obtenção do grau de Mestre na especialidade de Psicologia Social e das Organizações conforme o despacho da DGES, nº 19673/2006 publicado em Diário da República 2ª série de 26 de Setembro, 2006.

## **Agradecimentos**

Começo por expressar o meu apreço à Empresa onde recolhi a informação da amostra, assim como a todos os estagiários que se disponibilizaram a participar na investigação, por terem possibilitado a concretização deste trabalho.

Em segundo lugar ao Mestre Rui Bártolo Ribeiro, meu orientador, pelo acompanhamento no meu percurso académico, pelas diversas aulas relevantes para a minha formação, pelas oportunidades e desafios, pela discussão de ideias e pela orientação deste trabalho.

Agradeço também a todos os familiares e amigos, em especial à minha mãe e ao Hugo, por todo o apoio que sempre me deram e por toda a paciência.

Não posso deixar de agradecer ao meu pai, por tudo.

A ti pai, dedico este trabalho.

*“Can we predict traditional outcome measures better because of clearer understanding of relationships between predictors and criteria?”*

Sackett e Lievens (2008, p. 430).

## Resumo

A utilização de provas de avaliação psicológica num processo de selecção, especialmente as que avaliam capacidades cognitivas, é unanimemente considerada, na literatura de investigação, como fundamental, dada a sua elevada validade preditiva no que concerne ao desempenho.

A constatação de que, em 2008, num processo de selecção de estagiários para funções técnicas, esse instrumento não foi utilizado, aliada ao facto dos indivíduos seleccionados permanecerem na Empresa, durante um ano e serem sujeitos a um processo de avaliação de desempenho, criou as condições para os mesmos serem objecto do estudo, a que se propõe esta tese: demonstrar qual a capacidade preditiva da aplicação de provas de avaliação psicológica a esta população.

Foi, assim, efectuado esse exercício, suportado nos princípios teóricos existentes sobre esta matéria, quer no que se refere ao método de abordagem da questão, quer aos preditores usados.

A componente empírica deste trabalho resultou de dados reais recolhidos para uma população de 130 indivíduos.

Toda a informação recolhida foi sujeita a tratamento estatístico e não foi encontrada qualquer correlação entre os resultados dos testes e a avaliação de desempenho, o que conduziu à necessidade de analisar os pressupostos com que trabalhámos e encontrar explicações para este resultado, aparentemente contraditório.

A fragilidade real existente na variável critério e as condições em que foram recolhidos os dados das variáveis predictoras, explicam o resultado obtido e reforçam a necessidade de melhorias no processo, não estando de qualquer forma posta em causa, a importância de manter testes de avaliação psicológica nos processos de selecção.

Palavras-Chave: Selecção de Pessoal, Testes de Selecção, Validação, Predição.

## **Abstract**

The use of psychological testing, during the process of recruitment and selection, namely those used to evaluate cognitive functions is considered, throughout the literature as essential, given its predictive validity concerning performance.

Having found that in 2008, this type of testing had been neglected during the process of selecting new interns for technical jobs, and that, these interns were still employees at this company and were being subject to a performance evaluation, they were found to be ideal candidates for the present study: to demonstrate the predictive capabilities of psychological testing on this population.

This was the purpose behind this study, and its approach was based on the theoretical principles found on the subject, principles that also modeled the choice of predictors.

The empirical part of this study is based in real data, gathered from a population of 130 subjects.

All the available data was subject to statistical treatment, and no correlation was found between testing results and performance evaluation results, which lead to the need to return to the theoretical assumptions behind this study, in an attempt to better understand and explain our results, at first glance, contrary to what was found in the literature.

The fact that the variable “criteria” is inherently a fragile one, accompanied by the non-optimal information gathering conditions concerning predictors, seem to explain the present results, whilst also calling attention to the need to improve this process, without, however, at any point, questioning the role psychological testing should play, during recruitment and selection processes.

Keywords: Personnel Selection, Selection Testing, Validity, Predictive Capabilities.

## Índice

INTRODUÇÃO .....	1
SELECÇÃO E VALIDAÇÃO – Conceitos, Investigação e Resultados .....	3
O Contexto Histórico da Selecção de Pessoal .....	3
O Conceito de Validade e o Processo de Validação .....	7
Perspectiva Histórica .....	7
Validade de Critério, de Conteúdo e de Constructo .....	10
A Validação como Argumento de Avaliação .....	13
A Validade e Utilidade dos Métodos de Selecção de Pessoal – Implicações Teóricas e Práticas .....	15
O Papel da Validação no Processo de Construção dos Testes .....	16
As Fontes de Constructo na Elaboração dos Testes .....	17
Uma Visão Geral dos Indicadores Utilizados em Selecção de Pessoal .....	18
Os Determinantes do Valor Prático (Utilidade) dos Métodos de Selecção .....	21
Recrutamento e Selecção da Academia de Técnicos de Telecomunicações 2008 ...	23
História das Empresas .....	24
O Actual Processo de Recrutamento e Selecção .....	24
MÉTODO .....	28
Caracterização da Amostra .....	28
Variáveis e sua Operacionalização .....	30
Instrumentos .....	31
Teste de Atenção – d2 .....	31
Prova Mudanças .....	33

Bateria de Provas de Raciocínio Diferencial (BPRD).....	33
Procedimento .....	35
RESULTADOS .....	37
Estatística Descritiva .....	37
Estatística Descritiva – Análise Teste a Teste.....	38
Qualidades Métricas .....	39
Análise de Correlações.....	43
Análise de Regressões .....	46
Análise Discriminante .....	47
DISCUSSÃO .....	50
REFERÊNCIAS BIBLIOGRÁFICAS .....	57
ANEXOS.....	60
Anexo A: Outliers das Variáveis Predictoras .....	61
Anexo B: Estatística Descritiva dos Preditores e das Variáveis de Critério por Área de Trabalho .....	64
Anexo C: Estatística Descritiva – Análise Teste a Teste .....	65
Anexo D: Qualidades Métricas – Validade .....	66
Anexo E: Qualidades Métricas – Sensibilidade .....	80
Anexo F: Qualidades Métricas – Fidelidade .....	83
Anexo G: Correlações .....	87
Anexo H: Regressões .....	90
Anexo I: Análise Discriminante .....	92

## Lista de Figuras

<b>FIGURA 1:</b> Distribuição da Amostra por Idade .....	28
<b>FIGURA 2:</b> Distribuição da Amostra por Zona Geográfica .....	29
<b>FIGURA 3:</b> Distribuição da Amostra por Habilitações Literárias .....	29
<b>FIGURA 4:</b> Gráfico de Componentes do Mudanças após Rotação .....	40

## Lista de Tabelas

<b>TABELA 1:</b> Estatística Descritiva dos Preditores .....	37
<b>TABELA 2:</b> Estatística Descritiva das Variáveis Critério .....	38
<b>TABELA 3:</b> Resultados dos Testes KMO e de Esfericidade de Bartlett .....	39
<b>TABELA 4:</b> One-Sample Kolmogorov-Smirnov para o Teste d2 .....	41
<b>TABELA 5:</b> One-Sample Kolmogorov-Smirnov e Coeficientes de Assimetria e de Achatamento para o Mudanças e para as Provas SR, AR e MR da BPRD ..	42
<b>TABELA 6:</b> Consistência Interna para o Mudanças e para as Provas SR, AR e MR da BPRD .....	42
<b>TABELA 7:</b> Correlação de Pearson entre os Preditores .....	43
<b>TABELA 8:</b> One-Sample Kolmogorov-Smirnov para as Variáveis Critério .....	44
<b>TABELA 9:</b> Correlação das Variáveis Critério .....	45
<b>TABELA 10:</b> Correlação entre Preditores vs Critério .....	45
<b>TABELA 11:</b> Análise da Independência dos Erros .....	46
<b>TABELA 12:</b> Teste M-Box .....	48
<b>TABELA 13:</b> Anova para os Testes de Seleção – Análise Discriminante .....	48
<b>TABELA 14:</b> Média, Desvio Padrão e N nos diferentes grupos da Prova MR .....	49
<b>TABELA 15:</b> Função Discriminante .....	49
<b>TABELA 16:</b> Significância da Função Discriminante .....	49

## INTRODUÇÃO

A evolução, ao longo dos tempos, da forma como as Organizações encaram o papel dos recursos humanos na criação de valor, é bem reveladora das mudanças de paradigma a que vimos assistindo na Sociedade em geral. De uma economia sustentada em recursos físicos e tecnológicos e em recursos humanos abundantes e indiferenciados, compreensível em períodos de mudanças menos aceleradas, em que os papéis estavam bem definidos e em que o foco do valor era material, fomos evoluindo para uma economia com tecnologia massificada, mudanças aceleradas que exigem rapidez na resposta, diferenciação, capacidade para encontrar soluções rápidas para novos problemas. O foco passou a ser o conhecimento. Assim sendo, a competição das empresas pelos talentos é um dos traços mais marcantes das Organizações actuais, ou seja, a base para a estratégia de Organizações competitivas radica na singularidade dos seus recursos e capacidades.

Nesta perspectiva, as actividades de recrutamento e selecção, são claramente uma fonte de vantagem competitiva, uma vez que permitem atrair e fazer entrar na organização os candidatos que mais se adequam às suas necessidades. Para que tal se verifique, uma das questões centrais é a de saber em que medida os métodos e técnicas utilizados são preditores do sucesso dos candidatos, não só nas funções para que são recrutados mas também para o futuro. Importa seleccionar candidatos que maximizem a relação custo/ benefício da escolha. “Na generalidade, são utilizados dois grandes critérios de avaliação da validade preditiva: o desempenho profissional e o desempenho na aprendizagem. Ou seja, pretende-se saber em que medida esses métodos e técnicas ajudam a compreender como os candidatos (a) virão a desempenhar e (b) aprenderão a desenvolver as suas capacidades e competências, tanto para as funções de curto como para as de longo prazo” (Gomes, Cunha, Rego, Cunha, Cabral-Cardoso & Marques, 2008). É assim fundamental optar pelos métodos e critérios de selecção mais válidos ou seja, com maior poder preditivo do desempenho.

A procura por entender o que conduz a determinados resultados e a forma como os mesmos se encontram relacionados, é o que permite entender o grau de validade de um teste (Anastasi, 1986). A validade cruzada dos resultados obtidos na aplicação do teste, com as variáveis reais de desempenho é o que permite concluir se, efectivamente, estamos a utilizar o instrumento correcto de predição.

A escolha do tema da presente tese – Validação dos métodos de selecção da Academia de Técnicos de telecomunicações 2008 – tem a sua raiz na constatação de que muitas Empresas utilizam métodos e técnicas de selecção que raramente sujeitam, de forma organizada e

sistemática, a validação, que utilizam baterias de testes desenquadradas da população a que se destinam e que pouco revelam sobre o que se procura no candidato ou que seleccionam e não mais acompanham os escores de produtividade do trabalhador.

A importância da validação dos métodos usados é assim, essencial, para que se maximizem os resultados do investimento em recrutamento e selecção nas Organizações e se contribua para o aumento das suas vantagens competitivas.

A presente tese está desenvolvida em cinco grandes pontos: suporte teórico ao estudo, com relevância para o tema da validação de critério uma vez que é este o foco do trabalho prático desenvolvido; método – descrição do estudo; resultados obtidos; discussão e conclusão.

Procurou-se assim, através da aplicação de determinados testes de selecção e do cruzamento dos resultados com o desempenho real da amostra, inferir até que ponto deverão os mesmos passar a ser utilizados em futuros métodos de selecção, com vista a maximizar o valor investido pela Empresa na sua selecção de pessoal.

## SELECÇÃO E VALIDAÇÃO – Conceitos, Investigação e Resultados

### O Contexto Histórico da Selecção de Pessoal

A aplicação de testes psicológicos em selecção de recursos humanos tem uma história longa e um tanto controversa. A selecção de pessoal em geral, e o uso concomitante de variadas formas de testes psicológicos em particular, tem a sua origem no final do século XIX. Muito do desenvolvimento do trabalho nos métodos científicos de selecção teve a sua origem nos esforços dos primeiros psicólogos industriais, no apoio aos militares durante as duas guerras mundiais. Desde os conceitos de selecção natural, que formam os pressupostos da gestão científica de Frederic Taylor, através das técnicas informais de análise de características precoces, até à recente aplicação de instrumentos de selecção com base em análises estatísticas de fiabilidade e de validade, o uso de testes e de outras técnicas para a melhoria da selecção de pessoal e do seu desempenho nunca deixou de ser polémico – ou se discutia sobre o papel adequado dos profissionais que utilizam os testes, sobre o equilíbrio adequado na procura de eficiência de gestão e de equidade para os empregados ou sobre a utilidade dos próprios testes.

O artigo “Psychological Testing in Personnel Selection, Part I: A Century of Psychological Testing “ de Scroggins, Thomas e Morris (2008) explora a evolução dos testes de selecção de pessoal e descreve o impacto do desenvolvimento do trabalho dos primeiros psicólogos industriais.

As raízes dos testes psicológicos surgem com a origem da psicologia industrial no final do século XIX e início do século XX. Desde o início dos anos de 1880, autores como Henry R. Towne e Henry Metcalf propuseram que a gestão empresarial, vista como arte nos finais do século XIX, fosse pensada como ciência, beneficiando da profissionalização dos engenheiros, uma vez que tinha alicerces na ciência e propensão para a mesma (Van De Water, 1997).

Embora os psicólogos, como praticantes da tradicional disciplina académica de psicologia, resistissem à aplicação de modelos e teorias psicológicas direccionados aos problemas de gestão, Walter Dill Scott e Hugo Munsterberg fundaram o campo da psicologia industrial, quando começaram a explorar a aplicação dos princípios psicológicos a problemas na educação, na lei, no marketing e na gestão (Hearnshaw, 1987, cit. por Scroggins et al., 2008). Os anos seguintes viram um rápido crescimento na aplicação da psicologia industrial na área da psicologia do mercado, por profissionais que queriam resolver problemas complexos empresariais. Entre as ferramentas usadas contam-se testes psicológicos destinados a resolver o problema crescente da escolha dos melhores indivíduos, em termos de produtividade.

O desenvolvimento histórico da gestão, da ciência e da psicologia que levou à aceitação geral e à aplicação de testes psicológicos é pouco explorado. Um dos mais influentes pioneiros dos finais do século XIX e início do século XX foi Frederic W. Taylor. A influência de Taylor começou com a publicação, em 1895, "A Piece-Rate System, Being a Step Toward Partial Solution of the Labor Problem". O artigo abordava os problemas de eficiência industrial, ao analisar cientificamente comportamentos no trabalho, estabelecendo normas de desempenho e seleccionando trabalhadores usando métodos científicos. O modelo de Taylor da gestão científica permitiu aos gestores usar princípios científicos para resolver o problema dos militares e estabelecer o redesenho do trabalho e um sistema de motivação baseado em incentivos (Moorhead & Griffin, 1995).

Taylor também sugeriu que, uma justificação racional para as políticas de emprego, era tornar os salários dependentes, na sua maior parte, do desempenho dos trabalhadores. As normas propostas por Taylor eram baseadas em estudos de tempo e movimentos do desempenho óptimo no trabalho. A ideia era que a vinculação ao desempenho, acelerava a selecção natural sendo que, os indivíduos que eram mais adequados para uma tarefa ganhavam maiores compensações, enquanto aumentava a produtividade e se reduziam os custos com os trabalhadores (Taylor, 1916, cit. por Scroggins et al., 2008).

No sistema de Taylor de gestão de recursos humanos, os trabalhadores devem ser motivados por incentivos. Os gestores poderiam estabelecer recompensas contextuais que atingissem o estado mental interno do trabalhador e canalizá-las em produtividade. Assim, as raízes da aplicação dos princípios científicos à selecção e a outros aspectos da gestão dos trabalhadores foram estabelecidas, tanto na prática da gestão como na formação universitária de profissionais de recursos humanos no início de 1900 (Taylor, 1916, cit. por Scroggins et al., 2008).

A gestão científica tornou-se tão popular no início da décadas do século XX que o governo começou a usar os seus princípios no serviço militar. No entanto, a oposição cresceu relativamente à sua aplicação na gestão de recursos humanos. Por volta de 1911 a oposição sindical foi tão grande, que os trabalhadores denunciaram a gestão científica e apelaram à greve para combatê-la (Peterson, 1990, cit. por Scroggins et al., 2008).

Ao mesmo tempo que ocorria esta revolução na gestão enfatizando-se o uso da engenharia humana dentro do negócio/ empresa, os psicólogos começaram a aplicar os princípios científicos aos problemas da gestão. A primeira aplicação comercializável em psicologia foi o teste psicológico.

Os psicólogos, a fim de comercializarem eles próprios os testes, durante os primeiros anos de 1900 e de forma a ganharem maior credibilidade e suporte, auto-intitulavam-se de “engenheiros humanos”. Na prática o que pretendiam era a possibilidade de poderem utilizar testes científicos na selecção e avaliação dos empregados (Dewey, 1922, cit. por Scroggins et al., 2008).

Durante o ano de 1916 a Academia Nacional de Ciências criou o “National Research Council” (NRC) para organizar o suporte científico para o esforço de guerra eminente dos Estados Unidos. A subcomissão NRC, designada por “Committee of Psychology”, foi liderada por Robert Yerkes, que era, na altura, o presidente da “American Psychological Association”. Na primavera de 1917 os Estados Unidos entraram na Primeira Guerra Mundial e um grupo proeminente de psicólogos da universidade de Harvard, incluindo Yerkes, Thorndike, Thurstone e Otis, postularam que o esforço de guerra poderia ser ajudado por métodos psicológicos para seleccionar, categorizar e tomar decisões de atribuição e de formação das tropas (Driskell & Olmstead, 1989, cit. por Scroggins et al., 2008). O Comité de classificação de pessoal para o exército, liderado por Walter Dill Scott (primeiro académico Americano com o título de Professor em Psicologia Aplicada), desenvolveu 112 testes para colocar os indivíduos em 83 diferentes postos de trabalho militares e administrou os seus testes em 3,5 milhões de soldados.

Durante os anos entre a Primeira e a Segunda Guerra Mundial, o ambiente empresarial continuou a evoluir e a complexidade organizacional aumentou ao mesmo ritmo que o tamanho das Empresas. As pressões organizacionais sentidas em relação à concorrência e o aumento da regulamentação do trabalho ampliaram ainda mais o impulso para o desenvolvimento de sistemas de gestão racional e para a aplicação de métodos científicos, como forma de melhorar o desempenho.

Como os psicólogos usaram estudos experimentais e o método científico para desacreditar os instrumentos da concorrência e para estabelecer o valor dos seus próprios instrumentos, as normas para o desenvolvimento dos testes e para a sua utilização emergiram. Além disso, recomendações para a formação dos psicólogos industriais foram desenvolvidas e surgiram empresas que publicavam testes, de que é exemplo a “Psychological Corporation”. Cada vez mais, os psicólogos industriais traçavam uma clara distinção entre a psicologia industrial e a gestão científica (Van De Water, 1997). Os psicólogos enfatizavam a importância de factores humanos individuais como a personalidade e a inteligência, como determinantes do comportamento no trabalho, em contraste com o foco da gestão científica em factores contextuais como um sistema de incentivos (Viteles, 1932, cit. por Scroggins et al., 2008). Estudos sobre tempo e movimentos foram desacreditados pelos psicólogos industriais que viam

falhas na gestão científica ao considerarem o elemento humano no local de trabalho como um ponto fraco. Eles viam o desempenho no trabalho relacionado com diferenças individuais na satisfação, personalidade ou inteligência, as quais, podiam ser medidas através de técnicas psicológicas. Com essa mudança de paradigmas, os psicólogos tentaram agarrar o elevado fundamento científico de desenvolver, avaliar e validar a selecção de trabalhadores e introduziram técnicas e instrumentos para tal.

Durante a Segunda Guerra Mundial, a psicologia militar e os serviços psicológicos foram firmemente estabelecidos como essenciais para os esforços de defesa da nação. Por volta de 1940, os psicólogos foram capazes de avaliar e validar as técnicas de classificação e formação, e avanços significativos foram feitos na análise do papel dos factores humanos na concepção e operação de equipamentos, avaliação do desempenho no trabalho, testagem e formação tecnológica, e adaptação a ambientes especiais.

A capacidade dos testes psicológicos para encontrar e prever o desempenho foi bem documentada por psicólogos militares nos Estados Unidos e noutros países por volta de 1940 (Vernon, 1947; Flanagan, 1947, cit. por Scroggins et al., 2008).

Os psicólogos desenvolveram testes de atitudes de capacidades especiais, desenvolveram técnicas de *assessment centre* e criaram a base para o posterior desenvolvimento do Teste de Qualificação das Forças Armadas e a Bateria de Aptidão Profissional dos Serviços Armados (Driskell & Olmstead, 1989, cit. por Scroggins et al., 2008). Naquilo que se tornou o maior sistema de gestão de pessoal dos Estados Unidos – processando 800 mil recrutas anualmente – os testes psicológicos para a justificação de decisões de selecção, colocação e formação tornaram-se institucionalizados e aceites por volta dos anos 50.

A aceitação dos testes psicológicos e das formas de aplicação bem sucedidas, a questões organizacionais, não foram uniformes em todas as áreas. Enquanto alguns tipos de testes psicológicos ganharam ampla aceitação e suporte público, outros não. Por exemplo, a utilidade dos testes de capacidades cognitivas em selecção está bem estabelecida, dada a eficácia muito precisa desses testes e a sua capacidade de prever a aquisição de capacidades de trabalho e certos tipos de desempenho. O valor económico da utilização destes instrumentos de selecção tem sido bem estabelecido, com dados de investigação indicando uma relação elevadíssima entre candidatos seleccionados com elevadas pontuações nos testes e elevadas produtividades no desempenho da função – cerca de 48% acima da média – para cargos profissionais ou de direcção (Schmidt & Hunter, 1998).

Os testes de personalidade têm uma história um tanto diferente. O uso destes testes em selecção de trabalhadores é muito mais controverso. Em contraste aos testes cognitivos, a

opinião predominante dos testes de personalidade em selecção de pessoal é a de que eles não têm validade, que os testes são facilmente falsificados e que são geralmente inadequados para a selecção. Blinkhorn e Johnson concluíram que a baixa validade das medidas de personalidade e a problemática da falsificação dos resultados, torna difícil a recomendação dos mesmos como alternativa em selecção de pessoal (Blinkhorn & Johnson, 1990).

Estava lançada a questão da validade dos testes.

## **O Conceito de Validade e o Processo de Validação**

### ***Perspectiva Histórica***

Anastasi (1986), no seu estudo “*Envolving concepts of test validation*”, refere que, no início do surgimento dos testes padronizados, a validade era avaliada por uma diversidade de procedimentos e era designada por diversos nomes. O tipo de provas produzidas para demonstrar a validade de um teste variava com o propósito do mesmo, a orientação teórica do autor do teste e demasiadas vezes, com a disponibilidade imediata dos dados. Entre as primeiras abordagens empíricas para avaliar os itens do teste e seleccionar os mais válidos, estava o critério de diferenciação etário dos empregados de Binet e Simon (1908, cit. por Anastasi, 1986). Partindo do pressuposto de que as capacidades cognitivas constituintes da inteligência aumentam com a idade, os autores escolhiam tarefas cuja frequência da solução correcta aumentava com a idade. Atribuíam, assim, cada tarefa a um nível etário e a percentagem de crianças que a superava era enquadrada num determinado intervalo específico. Este foi também um dos principais procedimentos seguidos na construção da escala de Stanford-Binet e de outros testes individuais de inteligência, no período em que a inteligência era avaliada em termos de idade mental.

Mas rapidamente, os resultados obtidos nos testes foram analisados, não apenas atendendo à idade cronológica dos avaliados, mas também às suas efectivas realizações, de que são exemplos, a avaliação dos professores face aos resultados atingidos pelos alunos. Com o avanço da estatística, técnicas de análise foram desenvolvidas comparando variáveis internas e externas, testes factoriais e outros, introduzindo bastante confusão sobre o significado e método de validação. Diferentes investigadores e autores de testes empregavam o conceito de validação nas mais variadas acepções, validade intrínseca, validade lógica, validade empírica, factorial.

Em 1954, num esforço para organizar o estado caótico em que se encontravam os procedimentos de construção de testes, a “*American Psychological Association*” publicou, em colaboração com outras entidades, “*The Technical Recommendations for Psychological Tests and Diagnostic Techniques*”. Esta publicação introduziu a actual classificação dos conteúdos da

validação em preditivos, concorrentes e de constructo. Numa edição subsequente deste documento, "Standards for educational and psychological tests" de 1974, a validade preditiva e concorrente foram consideradas como subcategorias da validade de critério e esta divisão manteve-se até aos nossos dias.

Segundo Landy (1993), no seu estudo "The concept of validity and the validation process", a essência desta abordagem foi demonstrar um aspecto que poderia inferir alguma coisa sobre o real desempenho no trabalho através dos escores dos testes. A validade de constructo era mais ampla e menos claramente delimitada. Nesta abordagem, seria o demonstrar de inferências sobre o escore do teste, iluminando o significado do mesmo, através de análises que definiam o significado desse escore de teste. Este tipo de análise geralmente envolve evidências estatísticas sobre a medida em que o teste em questão foi, ou não, associado a outras medidas. Este padrão de associações foi designado por "nomological network". A validade relativa ao conteúdo foi baseada numa lógica do domínio da amostragem. A lógica era a que, se pudesse ser demonstrado que os itens do teste eram uma amostra razoável do trabalho real ou do domínio do trabalho, então, a relação entre os escores do teste e o desempenho era evidente. O ponto importante a ter-se em consideração, para cada um destes tipos de validade, é que eles foram pensados para ajudar a documentar a validade das inferências feitas sobre os escores dos testes mas não a validade do teste em si.

Estas três formas de validade tornaram-se uma taxionomia na prática, sustentada por uma série de acontecimento históricos que lhe deram vida. A Acta dos Direitos Civis de 1964 foi aprovada para garantir os direitos a cidadãos sob protecção. Uma das secções ou título dessa lei, título VII, trata especificamente da discriminação dos trabalhadores. Proibia a discriminação intencional. Alguns anos mais tarde, num marco do Supremo Tribunal, o Acórdão de Giggs, a discriminação não intencional foi também proibida. Assim, se um empregador usou um teste que teve o efeito de negar a igualdade de oportunidades de emprego a membros de um grupo protegido, esse empregador era chamado a demonstrar que o teste em questão estava relacionado com o trabalho. Noutras palavras, era necessário demonstrar que havia uma necessidade, para o negócio, de se usar o teste.

Foi assim que o tema dos tipos de validação surgiu. Em vários conjuntos de orientações administrativas que providenciavam normas para os empregadores no que respeita a uma demonstração da relação teste/ trabalho relacionado, os tipos de validação tornaram-se o foco de discussão. Num precoce conjunto de orientações, conhecidas como Orientações da Agência Executiva Federal (Federal Executive Agency Guidelines), a validação relativa ao critério foi identificada como o único método aceitável para demonstrar essa relação com o trabalho.

Os métodos de análise têm mudado, os testes e as medidas de critério têm-se tornado mais sofisticados, o conceito de generalização da validade veio substituir as anteriores noções da especificidade dos testes situacionais e várias meta-análises têm demonstrado a robustez das relações lineares entre os testes e o desempenho no trabalho. Como resultado, a Sociedade para a Psicologia Organizacional e Industrial (Society for Industrial and Organizational Psychology - SIOP) emitiu uma série de documentos ajudando a articular os princípios profissionais que deveriam guiar o desenvolvimento dos testes e o seu uso. Esses princípios lidam, com grande pormenor, com os tipos de validade e inferência e estão em desacordo com as Orientações Uniformes do governo federal (Uniform Guidelines of the federal government).

A relevância desta breve história dos tipos de validade é dupla. Primeiro, sugere que em vez de se colocar o estudo da validade num grande contexto, deve-se tornar familiar com o debate em relação aos tipos de validade e do propósito dos mesmos. Mas, mais importante, é que a história e o debate em relação aos tipos de validade sugere que qualquer tipo é tão forte quanto os seus elementos constituintes. A hipótese de trabalho, embutida no uso de um teste, é a de que as pessoas que são as melhores no teste são as melhores no trabalho. É o tipo de validade que determina se as hipóteses podem ser suportadas ou têm que ser rejeitadas. Adicionalmente, há ocasiões, em que é aparente que o tipo de validade não é igual à pergunta e que nenhuma resposta pode ser dada. As falhas no tipo de validade impedem a testagem da hipótese. A nossa tarefa será determinar se essa validade permite o teste da hipótese básica relacionando o desempenho no teste com o desempenho no trabalho. É importante reconhecer-se os vários tipos de validade uma vez que, ao fazê-lo estamos também a reconhecer a futilidade de aceitar os três elementos taxionómicos (critério, conteúdo e constructo) que se tornaram uma ortodoxia. Em vez disso iremos perceber que, como em qualquer design experimental, existem muitas variações no tema da validade e cada variação deve ser julgada pelos seus méritos independentemente do rótulo que se possa colocar no design genérico do estudo (eg. relacionado com o critério versus orientado para o conteúdo).

A perspectiva de Cronbach (1988) sobre o tema da validade é bem um exemplo desta abordagem diferenciada.

Em 1988, Lee Cronbach publicou o seu influente artigo intitulado "Five Perspectives on Validity Argument", em que a validação é considerada como um argumento persuasivo. Assim, de acordo com a validação, esta deve incluir um debate entre argumentos contra e a favor, de forma a defender o elemento persuasivo da escolha de um teste face a outro, o que na prática impõe clarificar o que uma determinada medição significa e as suas limitações. Para Cronbach (1988) a palavra-chave é clarificar. O grau de justificação está implícito na palavra limitação. Em

segundo lugar, surge a interpretação que pode assumir a forma de uma descrição, uma predição ou uma acção recomendada. Em terceiro, as inferências, medidas e interpretações de quem aplica os testes devem também ser analisadas. Em quarto lugar, há que ter em atenção que a tarefa de validação não é confirmar um teste, seja ele teórico ou prático. Idealmente quem faz trabalho de validação deve preparar-se como quem vai fazer um debate. Estudar o tópico sob todos os ângulos e dominar os argumentos, pró e contra, tão bem que lhe permita defender ambas as perspectivas.

É possível então, perceber-se no artigo de Cronbach (1988), um claro movimento em direcção à validação como argumento persuasivo.

### ***Validade de Critério, de Conteúdo e de Constructo***

A razão que levou os investigadores a agruparem a validade preditiva e concorrente e a designá-la por validade de critério (Cronbach & Meehl, 1955), reside no facto do padrão do estudo de orientação para o critério ser semelhante. O investigador é principalmente interessado num determinado critério que deseja predizer. Ele administra o teste, obtém uma medida de critério independente sobre o mesmo tema e calcula uma correlação. Se o critério é obtido algum tempo depois que o teste é devolvido, ele está a estudar a validade preditiva. Se a pontuação no teste e a pontuação do critério são determinadas essencialmente ao mesmo tempo, ele está a estudar a validade concorrente. A validade concorrente é estudada quando um teste é proposto como substituto de outro, ou quando é indicado que o teste se correlaciona com alguns critérios contemporâneos (eg. diagnóstico psiquiátrico).

A validade de conteúdo é estabelecida ao mostrar que os itens do teste são uma amostra do universo no qual o investigador está interessado. A validade de conteúdo é normalmente estabelecida dedutivamente, pela definição de um universo de itens e pela amostragem sistemática dentro deste universo para constituir o teste.

A validade de constructo está envolvida sempre que um teste deve ser interpretado como uma medida de algum atributo ou qualidade que não seja “operacionalmente definida”. O problema enfrentado pelo investigador é saber quais os constructos que contam para a variância do desempenho no teste. A validade de constructo não solicita nenhuma nova abordagem científica.

A validade de constructo não é para ser identificada exclusivamente por um processo investigativo particular, mas pela orientação do investigador. A validade orientada para o critério, tal como Bechtoldt (1951, cit. por Cronbach & Meehl, 1955) enfatiza (3, p.1245), “envolve a aceitação de um conjunto de operações como uma definição adequada de tudo o que está a ser

medido.” Quando um investigador acredita que nenhum critério disponível para ele é totalmente válido, ele torna-se necessariamente interessado na validade de constructo uma vez que é a única forma de evitar a “frustração infinita” relativa a cada critério. Na validade de conteúdo, a aceitação do universo do conteúdo como a definição de variáveis a medir, é essencial. A validade de constructo deve ser investigada quando nenhum critério ou universo de conteúdo é aceite como inteiramente adequado para definir a qualidade a ser medida. Determinar quais os constructos psicológicos que contam para o desempenho no teste, seria desejável para quase todos os testes.

Cook (1988), no capítulo intitulado “Validity and Criteria of Productivity. Does it Work? How can you tell?” do seu livro “Personnel Selection and Productivity”, refere que um teste válido é um teste que funciona, que mede o que se propõe a medir e que prediz qualquer coisa útil. Assim sendo, um teste aplicado em selecção de pessoal numa Empresa, só é válido se predizer a produtividade, uma vez que é esse o objectivo das organizações quando contratam empregados. Voltamos assim ao foco na validade de critério que é, aliás, o tipo de validade objecto do presente estudo. O princípio é: os indivíduos que têm elevados resultados nos testes, são mais produtivos – não é relevante o tipo de teste, as questões que coloca e o seu grau de plausibilidade. O que interessa é o critério – produtividade.

A validade de critério assume três formas: preditiva, concorrente e retrospectiva (Cook, 1988).

A validade preditiva suporta-se na afirmação: “o teste prediz quem vai produzir mais. Selecciona hoje e veja os resultados mais tarde”. É também conhecida como “follow-up” ou validade longitudinal.

Na validade concorrente o teste prediz quem produziu mais. É mais fácil e simples que a preditiva uma vez que os indivíduos são testados e os resultados da produtividade recolhidos simultaneamente. É também conhecida como validade cruzada.

Na validade retrospectiva o princípio é o de que testes antigos predizem a produtividade actual. Também é conhecida como “auto investigação” uma vez que os empregadores procuram validar ou melhorar a selecção utilizando a informação obtida quando os empregados foram recrutados mesmo que não necessariamente utilizada com o objectivo da selecção. Os estudos retrospectivos são muitas vezes inconclusivos porque o investigador tem de utilizar a informação disponível, em vez de decidir a informação a recolher.

O estudo que suporta a presente tese monográfica, está claramente associado à utilização da validade de critério concorrente, uma vez que como descrito anteriormente, foram passados testes à população objecto da amostra e simultaneamente recolhida informação sobre o seu grau

de performance, sendo ambos os resultados cruzados com o objectivo de estabelecer uma relação entre os escores obtidos nos testes e o seu desempenho real. Daqui será então possível inferir, se os testes escolhidos serão os mais adequados em futuras selecções, porque predizem níveis elevados de produtividade – validade preditiva.

Há, no entanto, que alertar para os seguintes riscos na utilização deste tipo de validade:

1) **Amostra Restritiva.** Como a amostra resulta já de um processo de selecção anterior, há sujeitos que desistiram e outros que foram dispensados pela organização logo, há percas de indivíduos que não podem ser estudados. Neste tipo de validação é argumentado que se perdem ambos os extremos da distribuição.

2) **Amostra pouco representativa.** Os actuais empregados não são os típicos candidatos.

3) **Alteração da amostra.** Os sujeitos estudados sofreram alterações ao exercerem a sua função, quer por via de formação, quer pelo próprio ambiente da Empresa em que estão inseridos e dos comportamentos exigidos a que se foram adaptando.

4) **Enviezamento dos resultados caso os sujeitos já tenham sido contratados.** Não é o caso da amostra do presente estudo uma vez que os sujeitos estão ainda em período formativo e de experiência sem qualquer garantia de emprego.

Estes argumentos conduzem a que a validade concorrente possa parecer menos conclusiva do que a preditiva mas, várias investigações concluem que, ambos os métodos de validação conduzem aos mesmos resultados (Barrett *et al.*, 1981, cit. por Cook, 1988). Bemis (1968, cit. por Cook, 1988), analisou 71 coeficientes de validade preditiva e 69 de validade concorrente para o “General Aptitude Test Battery” (GABT) e não encontrou diferenças. Também não encontrou diferenças na variância dos escores do GABT no que respeita às amostras, o que demonstra que os dados não foram restringidos com significância.

Porque motivo não diferem então, na prática, estes dois tipos de validação? (Cook, 1988).

1) Uma **amostra restritiva** reduz também os coeficientes de validação nas validações preditivas, nomeadamente porque os empregadores não se limitam a utilizar os testes seleccionados para a escolha dos candidatos, o que conduz necessariamente a redução na amostra, mesmo antes de aplicados os testes.

2) **Amostra pouco significativa.** Pode ser uma questão controversa se trabalhamos com um elevado número de sujeitos. Mesmo na validação preditiva, os resultados só serão significativos se a amostra for de dimensões consideráveis, sob pena de se inferirem resultados sem qualquer consistência.

Barrett *et al.*, (1981, cit. por Cook, 1988), apresentam dois estudos que provam que os escores obtidos nos testes de capacidade cognitiva, são resistentes aos efeitos da experiência “on

Job". Anastasi, (1981, cit. por Cook, 1988), refere que os efeitos do exercício da função e da formação em testes de habilidade existem, mas são suficientemente pequenos para serem considerados.

Os investigadores referem frequentemente o estudo das correlações entre preditor e critério a utilizar mas, se a relação entre eles não for linear, o estudo das correlações é enganador. Existem, no entanto, poucos exemplos de relações não lineares entre preditor e critério (Cook, 1988).

### ***A Validação como Argumento de Avaliação***

Entre os anos 70 e 80, o descontentamento relativamente ao programa de avaliação levou a uma revisão da concepção do papel da investigação na formação de políticas e aplicações práticas da mesma (Cook & Shadish, 1986; Cronbach et al., 1980, cit. por Cronbach, 1988). A Validação de um teste ou a sua utilização é a avaliação (Guion, 1980; Messick, 1980), assim sendo, o que Cronbach (1988) se propõe fazer no seu artigo "Five Perspectives on Validity Argument", é tornar extensível a todos os testes, os ensinamentos do programa de avaliação. Segundo Cronbach (1988), o que House (1977, cit. por Cronbach, 1988) designou por "lógica do argumento de avaliação" aplica-se como "argumento para a validade" em vez de argumento para a "validação da investigação".

Embora a quantidade de doutrinas e acórdãos que nos incutem sobre esta matéria seja caótica, uma nota determinante sobressai. O assunto validação interessa a uma audiência diversa e potencialmente crítica; assim sendo, o argumento tem de fazer a ligação de conceitos, evidências, consequências sociais e pessoais e valores (Cronbach, 1988).

A maioria dos teóricos afirma que os conteúdos e critérios de validação não são mais que fios ao longo de um cabo de argumentos de validação (Dunnette & Borman, 1979; Guion, 1980; Messick, 1980, cit. por Cronbach, 1988). Uma resposta favorável a uma ou duas questões, pode sustentar totalmente um teste apenas quando ninguém se preocupa muito em colocar mais questões. A procura por entender o que conduziu a determinados resultados, e como eles se encontram relacionados com os seus antecedentes e consequentes, bem como compreender também o contexto da aplicação dos testes, é o que permite aos profissionais decidir o que é efectivamente relevante mudar. Desenvolver uma interpretação implica ir mais fundo na validação uma vez que a confirmação, falsificação e revisão interferem. Especialmente quando um instrumento é novo, um inquérito clarifica possibilidades, anomalias e condições envolventes e é de grande validade.

Um argumento afirmativo deverá deixar claro e na medida do possível, de forma persuasiva, o princípio da construção da realidade e o valor implícito das ponderações de um teste e das suas aplicações. Para um argumento a favor ou contra ser plausível tem que encaixar nas crenças e valores prevaletentes ou derrubá-los com sucesso.

Segundo Cronbach (1988), as questões da validade relacionadas com os testes surgem através de cinco perspectivas: funcional, política, operacional, económica e exploratória.

A literatura sobre validação tem-se concentrado na veracidade das interpretações dos testes mas, a **perspectiva funcional** preocupa-se mais com o valor do que com a veracidade dos testes. Alguns autores afirmam que um teste é válido se medir “o que se propõe a medir”. Isto levanta, na sua forma primitiva, a questão do que é considerado verdade. Outros autores afirmam que um teste é válido se servir os propósitos para os quais é usado, o que coloca a questão do valor. A veracidade é uma componente do valor mas ambos não estão fortemente ligados.

A linha que não pode ser ultrapassada é a do princípio de que quem faz validação, tem a obrigação de ver se os resultados têm, na prática, as consequências apropriadas e principalmente precaver-se contra consequências adversas (Messick, 1980, cit. por Cronbach, 1988).

Segundo a **perspectiva política**, a aceitação ou rejeição de uma prática ou teoria dá-se porque a comunidade é persuadida. Mesmo os especialistas não julgam uma conclusão isoladamente, eles julgam a sua compatibilidade com uma rede de crenças prevaletentes (Fiske & Shweder, 1986; Lakatos & Musgrave, 1970; Suppe, 1977, cit. por Cronbach, 1988).

Os argumentos científicos e os argumentos políticos diferem em grau mais do que em tipo. Os científicos têm uma perspectiva temporal mais longa, participantes mais homogéneos e um raciocínio mais formal. Nenhum crítico na avaliação desvaloriza a excelência respeitante ao design experimental, à reprodutibilidade ou ao rigor estatístico. No entanto, essas mais valias têm um preço muito alto, quando restringem um inquérito ao que pode ser avaliado com maior certeza.

A validade dos argumentos contribui quando desenvolve factos e destaca incertezas dos factos ou das suas implicações. Argumentos informados muitas vezes não devem terminar em concordância uma vez que descobertas substanciais são equívocas e porque os participantes atribuem valores diferenciados. A unanimidade de crenças não é um requisito para o sistema político ter equilíbrio.

Adequar os testes a um determinado domínio de conhecimento é a ideia central dos “**operacionais**”. Para eles um procedimento é válido se recolher uma amostra adequada de um comportamento de determinado tipo. Cureton (1951, cit. por Cronbach, 1988), ao falar de critérios diz que a variável a ser medida deve ser definida através da especificação dos actos ou

operações que a compõem; se os actos observam séries bem definidas, obtém-se um conjunto de resultados, por definição, com relevância correcta. A especificação dos testes, incluindo as instruções respectivas para quem testa, é testado, ou apenas observa, são de extrema importância.

A linguagem operacional tem de especificar, se forma clara, o tipo de procedimentos a observar que tornam válida a investigação que está a ser feita. Em ciência a função da definição operacional é permitir replicar procedimentos, quer no objecto original, quer em objectos comparáveis. É também esta a sua função na aplicação de testes.

De forma a minimizar interpretações diferentes dos resultados, os “operacionais” põem de lado questões sobre a verdade e o valor. É como se as únicas questões válidas se limitassem à correcção dos procedimentos, face às definições estabelecidas.

A perspectiva de Campbell e Fiske (1959, cit. por Cronbach, 1988), que afirma que a medição de uma variável através de diferentes métodos é um passo chave na validação, vai contra esta perspectiva.

Segundo a **Perspectiva Económica**, nos testes de selecção de pessoal, a validade tem sido vista como empírica. A validação tradicional põe excessivo peso no critério e aí reside grande vulnerabilidade. A avaliação preditiva (bons resultados nos testes implicam bom desempenho no trabalho) nem sempre corresponde ao esperado. Assim sendo, se as decisões de pessoal estiverem associadas a um valor monetário determinado, a sua correcção é feita quase de forma automática.

Na óptica **exploratória**, existem, na prática, duas perspectivas: testes com margens para várias explicações – o que se chama de “programa forte” – aqueles em que a interpretação é mínima (o observador diz o que fez com a amostra, o que observou, quais os meios usados e apresenta os resultados); o “programa fraco” sujeito a variadas explicações dos resultados obtidos, é fortemente empirista e não testa correlações com outras variáveis.

Uma validação forte (programa forte) é a que se pode traduzir na frase – hipóteses plausíveis mas rivais. Não se trata de andar à procura de situações em que a nossa hipótese não sirva. O que se deve fazer é procurar explicações alternativas para os resultados. Esta ênfase é especialmente importante na validação, porque permite preparação para debates.

## **A Validade e Utilidade dos Métodos de Selecção de Pessoal – Implicações Teóricas e Práticas**

Segundo o artigo “Predictors used for personnel selection: An overview of constructs, methods and techniques” de Salgado, Viswesvaran, e Ones (2002), a selecção de pessoal é um

dos tópicos centrais no estudo do comportamento no trabalho (Guion, 1998, cit. por Salgado, Viswesvaran & Ones, 2002), e visa identificar os sujeitos que irão constituir a força de trabalho numa organização. “As pessoas fazem o lugar” (Schneider, 1987, cit. por Salgado et al., 2002), e seleccionar a pessoa certa para o trabalho certo constitui a fonte da vantagem competitiva para as Organizações. A congruência entre aptidões individuais, a recompensa e exigência organizacional tem sido uma preocupação para o desenvolvimento dos especialistas em recursos humanos.

A saúde financeira de uma Organização radica na óptima selecção e colocação dos trabalhadores (Hunter, Schmidt & Judiesch, 1990, cit. por Salgado et al., 2002),

### ***O Papel da Validação no Processo de Construção dos Testes***

Como é que um se constrói um teste válido? Quais são os ideais dos procedimentos da construção de um teste? Qual é o modelo geral de desenvolvimento de um teste que o autor do teste se esforça para se aproximar, dentro das dificuldades, das exigências de ordem prática e das limitações vida real?

O desenvolvimento de um teste válido requer múltiplos procedimentos, que são empregues sequencialmente em diferentes etapas da construção do teste (Jackson, 1970, 1973; Guion, 1983, cit. por Anastasi, 1986). A validade é, portanto, construída no teste desde o início ao invés de estar limitada aos últimos estádios de desenvolvimento do teste, como na tradicional validade relacionada com o critério. O processo de validação começa com a formulação do traço detalhado ou com as definições de constructo, derivadas de teoria psicológica, investigação prévia, ou da observação sistemática e análise do domínio do comportamento relevante. Os itens do teste são, então, preparados para encaixar nas definições de constructo. A análise empírica dos itens segue com a selecção dos itens mais eficazes (isto é, válidos) desde o primeiro conjunto de itens. Outra apropriada análise interna poderá então ser realizada., incluindo análise factorial do aglomerado dos itens ou de sub testes. A etapa final inclui validação e validação cruzada dos vários escores e combinações interpretativas dos escores através de análise estatística contra critérios externos da vida real.

Quase toda a informação recolhida no processo de desenvolvimento ou na utilização do teste é relevante para a sua validade. É relevante no sentido de que contribui para o nosso entendimento do que o teste mede. Certamente, os dados sobre a consistência interna e sobre a fiabilidade do reteste ajudam a definir a homogeneidade do constructo e a sua estabilidade temporal. As normas podem vir a fornecer outras especificações do constructo, especialmente se incluírem dados normativos separados para subgrupos classificados por idade, sexo, ou outras variáveis demográficas que afectem o desempenho no teste.

Se pensarmos na validade do teste em termos de compreender o que ele efectivamente mede, os dados obtidos virtual e empiricamente com o mesmo, representam uma potencial fonte de validade de informação. Após um teste ser comercializado para uso operacional, o significado interpretativo dos seus escores pode continuar a ser apurado, refinado e enriquecido através da progressiva acumulação de observações clínicas e através de especiais projectos de investigação.

### ***As Fontes de Constructo na Elaboração dos Testes***

Anastasi (1986), no seu estudo “*Involving Concepts of Test Validation*” pede-nos que olhemos mais de perto para as fontes do constructo empregue no desenvolvimento do teste. Como são esses constructos definidos pelo autor do teste? Esta questão realmente pertence à análise de critérios, isto é, uma análise sobre o que o autor pretende que o teste avalie. Independentemente da finalidade do teste, esta é uma questão do critério.

Entre as fontes mais comuns realmente utilizadas pelos criadores dos testes estão as teorias da personalidade, observações clínicas, factores analíticos das investigações das capacidades humanas e, mais recentemente, estudos de processamento de informações a partir da psicologia cognitiva.

Quando os testes são desenhados para uso em contextos especiais, os constructos relevantes são geralmente derivados de análises de conteúdo de domínios particulares do comportamento. Tais análises são muito variáveis na sua meticulosidade, plenitude e precisão.

Em testes ocupacionais, concebidos para a selecção e classificação de pessoal, a análise do critério é geralmente designada por análise de funções (*Job analysis*). Para ser efectiva, a análise de funções deve concentrar-se nos aspectos do desempenho que diferenciam, mais acentuadamente, os melhores e os piores trabalhadores. Em muitos empregos, trabalhadores de diferentes níveis de competência podem diferir na forma como executam grande parte do seu trabalho – apenas determinadas características do seu trabalho podem evidenciar as principais diferenças entre êxitos e falhas. No seu livro intitulado “*Aptitude Testing*”, Clark Hull (1928, cit. por Anastasi, 1986) salientou a importância destes diferentes aspectos do desempenho no trabalho designando-os por actividades críticas (“*critical part-activities*”). Mais tarde, este conceito foi novamente abordado por John Flanagan (1949,1954, cit. por Anastasi, 1986), sob o nome de requisitos críticos (“*critical requirements*”). Para implementar o conceito, propôs o nome de técnica dos incidentes críticos. Esta técnica é utilizada para descrições factuais de ocorrências específicas de características do comportamento no trabalho, quer em trabalhadores satisfeitos ou insatisfeitos. O foco nos requisitos críticos do trabalho conduziu ao desenvolvimento do método de trabalho para a construção de testes e para demonstrar a sua validade. Variantes deste procedimento têm

sido aplicadas a uma ampla diversidade de trabalhos, na indústria e no sector público federal, Estado e a níveis locais (McCormick et al. 1972, Primoff 1975, Menne et al. 1976, Tordy et al. 1976, Eyde et al. 1981, cit. por Anastasi, 1986).

Essencialmente, as componentes do trabalho são descritas e classificadas pelas pessoas que têm a seu cargo determinada função e pelos supervisores. Estas pessoas são escolhidas porque estão minuciosamente familiarizados com o trabalho. As componentes do trabalho referem-se aos comportamentos específicos no trabalho que diferenciam claramente os trabalhadores marginais, dos superiores. Confiando, em última instância, nas observações e julgamentos dos trabalhadores experientes, o método das componentes do trabalho providencia técnicas para a recolha sistemática e quantificação desses julgamentos. Embora várias adaptações deste método difiram nos detalhes procedimentais, todas fornecem a descrição das actividades do trabalho em termos de requisitos comportamentais específicos, a partir dos quais o teste do item pode ser formulado. As descrições do comportamento individual podem, por sua vez, ser agrupadas em categorias mais amplas ou constructos, tais como precisão computacional, visualização espacial, destreza manual, ou capacidade para trabalhar sob pressão. Existe um corpo crescente de investigação direccionada para o desenvolvimento de uma taxionomia geral do desempenho no trabalho em termos de constructos de desempenho relativamente amplos (Fleishman, 1975; Pearlman 1980, cit. por Anastasi, 1986). O método das componentes do trabalho contribui para este objectivo, o que facilita o uso efectivo de um teste em muitas funções com diferenças pouco acentuadas.

### ***Uma Visão Geral dos Indicadores Utilizados em Selecção de Pessoal***

Retomando o artigo "Predictors used for personnel selection: An overview of constructs, methods and techniques" de Salgado, Viswesvaran, e Ones (2002), os autores referem que a selecção de pessoal começa por identificar quais as características individuais que são prováveis de serem relacionadas com o desempenho no trabalho. Para se conseguir isto, dois tipos de fontes de informação são necessárias. Primeiramente, precisamos de entender o que é feito pelo sujeito num trabalho, sob que condições, e qual o propósito. Esta informação é obtida através da análise de funções (Job Analysis). Em segundo lugar, a informação é necessária para identificar o que é valorizado, isto é, devemos identificar o critério utilizado para julgar o desempenho no trabalho. Uma vez que sabemos quais as características necessárias para o trabalho e como vamos avaliar o desempenho, podemos especificar ligações entre essas características individuais e medidas de desempenho no trabalho. Quando as ligações são especificadas, é possível avançar para a medição real dessas características individuais e colocar questões sobre essa medida.

As diferentes características individuais usadas em selecção de pessoal são geralmente referidas como “diferenças individuais” variáveis. Tratam-se de variáveis em que os indivíduos diferem e essas diferenças estão relacionadas com diferenças no desempenho do trabalho. Essas características têm que ser medidas e diferentes métodos podem ser usados para tal. Por exemplo, a capacidade cognitiva é um traço que pode ser medido por diferentes métodos tais como testes de papel e lápis, assessment centre, certos itens de biodata e assim por diante. Para o objectivo a que nos propomos, referimo-nos a qualquer combinação de uma diferente característica individual e de um método de avaliação, como preditor.

Para cada preditor, vamos primeiro analisar a prevalência de seu uso. Resultado de inquéritos indicam quão amplamente o preditor é empregue em selecção de pessoal (e para que tipo de trabalhos e ocupações). Questões de medição e de validade de constructo, que são específicas para cada preditor (ex. capacidade geral vs. específica para capacidades cognitivas, estrutura das capacidades específicas, etc.) são revistas de seguida. A validade de critério para diferentes critérios é sumariada e seguida de uma discussão sobre a validade incremental que o preditor acrescenta, quando combinado com outros preditores.

Os Testes de Capacidade e de atitude têm sido frequentemente usados ao longo dos anos como ferramenta da selecção de pessoal. Nos últimos anos, vários questionários foram realizados com a finalidade de avaliar em que medida os instrumentos típicos de selecção de pessoal são usados. Por exemplo, Gowing e Slivinsky (1994, cit. por Salgado et al., 2002) relatam um levantamento efectuado nos Estados Unidos da América, no qual descobriram que 16% das empresas utilizam testes de capacidade cognitiva (capacidade mental geral) e 42% utilizam testes de atitudes específicas, para efeitos de selecção. Também relatam que 43% das empresas no Canadá usam testes de atitudes. Noutra pesquisa nos Estados Unidos, Marsden (1994, cit. por Salgado et al., 2002) relata que 7,4% das empresas usam a capacidade mental para decisões de contratação em funções que não são de gestão e 9,1% em funções de gestão. Um questionário efectuado em 12 países, membros da Comunidade Europeia, demonstrou que a percentagem de testes psicométricos (testes de capacidade mental geral) por empresas variou de 6% em empresas Alemãs para 74% em empresas Finlandesas (Dany & Torchy, 1994, cit. por Cook, 1998). A média desses 12 países era 34%. Vários questionários sobre o tema foram efectuados e da avaliação conjunta dos mesmos, concluímos que os testes de capacidade mental geral e os testes de atitudes têm sido usados, de forma geral, por todo o mundo.

A maioria dos testes de capacidade, medem mais do que uma capacidade isolada (Ree & Carretta, 1997, 1998, cit. por Salgado et al., 2002). Ademais, diferentes baterias de testes, têm sido desenvolvidas para medir capacidades cognitivas em configurações quer civis, quer militares.

Apesar dos múltiplos sub factores e capacidades medidas nessas baterias, a maior parte da variância dessas medidas deve-se a um factor geral designado por factor g ou Capacidade Mental Geral (Jensen, 1998; Ree & Carretta, 1997, cit. por Salgado et al., 2002).

Uma série de meta-análises foram desenvolvidas na validade relativa ao critério dos testes de capacidade mental geral. Por exemplo, Pearlman, Schmidt e Hunter (1980, cit. por Salgado et al., 2002), mostraram que a capacidade mental geral era um preditor válido do desempenho no trabalho, em trabalho de escritório e a sua validade é generalizada em empregos familiares e organizações.

Meta-análises da validade relativa ao critério em capacidades cognitivas, também foram exploradas em trabalhos específicos.

Na Europa, Salgado e Anderson (2001), fizeram uma meta-análise em estudos Britânicos e Espanhóis conduzidos com testes cognitivos e de capacidade mental geral. Nesta meta-análise, dois critérios foram usados: classificações do desempenho no trabalho e sucesso na formação. Os resultados mostram uma média da validade operacional de 0,44 para classificações de desempenho no trabalho e de 0,65 para o sucesso na formação. Salgado e Anderson (2001), concluíram também que, testes de capacidade mental geral e de atitudes, eram preditores válidos para vários trabalhos, incluindo funções de escritório, condução e comércio. As descobertas de níveis similares de validade generalizável da capacidade cognitiva no Reino Unido e em Espanha, é a primeira evidência multi-cultural, em grande escala, de que testes de capacidade retêm a validade em todos os trabalhos, organizações e contextos culturais.

A capacidade mental geral também prediz outros critérios como o "turnover", grau de realização, mudança de estatuto (promoções) e desempenho (Schmitt et al., 1984, cit. por Salgado et al., 2002). De um ponto de vista de um trabalho psicológico e organizacional, o mais interessante são as correlações positivas entre a capacidade mental geral e o estatuto profissional, o sucesso profissional, conhecimento prático e o salário, e as correlações negativas com o alcoolismo, a delinquência e ociosidade. Tendo em consideração o conjunto de todas estas descobertas, é possível concluir que o teste de capacidade mental geral é um dos mais válidos preditores em psicologia organizacional. Schmidt e Hunter (1998) sugeriram a mesma conclusão no seu artigo/revisão sobre os 85 anos de investigação em selecção de pessoal.

Uma série de estudos foram conduzidos para investigar se as capacidades cognitivas específicas mostram validade incremental ao longo da capacidade mental geral. As evidências actuais demonstram que capacidades específicas não incrementam a predição do desempenho no trabalho.

Apesar dos investigadores argumentarem que as capacidades específicas não acrescentam muito mais validade preditiva para além da capacidade mental geral, outros preditores (e.g. personalidade) foram encontradas para adicionar validade incremental à capacidade mental geral.

A taxonomia de Fleishman (Fleishman & Quaintance, 1984; Fleishman & Reilly, 1991, cit. por Salgado, Viswesvaran & Ones, 2002), agrupa as aptidões em quatro categorias: cognitiva, psicomotora, perceptiva e física. A maioria da investigação empírica conduzida no local de trabalho, tem incidido sobre os testes de capacidades cognitivas. Isto é, provavelmente, devido à aplicabilidade geral dos testes de capacidades cognitivas no local de trabalho. Os testes de capacidade cognitiva estão relacionados com o desempenho no trabalho em muitas mais funções, do que os testes de aptidão física. As aptidões cognitivas e físicas são apenas críticas num número limitado de trabalhos (e.g. bombeiros; segurança pessoal) comparativamente à vasta aplicabilidade dos testes de capacidades cognitivas.

Há poucos dados relativamente ao uso de testes de aptidão física em selecção de pessoal. Alguns inquéritos efectuados sugerem que os testes de aptidão física são usados em menos de 10% das empresas, sendo restritos a funções fisicamente exigentes como sendo os polícias. Não há dados sobre a prevalência de testes de aptidão psicomotora e perceptiva.

### ***Os Determinantes do Valor Prático (Utilidade) dos Métodos de Selecção***

No seu artigo "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings" os autores, Schmidt e Hunter (1998), referem que, do ponto de vista do valor prático, a propriedade mais importante do método de avaliação de pessoal é a validade preditiva: a capacidade de prever o futuro desempenho no trabalho. O coeficiente de validade preditiva é directamente proporcional ao valor económico prático (utilidade) do método de avaliação (Brogden, 1949; Schmidt, Hunter, McKenzie, & Muldrow, 1979, cit. por Schmidt & Hunter, 1998). A utilização de métodos de contratação com maior validade preditiva, levou a aumentos substanciais no desempenho e na aprendizagem dos empregados (Hunter, Schmidt, & Judiesch, 1990, cit. por Schmidt & Hunter, 1998).

A conclusão mais conhecida desta pesquisa é que, para contratar empregados sem experiência prévia no trabalho, o preditor mais válido do desempenho futuro e da aprendizagem, é a capacidade mental geral (Hunter & Hunter, 1984; Ree & Earles, 1992, cit. por Schmidt & Hunter, 1998). A capacidade mental geral pode ser medida usando testes comercialmente disponíveis. No entanto, muitas outras medidas podem contribuir para a validade global do processo de selecção. Isto inclui, por exemplo, medidas de conscienciosidade e de integridade

pessoal, entrevistas estruturadas e para trabalhadores experientes, testes de conhecimentos e *work sample tests* (amostras de trabalho).

A validade de um método de selecção é um determinante directo do seu valor prático, mas não o único determinante. Outro determinante directo é a variabilidade do desempenho no trabalho. Num extremo, se a variabilidade for zero, então todos os candidatos teriam exactamente o mesmo nível de desempenho no trabalho, caso fossem contratados. Neste caso, o valor prático da utilidade de todos os procedimentos de selecção seria zero. Em tal caso hipotético, não importa quem é contratado uma vez que todos os empregados são iguais. No outro extremo, se a variabilidade do desempenho é muito grande torna-se, então, importante a selecção do candidato com melhores resultados e a utilidade prática de métodos de selecção válidos é muito grande. Tal como acontece, este caso extremo aparenta ser a realidade para a maioria dos empregos.

A pesquisa entre 1983 e 1998 tem demonstrado que a variabilidade do desempenho e de resultados entre os trabalhadores é muito grande e seria maior ainda, se todos os candidatos fossem contratados ou seleccionados aleatoriamente (Hunter et al., 1990; Schmidt & Hunter, 1983; Schmidt et al., 1979, cit. por Schmidt & Hunter, 1998).

A variabilidade do desempenho do trabalho assalariado pode ser medida de várias formas, mas duas escalas têm sido tipicamente usadas: o da valorização da produção e a da variação da produção, face a uma percentagem média de saída. O desvio-padrão (SDy) do valor produzido em todos os indivíduos, constatou-se ser no mínimo 40 % da média salarial do emprego (Schmidt & Hunter, 1983; Schmidt et al., 1979; Schmidt, Mack, & Hunter, 1984, cit. por Schmidt & Hunter, 1998). Assim, se o salário médio de um emprego é de 40,000 dólares, então, o desvio-padrão (SDy) é, no mínimo, 16.000 dólares. Se o desempenho tem uma distribuição normal, então os trabalhadores do percentil 84 produzem 16.000 dólares a mais por ano do que a média dos trabalhadores (aqueles que estão no percentil 50). A diferença entre os trabalhadores no percentil 16 (abaixo da média) e aqueles que estão no percentil 84 é duas vezes superior: 32.000 dólares por ano. Essas diferenças são suficientemente grandes para serem valorizadas na saúde económica de uma Organização.

Outro determinante do valor prático dos métodos de selecção, é o rácio da selecção – a proporção de candidatos que são contratados. Num extremo, se uma organização deve contratar todos os que se candidatam ao trabalho, nenhum procedimento de contratação tem qualquer valor prático. No outro extremo, se a organização tem o luxo de contratar apenas 1% com melhores escores, o valor prático dos ganhos da selecção por pessoa contratada será extremamente grande. Mas são poucas as organizações que podem pagar para rejeitar 99% dos

candidatos. Actualmente, os rácios de selecção estão entre a faixa dos 0,30 e os 0,70, uma faixa que ainda produz uma utilidade prática substancial.

Segundo Schmidt e Hunter (1998) o objectivo do seu artigo, foi abordar três questões importantes: os ganhos do valor económico na melhoria dos métodos de contratação; esses ganhos são directamente proporcionais ao tamanho do aumento da validade quando se passa de métodos de selecção antigos para recentes; e nenhuma outra característica de uma medida pessoal é tão importante como a validade preditiva.

A adaptação deste conceito psicométrico de validade preditiva a uma linguagem de gestão, foi também feita por Cascio (1982), que propõe um sistema que permite avaliar a performance no trabalho em termos económicos: *Cascio-Ramos Estimate of Performance in Dollars* (CREPID). A avaliação do rendimento do sujeito é feita em termos do valor da sua actividade e não apenas através das avaliações de desempenho.

### **Recrutamento e Selecção da Academia de Técnicos de Telecomunicações 2008**

O presente estudo decorreu da necessidade de melhorar o processo de selecção dos candidatos para a Academia de Técnicos de Telecomunicações 2008, de uma operadora de telecomunicações. Esta população é constituída por sujeitos com habilitações literárias até ao 12º ano e conhecimentos de electrónica, telecomunicações e informática. O processo de selecção é partilhado com uma empresa prestadora de serviços do grupo, encarregue da fase inicial de selecção.

Esta necessidade foi percebida pela chefia do departamento do Recrutamento e Contratação da Empresa, face à constatação da existência de alguns pontos mais fracos no processo de selecção da Academia, nomeadamente a ausência de provas psicológicas, o que contraria as perspectivas de autores como Robertson e Smith (2000, cit. por Gomes, et al., 2008), baseados numa meta-análise de Schmidt e Hunter (1998), que referem serem, por exemplo, os testes de capacidades cognitivas, os que apresentam as mais elevadas validades preditivas no que concerne ao desempenho.

Efectivamente, e de acordo com Gomes et al. (2008), são várias as razões que levam a que as Organizações optem por procedimentos e testes tecnicamente apropriados: (1) maior eficácia e eficiência na escolha dos candidatos, com reflexos na produtividade; (2) poupança de custos, quer de formação quer de substituição; (3) fonte de atracção para candidatos com maior potencial.

## ***História das Empresas***

A Empresa de serviços partilhados, surgiu fruto de um processo de reestruturação do Grupo a que pertencia, para unificar os vários processos de negócio, sendo hoje a maior empresa de serviços partilhados em Portugal.

Inicialmente contavam com cerca de 150 profissionais mas hoje têm já cerca de 1600 colaboradores distribuídos por mais de 50 escritórios por todo o país.

O factor diferenciador, do modelo da presente empresa, e que o distingue de uma mera “centralização de operações”, é a logística e a dinâmica empresarial. O seu “core business” reside na actividade “não core” dos seus clientes, fazendo com que os processos de suporte (dos clientes) não sejam entendidos como actividades de Back-Office mas sim de Front-Office.

A Empresa onde efectivamente a amostra foi recolhida, a Empresa cliente, trata-se de uma operadora global de telecomunicações líder a nível nacional em todos os sectores em que actua. Assume-se como a entidade portuguesa com maior projecção nacional e internacional e dispõe de um portfólio de negócios diversificado, em que a qualidade e a inovação constituem aspectos determinantes, estando ao nível das mais avançadas empresas internacionais do sector.

Tem como missão prestar serviços de telecomunicações e multimédia para os seus clientes, através da permanente actualização tecnológica, de recursos qualificados e motivados, respeitando o ambiente e a sociedade.

A sua actividade abarca todos os segmentos do sector das telecomunicações: negócio fixo, móvel, multimédia, dados e soluções empresariais. Conta hoje com cerca de 38 milhões de clientes.

## ***O Actual Processo de Recrutamento e Selecção***

Em Fevereiro de 2008, no âmbito do processo de reestruturação de efectivos que a Empresa Cliente tem em curso desde 2003, surgiu a necessidade de rejuvenescimento das equipas de terreno, quer relativamente à estrutura etária, quer ao nível das competências técnicas.

Desde 2003 já foram eliminados cerca de 7000 postos de trabalho. Assim, devido à escassez de técnicos qualificados, a Empresa deu início a um programa de recrutamento e formação, a Academia de Técnicos de Telecomunicações 2008, para a função de técnicos de instalação e manutenção de rede.

A atracção dos candidatos foi feita ao longo do mês de Fevereiro de 2008, por intermédio da imprensa escrita e da Internet. O anúncio foi divulgado nos jornais Diário de Notícias Classificados, Global Notícias Classificados, Sol Caderno Emprego, Expresso Emprego, Metro

Classificados, Correio da Manhã Classificados e Jornal de Notícias Classificados e Via Internet, no Portal SAPO Emprego / Superemprego.

Para se candidatarem ao estágio os candidatos tinham que possuir habilitações literárias ao nível do 12º ano, preferencialmente, da área tecnológica (certificação nível III). Como requisitos mínimos, foram pedidos conhecimentos de informática; bom relacionamento interpessoal e facilidade de comunicação; dinamismo, forte sentido de responsabilidade e resistência à frustração; orientação para os resultados e para o cliente; carta de condução e disponibilidade para deslocações; assim como disponibilidade horária e imediata. Foram considerados factores preferências, o curso de especialização tecnológica de nível IV, preferencialmente na área de electrónica, telecomunicações, redes e informática.

Concorreram sujeitos de ambos os sexos.

O Processo de Recrutamento e Selecção teve duas fases distintas, cada uma realizada numa Empresa do Grupo. Uma primeira fase na Empresa de serviços partilhados e a segunda na Empresa cliente.

A primeira fase consistiu numa triagem inicial das candidaturas, enviadas voluntariamente pelos candidatos, ou de outras que as Empresas tinham guardado em sistema e que se adequavam ao perfil da função.

Todos os candidatos da lista final seleccionados, foram convocados, em grupos pequenos (limite máximo de 10/12), em alturas distintas, para se apresentarem nas instalações da Empresa. Preencheram um formulário de candidatura e assinaram, voluntariamente, uma declaração que permitia a futura divulgação dos Currículos a Empresas do Grupo. De seguida foi pedida a entrega de documentação relativa a *currículo vitae*, carta de condução e certificado de habilitações e feita uma pequena apresentação das Empresas e do Programa Academia. Posteriormente, efectuou-se uma primeira entrevista individual, em grupo, com posterior aplicação de um teste de conhecimentos, finalizando-se com uma Prova de Grupo (Abrigo Subterrâneo) com duração aproximada de 15 minutos.

No final desta fase de pré-selecção, a Empresa que estava a recrutar, enviou para a Empresa Cliente os relatórios de avaliação compostos por uma avaliação comportamental, por um parecer quanto à adequação do candidato à função (favorável, aceitável ou com reservas) e por um campo de observações onde constava um resumo do perfil e características do sujeito.

A selecção final foi realizada pela Empresa Cliente, através de prova de grupo semelhante à já utilizada e de nova entrevista.

Foram seleccionados 214 estagiários.

Os sujeitos seleccionados tiveram formação teórica de um mês e formação “On Job”, de onze meses, a partir do segundo mês até ao final do estágio, que foi de um ano.

Este estágio tinha dois objectivos essenciais:

- 1) Dar uma formação básica, essencial para futura integração;
- 2) Permitir uma avaliação do seu desempenho, de forma a contratar os sujeitos com melhor avaliação.

O processo de atracção dos candidatos, efectuado pela Empresa, adequou-se às características consideradas teoricamente correctas uma vez que, foi publicado anúncio a nível nacional onde foram bem especificadas as características da função e a amplitude geográfica abrangida. Já quanto ao processo de selecção, foram detectadas falhas uma vez que, apesar de ter sido feita uma análise de funções (que permite conhecer o trabalho a fazer, como deve ser feito e que meios e equipamentos o sujeito utilizará, assim como identificar as dimensões de carácter psicológico - aptidões e personalidade - que permitem distinguir o que é valorizado no desempenho óptimo da função) não foram feitas provas psicológicas.

De forma a melhorar o processo de selecção e a satisfazer os interesses da Empresa, procedeu-se à sugestão de reformulação do actual processo de selecção, com apresentação de alternativas.

Após trabalho conjunto com técnicos da Empresa, sugeriu-se que o processo para futuras selecções neste âmbito se baseasse, fundamentalmente, na aplicação de testes, prova de grupo e entrevista psicológica:

- Despiste de Daltonismo (Lâminas de Ishihara);
- Prova de Conhecimentos (Semelhante à já utilizada, com a diferença que terá uma área geral/comum mas específica e outra optativa consoante a área de especialização dos candidatos - telecomunicações ou energias);
- Teste de Atenção – d2;
- Prova Mudanças
- Bateria de Provas de Raciocínio Diferencial (BPRD) – Somente a Prova de Raciocínio Abstracto (Prova AR), a Prova de Raciocínio Espacial (Prova SR) e a Prova de Raciocínio Mecânico (Prova MR);
- Prova de Grupo (Tal como já existe mas alterando-se o tema para uma questão mais real de forma a promover discussão);
- Entrevista.

De forma a consolidar a adequação dos instrumentos propostos, desenvolveu-se um trabalho de aplicação/validação dos testes, a uma amostra da população seleccionada para o estágio:

1) Aplicação do Teste de Atenção (d2), da Prova de Raciocínio Abstracto (Prova AR), da Prova de Raciocínio Espacial (Prova SR), da Prova de Raciocínio Mecânico (Prova MR) da Bateria de Provas de Raciocínio Diferencial e da prova Mudanças;

2) Validação dos resultados obtidos na aplicação dos testes e comparação com a avaliação de desempenho efectuada pela Empresa.

Este estudo tem assim como objectivo primordial verificar a relação existente entre as variáveis predictoras (testes) e a variável critério (avaliação de desempenho), de forma a consolidar a necessidade de aplicação destes testes em futuros processos de selecção.

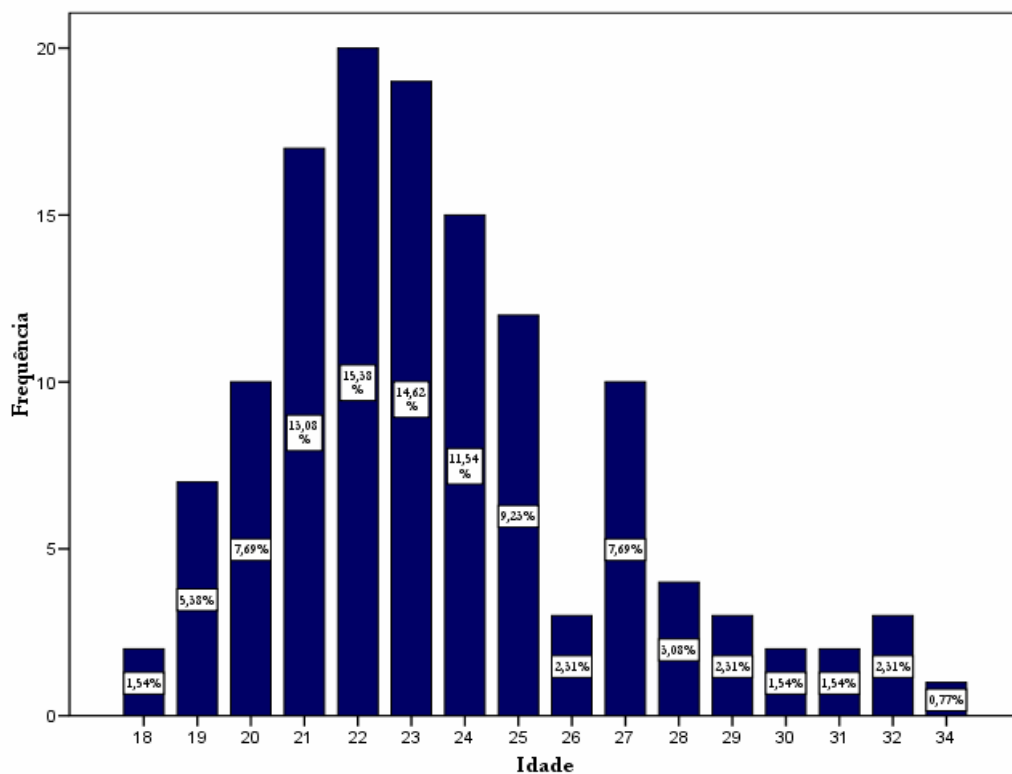
## MÉTODO

### Caracterização da Amostra

No presente estudo, a amostra é constituída por 130 sujeitos, todos do sexo masculino, com idades compreendidas entre os 18 e os 34 anos, de uma população de 214 seleccionados.

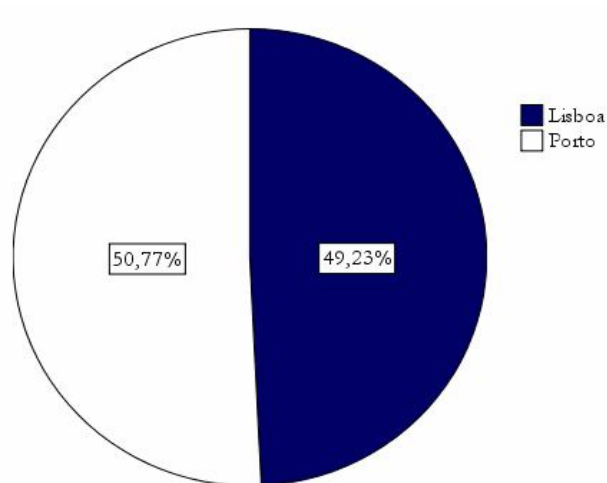
O tipo de modelo de amostragem utilizado foi a amostragem não aleatória ou não probabilística, dado que a probabilidade de um elemento da população fazer parte da amostra, não é igual à dos restantes elementos (Maroco & Bispo, 2003). O tipo de amostragem não aleatória do presente estudo foi acidental ou de conveniência uma vez que a amostra foi seleccionada em função do número de estagiários da Academia de Técnicos, presentes na formação do Porto e de Lisboa nos dias 13 e 15 de Fevereiro, respectivamente.

Relativamente à idade dos participantes pode-se verificar através da figura 1, que a moda se situa nos 22 anos.



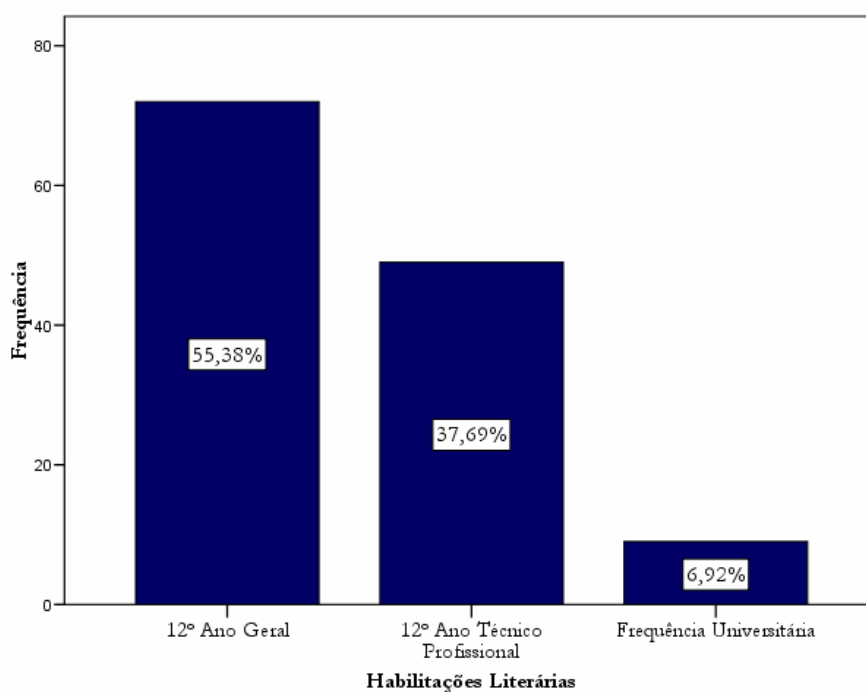
**Figura 1.** Distribuição da Amostra por Idade

A figura 2 descreve a distribuição do número de participantes por zona geográfica, zona de Lisboa ou zona do Porto, consoante o local onde se realizou a aplicação das provas e onde os sujeitos se encontram afectos na estrutura hierárquica da Direcção de Infra-estruturas. Pode-se verificar que 64 sujeitos (49,2%) prestaram provas em Lisboa e 66 sujeitos (50,8%) no Porto.



**Figura 2.** Distribuição da Amostra por Zona Geográfica

Através da figura 3, podemos observar a distribuição dos participantes por Habilitações Literárias. Todos os sujeitos têm o 12º ano de escolaridade concluído sendo que 72 sujeitos têm o 12º ano de carácter Geral (55,4%), 49 sujeitos o 12ºano Técnico Profissional (37,7%) e 9 sujeitos estão a frequentar a Universidade (6,9%).



**Figura 3.** Distribuição da Amostra por Habilitações Literárias

## Variáveis e sua Operacionalização

**a) Variável preditora:** Atenção selectiva e Capacidade de Concentração

**Operacionalização da variável preditora:** Resultados obtidos na Prova de Atenção – d2.

**b) Variável preditora:** Flexibilidade Cognitiva

**Operacionalização da variável preditora:** Resultados obtidos na Prova Mudanças

**c) Variável preditora:** Raciocínio Espacial; Raciocínio Mecânico; Raciocínio Abstracto

**Operacionalização da variável preditora:** Resultados obtidos na Prova de Raciocínio Abstracto (Prova AR), na Prova de Raciocínio Mecânico (Prova MR) e na Prova de Raciocínio Espacial (Prova SR) da Bateria de Provas de Raciocínio Diferencial (BPRD):

**d) Variável critério:** Desempenho dos sujeitos

**Operacionalização da variável critério:** Avaliação de desempenho (Nota Final de estágio).

Esta Nota Final, resulta de uma avaliação composta por dois factores, nota teórica com ponderação de 35% e nota operacional com ponderação de 65%.

- Nota Teórica (NT) – resulta da nota final da formação em sala de aula.
- Nota Operacional (NO) – resulta de duas avaliações (1º semestre e 2º semestre) com uma ponderação de respectivamente 40% e 60%.

A formação teórica iniciou-se com o acolhimento e teve quinze módulos diários (1. Redes de Telecomunicações; 2. Orientação para o cliente; 3. Visita à Central do Areeiro; 4. Rede ATM; 5. Tecnologias e Redes ADSL; Conceitos Ethernet; 7. Comunicação de dados; 8. Rede acesso Ethernet (Media Converters); 9. Conceitos IP; 10. Qualidade; 11. Configuração COE; 12. WI-FI; 13. Redes MPLS; 14. SST; 15. IPTV), terminando com a sessão de integração na Direcção Operacional de Infra-estruturas. A nota da formação (NT), reflecte a média das notas obtidas nos 15 módulos, numa escala de 1 a 6.

A nota operacional (NO), resulta da avaliação de desempenho “On Job”, sendo efectuada pelos orientadores de estágio. Esta avaliação é constituída por dois blocos de itens: Objectivos Smart, com um peso de 30 % e Avaliação Comportamental com um peso de 70%, sendo que a escala de avaliação é de natureza qualitativa. Com os Objectivos Smart, pretendeu-se avaliar a produtividade e o índice de reincidências de avarias (quanto se produz e com que qualidade). A

escala de avaliação é de 1 a 6 sendo 1=objectivo não atingido (<90%) e 6= objectivo atingido, tendo superada totalmente (>110%).

A Avaliação Comportamental é constituída por dez itens (Responsabilidade; Maturidade; Autonomia; Curiosidade Intelectual; Comunicação; Orientação para o Cliente; Espírito de Equipa; Dinamismo; Assiduidade e Pontualidade; e Orientação para os Resultados). A escala de avaliação é também de 1 a 6, sendo 1= Muito Insuficiente e 6= Excelente.

## **Instrumentos**

A avaliação psicológica dos candidatos que foi efectuada no âmbito deste estudo, baseia-se na aplicação de testes.

### ***Teste de Atenção – d2***

O d2 – Teste de Atenção de Brickenkamp, (2007) é uma prova que permite avaliar a atenção selectiva e a capacidade de concentração. Mede, igualmente, a velocidade de processamento da informação, a precisão nesse processamento e aspectos qualitativos relacionados com o desempenho.

É considerado como um teste de cancelamento ou de barragem devido ao tipo de tarefa realizada pelo sujeito que consiste em assinalar um estímulo específico. Na prática, cada uma das 14 linhas, que compõem o teste, apresenta 47 caracteres. Cada carácter é representado por uma letra (“d” ou “p”) que apresenta um, dois, três ou quatro traços por cima e/ou por baixo da letra. A tarefa do sujeito consiste em assinalar as letras “d”, de cada linha, da esquerda para a direita, que apresentem dois traços.

A aplicação demora, em média, aproximadamente entre 8 a 10 minutos, tendo em conta a leitura das instruções e a realização do exemplo, sendo o tempo cronometrado. O sujeito dispõe de 20 segundos por linha, para a realização da prova.

O teste pode ser aplicado de forma individual ou colectiva a sujeitos com mais de 8 anos. A folha de respostas é auto-corrigível.

Através da sua cotação é possível obter determinados resultados:

- Total de Caracteres (TC): Somatório do número de caracteres processados por linha, ao longo das 14 linhas. Possibilita avaliar a velocidade de processamento de informação assim como a quantidade de trabalho realizada.

- Total de Acertos (TA): Somatório do número de caracteres correctamente assinalados por linha, ao longo das 14 linhas. Possibilita avaliar a precisão/ eficácia do sujeito na execução da tarefa.
- Total de Eficácia (TC-E): Total de caracteres menos o número total de erros. Refere-se a uma medida de desempenho global.
- Índice de concentração (IC): Total de acertos menos o número total de erros tipo 2. Refere-se a uma medida da capacidade de concentração.
- Índice de Variabilidade (IV): Diferença entre o número máximo de caracteres processados e o número mínimo. Possibilita avaliar a estabilidade e a consistência do desempenho do sujeito ao longo da tarefa.
- Percentagem de Erros (E%): Percentagem de erros cometidos pelo sujeito ao longo do teste. Os erros podem ser de tipo 1 (E1), quando há omissão de caracteres relevantes, e de tipo 2 (E2), quando o sujeito marca os caracteres irrelevantes. Possibilita avaliar a precisão e a meticulosidade, aspectos qualitativos do desempenho.

Diversos estudos demonstram que o d2 é, ao nível dos países europeus, dos testes mais utilizados na avaliação psicológica. Na Alemanha foi considerado como um instrumento *standard* para a avaliação da atenção e da concentração.

Este teste poderá ser utilizado em diversos contextos, como por exemplo em recrutamento e selecção, contexto educacional, orientação escolar e profissional e em neuropsicologia.

Os estudos de fidelidade sugerem que o teste d2 apresenta coeficientes de fidelidade excelentes (*Alphas* de Cronbach superiores a 0.90), independentemente do tipo de amostra utilizada. Estudos sobre a estabilidade temporal do d2 revelam que este é útil para situações de diagnóstico e prognóstico.

O estudo da Validade facial do d2 indica-nos que este teste é percebido como sendo uma boa medida de atenção e de concentração sendo, portanto, um bom indicador das competências dos sujeitos (Brickenkamp, 2007).

O d2 demonstra excelentes indicadores de validade convergente e divergente. Quanto à validade convergente, a correlação entre os parâmetros do d2 e o Índice de Velocidade de Processamento da WISC-III é significativa ( $R > 0,34$ ). No que respeita à validade divergente, o estudo sobre correlação entre o d2 e o Mudanças, sugere que o bom desempenho nestas provas está relacionado com a capacidade de atenção e de concentração do sujeito. Os coeficientes mais

elevados e significativos ( $R > 0,4$ ) foram alcançados nos parâmetros TA (precisão) e IC (concentração).

### ***Prova Mudanças***

O Mudanças – Teste de Flexibilidade Cognitiva, de Seisdedos, (2005) é uma prova de tipo gráfico, concebida especialmente para avaliar a existência ou não, por parte do sujeito, de estratégias de actuação eficazes.

A tarefa a realizar exige concentração, uma vez que se têm de ter em consideração várias condições de mudança, bem como flexibilidade cognitiva, que permite analisar quase automaticamente se as diferentes mudanças se cumprem ou não. A prova é constituída por 27 itens e dois de treino, sendo cada item composto por três polígonos geométricos simples, em relação aos quais podem haver mudanças relativas ao aumento ou diminuição de uma das suas características, como sendo o número de lados, o tamanho e a densidade da rede interior. O sujeito deverá, primeiramente, averiguar se na passagem do primeiro polígono para o segundo são cumpridas as mudanças designadas pelo círculo A. De seguida, deverá avaliar se as mudanças pedidas no círculo B são cumpridas em relação à passagem do segundo para o terceiro polígono.

Esta prova tem uma duração limite de realização de 7 minutos. A sua cotação é efectuada através de um programa informático próprio para o efeito. A pontuação directa máxima é de 27 pontos.

Esta prova é especialmente direccionada para níveis superiores, nomeadamente sujeitos com escolaridade igual ou superior ao 12º ano.

Os estudos de fidelidade sugerem que o teste Mudanças apresenta um coeficiente de *Alpha* de Cronbach de 0,89 e um Coeficiente de *Split-Half* de 0,93 o que indica uma boa consistência interna do teste.

Relativamente ao estudo da validade, na adaptação Portuguesa da prova procurou-se testar a validade de constructo e a validade concorrente através da correlação do Mudanças com outras provas.

### ***Bateria de Provas de Raciocínio Diferencial (BPRD)***

Bateria de Provas de Raciocínio Diferencial, de Leandro S. Almeida, (2006) é composta por cinco testes que avaliam cinco aptidões, diferenciadas no conteúdo dos respectivos itens – Raciocínio Numérico, Raciocínio Verbal, Raciocínio Espacial, Raciocínio Abstracto e Raciocínio Mecânico. Das cinco provas, apenas a prova de raciocínio numérico requer uma elaboração por

parte do sujeito uma vez que nas restantes provas são facultadas alternativas. Não é necessário a aplicação global da totalidade da bateria.

A sua cotação é efectuada através de um programa informático sendo a nota obtida em cada prova analisada no quadro das normas respectivas.

Esta bateria destina-se, principalmente, a populações de faixas escolares do 9º e 12º ano e a população profissional com um nível de escolaridade baixo.

No presente estudo foram aplicadas as três seguintes provas:

#### 1) Prova de Raciocínio Espacial (Prova SR)

Prova, constituída por 30 itens, que permite avaliar a capacidade de reconhecimento ou de visualização de elementos figurativos. O sujeito tem que identificar o movimento de rotação de um cubo e assim continuar a sequência de movimentos, escolhendo para tal, entre uma das cinco alternativas de resposta, o cubo que completa a sequência.

O sujeito dispõe de, sensivelmente, 16 minutos para a execução desta prova. Trata-se de uma prova particularmente importante para sujeitos com funções técnicas, mecânicas, e artísticas devido à capacidade de reflexão, concentração e resistência à fadiga necessária para a sua realização.

#### 2) Prova de Raciocínio Abstracto (Prova AR)

Prova que permite avaliar a capacidade de raciocínio dos sujeitos, independentemente dos factores culturais, constituída por 35 itens figurativos ou de conteúdo abstracto em termos de significado. O sujeito tem que escolher uma de cinco alternativas, de forma a completar a sequência de pares de figuras.

Esta prova tem uma duração limite de 9 minutos. Aproxima-se dos testes de factor g (inteligência geral) sendo, portanto, uma boa prova para avaliar a capacidade de raciocínio dos sujeitos.

#### 3) Prova de Raciocínio Mecânico (Prova MR)

Prova formada por 40 itens onde são apresentadas situações/ problemas, geralmente do quotidiano, de índole prática, perceptiva e físico-mecânica. Implica conhecimentos básicos de física e de mecânica. A tarefa do sujeito consiste em assinalar a resposta que considera mais adequada para o problema, de entre quatro respostas possíveis. Trata-se de uma prova que requer atenção assim como percepção, em simultâneo, do conteúdo do texto e da figura de cada exercício.

É uma prova com uma duração limite de realização de 15 minutos, sendo adequada a sujeitos ligados a áreas técnicas e práticas.

Os estudos de fidelidade sugerem que as cinco provas da Bateria de Testes de Raciocínio Diferencial apresentam uma média de coeficientes de estabilidade (teste-reteste) entre 0,66 e 0,87. Relativamente à consistência interna (homogeneidade dos itens) os coeficientes situam-se ente 0,75 e 0,92. Os valores obtidos são satisfatórios podendo-se afirmar a precisão ou a confiança na medida.

### **Procedimento**

O presente trabalho de investigação é quantitativo e de carácter correlacional uma vez que pretende verificar o grau de associação entre as variáveis predictoras e a variável critério (Robson, 1999 cit. por D`Oliveira, 2007).

O pedido de autorização para a aplicação das provas foi feito à Empresa cliente, assim como o pedido das notas finais do estágio de cada participante, uma vez que é esta a Empresa que avalia os estagiários.

No que respeita aos testes de selecção, a aplicação das três provas (Prova de Atenção - d2, Prova Mudanças e as três provas da Bateria de Provas de Raciocínio Diferencial) requereu a ocupação da tarde do dia 13 de Janeiro, no Porto, e do dia 15 de Janeiro, em Lisboa, no final do último *workshop* realizado no âmbito do estágio, sendo feita portanto em dois grupos distintos, cada um constituído por 99 e 92 sujeitos, respectivamente.

Em virtude da morosidade do tempo requerido pelas instruções de cada prova, com explicação dos respectivos exemplos e aplicação das mesmas, factores potencialmente geradores de algum desinvestimento e/ou fadiga por parte dos sujeitos, definiu-se uma sequência uniforme de aplicação das Provas – Teste de Atenção d2, Prova Mudanças e, por fim, os três testes da Bateria de Provas de Raciocínio Diferencial – optando-se por não se efectuar a Prova de Raciocínio Numérico e a Prova de Raciocínio Verbal, uma vez que não eram essenciais para o estudo em questão. Esta sequência procurou contrabalançar provas de diferentes níveis de dificuldade.

Em ambos os dias, a aplicação das provas foi feita pela mesma aplicadora, auxiliada por três colegas, devidamente treinados para a aplicação das mesmas. A aplicação foi feita de forma colectiva.

Após uma breve introdução explicando sucintamente os objectivos do estudo e solicitando-se a colaboração dos sujeitos, garantindo a confidencialidade dos dados, procedeu-se

à aplicação das provas iniciando-se, cada uma delas, com o preenchimento dos dados de identificação e com a leitura cuidada das instruções e respectiva realização e correcção do(s) item(s) de treino e posterior esclarecimento de dúvidas. Refira-se que todas as provas têm as instruções transcritas, ou na folha de respostas, ou no respectivo caderno.

No Teste de Atenção d2 o cronómetro mantém-se ligado por um período de 4 minutos e 40 segundos, correspondente ao tempo de realização da tarefa. O tempo de resposta por linha, é de 20 segundos, fim dos quais, é dada a instrução para passar à linha seguinte, num total de 14 linhas. No final da prova, o examinador deverá dizer: "Terminou. Deixem a caneta em cima da mesa e voltem a folha". Recolhem-se as folhas de resposta.

Esta prova não suscitou grandes questões/ problemas de compreensão sendo as instruções e o exemplo suficientes para a compreensão dos participantes.

A Prova Mudanças tem uma duração de 7 minutos cronometrados, fim dos quais é dito ao sujeito: "Atenção, terminou o tempo! Deixe a caneta sobre a mesa e feche o exemplar". Recolhem-se os cadernos de teste.

Esta prova suscitou algumas dúvidas pelo que se sugere a utilização, quando possível, de qualquer meio para a transcrição e explicação do exemplo, de forma a ser explicado de forma mais detalhada.

As três provas utilizadas da Bateria de Provas de Raciocínio Diferencial não diferem da generalidade das provas de papel e lápis. Iniciou-se com a Prova de Raciocínio Espacial, seguindo-se a Prova de Raciocínio Abstracto, finalizando-se com a Prova de Raciocínio Mecânico. Em todas as provas o cronómetro manteve-se ligado por um período de 16, 9 e 15 minutos, respectivamente, correspondente ao tempo de realização da tarefa. Das três provas, a Prova de Raciocínio Espacial foi a que suscitou mais dúvidas e alguma dificuldade ao longo da sua resolução.

Com os dados recolhidos, quer da avaliação de desempenho, de acordo com metodologia atrás referida, quer das notas obtidas nas provas, procedeu-se à elaboração de uma base de dados. Os dados foram inseridos e tratados estatisticamente através da versão 15.0 para *windows* do SPSS (*Statistical Package for Social Sciences*).

## RESULTADOS

Os resultados foram obtidos a partir de análises estatísticas, realizadas no SPSS (versão 15.0), considerando o nível de significância de  $\alpha = 0,05$ .

A amostra inicial é constituída por 130 sujeitos mas, para efeitos da análise estatística, foram retirados 16 outliers identificados na análise de resíduos, 9 do Teste d2, 2 da Prova Mudanças e 6 do BPRD (anexo A).

Iniciou-se o tratamento estatístico, com uma fase descritiva dos dados, seguida de uma análise às qualidades métricas dos testes, nomeadamente a Validade, Sensibilidade e Fidelidade. Procedeu-se, seguidamente, aos tratamentos estatísticos que permitiram observar as relações existentes entre as variáveis predictoras e critério.

### Estatística Descritiva

A tabela 1 apresenta a caracterização das variáveis Predictoras e a tabela 2 apresenta a caracterização das variáveis de Critério, quanto ao tamanho da amostra (N), valores mínimos e máximos, média e desvio-padrão. As variáveis tratadas são as notas dos três testes (D2, Mudanças e BPRD) e as notas obtidas na avaliação de desempenho (Nota Operacional do 1º e 2º semestre, Nota Teórica e Nota Final).

**Tabela 1.** Estatística Descritiva dos Predictores

	N	Mínimo	Máximo	Média	Desvio Padrão
Total Caracteres do d2	114	258	649	445,36	72,26
Total Acertos do d2	114	91	238	170,76	29,47
Total de Eficácia do d2 (TC-E)	114	256	584	427,15	67,28
Índice de Concentração do d2	114	87	238	170,36	29,55
Índice de Variabilidade do d2	114	4	28	15,46	5,02
% de Erros do d2	114	0	16	4,00	3,37
Mudanças	114	4	26	15,70	5,05
Prova MR da BPRD	114	11	34	22,35	4,74
Prova SR da BPRD	114	4	27	17,68	4,90
Prova AR da BPRD	114	16	30	23,11	3,18

**Tabela 2.** Estatística Descritiva das Variáveis Critério

	N	Mínimo	Máximo	Média	Desvio Padrão
Nota Teórica	114	3,66	4,53	4,15	,21
Nota Operacional 1º Semestre	114	,98	4,63	2,84	,65
Nota Operacional 2º Semestre	114	1,21	4,93	3,28	,72
Nota Final	114	2,28	4,50	3,47	,43

Fez-se idêntica análise estatística, por área de trabalho (anexo B), agrupando os candidatos a norte de Fátima, na Zona Porto e a sul, na Zona Lisboa, uma vez que o tratamento da informação de forma mais dispersa, não tinha significância estatística e porque os orientadores de estágio, se encontram alocados por essas zonas.

### **Estatística Descritiva – Análise Teste a Teste**

A análise descritiva dos resultados obtidos nos diferentes testes, incluindo informação relativa aos respectivos grupos normativos constantes nos diferentes manuais, é apresentada no anexo C.

Na tabela relativa ao teste d2, a população está dividida por grupos etários – tabela de normas para adolescentes dos 17 aos 18 anos (N=161); tabela de normas para jovens adultos dos 19 aos 25 anos (N=248); e tabela de normas para adultos dos 26 aos 35 anos (N=173). Estão destacados, os valores inferiores aos do grupo normativo respectivo.

Na tabela relativa ao teste Mudanças, o grupo normativo é o respeitante aos sujeitos com habilitações ao nível do Ensino Secundário. Podemos verificar, através da tabela ilustrativa, que a média e o desvio-padrão dos resultados obtidos no teste Mudanças são ligeiramente superiores às do grupo normativo.

Na tabela relativa ao BPRD, o grupo normativo é relativo a população profissional com o 11º/12º anos. A média dos resultados obtidos nas três provas é inferior à do grupo normativo, sendo o desvio-padrão sempre superior.

### Qualidades Métricas

Foram analisadas as qualidades métricas dos testes, nomeadamente a validade, sensibilidade e fidelidade.

No que respeita à Validade efectuou-se o teste do KMO e o teste de Bartlett, a fim de se verificar se se podia fazer a análise factorial.

**Tabela 3:** Resultados dos Testes KMO e de Esfericidade de Bartlett

	d2	Mudanças	BPRD		
			SR	AR	MR
Kaiser-Meyer-Olkin Medida de adequação da amostra	0,59	0,86	0,68	0,53	0,51
Teste de esfericidade de Bartlett sig.	0,00	0,00	0,00	0,00	0,00

Os valores obtidos pelo teste KMO são todos superiores a 0,5 o que indica que a análise de componentes principais pode ser feita. Com o teste de Esfericidade de Bartlett, pela análise do nível de significância (sig. = 0,000 <0,05) conclui-se que as variáveis são correlacionáveis.

Esta informação permitiu-nos testar a Validade de Constructo, procedendo-se à análise factorial.

Através da Matriz de Componentes Principais após rotação do Teste d2 (anexo D), verifica-se que há duas dimensões. Efectuou-se a rotação apenas para este teste, com o objectivo de dividir o conjunto inicial de variáveis em subconjuntos com o maior grau de independência possível, uma vez que o Teste Mudanças e os 3 Testes do BPRD, não têm diversas variáveis.

Deve-se ainda ter em conta três critérios para determinar o número de componentes a incluir na análise factorial: Eigenvalues, Variância Explicada e Scree Plot (Pestana & Gageiro, 2008).

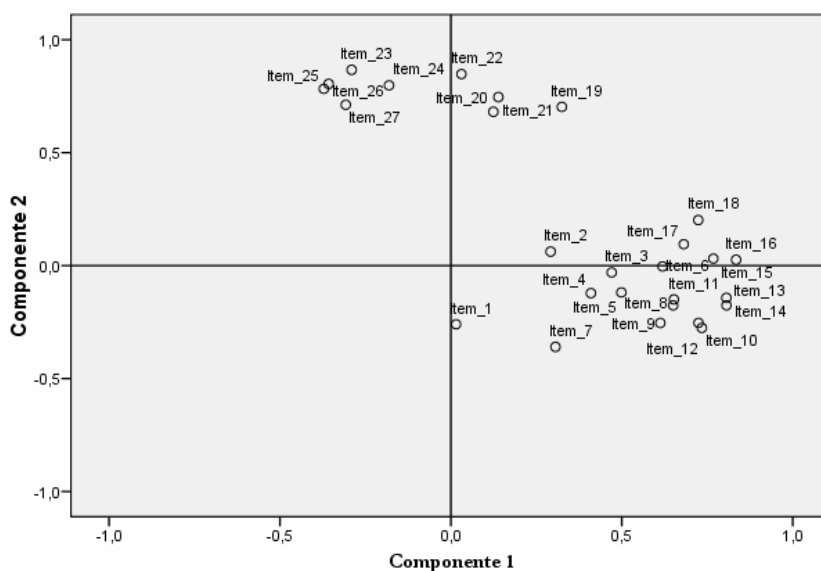
Através da análise das tabelas (anexo D), referentes à análise factorial de cada teste, conclui-se serem as seguintes componentes as que apresentam eigenvalues superiores a 1,00, bem como a respectiva percentagem de variância dos dados iniciais: teste d2, duas componentes que explicam 85,9%; teste Mudanças, cinco componentes que explicam 68,3%; prova SR, onze componentes que explicam 69,1%; prova AR, doze componentes que explicam 65,2%; e na prova MR, quinze componentes que explicam 67,5%.

Através da análise das figuras (anexo D), referentes ao Scree Plot (gráfico dos valores próprios de cada componente) de cada teste, conclui-se haver opções diferentes, quanto ao

número de componentes, para as respectivas análises factoriais: teste d2, uma ou três componentes; teste Mudanças, duas componentes; prova SR, duas componentes; prova AR, uma componente; e prova MR, uma componente.

Importa ainda referir que a opção apurada através do scree plot para se efectuar a análise factorial, do teste d2 a três factores não foi realizada uma vez que, quer através da matriz rodada, quer através dos eigenvalues, se identifica uma existência clara de apenas dois factores.

Para o teste Mudanças e para a prova SR, a análise do Scree Plot dá-nos, também, a opção de fazer a análise factorial a duas dimensões dado o declive acentuado da primeira dimensão para a segunda e da segunda para a terceira. Uma vez que o teste Mudanças mede a flexibilidade cognitiva e a prova SR, o raciocínio espacial, tomou-se a opção de fazer a análise factorial a um factor que explica respectivamente 32,41% e 15,99% da variância (anexo D). No entanto, considerando que o Manual do Mudanças diz que a partir do item 19, inclusive, os valores na Matriz Factorial apresentam escores negativos, decidiu-se forçar a análise factorial exploratória a dois factores (anexo D), que nos revelou um afastamento claro entre os dois conjuntos de itens (Figura 4), pelo que foi decidido testar a capacidade preditiva de cada uma destas novas componentes. Dado os resultados obtidos na correlações posteriormente efectuadas, abandonou-se esta via mantendo para efeitos deste estudo a análise a uma componente.



**Figura 4.** Gráfico de Componentes do Mudanças após Rotação

A análise do scree plot da prova AR e MR, dá-nos a opção de fazer a análise factorial a uma dimensão dado o declive acentuado da primeira dimensão para a segunda, apesar de explicar apenas, respectivamente 10,31% e 10,70% da variância (anexo D).

Não se realizou o estudo da validade de conteúdo uma vez que esta serve essencialmente para provas de conhecimento e, não havendo coeficiente de validade de conteúdo, esta só pode ser garantida após a análise de um perito.

Para se efectuar a análise da Sensibilidade procedeu-se ao teste One-Sample Kolmogorov-Smirnov, para todas as provas.

No caso do teste d2, é necessário testar a normalidade de cada parâmetro, para se poder efectuar estudos posteriores de correlação, de forma a decidir quais os parâmetros a usar nas análises seguintes.

**Tabela 4.** One-Sample Kolmogorov-Smirnov para o Teste d2

	Total de Caracteres do d2	Total de Acertos do d2	Total de Eficácia do d2 (TC-E)	Índice de Concentração do d2	Índice de Variabilidade do d2	% de erros do d2
Kolmogorov-Smirnov (sig.)	0,93	0,84	1,00	0,89	0,23	0,00

Através da análise da tabela 4, One-Sample Kolmogorov-Smirnov, concluiu-se que apenas a Percentagem de Erros do d2, não apresenta distribuição normal ( $p < 0,05$ ).

Seguidamente, foram obtidas as correlações entre os diferentes parâmetros do Teste d2, com excepção do parâmetro Percentagem de Erros, por não apresentar distribuição normal (anexo E).

Através dos resultados da correlação de Pearson (anexo E), concluiu-se que o Total de Acertos e o índice de Concentração têm uma correlação perfeita positiva ( $R=1$ ), sendo que, quando um aumenta, o outro também aumenta em média num valor proporcional.

Importa referir que a variável Total de Eficácia é a que se correlaciona mais fortemente com todas as variáveis, não se correlacionando com o Índice de Variabilidade. Por outro lado, a variável Índice de Variabilidade, é a que menos se correlaciona com as outras variáveis.

Optou-se então por ter em conta para efeito das análises seguintes, o Total de Eficácia (que corresponde a uma medida do desempenho global) e o Índice de Variabilidade (que permite avaliar a estabilidade e a consistência do desempenho do sujeito ao longo da tarefa).

Procedeu-se de seguida à análise dos Coeficiente de Assimetria (-0,61) e de Achatamento (-0,72) do Total de Eficácia, cujos resultados nos indicam, que a distribuição é normal. Procedeu-se também à análise dos Coeficiente de Assimetria (1,72) e de Achatamento (-0,13) do Índice de Variabilidade, cujos resultados, também, enquadram a distribuição como normal (anexo E).

Para a análise de sensibilidade do teste Mudanças, SR, AR e MR, efectuou-se o teste One-Sample Kolmogorov-Smirnov, e apuraram-se os coeficientes de assimetria e de achatamento.

**Tabela 5.** One-Sample Kolmogorov-Smirnov e Coeficientes de Assimetria e de Achatamento para o Mudanças e para as Provas SR, AR e MR da BPRD

	Mudanças	BPRD		
		SR	AR	MR
Kolmogorov-Smirnov (sig.)	0,51	0,87	0,30	0,12
Coeficiente de assimetria	-1,45	-1,23	-0,47	-0,33
Coeficiente de achatamento	-0,44	0,10	-1,00	0,17

Através da análise da tabela 5, concluiu-se que todas as provas apresentam distribuição normal segundo o K-S ( $p > 0,05$ ), bem como no que respeita aos valores dos coeficientes de assimetria e de achatamento, que se encontram dentro do intervalo  $]-1,96 ; + 1,96[$ .

No que respeita à análise da fidelidade, esta não foi efectuada para o teste d2, por este ser uma prova de rapidez.

De forma a testar a consistência interna das restantes provas, procedeu-se ao cálculo do Alpha de Cronbach.

**Tabela 6.** Consistência Interna para o Mudanças e para as Provas SR, AR e MR da BPRD

	Mudanças	BPRD		
		SR	AR	MR
Alpha de Cronbach	0,82	0,81	0,63	0,70
Nº de itens	27	30	35	40

Através da análise da tabela 6, concluiu-se que todos os testes, com excepção das provas AR e MR, apresentam fidelidade ( $\alpha > 0,8$ ). A análise da consistência interna das várias

provas, quando se retiram itens, revela que não existe qualquer alteração significativa ao valor de fidelidade (anexo F).

### Análise de Correlações

Esta análise tem por objectivo avaliar o tipo de relação existente entre as variáveis em estudo (preditores e critério), quer em termos de magnitude, quer em termos da direcção ou correlação existente entre as variáveis. Para este efeito, foram estudadas em primeiro lugar as correlações entre variáveis predictoras e entre variáveis critério (intra correlações) e posteriormente realizou-se a análise entre as variáveis predictoras e critério (inter correlações).

Procedeu-se à correlação entre os preditores de forma a perceber até que ponto são suficientemente independentes entre si (tabela 7).

**Tabela 7.** Correlação de Pearson entre os Preditores

		Total de Eficácia do d2 (TC-E)	Índice de Variabilidade do d2	Mudanças	Prova SR da BPRD	Prova AR da BPRD	Prova MR da BPRD
Total de Eficácia do d2 (TC-E)	Correlação de Pearson	1	,022	,343**	,255**	,239*	,172
	Sig. (2-tailed)		,815	,000	,006	,010	,067
	N	114	114	114	114	114	114
Índice de Variabilidade do d2	Correlação de Pearson	,022	1	,004	-,108	-,164	-,125
	Sig. (2-tailed)	,815		,969	,252	,081	,184
	N	114	114	114	114	114	114
Mudanças	Correlação de Pearson	,343**	,004	1	,506**	,450**	,345**
	Sig. (2-tailed)	,000	,969		,000	,000	,000
	N	114	114	114	114	114	114
Prova SR da BPRD	Correlação de Pearson	,255**	-,108	,506**	1	,571**	,416**
	Sig. (2-tailed)	,006	,252	,000		,000	,000
	N	114	114	114	114	114	114
Prova AR da BPRD	Correlação de Pearson	,239*	-,164	,450**	,571**	1	,363**
	Sig. (2-tailed)	,010	,081	,000	,000		,000
	N	114	114	114	114	114	114
Prova MR da BPRD	Correlação de Pearson	,172	-,125	,345**	,416**	,363**	1
	Sig. (2-tailed)	,067	,184	,000	,000	,000	
	N	114	114	114	114	114	114

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Através dos resultados da correlação de Pearson, pode-se concluir que não existe correlação entre o Índice de Variabilidade do d2 e as restantes variáveis.

A análise prova a prova, indica-nos que o Total de Eficácia, não está correlacionado com a Prova MR e apresenta uma correlação positiva, com as restantes variáveis. O Mudanças, correlaciona-se positivamente com todas as variáveis, assim como a Prova SR e a Prova AR. A

Prova MR, não se correlaciona com o Total de Eficácia, no entanto, apresenta uma correlação positiva com as restantes variáveis.

De forma a tentar perceber quantas dimensões estamos a avaliar, fez-se uma análise factorial com todos os preditores, após se efectuar o teste KMO (0,776) e de Esfericidade de Bartlett ( $\text{sig}=0,000 < 0,05$ ), que nos permitiu concluir que as variáveis são correlacionáveis (anexo G).

Através da Matriz de Componentes Principais após rotação dos Preditores (anexo G), verifica-se que há duas dimensões, a primeira composta pelos dois parâmetros do d2 (Total de eficácia e Índice de Variabilidade) e a segunda por todos os restantes testes, de carácter mais cognitivo.

Através da análise factorial dos Preditores (anexo G), observamos que apenas as 2 primeiras componentes têm Eigenvalues superiores a 1,00 e que estas explicam 59,95% da variância dos dados iniciais.

A análise referente ao Scree Plot (anexo G), dá-nos a opção de fazer a análise factorial a 5 dimensões, não tendo a mesma sido efectuada neste estudo.

Uma vez analisadas as correlações entre os preditores, efectuou-se idêntica análise para as variáveis critério, de forma a encontrar explicações para as notas finais obtidas pelos estagiários.

No que respeita às Notas obtidas pelos estagiários ao longo do estágio e no final do mesmo (Nota Final), procedeu-se ao teste de Kolmogorov-Smirnov (K-S).

**Tabela 8.** One-Sample Kolmogorov-Smirnov para as Variáveis Critério

	Nota Operacional 1º semestre	Nota Operacional 2º semestre	Nota Teórica	Nota Final
Kolmogorov-Smirnov (sig.)	0,05	0,38	0,80	0,96

Verifica-se, através da tabela 8, que todas as variáveis apresentam distribuição normal ( $\text{sig} > 0,05$ ), com excepção da Nota Operacional do 1º semestre .

Seguidamente, procedeu-se às correlações das diferentes Notas no sentido de se perceber a que melhor explica a Nota Final.

**Tabela 9.** Correlação das Variáveis Critério

		Nota Teórica	Nota Operacional 1º Semestre	Nota Operacional 2º Semestre	Nota Final (NO+NT)
Nota Teórica	Correlação de Pearson	1	,130	,052	,252**
	Sig. (2-tailed)		,168	,579	,007
	N	114	114	114	114
Nota Operacional 1º Semestre	Correlação de Pearson	,130	1	,691**	,870**
	Sig. (2-tailed)	,168		,000	,000
	N	114	114	114	114
Nota Operacional 2º Semestre	Correlação de Pearson	,052	,691**	1	,931**
	Sig. (2-tailed)	,579	,000		,000
	N	114	114	114	114
Nota Final (NO+NT)	Correlação de Pearson	,252**	,870**	,931**	1
	Sig. (2-tailed)	,007	,000	,000	
	N	114	114	114	114

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Através dos resultados da correlação de Pearson, pode-se concluir que a Nota Teórica apenas se correlaciona com a Nota Final, sendo esta uma associação baixa. Relativamente à Nota Final conclui-se que se correlaciona positivamente, com todas as variáveis, tendo uma associação alta com a Nota Operacional do 1º Semestre e muito alta com a Nota Operacional do 2º Semestre.

Procedeu-se por fim à análise de correlação entre preditores e variáveis critério.

**Tabela 10.** Correlação entre Preditores vs Critério

		Nota Teórica	Nota Operacional 1º Semestre	Nota Operacional 2º Semestre	Nota Final (NO+NT)
Total de Eficácia do d2 (TC-E)	Correlação de Pearson	,191*	-,177	-,079	-,089
	Sig. (2-tailed)	,042	,060	,403	,346
	N	114	114	114	114
Índice de Variabilidade do d2	Correlação de Pearson	-,003	-,111	-,133	-,112
	Sig. (2-tailed)	,978	,240	,159	,234
	N	114	114	114	114
Mudanças	Correlação de Pearson	,163	-,069	-,071	-,055
	Sig. (2-tailed)	,084	,469	,454	,562
	N	114	114	114	114
Prova SR da BPRD	Correlação de Pearson	,105	,051	,004	,033
	Sig. (2-tailed)	,266	,591	,970	,725
	N	114	114	114	114
Prova AR da BPRD	Correlação de Pearson	,086	-,004	-,012	-,006
	Sig. (2-tailed)	,362	,969	,899	,950
	N	114	114	114	114
Prova MR da BPRD	Correlação de Pearson	,081	,163	,129	,154
	Sig. (2-tailed)	,392	,083	,170	,102
	N	114	114	114	114

\*. Correlation is significant at the 0.05 level (2-tailed).

Através da análise dos resultados da correlação de Pearson, pode-se concluir que não há qualquer correlação entre as variáveis preditoras e a variável critério, com excepção da que se verifica entre o Total de Eficácia do d2 e a Nota Teórica, sendo esta uma associação baixa.

Relativamente à tentativa exploratória atrás referida de análise da capacidade preditiva do Teste Mudanças a duas componentes, concluiu-se também pela inexistência de correlação com as variáveis critério (anexo G).

### Análise de Regressões

De modo a verificar a capacidade preditiva dos preditores (testes) foi realizada uma regressão linear múltipla tendo como variáveis independentes os 5 Testes, e como variável dependente a Nota Final (Nota Operacional + Nota Teórica) obtida no final dos 12 meses de estágio. Foi utilizado o método *Enter*, pois todas as variáveis independentes são consideradas na análise.

Para que a regressão linear seja correctamente aplicada deve satisfazer os seguintes pressupostos: erros (ou resíduos) aleatórios e independentes, erros com distribuição normal de média zero e variância constante; e variáveis independentes (quando se trata de uma regressão linear múltipla) não colineares (Maroco, 2003).

A análise aos pressupostos foi assim iniciada com a verificação do pressuposto de independência dos erros, através da estatística Durbin-Watson (d), de acordo com a qual os erros são independentes e aleatórios (auto correlação nula entre resíduos) se  $d=2, \pm 0,2$ .

**Tabela 11.** Análise da Independência dos Erros

Model	R	R <sup>2</sup>	R <sup>2</sup> Ajustado	Erro Padrão	Durbin-Watson
1	,231 <sup>a</sup>	,054	,001	,425	1,623

a. Preditores: (Constante), Prova MR da BPRD, Índice de Variabilidade do d2, Total de Eficácia do d2 (TC-E), Prova AR da BPRD, Mudanças, Prova SR da BPRD

Através da análise da tabela 11, verifica-se que os erros são independentes ( $d=1,623$ ).

O pressuposto de normalidade dos erros pode ser testado através de um gráfico de probabilidade normal (gráfico de frequências esperadas vs. frequências observadas), sendo que se

os valores se distribuírem mais ao menos ao longo da diagonal principal pode considerar-se que possuem distribuição normal.

O pressuposto de ausência de multicolinearidade entre as variáveis independentes refere-se à ausência de correlação entre as mesmas. Considera-se que existem problemas de colinearidade quando o valor da Tolerância  $<0,1$  ou  $VIF > 10$ .

Através da análise de normalidade sem outliers, verifica-se também o pressuposto de normalidade dos erros (pois distribuem-se em torno da diagonal principal) e o de ausência de multicolinearidade (valores de tolerância  $T > 0,1$  e  $VIF < 10$ ) (anexo H).

A tabela 11 apresenta também o modelo de regressão. De modo a verificar a percentagem de variabilidade da variável dependente explicada pelas variáveis independentes, foi utilizado o coeficiente de determinação ( $R^2$ ) ajustado uma vez este tem em conta o número de variáveis independentes incluídas na análise. A análise da tabela permite verificar que os Testes explicam 0,1% da variabilidade da Nota.

### **Análise Discriminante**

De modo a analisar se os Testes de Selecção são bons preditores do desempenho futuro dos candidatos foi também utilizada a técnica de análise discriminante, através do método Enter uma vez que neste método todas as variáveis predictoras (ou independentes) entram na análise. A variável critério (ou dependente) é a Nota Final.

Uma vez que a análise discriminante exige que a variável dependente seja discreta, a variável original Nota Final (utilizada na análise de regressão) foi codificada em três grupos: grupo1 – média inferior (sujeitos com média entre 2,28 e 3,22 valores), grupo 2 – média moderada (média entre 3,23 e 3,72 valores), e grupo 3 - média superior (média entre 3,73 e 4,50 valores). Esta categorização teve subjacente a divisão da amostra com base nos percentis (anexo I), sendo que o grupo 1 corresponde ao percentil 25 (25% da amostra), o grupo 3 ao percentil 75 (também 25% da amostra), e o grupo 2 à diferença entre os dois (50% da amostra).

O primeiro grupo contém 30 sujeitos, o segundo 56 e o terceiro 28.

A aplicação correcta da análise discriminante pressupõe o cumprimento de três pressupostos: distribuição normal multivariada das variáveis, homocedasticidade (i.e. homogeneidade de matrizes de variância-covariâncias), e ausência de multicolinearidade entre variáveis (Maroco, 2003).

O pressuposto de distribuição normal multivariada é avaliado através dos gráficos de resíduos de probabilidade normal (análise de normalidade da distribuição das variáveis). O segundo pressuposto, de homogeneidade de matrizes de variância-covariâncias pode ser testado através da estatística M-Box (fornecida na pela análise discriminante), e o terceiro pressuposto, ausência de multicolinearidade entre variáveis deve ser avaliado através das medidas Tolerância ou VIF.

Na análise discriminante realizada foi testado o pressuposto de homogeneidade de matrizes de variância-covariâncias, uma vez que os outros dois pressupostos já foram testados nas análises de regressão. Estas análises revelam que as variáveis não são colineares e apresentam distribuição normal.

**Tabela 12.** Teste M-Box

Box's M		1,133
F	Approx.	,551
	df1	2
	df2	5032,788
	Sig.	,577

Através da análise da tabela 12 pode verificar-se a validação do pressuposto de homogeneidade das matrizes de variância-covariância uma vez que  $p > 0,05$ , não se rejeitando a hipótese nula de homogeneidade das matrizes de variância-covariância.

Uma vez validados todos os pressupostos, realizou-se a análise discriminante através do método Enter utilizando como variáveis preditoras os Testes de Selecção.

**Tabela 13.** Anova para os Testes de Selecção – Análise Discriminante

	Wilks' Lambda	F	df1	df2	Sig.
Total de Eficácia do d2 (TC-E)	,998	,138	2	111	,871
Índice de Variabilidade do d2	,981	1,096	2	111	,338
Mudanças	,991	,483	2	111	,618
Prova SR da BPRD	,994	,355	2	111	,702
Prova AR da BPRD	,993	,385	2	111	,681
Prova MR da BPRD	,920	4,856	2	111	,010

Observando a tabela 13 e utilizando o  $\alpha$  corrigido,  $\alpha = 0,016$  ( $0,10/6 = 0,016$ ), pois existem seis variáveis preditoras (Testes de Selecção), pode verificar-se que não existem diferenças de médias significativas entre os três grupos de Notas quanto aos diferentes Testes. Isto é, os

resultados desta análise são semelhantes aos anteriormente obtidos, confirmando que nenhum Teste (quando considerado em separado dos outros Testes de Selecção) diferencia a categoria de Nota dos sujeitos.

Pode verificar-se que apenas a variável Nota Bruta da Prova MR ( $p < 0,016$ ) possui pelo menos um grupo com uma média significativamente diferente da média dos restantes grupos.

**Tabela 14.** Média, Desvio Padrão e N nos diferentes grupos da Prova MR

	Média	Desvio Padrão	N
Prova MR da BPRD			
Grupo 1	21,62	4,96	13
Grupo 2	21,91	4,65	88
Grupo 3	26,08	3,71	13
Total	22,35	4,75	114

Através da análise da tabela 14, referente à prova MR, verifica-se que ao longo dos três grupos (média inferior, média moderada e média superior), a média vai aumentando e o desvio padrão diminuindo. Pode-se assim concluir que o grupo 3 é significativamente diferente dos restantes.

**Tabela 15.** Função Discriminante

Function	Eigenvalue	% de Variância	% Cumulativa	Correlação Canónica
1	,088	100,0	100,0	,284

**Tabela 16.** Significância da Função Discriminante

Teste às Funções	Wilks' Lambda	Qui-Quadrado	df	Sig.
1	,920	9,311	2	,010

Pela análise da tabela 15 pode verificar-se que se pode obter assim uma função discriminante, que explica 100% da variância das diferenças entre os três grupos (de Notas Finais). A tabela 16 apresenta a significância da função discriminante (através do Lambda de Wilks), podendo concluir-se que discrimina significativamente as diferenças entre os grupos ( $p < 0,05$ ).

## DISCUSSÃO

Clark Hull (Hull, 1928, cit. por Cook, 1990), foi o primeiro psicólogo a estudar as diferenças de produtividade entre trabalhadores e a descobrir o princípio de que o melhor, é duas vezes mais produtivo que o pior. Schmidt e Hunter (1983, cit. por Cook, 1990), comprovaram este princípio e estabeleceram que os melhores 5% de trabalhadores produzem, normalmente, o dobro do que os piores 5%. Assim sendo, se uns trabalhadores produzem mais que outros, o resultado de um processo de selecção, tem forte impacto nos resultados de uma Empresa.

Mesmo que a quantificação monetária do impacto de um processo de selecção seja difícil, concluímos, da investigação efectuada, que a comunidade científica aceitou o paradigma da predição como o único ajustável à selecção de pessoal (Wolf, 1993, cit. por Ribeiro, 1996).

Foi com base neste pressuposto que este trabalho foi desenvolvido. Propusemo-nos avaliar como e em que grau, se correlaciona um conjunto de dimensões avaliadas em testes psicométricos e a avaliação de desempenho, efectuada pelas chefias, uma vez que os testes utilizados são preditores de desempenhos futuros, em população com idênticas características da amostra utilizada. Efectivamente, os testes psicométricos, especificamente os que avaliam as aptidões mentais, possuem valores de validade que oscilam entre 0,25 e 0,53. Estudos recentes (Ree & Earles, 1992; Ree, Earles & Teachout, 1994, cit. por Ribeiro, 2007), indicam o factor g como o principal preditor do desempenho profissional.

Os testes aplicados medem constructos (flexibilidade cognitiva, raciocínio e atenção), subjacentes ao exercício de um alargado número de actividades, com repercussões provadas no funcionamento intelectual, emocional e social dos sujeitos e respeitam elevados critérios de avaliação da sua qualidade (fundamentação psicométrica, amplitude de utilização, carácter prático e longevidade). Quanto à fidelidade e com excepção da prova de raciocínio mecânico, que apresenta coeficientes de consistência interna mais baixos, podendo isso decorrer de uma maior interferência do efeito aprendizagem no conteúdo dos respectivos itens, todos os restantes testes têm uma consistência interna elevada. Os estudos sobre validade indicam a existência de correlação entre o Mudanças e os restantes testes e uma validade preditiva com o desempenho que, apesar de variar nos diferentes testes, é indicativa de serem os mesmos preditores de comportamentos futuros.

Cabe a este propósito referir que, para que esta validade preditiva funcione, o conceito e operacionalização do que é entendido como desempenho é um factor essencial. Efectivamente, a medida de desempenho a utilizar, tem de reflectir da forma mais aproximada possível a associação entre o que o sujeito faz no seu dia-a-dia de trabalho e as competências a avaliar. Sem

que este pressuposto seja cumprido, dificilmente a capacidade preditiva de qualquer teste funcionará.

Não foram utilizados neste estudo, testes de aptidões sensoriais e psicomotores e de personalidade. Os primeiros, com excepção do teste de despiste de acuidade visual, que foi efectuado mas a que não tivemos acesso, não são considerados essenciais para o tipo de trabalho a executar pelos sujeitos objecto da amostra. Quanto aos testes de personalidade, não houve oportunidade para proceder à sua aplicação mas, em contraste aos testes cognitivos, a opinião predominante do seu uso em selecção de pessoal é a de que eles não têm validade, que são facilmente falsificados e inadequados para a selecção. Blinkorn e Johnson (1990, cit. por Scroggins, Thomas & Morris, 2008) concluíram que a baixa validade das medidas de personalidade e a problemática da falsificação dos resultados, torna difícil a recomendação dos mesmos como alternativa em selecção de pessoal. A sua utilização deverá ser sempre complementar.

Escolhidos os testes e recolhidos os dados, procedeu-se ao seu tratamento estatístico uma vez que qualquer análise de validação implica não só a existência de uma amostra com significância estatística, como a escolha do método a usar para efectuar a análise. É assim de extrema relevância a aferição da qualidade da informação de que dispomos.

Foi iniciado o tratamento estatístico com o apuramento de medidas de localização e de dispersão para as variáveis predictoras e critério, globalmente e divididas pelas áreas de Lisboa e Porto. Esta última análise justificou-se uma vez que as avaliações de desempenho dos sujeitos da amostra foram efectuadas pelos seus orientadores de estágio, nas suas áreas de actividade. Constatou-se que a média obtida nos testes é similar nas duas zonas, com resultados ligeiramente mais elevados em Lisboa, com excepção da Percentagem de Erros do d2 e da Prova MR do BPRD, tendo a nota final da avaliação de desempenho um comportamento inverso. Na análise teste a teste, e da comparação com os valores do grupo normativo de referência, importa assinalar, no Mudanças resultados médios ligeiramente superiores e nas provas do BPRD, a existência de valores médios inferiores.

Quanto à validade, verificou-se que todos cumprem os requisitos exigidos, necessários para continuar a análise: KMO superior a 0,5 o que permitiu avançar com a análise das componentes principais e nível de significância de Bartlett menor que 0,05 que nos diz serem as variáveis correlacionáveis. Há a destacar a opção tomada para o teste do d2 de aplicar a Matriz de duas componentes rodadas, uma vez que este era o único teste que media várias variáveis, sendo necessário encontrar as que possuíam menor grau de independência. Para os restantes testes, a análise factorial e o Scree Plot indicaram um número de dimensões diferente, tendo-se optado

pelo resultado deste último, com ajustes no Mudanças e na Prova SR, uma vez que os declives das rectas conduziam a esta opção.

O número de factores seleccionados quer na prova SR, quer no Mudanças, apenas um, não respeitou os resultados de qualquer um dos critérios aplicados (o Scree Plot identificava dois factores em ambos os testes), por motivos que se prenderam com o conhecimento do conteúdo das provas e que conduziram a esta opção.

Quanto ao Mudanças e tal como vem referido no respectivo manual de aplicação, o primeiro vector agrupa as quatro variáveis primárias de inteligência. As páginas dois e três do teste apresentam elevadas saturações deste factor. A segunda dimensão reflecte os traços aptitudinais necessários para realizar com êxito os itens da primeira página do teste e alguns da segunda, acompanhando-se de um peso elevado de aptidão espacial. Este segundo factor tem um peso menor do que as aptidões necessárias para responder às outras duas páginas do Mudanças e para obter uma boa pontuação. Refira-se a propósito os resultados obtidos na análise factorial deste teste a um factor, que nos indica, que do item 19 até ao item 27, todos os valores são negativos – trata-se de uma prova com tempo limite e com índice de dificuldade crescente, o que vai ao encontro de estudos feitos que nos indicam que, quanto à percentagem de respostas certas, o valor médio do item 19 ao 27 é de 16,2% enquanto nos dois primeiros blocos (do item 1 ao 9 e do item 10 ao 18) é de respectivamente 81,2% e 61,4%.

A opção de análise a um factor no caso da prova SR, apesar desta prova avaliar duas componentes associadas ao factor espacial – capacidade de reconhecimento ou de visualização dos elementos figurativos e a capacidade de rotação ou de acompanhar os movimentos das figuras – resultou do entendimento de não existir interesse para o objectivo deste trabalho trabalhar a dois factores, uma vez que numa tentativa exploratória de correlação com a variável critério, os resultados foram idênticos.

O último passo seguido antes de se efectuar os estudos de correlação e regressão, respeitou a análise de sensibilidade e de consistência interna. Quanto à sensibilidade, importa referir que, relativamente ao teste d2 e após se testar a normalidade de cada um dos seus parâmetros, concluiu-se que apenas a Percentagem de Erros não apresenta distribuição normal. Através dos resultados da correlação de Pearson, optou-se por trabalhar nas análises posteriores, com o Total de Eficácia e o Índice de Variabilidade uma vez que para além de explicarem factores independentes, são, respectivamente, medidas de desempenho global e estabilidade e consistência do desempenho ao longo da tarefa. Relativamente à sensibilidade nos restantes testes, todos apresentaram distribuição normal, concluindo-se que os testes utilizados são sensíveis.

No que respeita à análise de Consistência Interna, a mesma não foi efectuada para o Teste d2 por este ser uma prova de rapidez, mas relativamente às restantes provas, com excepção da AR, os resultados do Alpha de Cronbach, são satisfatórios possibilitando afirmar-se que os testes possuem consistência interna, não havendo necessidade de retirar itens.

Iniciou-se a análise de correlações pelas variáveis preditoras, sendo de destacar a correlação positiva do Mudanças com os restantes testes, com excepção do Índice de Variabilidade do d2. Comprova-se, tal como é referido nos estudos que constam no manual desta prova, que o desempenho está mais relacionado com componentes de tipo Raciocínio e Espacial (as correlações mais elevadas são com as provas do BPRD). É por isso uma boa prova a aplicar mesmo que isoladamente, a populações com as características da amostra.

Quanto às dimensões avaliadas pela globalidade dos preditores, foram identificadas duas: a primeira composta pelos parâmetros seleccionados do d2 e a segunda, por todos os restantes testes. Idêntica análise foi efectuada à variável critério, com o objectivo de perceber qual das suas componentes explica melhor a nota final, tendo-se concluído ser a nota operacional do 2º semestre, o que reflecte o peso com que a mesma entra no apuramento da nota final de desempenho.

A correlação final efectuada entre preditores e variáveis critério revelou não existir entre elas um grau de associação estatisticamente significativo. Atendendo a que a correlação estuda o grau de associação entre as variáveis, cuja relação pode ou não ser de dependência funcional e a regressão estuda a relação funcional entre as mesmas, efectuou-se esta última análise testados que foram os pressupostos a cumprir para a mesma se poder efectuar, tendo o resultado final indicado que os resultados dos testes explicam 0,1% da variabilidade da nota final ou seja, uma percentagem sem qualquer significância.

Através da análise discriminante obtiveram-se resultados semelhantes aos anteriores, confirmando que nenhum teste diferencia a categoria da nota final dos sujeito, com excepção da Prova MR da BPRD, que revelou ser um bom preditor dos sujeitos com resultados mais elevados. Atendendo a que a literatura (Robertson & Smith, 1988; Hunter & Hunter, 1984; Ribeiro *et al.* 1992) assinala o facto de que as variáveis preditoras resultantes da combinação simples ou ponderada predizem melhor comportamentos futuros, foi ainda feita uma última tentativa de correlacionar os resultados da variável critério com uma nota global de todos os testes, apurada por um simples somatório das notas parciais de cada um, apontando os resultados uma vez mais para a ausência da correlação.

Este resultado, pelos motivos que vamos referir, não permite de forma alguma concluir pela baixa capacidade preditora dos testes, apenas nos permite concluir que, tal como em

qualquer outra investigação científica, os pressupostos com que trabalhámos influenciam, de forma determinante os resultados obtidos e têm de ser melhorados.

Cabe, em primeiro lugar, referir que, a avaliação de desempenho efectuada não reflecte o real desempenho dos sujeitos, se considerarmos como desempenho a aplicação na prática dos conhecimentos possuídos, com tradução em efectivas realizações mensuráveis. Efectivamente os resultados desta variável, referem-se apenas a um curto período de tempo (12 meses), os indivíduos estiveram, boa parte do mesmo, em formação teórica e o restante em formação prática com o seu orientador, os dois indicadores quantitativos da nota final têm uma escala de avaliação qualitativa e as competências avaliadas não têm comportamentos tipo associados, o que torna a avaliação bastante subjectiva. Acresce a isto o facto da escala de avaliação ser de 1 a 6, tendo a maioria da população ficado posicionada no valor mediano da distribuição, uma vez que as chefias não usaram as notas dos extremos da escala.

No ensaio efectuado com a nota global dos testes foi apurado que, abaixo da nota média onde se situa 50% da população, o valor médio da avaliação de desempenho foi de 3,39 sendo de 3,55 para os sujeitos com um valor médio dos testes acima da média. O valor médio da avaliação de desempenho é de 3,47. Esta concentração das notas, pode dever-se à conjugação de vários factores: o facto da amostra ser já resultado de um processo de selecção; a necessidade de reforço de efectivos que as áreas envolvidas têm, sendo a nota da avaliação de desempenho o indicador chave para a contratação dos sujeitos; e por fim a grande subjectividade da própria avaliação, dada a inexistência de comportamentos tipo associados à escala. Este último factor ajuda a que o segundo factor recebido ganhe ainda mais peso.

Efectivamente, as escalas gráficas ou de atributos têm como vantagem a sua facilidade de utilização e os baixos custos envolvidos, se comparadas com outros métodos, mas têm como grandes desvantagens e riscos, a interpretação de cada atributo poder variar consoante o avaliador e o facto de alguns critérios poderem ser ambíguos ou estarem mal definidos (Gomes *et al.* 2008).

Será assim, de grande importância, melhorar a construção deste método de avaliação em dois sentidos: quantificando os indicadores quantitativos (definir um valor objectivo e apurar um realizado) e, caso se pretenda manter uma escala de atributos, especificar os comportamentos associados à escala. Outra sugestão a fazer e com possibilidade clara de obtenção de resultados de avaliação mais consentâneos com o real desempenho dos sujeitos, será refazer a ficha de avaliação, seleccionando os comportamentos a avaliar através do método dos Incidentes Críticos, proposto pela primeira vez por Flanagan (1954). Efectivamente e tal como referido por Gomes *et al.* (2008) “em vez de definir escalas gráficas em termos de critérios gerais padronizados e ou descontextualizados obtêm-se incidentes críticos que permitem identificar os aspectos do

trabalho sobre os quais devem incidir a avaliação". Anastasi (1986), refere, também, a este propósito, a importância da identificação de constructos suportados numa análise de funções, quer para a escolha dos testes a seleccionar, quer para a avaliação de desempenho, quando o objectivo é investigar a capacidade preditiva dos testes. A opção por este método exige bastante tempo da parte do avaliador, uma vez que implica que, durante o período de avaliação, sejam anotados para cada uma das competências previamente seleccionadas, os comportamentos que indicam desempenhos muito bons ou muito fracos, associados à função em causa, devendo esta questão ser referida porque a escolha do critério tem de ser consentânea com o objectivo da própria avaliação.

Refira-se ainda, a propósito desta questão da conceptualização da variável critério, que a investigação continua a orientar-se para a procura de um modelo de avaliação de performance, que abandone esta ideia de um constructo de medição suportado apenas em variáveis relacionadas com a função, avançando para um modelo diferenciado, que abarque outras perspectivas de análise (Sackett & Lievens, 2008).

Rotundo e Sackett (2002, cit. por Sackett & Lievens, 2008) apontam a necessidade de medir, para além da tradicional performance associada ao trabalho, mais duas outras dimensões, a performance relativa à cidadania (comportamentos que contribuam para o ambiente social e psicológico da Organização) e os comportamentos contraproducentes ao trabalho (comportamentos que prejudicam a Organização e os seus membros). Nesta perspectiva, a introdução no processo de selecção de variáveis predictoras de cariz não cognitivo, seria indispensável e de bastante utilidade. Tal como referem Chamorro, Furnham e Ackerman, (2006), no artigo "Incremental Validity of the Typical Intellectual Engagement Scale As Predictors of Different Academic Performance Measures", enquanto os testes de capacidade cognitiva são um importante indicador do que um sujeito pode fazer (performance máxima), a avaliação de factores não cognitivos fornece importante informação do que um sujeito fará (performance normal).

No que respeita às variáveis predictoras, importa referir alguns factores que podem também ter contribuído para que os resultados obtidos não reflectam, eventualmente, as reais capacidades dos sujeitos. Os testes foram passados no final de uma sessão de formação que durou cerca de duas horas, as sessões tinham cerca de 60 sujeitos em sala, com pouco espaço entre eles e os sujeitos foram alertados para o facto do resultado dos testes não ter qualquer influência no processo final de selecção, o que, do ponto de vista do empenhamento e interesse é bastante decisivo.

Acresce a estas questões relativas à fragilidade real existente nas variáveis critério e predictoras, uma outra que se prende também com restrições aos coeficientes de validade, já abordada superficialmente, mas que importa reforçar com a literatura e que se prende com o facto da amostragem se centrar num conjunto de sujeitos já pré-seleccionado. Neste caso, os resultados de qualquer avaliação tenderão a ser bastante homogéneos, originando baixos coeficientes de validade (Ribeiro, 2007).

Importa por fim referir que, a questão dos futuros níveis de rendimento dos sujeitos (na prática é disso que se trata), objecto deste estudo, está também fortemente relacionada com outro tipo de variáveis, que não foi possível considerar, nomeadamente a relação do sujeito com a cultura da organização, quer no que respeita às características da equipa onde esta inserido, quer da própria empresa (Ribeiro, 1996). Prieto (1993, cit. por Ribeiro, 1996), refere a este propósito, a existência de cinco factores que permitem ao recém admitido contribuir para uma maior proficiência da equipa: (1) saber ganhar a aceitação do grupo; (2) aumentar a solidariedade do grupo; (3) ter conhecimento da consciência do grupo; (4) sentir-se identificado com o grupo; e (5) saber gerir as impressões que os outros formam dele. Esta avaliação de factores psicológicos resultantes da interacção grupal, poderá ser efectuada através de simulações, entrevistas colectivas, e outros métodos, e é de grande utilidade, dada a influência que tem no desempenho individual.

Seria assim, de grande interesse teórico e de elevada utilidade para a Empresa, o acompanhamento dos sujeitos objecto desta amostra, por algum período de tempo, efectuando-se sobre a mesma nova análise de desempenho e uma avaliação psicológica.

Concluo assim que os resultados obtidos não só não invalidam a necessidade de manter testes de selecção para a entrada de novos indivíduos na Organização, como reforçam a necessidade de dispor de mecanismos de avaliação de desempenho que reflectam o valor que efectivamente cada indivíduo aporta à mesma. Termino citando Cronbach:

A tarefa de validação não é confirmar um teste, seja ele teórico ou prático. Quem faz trabalho de validação deve preparar-se como quem vai fazer um debate. Estudar o tópico sobre todos os ângulos e dominar os argumentos pró e contra tão bem, que lhe permita defender ambas as perspectivas (Cronbach, 1988 in "Five Perspectives on Validity Argument").

## REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, L.S. & Martins, T. (1996). A validade de critério em provas cognitivas: Contributos para um banco de dados na população portuguesa. In *Actas da IV Conferência Internacional "Avaliação Psicológica: Formas e Contextos"* (pp.577-584). Braga: Associação dos Psicólogos Portugueses.
- Almeida, L.S. (2006). *BPRD – Bateria de Provas de Raciocínio Diferencial*. Vila Nova de Gaia: EDIPSICO.
- Anastasi, A. (1986). Envolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Anastasi, A. & Urbina, S. (2000). *Testagem Psicológica (7ª Edição)*. Porto Alegre: Artmed Editora.
- Blinkhorn, S., & Johnson, C. (1990). The insignificance of personality testing. *Nature*, 348(6303), 671-672.
- Caetano, C. & Vala, J. (2007). *Gestão de Recursos Humanos: Contextos, Processos e técnicas*. Lisboa: Rh Editora.
- Cascio, W.F. (1982). *Costing human resources: the financial impact of behavior in organisations*. Kent, Boston, M.A.
- Chamorro, T., Furnham, A. & Ackerman, P. (2006). Incremental Validity of the Typical Intellectual Engagement Scale As Predictors of Different Academic Performance Measures. *Journal of Personality Assessment*, 83(3), 261-268.
- Cowling, A. & Mailer, C. (2003). *Gerir os Recursos Humanos (2ª ed.)*. Lisboa: Publicações Dom Quixote
- Brickenkamp, R. (2007). *Teste d2. Teste de Atenção*. Lisboa: CEGOC-TEA.
- Centro de Documentação do ISPA (2003). *Guia prático para a elaboração das citações e referências Bibliográficas de acordo com a normativa da APA/2001*. Lisboa: Instituto Superior de Psicologia Aplicada.
- Cook, M. (1988). *Personnel selection and productivity*. New York: John Wiley & Sons.
- Cook, M. (1990). *Personnel selection and productivity*. New York: John Wiley & Sons.

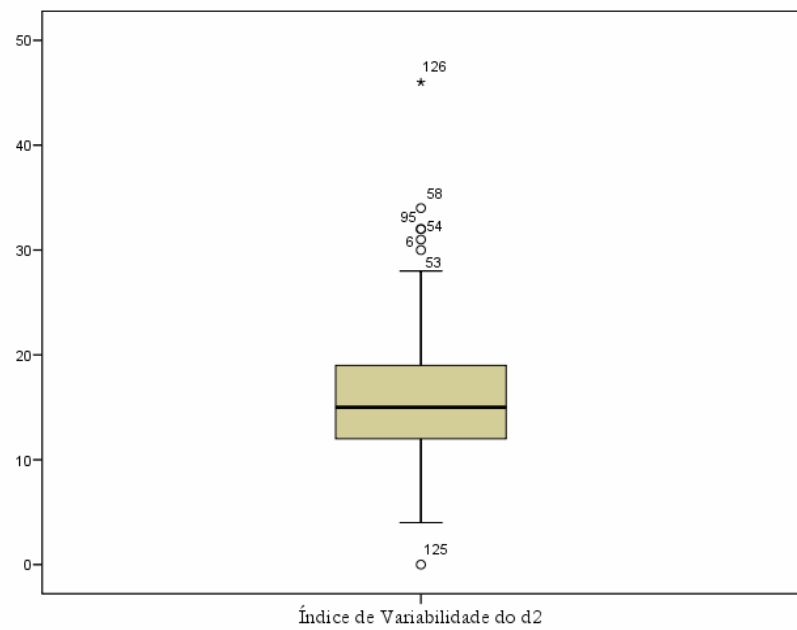
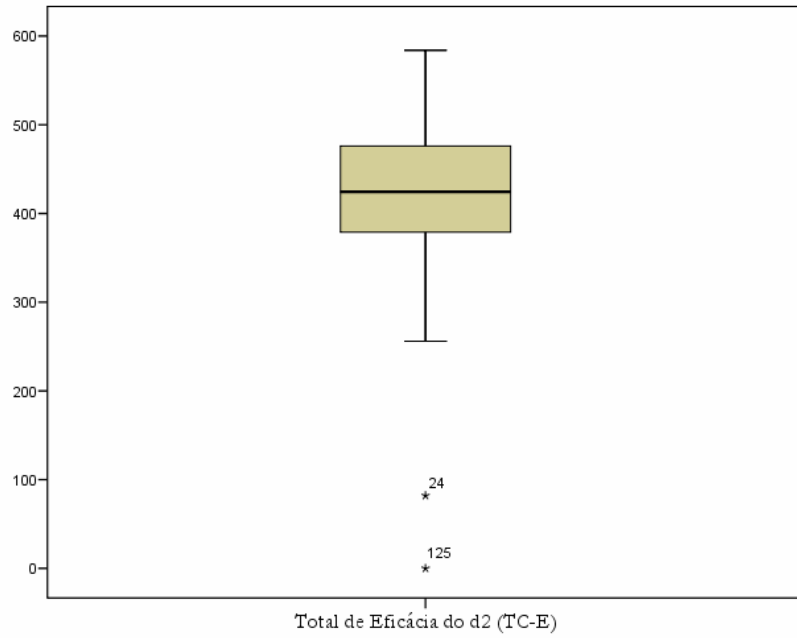
- Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer & H. Brown (Eds.). *Test Validity*. Hillsdale, Nj: Erlbaum.
- Cronbach, L. J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Faria, A., Pereira, J.V., & Almeida, L.S. (1992). A bateria de provas de raciocínio diferencial: um estudo de validade preditiva em contexto de formação. *Revista de Psicologia Militar*, N<sup>o</sup>. Especial: 303-312.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-58.
- Gomes, J.F., Cunha, M.P., Rego, A., Cunha, R.C., Cabral-Cardoso, C. & Marques, C.A. (2008). *Manual de gestão de pessoas e do capital humano* (1<sup>a</sup> Ed.). Lisboa: Edições Sílabo, Lda.
- Landy, F. J. (1993). The concept of validity and the validation process. *Journal of Business and Psychology*, 7, 369-371.
- Maroco, J. & Bispo, R. (2003). *Estatística aplicada às ciências sociais e humanas* (1<sup>a</sup> Ed.). Lisboa: Climepsi Editores.
- Moorhead, G., & Griffin, R. W. (1995). *Organizational behaviour: Managing people and organizations* (4<sup>th</sup> ed.). New York: Houghton Mifflin Company.
- Oliveira, T. (2007). Teses e Dissertações: *Recomendações para a elaboração e estruturação de trabalhos científicos*. (2<sup>a</sup> Ed.). Lisboa: RH Editores.
- Pestana, M. H. & Gageiro, J. N. (2008). *Análise de dados para ciências sociais – A complementaridade do SPSS* (5<sup>a</sup> Ed.). Lisboa: Edições Sílabo, Lda.
- Primi, R., Almeida, L.S. & Lucarelli, M.D.M. (1996). Bateria de provas de raciocínio diferencial (BPRD): Resultados numa amostra de adolescentes de são paulo. In *Actas da IV Conferência Internacional "Avaliação Psicológica: Formas e Contextos"* (pp.385-392). Braga: Associação dos Psicólogos Portugueses.
- Ribeiro, R.B. (1992). Predição de Performance em Psicologia Aeronáutica: Validação de uma Bateria de Selecção. *Análise Psicológica*, 3, 353-365.

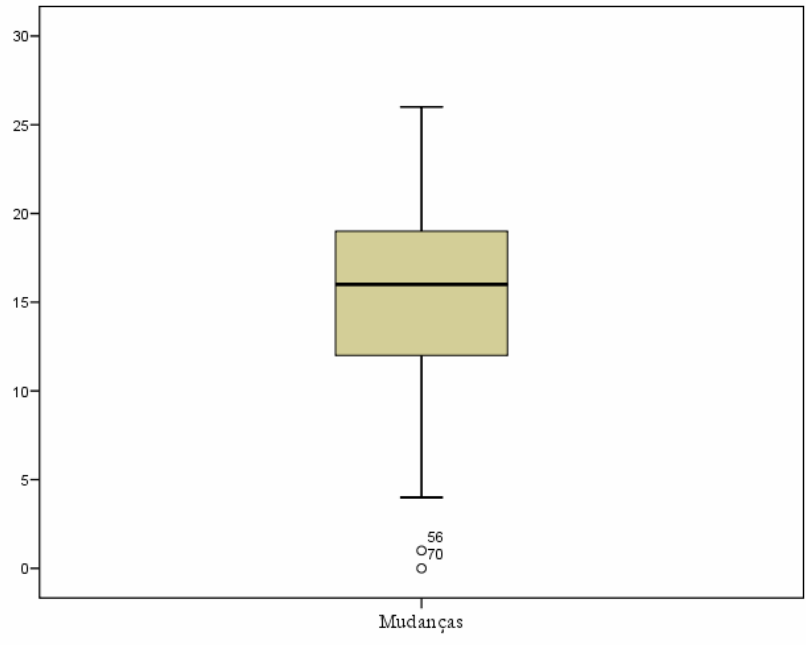
- Ribeiro, R.B., Martins, A., Viçoso, A.C., Carpinteiro, M.J. & Estrela, R.M. (1992). A validação da capacidade preditiva dos testes utilizados na selecção de pilotos. *Revista de Psicologia Militar*, N.º. Especial: 271-280.
- Ribeiro, R. B. (1996). Selecção de pessoal métodos e aplicações. In C. A. Marques & M. P. Cunha (Orgs.), *Comportamento organizacional e gestão de empresas* (pp. 93-113). Lisboa: Publicações D. Quixote.
- Ribeiro, R. B. (2007). Recrutamento e Selecção. In A. Caetano e J. Vala (Orgs). *Gestão de recursos humanos* (3ª Ed.). Lisboa: Editora RH.
- Sackett, P.R. & Lievens, F. (2008). Personnel Selection. *Annual Review of Psychology*, 2008(59), 419-450.
- Salgado, J.F., & Anderson, N. (2001). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance*, 14.
- Salgado, J.F., Viswesvaran, C. & Ones, D.S. (2002). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D.S. Ones, H.K. Sinangil and C. Viswesvaran (Eds.), *International Handbook of Industrial, Work and Organizational Psychology* Vol. 1 (pp.165-199). London: Sage.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.
- Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2008). Psychological Testing in Personnel Selection, Part I: A Century of Psychological Testing. *Public Personnel Management*, 37 (1), 99-108.
- Seisdedos, N. (2005). *Mudanças – Teste de Flexibilidade Cognitiva*. Lisboa: CEGOC-TEA.
- Thurstone, L. L. The criterion problem in personality research. *Psychometric Lab. Rep.*, No. 78. Chicago: Univer. of Chicago, 1952.
- Van De Water, T. L. (1997). Psychology's entrepreneurs and the marketing of the industrial psychology. *Journal of Applied Psychology*, 82(4), 486-499.

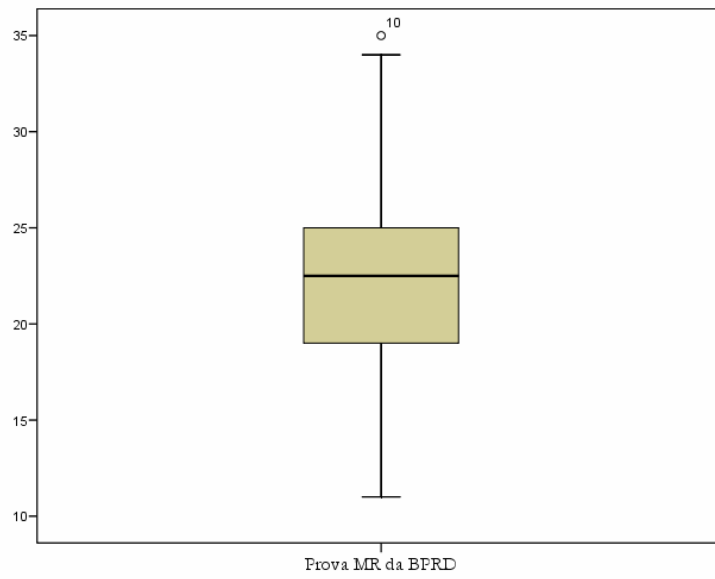
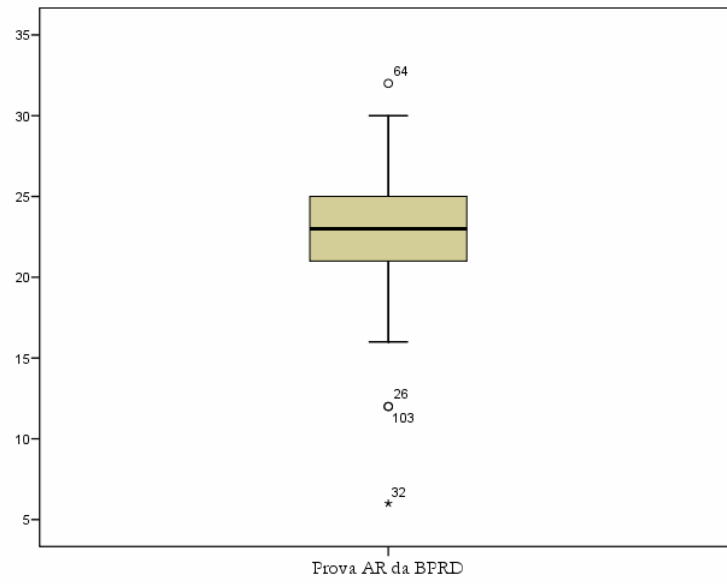
# **ANEXOS**

## Anexo A

### Outliers das Variáveis Preditoras







## Anexo B

### Estatística Descritiva dos Preditores e das Variáveis de Critério por Área de Trabalho

#### Case Summaries

	Área de Trabalho									
	Lisboa					Porto				
	Total					Total				
	N	Minimum	Maximum	Mean	Std. Deviation	N	Minimum	Maximum	Mean	Std. Deviation
Total Caracteres do d2	53	354	649	465,00	63,592	61	258	590	428,30	75,440
Total Acertos do d2	53	131	238	183,49	26,692	61	91	224	159,70	27,409
Total de Eficácia do d2 (TC-E)	53	348	584	449,89	59,246	61	256	555	407,39	68,040
Índice de Concentração do d2	53	131	238	183,00	26,622	61	87	223	159,38	27,692
Índice de Variabilidade do d2	53	4	27	16,55	4,967	61	7	28	14,51	4,915
%Erros do d2	53	,0	10,1	3,119	2,6166	61	,6	16,0	4,757	3,7661
Nota Bruta do Mudanças	53	4	26	16,74	5,054	61	4	24	14,80	4,905
Nota Bruta da Prova SR da BPRD	53	4	27	17,91	5,411	61	5	27	17,48	4,452
Nota Bruta da Prova AR da BPRD	53	16	30	23,38	465,00	61	16	29	22,89	3,012
Nota Bruta da Prova MR da BPRD	53	11	33	22,08	183,49	61	11	34	22,59	4,566
Nota Operacional 1º Semestre	53	,98	3,21	2,4751	449,89	61	2,00	4,63	3,1569	,65379
Nota Operacional 2º Semestre	53	1,21	4,93	3,0600	183,00	61	1,65	4,78	3,4630	,76286
Nota Teórica	53	3,68	4,51	4,1147	16,55	61	3,66	4,53	4,1803	,20829
Nota Final (NO+NT)	53	2,28	4,16	3,2845	3,119	61	2,72	4,50	3,6348	,43160

## Anexo C

### Estatística Descritiva – Análise Teste a Teste

#### Média, Mediana e Desvio-padrão do d2

		Total Caracteres do d2		Total Acertos do d2		Total de Eficácia do d2 (TC-E)		Índice de Concentração do d2		Índice de Variabilidade do d2		% Erros do d2	
		Manual	Estudo	Manual	Estudo	Manual	Estudo	Manual	Estudo	Manual	Estudo	Manual	Estudo
Idade <= 18 Total	N	161	2	161	2	161	2	161	2	161	2	161	2
	Média	415,90	419,5	154,47	163,00	392,34	398,50	152,68	162,50	18,52	14,50	4,81	4,20
	Mediana	410,00	419,5	153,00	163,00	394,00	398,50	152,00	162,50	18,00	14,50	3,80	4,20
	Desvio Padrão	80,25	160,51	29,50	39,60	68,59	136,47	28,60	40,31	7,15	0,71	3,85	4,10
19 - 25 Total	N	248	88	248	88	248	88	248	88	248	88	248	88
	Média	419,38	440,57	151,32	169,00	390,13	422,42	148,30	168,53	15,95	15,31	6,56	4,02
	Mediana	416,00	435,50	150,00	167,00	388,00	420,50	150,00	167,00	14,00	15,00	4,60	3,10
	Desvio Padrão	76,25	74,62	34,14	30,75	71,72	69,98	33,49	30,85	6,63	4,99	5,80	3,40
26 - 35 Total	N	173	24	173	24	173	24	173	24	173	24	173	24
	Média	421,86	465,08	154,09	177,88	394,09	446,88	152,22	177,71	15,46	16,08	6,68	3,90
	Mediana	412,00	451,50	151,00	174,50	385,00	436,50	149,00	174,50	15,00	15,50	4,70	2,95
	Desvio Padrão	77,61	54,02	37,45	23,52	74,74	48,37	35,24	23,35	5,53	5,39	6,77	3,36
Total	N	494	114	494	114	494	114	494	114	494	114	494	114
	Média	412,82	445,36	147,37	170,76	381,69	427,15	144,28	170,36	16,30	15,46	7,01	4,00
	Mediana	410,00	442,00	147,00	169,00	379,00	425,00	146,00	167,00	15,00	15,00	4,80	3,10
	Desvio Padrão	81,01	72,26	37,98	29,47	76,74	67,28	37,94	29,55	6,72	5,02	6,71	3,37

#### Média e Desvio-padrão do Mudanças

##### Descriptive Statistics

	Manual	Nota Bruta do Mudanças
N	184	114
Mean	13,67	15,70
Std. Deviation	5,050	5,047

#### Média e Desvio-padrão das Provas SR, AR e MR do BPRD

	Prova SR da BPRD		Prova AR da BPRD		Prova MR da BPRD	
	Manual	Estudo	Manual	Estudo	Manual	Estudo
N	612	114	612	114	612	114
Média	18,80	17,68	23,70	23,11	25,20	22,35
Desvio Padrão	4,20	4,90	3,56	3,18	4,65	4,75

## Anexo D

### Qualidades Métricas – Validade

Resultados dos Testes KMO e de Esfericidade de Bartlett para o d2

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,592
Bartlett's Test of Sphericity	Approx. Chi-Square	1835,566
	df	15
	Sig.	,000

Resultados dos Testes KMO e de Esfericidade de Bartlett para o Teste Mudanças

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,862
Bartlett's Test of Sphericity	Approx. Chi-Square	2180,487
	df	351
	Sig.	,000

Resultados dos Testes KMO e de Esfericidade de Bartlett para a Prova SR da BPRD

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,675
Bartlett's Test of Sphericity	Approx. Chi-Square	1071,259
	df	435
	Sig.	,000

Resultados dos Testes KMO e de Esfericidade de Bartlett para a Prova AR da BPRD

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,525
Bartlett's Test of Sphericity	Approx. Chi-Square	899,406
	df	561
	Sig.	,000

## Resultados dos Testes KMO e de Esfericidade de Bartlett para a Prova MR da BPRD

### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,478
Bartlett's Test of Sphericity	Approx. Chi-Square	1223,043
	df	780
	Sig.	,000

## Matriz de 2 Componentes Rodadas do Teste d2

### Rotated Component Matrix<sup>a</sup>

	Component	
	1	2
Total Caracteres do d2	,953	,238
Total Acertos do d2	,975	-,183
Total de Eficácia do d2 (TC-E)	,993	,055
Índice de Concentração do d2	,975	-,189
Índice de Variabilidade do d2	,021	,688
%Erros do d2	-,065	,865

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

## Análise Factorial do Teste d2

### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,807	63,456	63,456	3,807	63,456	63,456	3,801	63,353	63,353
2	1,344	22,400	85,856	1,344	22,400	85,856	1,350	22,503	85,856
3	,813	13,550	99,407						
4	,034	,567	99,974						
5	,001	,020	99,994						
6	,000	,006	100,000						

Extraction Method: Principal Component Analysis.

## Análise Factorial do Teste Mudanças

### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8,752	32,414	32,414	8,752	32,414	32,414	5,852	21,675	21,675
2	4,904	18,162	50,577	4,904	18,162	50,577	5,277	19,545	41,220
3	2,153	7,973	58,550	2,153	7,973	58,550	3,880	14,369	55,588
4	1,488	5,511	64,061	1,488	5,511	64,061	2,233	8,271	63,860
5	1,157	4,286	68,347	1,157	4,286	68,347	1,212	4,488	68,347
6	,979	3,627	71,974						
7	,857	3,173	75,147						
8	,717	2,655	77,802						
9	,649	2,405	80,207						
10	,639	2,367	82,574						
11	,583	2,159	84,733						
12	,512	1,895	86,628						
13	,441	1,635	88,262						
14	,419	1,554	89,816						
15	,399	1,478	91,295						
16	,351	1,299	92,594						
17	,322	1,193	93,787						
18	,282	1,045	94,832						
19	,246	,911	95,743						
20	,221	,818	96,561						
21	,194	,720	97,281						
22	,175	,647	97,928						
23	,152	,564	98,492						
24	,137	,508	99,000						
25	,112	,416	99,416						
26	,106	,392	99,808						
27	,052	,192	100,000						

Extraction Method: Principal Component Analysis.

## Análise Factorial da Prova SR

### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,989	16,629	16,629	4,989	16,629	16,629	3,509	11,696	11,696
2	3,405	11,351	27,980	3,405	11,351	27,980	2,224	7,413	19,109
3	2,017	6,725	34,705	2,017	6,725	34,705	2,069	6,898	26,006
4	1,875	6,251	40,957	1,875	6,251	40,957	1,905	6,350	32,357
5	1,440	4,799	45,756	1,440	4,799	45,756	1,857	6,190	38,546
6	1,331	4,437	50,193	1,331	4,437	50,193	1,850	6,165	44,712
7	1,264	4,214	54,407	1,264	4,214	54,407	1,732	5,772	50,484
8	1,210	4,035	58,442	1,210	4,035	58,442	1,499	4,995	55,479
9	1,139	3,797	62,239	1,139	3,797	62,239	1,387	4,622	60,101
10	1,048	3,493	65,732	1,048	3,493	65,732	1,377	4,589	64,690
11	1,008	3,359	69,091	1,008	3,359	69,091	1,320	4,401	69,091
12	,922	3,074	72,165						
13	,870	2,898	75,063						
14	,811	2,703	77,767						
15	,721	2,405	80,171						
16	,667	2,224	82,395						
17	,629	2,097	84,492						
18	,579	1,929	86,420						
19	,534	1,780	88,200						
20	,500	1,667	89,867						
21	,482	1,605	91,472						
22	,381	1,269	92,742						
23	,355	1,182	93,924						
24	,344	1,148	95,072						
25	,331	1,102	96,173						
26	,299	,996	97,170						
27	,279	,931	98,100						
28	,234	,781	98,881						
29	,222	,740	99,621						
30	,114	,379	100,000						

Extraction Method: Principal Component Analysis.

## Análise Factorial da Prova AR

### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,504	10,305	10,305	3,504	10,305	10,305
2	2,440	7,175	17,481	2,440	7,175	17,481
3	2,234	6,572	24,053	2,234	6,572	24,053
4	2,105	6,191	30,244	2,105	6,191	30,244
5	1,896	5,575	35,819	1,896	5,575	35,819
6	1,787	5,256	41,075	1,787	5,256	41,075
7	1,558	4,582	45,658	1,558	4,582	45,658
8	1,507	4,432	50,089	1,507	4,432	50,089
9	1,354	3,983	54,072	1,354	3,983	54,072
10	1,328	3,906	57,978	1,328	3,906	57,978
11	1,254	3,687	61,665	1,254	3,687	61,665
12	1,187	3,492	65,157	1,187	3,492	65,157
13	,960	2,822	67,979			
14	,936	2,752	70,731			
15	,918	2,700	73,431			
16	,800	2,354	75,785			
17	,777	2,286	78,071			
18	,705	2,074	80,146			
19	,678	1,995	82,140			
20	,655	1,927	84,067			
21	,640	1,883	85,950			
22	,568	1,671	87,621			
23	,558	1,641	89,262			
24	,486	1,429	90,691			
25	,471	1,385	92,076			
26	,421	1,237	93,313			
27	,395	1,162	94,475			
28	,346	1,019	95,494			
29	,335	,985	96,480			
30	,320	,941	97,421			
31	,265	,780	98,201			
32	,259	,763	98,964			
33	,184	,542	99,505			
34	,168	,495	100,000			

Extraction Method: Principal Component Analysis.

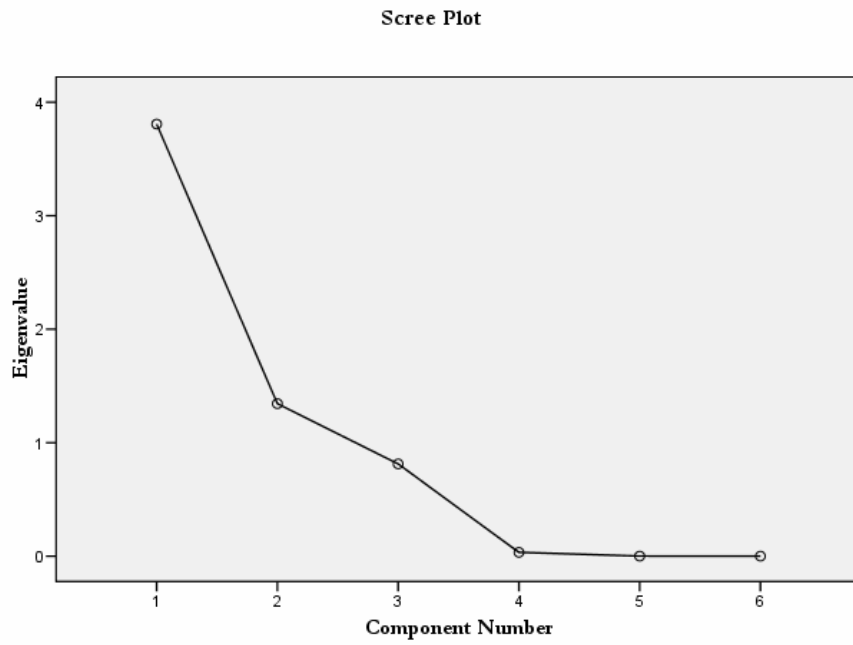
## Análise Factorial da Prova MR

### Total Variance Explained

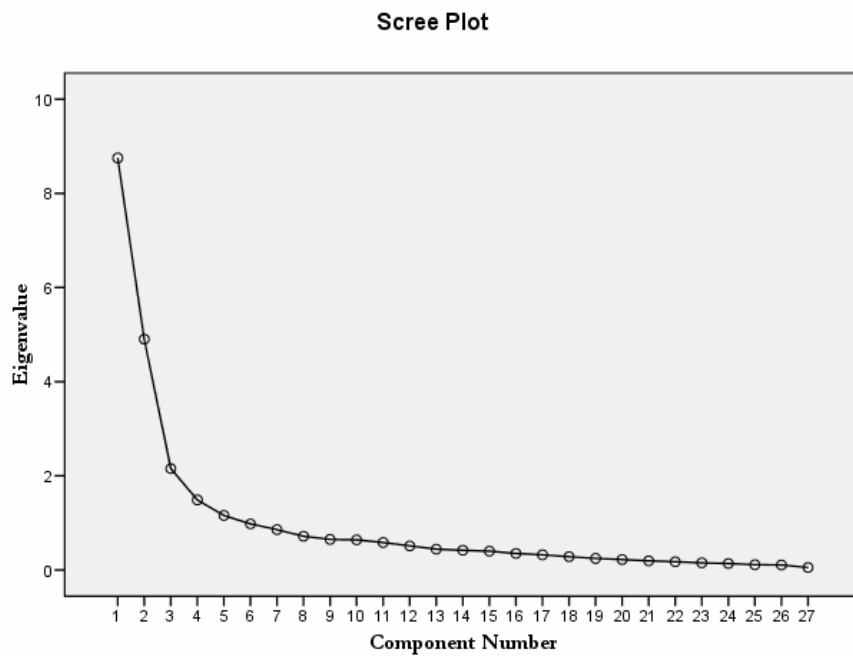
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,278	10,695	10,695	4,278	10,695	10,695
2	2,367	5,918	16,613	2,367	5,918	16,613
3	2,215	5,538	22,151	2,215	5,538	22,151
4	1,999	4,999	27,150	1,999	4,999	27,150
5	1,928	4,821	31,971	1,928	4,821	31,971
6	1,822	4,556	36,526	1,822	4,556	36,526
7	1,702	4,256	40,782	1,702	4,256	40,782
8	1,677	4,192	44,974	1,677	4,192	44,974
9	1,524	3,811	48,785	1,524	3,811	48,785
10	1,510	3,774	52,559	1,510	3,774	52,559
11	1,351	3,377	55,937	1,351	3,377	55,937
12	1,283	3,208	59,144	1,283	3,208	59,144
13	1,201	3,003	62,148	1,201	3,003	62,148
14	1,105	2,763	64,911	1,105	2,763	64,911
15	1,036	2,589	67,499	1,036	2,589	67,499
16	,970	2,424	69,923			
17	,959	2,397	72,320			
18	,915	2,287	74,608			
19	,855	2,136	76,744			
20	,801	2,003	78,747			
21	,763	1,908	80,655			
22	,743	1,857	82,512			
23	,694	1,735	84,247			
24	,671	1,678	85,925			
25	,643	1,607	87,531			
26	,588	1,469	89,001			
27	,547	1,368	90,369			
28	,524	1,309	91,678			
29	,442	1,105	92,784			
30	,422	1,054	93,838			
31	,399	,997	94,835			
32	,368	,919	95,754			
33	,319	,798	96,552			
34	,278	,694	97,246			
35	,249	,622	97,868			
36	,220	,551	98,419			
37	,184	,460	98,879			
38	,173	,433	99,313			
39	,148	,370	99,683			
40	,127	,317	100,000			

Extraction Method: Principal Component Analysis.

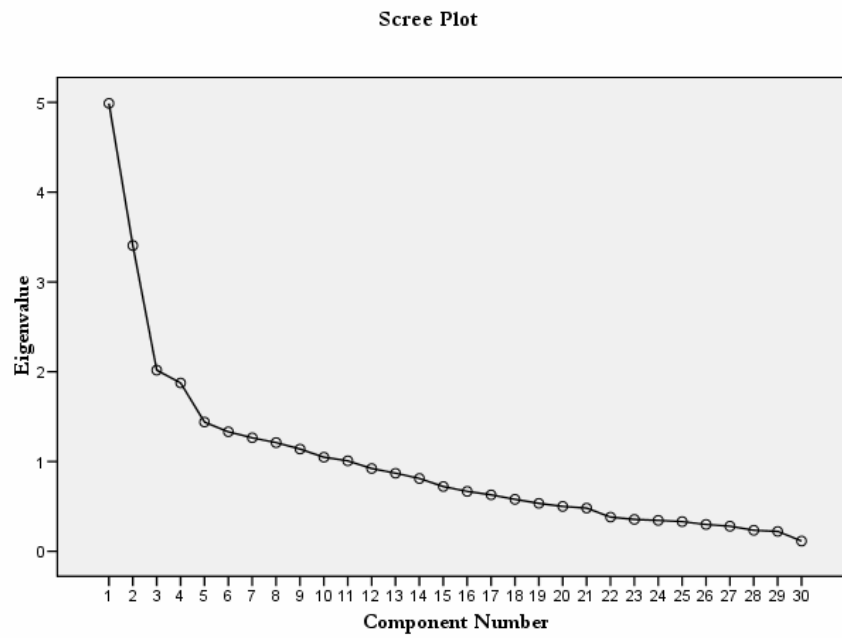
### Scree Plot para a Análise Factorial do Teste d2



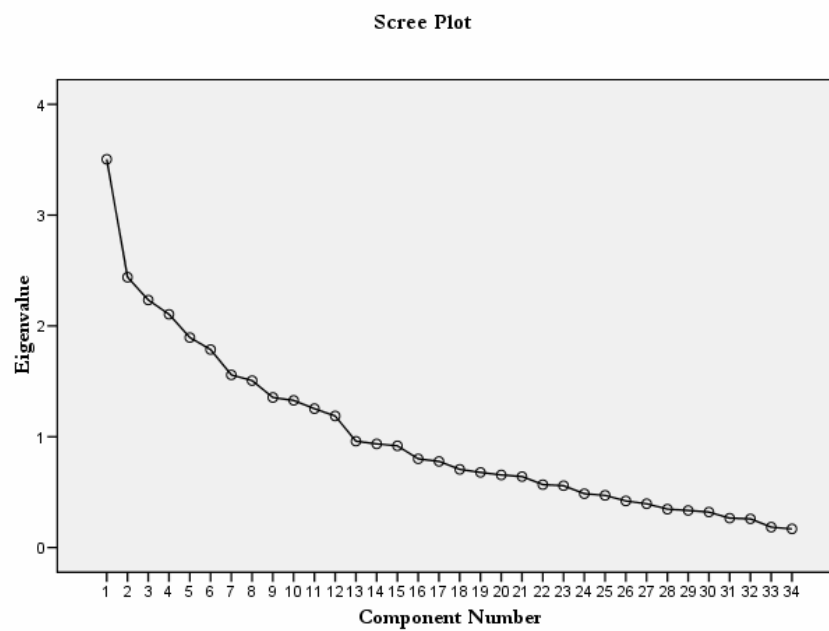
### Scree Plot para a Análise Factorial do Teste Mudanças



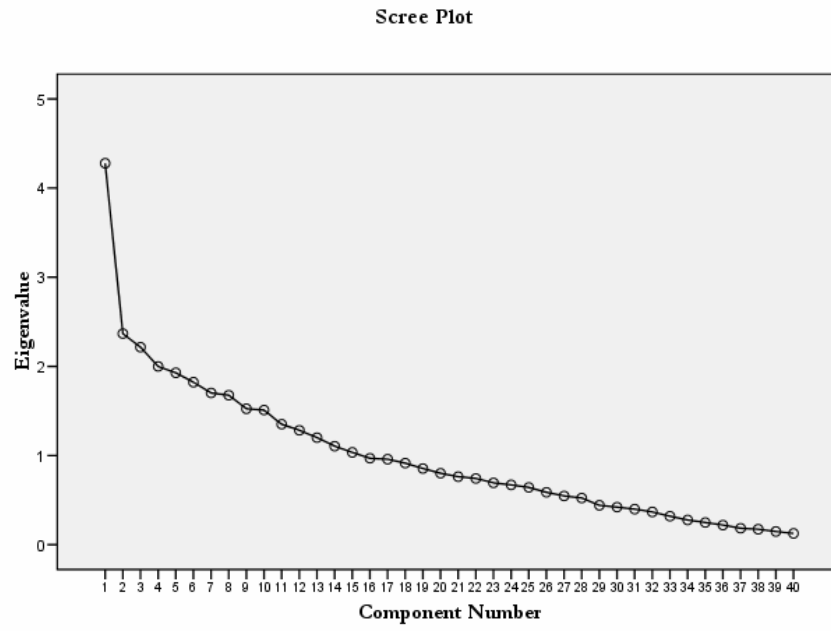
## Scree Plot para a Análise Factorial da Prova SR



## Scree Plot para a Análise Factorial da Prova AR



## Scree Plot para a Análise Factorial da Prova MR



## Análise Factorial do Teste Mudanças a 1 Factor

**Component Matrix<sup>a</sup>**

	Component
	1
Item_1	,150
Item_2	,215
Item_3	,416
Item_4	,412
Item_5	,486
Item_6	,528
Item_7	,450
Item_8	,645
Item_9	,654
Item_10	,769
Item_11	,633
Item_12	,749
Item_13	,778
Item_14	,760
Item_15	,635
Item_16	,695
Item_17	,528
Item_18	,508
Item_19	-,095
Item_20	-,254
Item_21	-,276
Item_22	-,422
Item_23	-,704
Item_24	-,575
Item_25	-,729
Item_26	-,728
Item_27	-,637

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

## Análise Factorial da Prova SR a 1 Factor

Component Matrix<sup>a</sup>

	Component
	1
Item 1 Prova SR	,424
Item 2 Prova SR	,110
Item 3 Prova SR	,286
Item 4 Prova SR	,289
Item 5 Prova SR	,387
Item 6 Prova SR	,536
Item 7 Prova SR	,503
Item 8 Prova SR	,477
Item 9 Prova SR	,470
Item 10 Prova SR	,381
Item 11 Prova SR	,457
Item 12 Prova SR	,502
Item 13 Prova SR	,379
Item 14 Prova SR	,424
Item 15 Prova SR	,232
Item 16 Prova SR	,493
Item 17 Prova SR	,595
Item 18 Prova SR	,412
Item 19 Prova SR	,331
Item 20 Prova SR	,371
Item 21 Prova SR	,299
Item 22 Prova SR	,433
Item 23 Prova SR	,556
Item 24 Prova SR	,495
Item 25 Prova SR	,200
Item 26 Prova SR	,415
Item 27 Prova SR	,408
Item 28 Prova SR	,452
Item 29 Prova SR	,244
Item 30 Prova SR	-,146

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

## Análise Factorial do Teste Mudanças a 2 Factores

**Rotated Component Matrix<sup>a</sup>**

	Component	
	1	2
Item_1	,015	-,260
Item_2	,292	,062
Item_3	,470	-,030
Item_4	,410	-,122
Item_5	,499	-,119
Item_6	,619	-,003
Item_7	,306	-,360
Item_8	,650	-,176
Item_9	,613	-,254
Item_10	,734	-,276
Item_11	,652	-,150
Item_12	,724	-,253
Item_13	,806	-,176
Item_14	,806	-,143
Item_15	,768	,031
Item_16	,834	,026
Item_17	,681	,095
Item_18	,724	,202
Item_19	,325	,703
Item_20	,124	,681
Item_21	,139	,746
Item_22	,031	,848
Item_23	-,290	,867
Item_24	-,181	,798
Item_25	-,372	,782
Item_26	-,358	,804
Item_27	-,308	,712

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

## Análise Factorial da Prova AR a 1 Factor

**Component Matrix<sup>a</sup>**

	Component
	1
Item 2 Prova AR	,027
Item 3 Prova AR	,089
Item 4 Prova AR	,222
Item 5 Prova AR	,136
Item 6 Prova AR	,146
Item 7 Prova AR	,143
Item 8 Prova AR	,203
Item 9 Prova AR	,198
Item 10 Prova AR	,016
Item 11 Prova AR	,113
Item 12 Prova AR	,246
Item 13 Prova AR	,132
Item 14 Prova AR	,078
Item 15 Prova AR	,262
Item 16 Prova AR	,322
Item 17 Prova AR	-,079
Item 18 Prova AR	-,102
Item 19 Prova AR	-,119
Item 20 Prova AR	,205
Item 21 Prova AR	,217
Item 22 Prova AR	-,070
Item 23 Prova AR	,248
Item 24 Prova AR	,267
Item 25 Prova AR	,192
Item 26 Prova AR	,285
Item 27 Prova AR	,524
Item 28 Prova AR	,585
Item 29 Prova AR	,675
Item 30 Prova AR	,736
Item 31 Prova AR	,690
Item 32 Prova AR	,492
Item 33 Prova AR	,478
Item 34 Prova AR	,227
Item 35 Prova AR	,218

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

## Análise Factorial da Prova MR a 1 Factor

**Component Matrix<sup>a</sup>**

	Component
	1
Item 1 Prova MR	,050
Item 2 Prova MR	,146
Item 3 Prova MR	-,137
Item 4 Prova MR	,047
Item 5 Prova MR	-,110
Item 6 Prova MR	,132
Item 7 Prova MR	,305
Item 8 Prova MR	,169
Item 9 Prova MR	,149
Item 10 Prova MR	,253
Item 11 Prova MR	,468
Item 12 Prova MR	,219
Item 13 Prova MR	,329
Item 14 Prova MR	-,035
Item 15 Prova MR	,091
Item 16 Prova MR	,089
Item 17 Prova MR	,305
Item 18 Prova MR	,218
Item 19 Prova MR	,262
Item 20 Prova MR	,274
Item 21 Prova MR	,171
Item 22 Prova MR	,195
Item 23 Prova MR	,092
Item 24 Prova MR	,074
Item 25 Prova MR	,265
Item 26 Prova MR	-,110
Item 27 Prova MR	,332
Item 28 Prova MR	,395
Item 29 Prova MR	,532
Item 30 Prova MR	,165
Item 31 Prova MR	,505
Item 32 Prova MR	,575
Item 33 Prova MR	,381
Item 34 Prova MR	,427
Item 35 Prova MR	,612
Item 36 Prova MR	,576
Item 37 Prova MR	,597
Item 38 Prova MR	,537
Item 39 Prova MR	,387
Item 40 Prova MR	,407

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

## Anexo E

### Qualidades Métricas – Sensibilidade

#### One-Sample Kolmogorov-Smirnov para o Teste d2

One-Sample Kolmogorov-Smirnov Test

	Total Caracteres do d2	Total Acertos do d2	Total de Eficácia do d2 (TC-E)	Índice de Concentração do d2	Índice de Variabilidade do d2	%Erros do d2	
N	114	114	114	114	114	114	
Normal Parameters <sup>a,b</sup>	Mean	445,36	170,76	427,15	170,36	15,46	3,996
	Std. Deviation	72,256	29,475	67,280	29,553	5,022	3,3698
Most Extreme Differences	Absolute	,051	,058	,037	,054	,098	,169
	Positive	,051	,058	,030	,054	,098	,169
	Negative	-,046	-,047	-,037	-,044	-,044	-,121
Kolmogorov-Smirnov Z	,546	,621	,392	,577	1,042	1,807	
Asymp. Sig. (2-tailed)	,927	,835	,998	,893	,228	,003	

a. Test distribution is Normal.

b. Calculated from data.

#### Correlação dos parâmetros do Teste d2

Correlations

	Total Caracteres do d2	Total Acertos do d2	Total de Eficácia do d2 (TC-E)	Índice de Concentração do d2	Índice de Variabilidade do d2
Total Caracteres do d2	Pearson Correlation	1	,865**	,974**	,864**
	Sig. (2-tailed)		,000	,000	,000
	N	114	114	114	114
Total Acertos do d2	Pearson Correlation	,865**	1	,946**	1,000**
	Sig. (2-tailed)	,000		,000	,000
	N	114	114	114	114
Total de Eficácia do d2 (TC-E)	Pearson Correlation	,974**	,946**	1	,946**
	Sig. (2-tailed)	,000	,000		,000
	N	114	114	114	114
Índice de Concentração do d2	Pearson Correlation	,864**	1,000**	,946**	1
	Sig. (2-tailed)	,000	,000	,000	
	N	114	114	114	114
Índice de Variabilidade do d2	Pearson Correlation	,063	-,045	,022	-,051
	Sig. (2-tailed)	,508	,638	,815	,590
	N	114	114	114	114

\*\* . Correlation is significant at the 0.01 level (2-tailed).

## Estatística Descritiva do Total de Eficácia e do Índice de Variabilidade

### Descriptives

			Statistic	Std. Error
Total de Eficácia do d2 (TC-E)	Mean		427,15	6,301
	95% Confidence Interval for Mean	Lower Bound	414,67	
		Upper Bound	439,63	
	5% Trimmed Mean		427,81	
	Median		425,00	
	Variance		4526,535	
	Std. Deviation		67,280	
	Minimum		256	
	Maximum		584	
	Range		328	
	Interquartile Range		91	
	Skewness		-,138	,226
	Kurtosis		-,323	,449
	Índice de Variabilidade do d2	Mean		15,46
95% Confidence Interval for Mean		Lower Bound	14,52	
		Upper Bound	16,39	
5% Trimmed Mean			15,32	
Median			15,00	
Variance			25,224	
Std. Deviation			5,022	
Minimum			4	
Maximum			28	
Range			24	
Interquartile Range			6	
Skewness			,388	,226
Kurtosis			-,058	,449

## One-Sample Kolmogorov-Smirnov para o Mudanças e para as provas SR, AR e MR do BPRD

### One-Sample Kolmogorov-Smirnov Test

		Nota Bruta do Mudanças	Nota Bruta da Prova SR da BPRD	Nota Bruta da Prova AR da BPRD	Nota Bruta da Prova MR da BPRD
N		114	114	114	114
Normal Parameters <sup>a,b</sup>	Mean	15,70	17,68	23,11	22,35
	Std. Deviation	5,047	4,903	3,176	4,745
Most Extreme Differences	Absolute	,077	,056	,091	,111
	Positive	,045	,049	,066	,078
	Negative	-,077	-,056	-,091	-,111
Kolmogorov-Smirnov Z		,821	,599	,971	1,184
Asymp. Sig. (2-tailed)		,510	,865	,302	,121

a. Test distribution is Normal.

b. Calculated from data.

Estatística Descritiva do Mudanças e das provas SR, AR e MR do BPRD

**Descriptives**

		Statistic	Std. Error	
Nota Bruta do Mudanças	Mean	15,70	,473	
	95% Confidence Interval for Mean	Lower Bound	14,77	
		Upper Bound	16,64	
	5% Trimmed Mean	15,79		
	Median	16,00		
	Variance	25,468		
	Std. Deviation	5,047		
	Minimum	4		
	Maximum	26		
	Range	22		
	Interquartile Range	6		
	Skewness	-,327	,226	
	Kurtosis	-,196	,449	
	Nota Bruta da Prova SR da BPRD	Mean	17,68	,459
95% Confidence Interval for Mean		Lower Bound	16,77	
		Upper Bound	18,59	
5% Trimmed Mean		17,81		
Median		18,00		
Variance		24,044		
Std. Deviation		4,903		
Minimum		4		
Maximum		27		
Range		23		
Interquartile Range		7		
Skewness		-,279	,226	
Kurtosis		,046	,449	
Nota Bruta da Prova AR da BPRD		Mean	23,11	,297
	95% Confidence Interval for Mean	Lower Bound	22,52	
		Upper Bound	23,70	
	5% Trimmed Mean	23,14		
	Median	23,00		
	Variance	10,084		
	Std. Deviation	3,176		
	Minimum	16		
	Maximum	30		
	Range	14		
	Interquartile Range	4		
	Skewness	-,106	,226	
	Kurtosis	-,452	,449	
	Nota Bruta da Prova MR da BPRD	Mean	22,35	,444
95% Confidence Interval for Mean		Lower Bound	21,47	
		Upper Bound	23,23	
5% Trimmed Mean		22,36		
Median		23,00		
Variance		22,513		
Std. Deviation		4,745		
Minimum		11		
Maximum		34		
Range		23		
Interquartile Range		6		
Skewness		-,075	,226	
Kurtosis		,078	,449	

## Anexo F

### Qualidades Métricas – Fidelidade

#### Consistência Interna do Mudanças

##### Reliability Statistics

Cronbach's Alpha	N of Items
,821	27

#### Consistência Interna do Teste Mudanças se um item for retirado

##### Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item_1	15,07	25,446	-,044	,831
Item_2	14,88	24,303	,270	,818
Item_3	14,86	24,033	,362	,815
Item_4	14,81	24,370	,330	,816
Item_5	14,82	24,187	,375	,815
Item_6	15,00	22,991	,514	,808
Item_7	15,22	24,615	,121	,825
Item_8	14,90	23,539	,452	,811
Item_9	15,00	23,540	,385	,813
Item_10	14,89	23,659	,437	,812
Item_11	14,98	23,522	,398	,813
Item_12	14,96	23,281	,473	,810
Item_13	14,95	23,059	,535	,808
Item_14	14,99	22,911	,539	,807
Item_15	15,07	22,632	,564	,805
Item_16	15,06	22,359	,631	,802
Item_17	15,18	22,806	,503	,808
Item_18	15,21	22,362	,601	,803
Item_19	15,18	22,819	,500	,808
Item_20	15,25	23,643	,324	,816
Item_21	15,30	23,415	,380	,814
Item_22	15,36	23,684	,336	,815
Item_23	15,40	24,862	,086	,825
Item_24	15,46	24,498	,183	,821
Item_25	15,47	25,260	,007	,827
Item_26	15,46	25,189	,022	,827
Item_27	15,52	25,048	,069	,825

Consistência Interna da Prova SR

**Reliability Statistics**

Cronbach's Alpha	N of Items
,808	30

Consistência Interna da Prova SR se um item for retirado

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item 1 Prova SR	16,76	22,737	,346	,802
Item 2 Prova SR	16,73	23,625	,058	,810
Item 3 Prova SR	16,99	22,705	,204	,807
Item 4 Prova SR	16,81	22,938	,221	,806
Item 5 Prova SR	16,81	22,688	,294	,803
Item 6 Prova SR	16,89	21,971	,423	,798
Item 7 Prova SR	16,79	22,472	,391	,800
Item 8 Prova SR	16,80	22,485	,375	,800
Item 9 Prova SR	16,84	22,296	,381	,800
Item 10 Prova SR	16,87	22,545	,292	,803
Item 11 Prova SR	16,83	22,409	,357	,801
Item 12 Prova SR	16,89	22,006	,414	,798
Item 13 Prova SR	17,04	22,275	,289	,804
Item 14 Prova SR	16,86	22,480	,315	,802
Item 15 Prova SR	17,09	22,814	,166	,810
Item 16 Prova SR	16,94	21,898	,413	,798
Item 17 Prova SR	16,88	21,621	,522	,794
Item 18 Prova SR	16,95	22,158	,346	,801
Item 19 Prova SR	17,05	22,301	,282	,804
Item 20 Prova SR	17,12	22,192	,299	,803
Item 21 Prova SR	17,22	22,442	,247	,806
Item 22 Prova SR	17,24	21,916	,365	,800
Item 23 Prova SR	17,27	21,465	,473	,795
Item 24 Prova SR	17,42	21,997	,423	,798
Item 25 Prova SR	17,42	23,014	,163	,808
Item 26 Prova SR	17,44	22,231	,375	,800
Item 27 Prova SR	17,43	22,212	,373	,800
Item 28 Prova SR	17,46	22,179	,405	,799
Item 29 Prova SR	17,50	22,984	,215	,806
Item 30 Prova SR	17,62	24,059	-,132	,812

Consistência Interna da Prova AR

**Reliability Statistics**

Cronbach's Alpha	N of Items
,633	35

Consistência Interna da Prova AR se um item for retirado

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item 1 Prova AR	22,14	10,768	,000	,633
Item 2 Prova AR	22,17	10,713	,026	,634
Item 3 Prova AR	22,19	10,638	,051	,634
Item 4 Prova AR	22,22	10,340	,198	,624
Item 5 Prova AR	22,17	10,645	,071	,632
Item 6 Prova AR	22,17	10,676	,060	,632
Item 7 Prova AR	22,15	10,700	,093	,632
Item 8 Prova AR	22,19	10,379	,228	,624
Item 9 Prova AR	22,20	10,533	,109	,630
Item 10 Prova AR	22,17	10,664	,056	,633
Item 11 Prova AR	22,18	10,614	,080	,632
Item 12 Prova AR	22,24	10,239	,226	,622
Item 13 Prova AR	22,39	10,167	,146	,629
Item 14 Prova AR	22,34	10,189	,162	,627
Item 15 Prova AR	22,25	10,299	,183	,625
Item 16 Prova AR	22,28	9,905	,341	,611
Item 17 Prova AR	22,32	10,720	-,041	,646
Item 18 Prova AR	22,37	10,457	,049	,639
Item 19 Prova AR	22,72	10,465	,018	,646
Item 20 Prova AR	22,28	10,428	,093	,633
Item 21 Prova AR	22,28	10,391	,109	,631
Item 22 Prova AR	22,43	10,322	,081	,637
Item 23 Prova AR	22,62	9,978	,170	,628
Item 24 Prova AR	22,28	10,076	,251	,619
Item 25 Prova AR	22,51	10,011	,169	,628
Item 26 Prova AR	22,83	10,139	,141	,630
Item 27 Prova AR	22,70	9,417	,360	,604
Item 28 Prova AR	22,87	9,409	,426	,598
Item 29 Prova AR	22,79	9,261	,439	,594
Item 30 Prova AR	22,88	9,310	,472	,592
Item 31 Prova AR	23,01	9,880	,367	,609
Item 32 Prova AR	23,05	10,174	,276	,619
Item 33 Prova AR	23,07	10,291	,264	,621
Item 34 Prova AR	23,10	10,647	,069	,632
Item 35 Prova AR	23,10	10,647	,069	,632

Consistência Interna da Prova MR

**Reliability Statistics**

Cronbach's Alpha	N of Items
,703	40

Consistência Interna da Prova MR um item for retirado

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item 1 Prova MR	21,46	21,512	,030	,703
Item 2 Prova MR	21,44	21,474	,132	,702
Item 3 Prova MR	21,47	21,765	-,108	,708
Item 4 Prova MR	22,00	21,045	,067	,707
Item 5 Prova MR	21,49	21,549	-,004	,705
Item 6 Prova MR	21,54	21,187	,104	,702
Item 7 Prova MR	21,54	20,629	,296	,693
Item 8 Prova MR	21,49	21,099	,197	,698
Item 9 Prova MR	21,68	20,779	,159	,700
Item 10 Prova MR	21,85	20,364	,221	,696
Item 11 Prova MR	21,89	19,682	,375	,685
Item 12 Prova MR	21,48	21,117	,207	,698
Item 13 Prova MR	21,83	20,466	,200	,697
Item 14 Prova MR	22,04	21,476	-,026	,713
Item 15 Prova MR	21,63	21,135	,080	,704
Item 16 Prova MR	21,66	21,145	,070	,705
Item 17 Prova MR	21,74	20,284	,262	,693
Item 18 Prova MR	21,78	20,445	,214	,696
Item 19 Prova MR	21,87	20,171	,263	,693
Item 20 Prova MR	21,85	20,328	,229	,695
Item 21 Prova MR	21,62	21,067	,105	,702
Item 22 Prova MR	21,85	20,454	,200	,697
Item 23 Prova MR	21,87	20,892	,101	,704
Item 24 Prova MR	22,00	21,099	,055	,707
Item 25 Prova MR	21,79	20,332	,238	,695
Item 26 Prova MR	21,79	21,660	-,066	,715
Item 27 Prova MR	21,67	20,187	,318	,690
Item 28 Prova MR	21,87	19,829	,343	,687
Item 29 Prova MR	21,88	19,401	,442	,680
Item 30 Prova MR	22,20	20,934	,125	,702
Item 31 Prova MR	22,15	19,950	,341	,688
Item 32 Prova MR	22,00	19,820	,346	,687
Item 33 Prova MR	22,22	20,715	,194	,697
Item 34 Prova MR	22,21	20,363	,281	,692
Item 35 Prova MR	22,21	19,930	,400	,685
Item 36 Prova MR	22,31	20,433	,365	,690
Item 37 Prova MR	22,29	20,228	,407	,687
Item 38 Prova MR	22,30	20,520	,322	,692
Item 39 Prova MR	22,38	21,031	,275	,696
Item 40 Prova MR	22,37	21,045	,222	,697

## Anexo G

### Correlações

**Tabela 28.** Correlação de Pearson entre os preditores

		Correlations					
		Total de Eficácia do d2 (TC-E)	Índice de Variabilidade do d2	Nota Bruta do Mudanças	Nota Bruta da Prova SR da BPRD	Nota Bruta da Prova AR da BPRD	Nota Bruta da Prova MR da BPRD
Total de Eficácia do d2 (TC-E)	Pearson Correlation	1	,022	,343**	,255**	,239*	,172
	Sig. (2-tailed)		,815	,000	,006	,010	,067
	N	114	114	114	114	114	114
Índice de Variabilidade do d2	Pearson Correlation	,022	1	,004	-,108	-,164	-,125
	Sig. (2-tailed)	,815		,969	,252	,081	,184
	N	114	114	114	114	114	114
Nota Bruta do Mudanças	Pearson Correlation	,343**	,004	1	,506**	,450**	,345**
	Sig. (2-tailed)	,000	,969		,000	,000	,000
	N	114	114	114	114	114	114
Nota Bruta da Prova SR da BPRD	Pearson Correlation	,255**	-,108	,506**	1	,571**	,416**
	Sig. (2-tailed)	,006	,252	,000		,000	,000
	N	114	114	114	114	114	114
Nota Bruta da Prova AR da BPRD	Pearson Correlation	,239*	-,164	,450**	,571**	1	,363**
	Sig. (2-tailed)	,010	,081	,000	,000		,000
	N	114	114	114	114	114	114
Nota Bruta da Prova MR da BPRD	Pearson Correlation	,172	-,125	,345**	,416**	,363**	1
	Sig. (2-tailed)	,067	,184	,000	,000	,000	
	N	114	114	114	114	114	114

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

### Resultado dos Testes KMO e de Esfericidade de Bartlett para os Preditores

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,776
Bartlett's Test of Sphericity	Approx. Chi-Square	128,766
	df	15
	Sig.	,000

## Matriz de Componentes Principais após rotação dos Preditores

**Rotated Component Matrix<sup>a</sup>**

	Component	
	1	2
Total de Eficácia do d2 (TC-E)	,491	,459
Índice de Variabilidade do d2	-,196	,865
Nota Bruta do Mudanças	,751	,256
Nota Bruta da Prova SR da BPRD	,812	-,031
Nota Bruta da Prova AR da BPRD	,778	-,126
Nota Bruta da Prova MR da BPRD	,648	-,165

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

## Análise Factorial dos Preditores

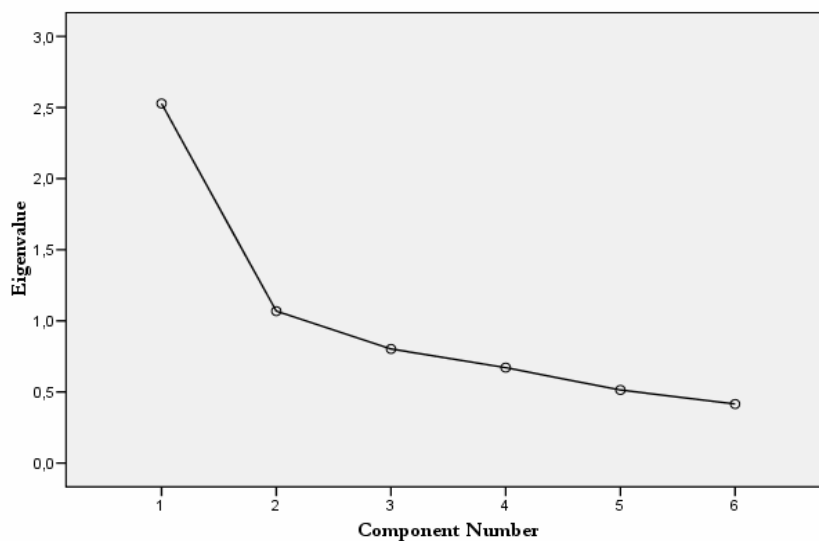
**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,528	42,140	42,140	2,528	42,140	42,140	2,528	42,136	42,136
2	1,069	17,809	59,949	1,069	17,809	59,949	1,069	17,813	59,949
3	,802	13,375	73,324						
4	,671	11,185	84,508						
5	,514	8,567	93,075						
6	,415	6,925	100,000						

Extraction Method: Principal Component Analysis.

## Scree Plot para a Análise Factorial dos Preditores

**Scree Plot**



## One-Sample Kolmogorov-Smirnov para as Variáveis Critério

**One-Sample Kolmogorov-Smirnov Test**

		Nota Operacional 1º Semestre	Nota Operacional 2º Semestre	Nota Teórica	Nota Final (NO+NT)
N		114	114	114	114
Normal Parameters <sup>a,b</sup>	Mean	2,8399	3,2756	4,1498	3,4719
	Std. Deviation	,65045	,71637	,20636	,42519
Most Extreme Differences	Absolute	,127	,086	,061	,047
	Positive	,127	,055	,044	,035
	Negative	-,078	-,086	-,061	-,047
Kolmogorov-Smirnov Z		1,357	,913	,647	,506
Asymp. Sig. (2-tailed)		,050	,375	,796	,960

a. Test distribution is Normal.

b. Calculated from data.

## Correlação do Mudanças a duas componentes com as Variáveis Critério

**Correlations**

		MudancasF1	MudancasF2
Nota Operacional 1º Semestre (40%)	Pearson Correlation	-,128	,115
	Sig. (2-tailed)	,174	,221
	N	114	114
Nota Operacional 2º Semestre (60%)	Pearson Correlation	-,067	-,009
	Sig. (2-tailed)	,478	,921
	N	114	114
Nota Teórica (35%)	Pearson Correlation	,136	,071
	Sig. (2-tailed)	,149	,456
	N	114	114
Nota Final (NO+NT)	Pearson Correlation	-,077	,044
	Sig. (2-tailed)	,413	,640
	N	114	114

## Anexo H

### Regressões

#### Análise da Independência dos Erros

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,231 <sup>a</sup>	,054	,001	,42508	,054	1,010	6	107	,423	1,623

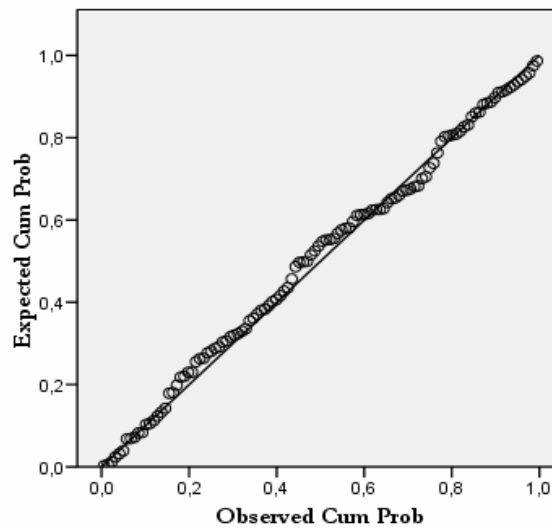
a. Predictors: (Constant), Nota Bruta da Prova MR da BPRD, Índice de Variabilidade do d2, Total de Eficácia do d2 (TC-E), Nota Bruta da Prova AR da BPRD, Nota Bruta do Mudanças, Nota Bruta da Prova SR da BPRD

b. Dependent Variable: Nota Final (NO+NT)

#### Análise de Normalidade sem Outliers

**Normal P-P Plot of Regression Standardized Residual**

Dependent Variable: Nota Final (NO+NT)



## Análise de Multicolinearidade sem Outliers

**Coefficients**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	3,664	,408		8,972	,000	2,854	4,473					
	Total de Eficácia do d2 (TC-E)	-,001	,001	-,088	-,869	,387	-,002	,001	-,089	-,084	-,082	,867	1,153
	Índice de Variabilidade do d2	-,008	,008	-,091	-,941	,349	-,024	,008	-,112	-,091	-,088	,954	1,048
	Nota Bruta do Mudanças	-,007	,010	-,087	-,743	,459	-,027	,012	-,055	-,072	-,070	,648	1,542
	Nota Bruta da Prova SR da BPRD	,004	,011	,042	,332	,741	-,018	,025	,033	,032	,031	,566	1,768
	Nota Bruta da Prova AR da BPRD	-,007	,016	-,053	-,444	,658	-,039	,025	-,006	-,043	-,042	,614	1,630
	Nota Bruta da Prova MR da BPRD	,017	,010	,190	1,782	,078	-,002	,036	,154	,170	,168	,782	1,278

a. Dependent Variable: Nota Final (NO+NT)

## Anexo I

### Análise Discriminante

#### Distribuição percentílica da Nota Final

##### Statistics

Nota Final (NO+NT)		
N	Valid	114
	Missing	0
Percentiles	25	3,220
	50	3,475
	75	3,723

## Distribuição percentil da Nota Final

Nota Final (NO+NT)				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2,28	1	,9	,9
	2,34	1	,9	1,8
	2,39	1	,9	2,6
	2,65	1	,9	3,5
	2,72	1	,9	4,4
	2,74	1	,9	5,3
	2,80	1	,9	6,1
	2,87	1	,9	7,0
	2,90	1	,9	7,9
	2,91	1	,9	8,8
	2,94	1	,9	9,6
	2,96	1	,9	10,5
	2,98	1	,9	11,4
	3,00	1	,9	12,3
	3,03	2	1,8	14,0
	3,05	2	1,8	15,8
	3,06	1	,9	16,7
	3,10	2	1,8	18,4
	3,16	2	1,8	20,2
	3,17	1	,9	21,1
	3,19	1	,9	21,9
	3,20	1	,9	22,8
	3,21	1	,9	23,7
	3,22	3	2,6	26,3
	3,25	2	1,8	28,1
	3,26	3	2,6	30,7
	3,27	2	1,8	32,5
	3,29	1	,9	33,3
	3,30	2	1,8	35,1
	3,31	1	,9	36,0
	3,32	2	1,8	37,7
	3,34	2	1,8	39,5
	3,37	1	,9	40,4
	3,39	3	2,6	43,0
	3,41	3	2,6	45,6
	3,43	1	,9	46,5
	3,45	1	,9	47,4
	3,46	1	,9	48,2
	3,47	2	1,8	50,0
	3,48	1	,9	50,9
	3,49	2	1,8	52,6
	3,50	1	,9	53,5
	3,51	2	1,8	55,3
	3,52	2	1,8	57,0
	3,53	1	,9	57,9
	3,54	1	,9	58,8
	3,55	1	,9	59,6
	3,56	1	,9	60,5
	3,57	1	,9	61,4
	3,59	2	1,8	63,2
	3,60	1	,9	64,0
	3,62	1	,9	64,9
	3,64	1	,9	65,8
	3,65	2	1,8	67,5
	3,66	2	1,8	69,3
	3,67	2	1,8	71,1
	3,68	1	,9	71,9
	3,70	1	,9	72,8
	3,71	2	1,8	74,6
	3,72	1	,9	75,4
	3,73	1	,9	76,3
	3,77	1	,9	77,2
	3,80	2	1,8	78,9
	3,82	1	,9	79,8
	3,87	1	,9	80,7
	3,88	3	2,6	83,3
	3,89	1	,9	84,2
	3,93	1	,9	85,1
	3,96	3	2,6	87,7
	3,99	1	,9	88,6
	4,02	1	,9	89,5
	4,04	1	,9	90,4
	4,08	1	,9	91,2
	4,11	1	,9	92,1
	4,15	1	,9	93,0
	4,16	3	2,6	95,6
	4,18	2	1,8	97,4
	4,30	1	,9	98,2
	4,31	1	,9	99,1
	4,50	1	,9	100,0
Total		114	100,0	

## Dispersões existentes nos grupos

**Log Determinants**

Nota Final (NO+NT)	Rank	Log Determinant
2,00	1	3,202
3,00	1	3,073
4,00	1	2,621
Pooled within-groups	1	3,048

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

## Média, Desvio Padrão e N nos diferentes grupos de Notas Finais

**Group Statistics**

Nota Final (NO+NT)		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
2,00	Total de Eficácia do d2 (TC-E)	429,46	62,961	13	13,000
	Índice de Variabilidade do d2	17,38	5,363	13	13,000
	Nota Bruta do Mudanças	14,77	5,790	13	13,000
	Nota Bruta da Prova SR da BPRD	16,77	5,615	13	13,000
	Nota Bruta da Prova AR da BPRD	22,38	3,709	13	13,000
	Nota Bruta da Prova MR da BPRD	21,62	4,959	13	13,000
3,00	Total de Eficácia do d2 (TC-E)	428,17	69,619	88	88,000
	Índice de Variabilidade do d2	15,24	4,729	88	88,000
	Nota Bruta do Mudanças	15,95	5,108	88	88,000
	Nota Bruta da Prova SR da BPRD	17,70	4,920	88	88,000
	Nota Bruta da Prova AR da BPRD	23,22	3,049	88	88,000
	Nota Bruta da Prova MR da BPRD	21,91	4,648	88	88,000
4,00	Total de Eficácia do d2 (TC-E)	417,92	58,443	13	13,000
	Índice de Variabilidade do d2	15,00	6,481	13	13,000
	Nota Bruta do Mudanças	14,92	3,883	13	13,000
	Nota Bruta da Prova SR da BPRD	18,38	4,234	13	13,000
	Nota Bruta da Prova AR da BPRD	23,15	3,625	13	13,000
	Nota Bruta da Prova MR da BPRD	26,08	3,707	13	13,000
Total	Total de Eficácia do d2 (TC-E)	427,15	67,280	114	114,000
	Índice de Variabilidade do d2	15,46	5,022	114	114,000
	Nota Bruta do Mudanças	15,70	5,047	114	114,000
	Nota Bruta da Prova SR da BPRD	17,68	4,903	114	114,000
	Nota Bruta da Prova AR da BPRD	23,11	3,176	114	114,000
	Nota Bruta da Prova MR da BPRD	22,35	4,745	114	114,000

## Função discriminante

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,088 <sup>a</sup>	100,0	100,0	,284

a. First 1 canonical discriminant functions were used in the analysis.