

A CRÍTICA DESINFORMADA AOS TESTES DE INTELIGÊNCIA

THE UNSUPPORTED CRITICISM TO THE INTELLIGENCE TESTS

Carmen E. FLORES-MENDOZA¹
Elizabeth do NASCIMENTO
Adail Victorino CASTILHO

RESUMO

Prestes a comemorar um século da criação dos testes psicológicos e, com ela, a sofisticação da avaliação psicológica, o presente artigo pretende apresentar alguns dos pontos polêmicos que acompanham a produção e aplicação dos instrumentos psicológicos, principalmente aqueles relacionados à medição da capacidade intelectual. Discute-se que o desprestígio dos testes de inteligência no país deve-se, em primeiro lugar, ao desconhecimento dos princípios básicos sobre construção de testes, o que permite freqüentemente que se exija dos testes o que não pode ser esperado (ex. predição perfeita) e, em segundo lugar, a exacerbação de crenças políticas pretensamente científicas. Conclui-se que os testes que, inicialmente, indicaram a existência de grandes diferenças intelectuais entre grupos humanos conforme a etnia, faixa etária e gênero sexual são os mesmos que atualmente indicam que essas diferenças estão diminuindo. Os testes continuam sendo os mesmos, porém mudaram as condições sociais. Portanto, os testes não criam diferenças intelectuais, apenas as retratam.

Palavras-chave: Testes psicológicos, medição do comportamento, crítica da medição.

ABSTRACT

Now, when psychologists can celebrate a century since the creation of psychological tests and the sophistication of psychological measurements,

⁽¹⁾ Dra. da Universidade Federal de Minas Gerais. Endereço para correspondência: Av. Antônio Carlos, 6627. Departamento de Psicologia Sala 4042 Universidade Federal de Minas Gerais - Belo Horizonte/MG
Tels. (031) 3499-6277 (Sala) (031) 3492-1077 (Res.) - E-mail: carmenflor@uol.com.br

this paper presents some polemical viewpoints that comes along with the uses and applications of psychological tests, mainly those concerned with assessment intellectual ability. It says the discre about tests, in our country, at first because of lack of acquaintance with its basic construction principles that often allows unreal expectations as for example perfect predictions and, at second, because political and ideological beliefs intended as scientific itelves. The psychological tests that years ago pointed to the existence of large intellectual differences among human groups concerning ethnical, age and gender variables are the same that, currently tell us these differences are showing diminution. Psychological tests continue to be the same, notwithstanding social conditions changed. So, psychological tests do not create intellectual differences, they just show it.

Keywords: *Psychological tests, behavior measurement, measurement criticism.*

INTRODUÇÃO

Em 2005, a área de avaliação psicológica estará comemorando um século da criação de testes desde que o psicólogo francês Alfred Binet, no início do século XX, criou o primeiro teste de avaliação da inteligência.

No decorrer dos anos, muitos testes psicológicos foram criados e hoje em dia pode-se encontrar desde aqueles que pretendem avaliar comportamentos específicos como ansiedade, crenças irracionais, liderança e outros, até aqueles que pretendem avaliar aspectos mais gerais do comportamento, como inteligência ou personalidade; também os modelos estatísticos aplicados à medição psicológica foram aperfeiçoados e diversificados. A psicometria aliou-se à psicologia cognitiva e, assim, a técnica correlacional juntou-se ao método experimental. Também, consoante com a era da informação, a medição psicológica pouco a pouco deixa de se apresentar apenas em mídia impressa e cada vez mais utiliza a mídia eletrônica (*software*).

Tal produção permitiu um acúmulo de informações valiosas a respeito do

comportamento humano, particularmente, no que se refere à inteligência humana e seus correlatos sociais. No entanto, muitas das informações produzidas, principalmente na metade do século passado, criaram certo mal-estar em alguns segmentos acadêmicos e não sem razão. Houve no passado alguns fatos, no mínimo suspeitos, de prática científica fraudulenta e crenças políticas antipopulares de alguns pioneiros da psicometria; o que contribuiu para reforçar uma atitude de descrença em relação aos testes psicológicos em geral. Lamentavelmente, a descrença generalizada propiciou o surgimento de críticas sem fundamento científico, opiniões baseadas na desinformação, atitudes emotivas e crenças irracionais, apesar dos testes terem permitido um melhor conhecimento da inteligência do que outros construtos psicológicos (Wechsler, 1971).

A atitude de repúdio aos testes agrava-se quando se veiculam em revistas acadêmicas alguns artigos que se posicionam contra os testes. Se bem fundamentados, os artigos deveriam servir de reflexão para aqueles que se preocupam com a medição psicológica,

porém o que se lê, em geral, é um conjunto de afirmações baseadas no terreno político-ideológico e não na análise do conhecimento teórico e técnico, que são as fontes da construção dos testes. Tais artigos, absorvidos por profissionais, alunos e leigos em avaliação psicológica, fazem ecoar as vozes contrárias à medição da conduta humana. Pretende-se aqui, portanto, discutir algumas dessas críticas e levantar os alcances e limites da prática de medição psicológica, especificamente no que se refere à inteligência humana.

Crítica 1: “Os testes reforçam a inferioridade dos segmentos sociais desfavorecidos e das minorias étnicas”.

O argumento é um dos preferidos da crítica aos testes. Um dos fatos históricos mais lembrados é o estudo de Henry Goddard efetuado nos Estados Unidos, no início do século passado, com imigrantes do sul e do leste da Europa. Em nota, Roazzi, Spinillo e Almeida (1991) comentam que o estudo de Goddard, realizado com 178 imigrantes, teria mostrado que 83% dos húngaros, 79% dos italianos e 87% dos russos apresentavam resultados muito baixos, o que reforçaria a crença do baixo nível intelectual dos imigrantes. Segundo os autores “*as pesquisas de Goddard tiveram um impacto significativo na implementação de leis de restrição à imigração nos anos 20*” (pág. 31). Esse mesmo episódio foi enfaticamente descrito nas obras de Kamin (1974) e Gould (1996), dois autores americanos, um psicólogo social e o outro paleontobiólogo, fartamente evocado pelos críticos dos testes psicológicos.

A respeito desse fato, Colom (2000) afirma que os sujeitos estudados por Goddard eram sujeitos pré-selecionados da Ilha de Ellis de Nova York e vistos pela comunidade como portadores de retardo mental. O objetivo de Goddard era apenas verificar a sensibilidade da escala métrica de Binet para identificar pessoas com retardo mental tanto dentro da população norte-americana quanto da dos

imigrantes. Segundo Colom (2000), em nenhum momento Goddard teria dito que 80% dos imigrantes eram retardados mentais ou que as amostras por ele estudadas eram representativas dos imigrantes que desembarcavam na Ilha. Portanto, a totalidade dos imigrantes nunca fora estudada utilizando os testes psicológicos. Tampouco o Congresso dos Estados Unidos soubera dos resultados dessas avaliações e nem Goddard fora chamado a depor sobre seus resultados. Segundo Rushton (1994), a responsabilidade de Goddard na lei de restrição à imigração de 1924 parece ser mais um “mito” do que um fato histórico.

Quando se reproduz o mito de que os testes psicológicos teriam fornecido as bases para a lei de restrição à imigração, no período entre 1910 e 1930, pouco se diz do clima político que se respirava na Universidade de Harvard quando tal crença se disseminou no mundo acadêmico na década de 70.

Um retrato dessa época foi oferecido pelo professor Edward Wilson, um pesquisador da sociobiologia, no seu artigo autobiográfico intitulado “Ciência e Ideologia”, apresentado originalmente em 1994 na reunião da Associação Nacional de Acadêmicos na Universidade de Cambridge. Ele descreve o ambiente acadêmico da Universidade de Harvard nos anos 70 e 80, no qual era freqüente assistir a calorosos debates políticos entre cientistas que se autodenominavam “esquerda” e os que eram apelidados de “direita”. Como o professor Wilson pesquisava as bases biológicas da conduta social, ele foi classificado como de “direita” pelo “*Grupo de Estudo sobre a Sociobiologia*” - um grupo presidido por Richard Lewontin, então diretor do programa no qual Wilson trabalhava. Lewontin, junto com Steven Rose e Leon Kamin, liderava o que eles mesmos chamaram de “filosofia da ciência radical”. Essa filosofia teria como objetivo fazer oposição a uma parte da ciência que, interessada em preservar uma classe dominante, não estaria permitindo a construção

de uma sociedade igualitária e socialista. Supostamente Wilson ao afirmar que alguns comportamentos sociais teriam sua base biológica, ele faria parte desse tipo de ciência. Nesse sentido, uma das medidas adotadas pelo “Grupo de Estudo sobre a Sociobiologia” foi a condenação da obra principal de Wilson (“Sociobiologia: a nova síntese”), por meio de uma carta enviada e publicada no *New York Review of Books*, em 13 de novembro de 1975. Na carta vinculava-se Wilson a teorias eugenicistas e políticas nazistas. Foi nessa carta que se afirmou que teorias que mostrassem uma justificativa genética dos privilégios existentes em uma determinada classe ou raça teriam permitido a aprovação, no início do século XX, da famosa lei de restrição à imigração nos Estados Unidos. Seguiram-se a essa carta numerosos manifestos contra a sociobiologia humana. E, como ocorrera com Arthur Jensen (citado mais à frente), as aulas de Wilson foram interrompidas, palestras foram palco de agressão física e, numa determinada conferência, um contingente policial teve que ser acionado para preservar a integridade física do professor. Passados 20 anos, 90% das obras, dentre as 200 que se seguiram à obra de Edward Wilson, aceitaram os pressupostos contidos no livro “Sociobiologia: a nova síntese” (Wilson, 1995).

Outra “prova” considerada crucial pelos críticos dos testes psicológicos é o efeito “pigmaleão”. Trata-se do efeito das expectativas dos pesquisadores sobre suas investigações. Sem querer, os pesquisadores trabalhavam inconscientemente para que os resultados de experimentos se apresentassem na direção esperada por eles. Essas evidências foram transpostas ao ensino. Se as expectativas dos professores forem boas então os alunos terão bom desempenho; se, pelo contrário, elas forem más, os alunos apresentarão baixo rendimento. É como se os professores trabalhassem inconscientemente para que a profecia se cumprisse. Chamou-se esse

fenômeno de “efeito pigmaleão” e foi apresentado pela primeira vez ao público por Rosenthal e Jacobson por meio do artigo “*Teacher’s expectancies: Determinants of pupils’ IQ gains*” (Rosenthal & Jacobson, 1966) e depois com a obra “*Pygmalion in the classroom: Teacher expectation and pupils’ intellectual development*” (Rosenthal & Jacobson, 1968).

As observações de Rosenthal e Jacobson interessavam muito, posto que, se a inteligência das crianças fosse tão maleável a ponto de as opiniões dos professores determinarem o aumento ou diminuição dos seus níveis intelectuais, então as reivindicações dos ambientalistas radicais estariam corretas. O estudo consistiu em fornecer aos professores de 18 classes, da primeira à sexta série, três classes por série, uma lista de alunos (grupo experimental) que supostamente teriam alto potencial para as atividades acadêmicas, se comparados ao restante da classe (grupo controle), segundo resultados do TOGA, um teste de inteligência não verbal, desenvolvido pela Universidade de Harvard. Meses mais tarde, novas testagens dos alunos (três ao todo) apontaram um QI mais elevado do grupo experimental do que dos alunos do grupo de controle. Os professores, portanto, tinham aumentado o QI dos alunos trabalhando segundo suas expectativas. O efeito pigmaleão teve grande repercussão na mídia e na sociedade americana, a ponto de influenciar decisões judiciais relacionadas ao ensino, como é o caso da decisão do Estado da Califórnia que proibiu o uso dos testes de inteligência para encaminhamento de crianças suspeitas de retardo mental leve. A alegação era que as situações de testagem sofriam o efeito pigmaleão.

A reedição americana de 1992 do “*Pygmalion in the Classroom*”, sem nenhuma mudança no texto ou informação sobre as fortes críticas à metodologia usada, motivou Spitz (1999) a realizar um levantamento dos aspectos frágeis do estudo de Rosenthal e

Jacobson, primeiramente observados por renomados estudiosos como Thorndike e Snow. Entre esses podem ser citados:

- O aumento de QI observado no grupo experimental referia-se apenas às séries iniciais (1 e 2), sendo insignificante na terceira e quarta série. Na quinta e sexta série, o QI do grupo controle foi ligeiramente maior que o do grupo experimental. Na escala verbal do TOGA, o grupo experimental (considerando sempre a 1ª e 2ª série) teria ganho 14,5 pontos e o grupo controle teria ganho 4,5 pontos de QI. Surpreendentemente, na escala de raciocínio do TOGA, o grupo experimental teria ganho nada menos do que 40 pontos de QI e o grupo de controle teria ganho “apenas” 27 pontos (quase dois desvios padrão). A variabilidade interescalar e a extensão de ganhos em pontos de QI evidenciavam que o TOGA não apresentava normas adequadas para aplicação em crianças muito pequenas.
- Resultados individuais mostravam casos de crianças que depois de 8 meses teriam ganhos de QI, no mínimo, suspeitos: de 133 a 202, de 61 a 106, de 88 a 128, de 60 a 97. Tais resultados significavam que crianças com pontuação típica de deficientes mentais (QI 61) teriam alcançado, em oito meses, pontuações normais (QI 106), assim como crianças de inteligência superior (QI 133) teriam alcançado pontuações típicas de genialidade (QI 202).
- Parte da amostra era composta de 63 crianças que recentemente tinham entrado na escola. O QI médio de raciocínio dessa amostra era de 58, ou seja, todas apresentavam deficiência mental. Para obter resultado tão baixo as crianças teriam que ter acertado apenas duas questões do teste de inteligência, ou seja, menos do

necessário para obter um escore por acaso.

- Houve exagerada variabilidade de QIs: uma criança de QI 17 no pré-teste obteve nas retestagens QIs de 148, 110 e 112; assim como uma outra com QI de 18 obteve nos pós-testes QIs de 44, 122 e 98. E houve uma terceira que obteve QI Verbal de 183 e depois QIs de 166, 221 e 168. Os dados caóticos ficam mais absurdos quando se sabe que as normas do TOGA iam de QI 60 a QI 160, sendo que o manual do teste recomendava não extrapolar tal amplitude. Portanto, alguns dados simplesmente parecem não ter validade.
- De 18 estudos produzidos entre 1969 e 1974, tentando replicar o estudo do efeito pigmaleão, apenas um estudo mostrou um ganho de cerca de cinco pontos de QI no grupo experimental.

Surpreende, pois, que um estudo com sérios problemas metodológicos possa ter recebido tanta atenção.

Enquanto se discutia o efeito pigmaleão, em 1969 surgia um dos artigos mais polêmicos. Trata-se do artigo produzido por Jensen intitulado “*How Much Can We Boost IQ and Scholastic Achievement?*” (Jensen, 1969), publicado pela *Harvard Educational Review*. Trata-se de um documento de 125 páginas, que foi reproduzido e comentado, parcialmente, nas principais publicações jornalísticas americanas, assim como na televisão e no rádio. A repercussão do artigo provocou uma onda de protestos que iam desde a agressão verbal até a ameaça física; pedia-se inclusive a cabeça de Jensen. Devido a essas ameaças, tanto na escola quanto na sua residência, Jensen (1998b) narrou a necessidade, naquela época, de estar revisando correspondências suspeitas de conterem explosivos. Dois anos mais tarde, em 1971, o psicólogo alemão, naturalizado inglês, Hans Eysenck publicou

“*Race, Intelligence, and Education*”, no qual reforçava as afirmações de Jensen. Em visita à *London School of Economics* em 1973, Eysenck receberia um ataque físico de estudantes maoístas (Rushton, 1998). Anos mais tarde, a polêmica ainda continuava e, em 1977, um contingente de 50 policiais teve que resguardar Jensen de uma multidão irada que tinha ido à Universidade de Melbourne para agredi-lo. Que informações continha o artigo que provocou tamanha reação social? Nas palavras do próprio Jensen (1998b):

“...foram as poucas páginas sobre diferenças de QI e alcance acadêmico entre as raças (cerca de 5% do artigo) que causaram tanto barulho e fúria... . Eis o que escrevi a respeito das diferenças de QI entre brancos e negros: ‘a preponderância da evidência é, na minha opinião, menos consistente com a hipótese estritamente ambiental do que com a hipótese genética, a qual por certo não exclui a influência do ambiente ou de sua interação com os fatores genéticos’...” (pág. 197).

Durante anos Jensen tentou explicar que o que tinha afirmado na época era a necessidade de verificar uma hipótese, no caso, a influência genética nas diferenças individuais e não que as pessoas da raça negra nasciam intelectualmente inferiores. Jensen (1998b) afirmou que seria tolice transpor as evidências genéticas observadas nas diferenças intelectuais **dentro** de uma raça para explicar as diferenças **entre** as raças.

A investigação das causas que determinam as diferenças individuais de uma determinada população no fator *g* considera a variância genética e a variância ambiental. Porém, a variância ambiental pode advir de duas fontes: o ambiente compartilhado (ambiente que compartilham os membros de uma família) e o ambiente não compartilhado (ambientes específicos, por exemplo, escola, nos quais interagem pessoas de outras famílias).

O estudo de gêmeos univitelinos criados dentro e fora de seus lares de origem permite estimar a influência do ambiente compartilhado. Convencionou-se que o grau de hereditariedade fosse representado pelo símbolo h^2 . Um h^2 igual ou próximo de zero significa que a contribuição dos fatores genéticos é quase nula. Por outro lado, um h^2 igual ou próximo de 1,00 significa que a contribuição genética é quase total. Pois bem, é um consenso científico que **dentro** da população branca americana o h^2 para medidas como o QI oscila entre 0,40 e 0,80 (Neisser, 1998). Esses valores indicam que a influência genética é bastante forte (a oscilação dos valores é devida ao fator idade; em geral, quanto maior a idade maior a influência genética no QI), mas indicam também uma participação considerável do ambiente ($1-h$) nas diferenças individuais com relação à inteligência. Especificamente, quais fatores do ambiente são responsáveis pelas diferenças intelectuais na população branca? Ninguém sabe ao certo. Agora imagine realizar um estudo para investigar as diferenças intelectuais entre as raças. As diferenças de fatores ambientais entre os grupos não são apenas de natureza quantitativa, mas também qualitativa. O ambiente da raça branca não é semelhante ao da raça negra. É sabido que este último grupo tem enfrentado maiores obstáculos sociais que o primeiro. Portanto, não faz sentido pensar que as mesmas influências genéticas e ambientais responsáveis pelas diferenças intelectuais dentro de um grupo sejam as mesmas responsáveis pelas diferenças entre os grupos.

A obra “*The Bell Curve*” (1994) talvez represente a retomada das discussões em torno da medida psicológica. Seus autores, os psicólogos americanos Herrnstein e Murray (o primeiro falecido no mesmo ano da publicação), apresentaram numerosas evidências dos correlatos físicos e sociais da inteligência. A obra contém duas partes: a primeira, na qual analisam numerosos trabalhos psicométricos

e a segunda, na qual os autores apresentam suas posições sociais e políticas. Uma das conclusões foi a de que a sociedade moderna está se mobilizando em função das diferenças intelectuais e não mais do *status* ou da classe social de origem. Assim sendo, estaria em formação uma elite intelectual e uma subclasse cognitiva, sendo que a primeira, na previsão dos autores, imporá normas à segunda. Porém, a parte mais polêmica, e pela qual a obra se tornou a mais conhecida e debatida tanto nos meios acadêmicos quanto nos governamentais, foi a apresentação de dados que mostram novamente a questão das diferenças de desempenho cognitivo entre as raças branca e negra. Os críticos acadêmicos e da imprensa em geral se dividiram, uns afirmando que se trata da obra mais antidemocrática de que se teve notícia e outros afirmando que a obra merece respeito, haja vista que os autores não inventaram os dados, apenas os compilaram (Colom, 1995).

Devido à extrema polêmica que a obra suscitou na sociedade americana, graças à sua difusão nos meios de comunicação, o *Wall Street Journal*, em 13 de dezembro de 1994, resolveu publicar um relatório de 25 pontos, assinado por 52 renomados cientistas e pesquisadores da área da inteligência. Pouco tempo depois, a *American Psychological Association-APA* tomou a iniciativa de publicar um informe mais detalhado a respeito do que a psicologia, enquanto ciência, conhece sobre o construto chamado inteligência. Um grupo de seletos investigadores da área da cognição humana, como Sternberg, Brody, Bouchard, Ceci, Loehlin e outros, redigiram e assinaram um relatório, em nome da *APA*, que foi publicado na revista *American Psychologist* (Neisser et al., 1996). O informe é bastante prudente nas suas conclusões, mas nada diferente do que já foi apresentado até o momento. No que diz respeito às diferenças intelectuais, considerando os diversos grupos étnicos, diz-se que, de fato, há diferenças de QIs entre os grupos étnicos, porém no que se

refere a brancos e negros americanos, as diferenças estão diminuindo. As diferenças não se devem a um possível viés dos testes contra determinados grupos étnicos e tampouco refletem diferenças socioeconômicas. Na realidade, se desconhece a causa dessas diferenças.

Neisser (1998) organizou uma obra intitulada "*The Raising Curve-Long-Term Gains in IQ and Related Measures*", na qual escrevem reconhecidos pesquisadores como James Flynn, Richard Lynn, Stephen Ceci. A obra discute as duas boas notícias que estão sendo comentadas hoje em dia: 1) no mundo industrializado os escores nos testes de inteligência estão aumentando e não caindo e 2) a disparidade de alcance acadêmico entre crianças negras e brancas tem diminuído nos últimos anos. Essas notícias têm importante repercussão social e se contrapõem à dose pessimista de Herrnstein e Murray (1994).

A discussão sobre o aumento do QI no mundo teve início a partir da meta-análise de estudos provenientes de 14 nações, realizada por Flynn (1987). O estudo apontava um ganho, em apenas uma geração, entre 5 e 25 pontos de QI em testes de inteligência fluida (Gf). Em média, o ganho teria sido de 15 pontos, ou seja, o correspondente a um desvio padrão. Contudo, o ganho obtido não está muito claro em 10 países, mas sim em apenas quatro (Bélgica, Noruega, Nova Zelândia e Holanda) dos 14 países estudados. Pouco mais de uma década depois, Flynn (1998) continua afirmando o que dizia em 1987: dado que é improvável que os habitantes dos quatro países, onde os ganhos de QI parecem ser contundentes (ganhos de 15 pontos em média), tivessem em 1896, portanto duas gerações atrás, um QI de 70, seria pouco provável que os testes meçam exatamente inteligência.

Na Espanha, Colom, Andrés-Pueyo e Juan-Espinosa (1998) analisaram, utilizando a metodologia de Flynn, os dados brutos das duas normatizações do teste Raven (Escala Geral e Avançada), a primeira feita entre 1963

e 1968, e a segunda entre 1991 e 1996. Os resultados apontaram um ganho de 19,2 pontos de QI na Escala Geral e 6,75 pontos de QI na Escala Avançada. O ganho, portanto, parece ser maior na população heterogênea (Escala Geral) do que naquela mais homogênea (Escala Avançada). Um outro estudo realizado por Abad *et al.* (2001), com base nos dados da adaptação e padronização para a população espanhola do WISC-III, mostrou que o fator *g* explica menos a variância dos grupos mais instruídos do que dos menos instruídos.¹ Por outro lado, fatores de grupo explicam mais a variância de desempenho dos grupos com maior educação do que a variação de desempenho dos grupos menos instruídos. Essa hipótese, em realidade, já havia sido levantada por Detterman e Daniel (1989). Os autores afirmam que quanto maior for o nível intelectual da pessoa, menos relevante é a inteligência geral (fator *g*), pois a inteligência se especializa. É o caso dos resultados de Abad *et al.* (2001). A variância de desempenho dos grupos mais instruídos, que explica os fatores de grupo, foi maior. No caso dos ganhos de QI nos países industrializados, deve-se considerar que os estudos utilizaram amostras heterogêneas. Os ganhos, portanto, devem ser mesmo de inteligência geral (fator *g*). Mas como então explicar os ganhos de inteligência geral? Uma possível resposta é que, se as pessoas nos últimos 50 anos ficaram mais altas, melhor alimentadas, passaram a freqüentar em maior número as universidades e passaram a ter maior acesso à tecnologia, é de se esperar, portanto, uma melhora tanto das capacidades fluidas (*Gf*), aquelas que dependem pouco da educação formal, quanto das cristalizadas (*Gc*), aquelas que dependem em maior extensão da educação formal. Contudo, os ganhos observados foram maiores nas capacidades fluidas (*Gf*), isto é, naquelas capacidades que são mais vulneráveis a componentes biológicos (e, portanto, mais

próximas do fator *g*) do que nas capacidades relacionadas à aprendizagem.

Embora ainda não esteja claro porque os fatores ambientais (educação, estimulação) e os fatores biológicos (nutrição) afetam mais a capacidade fluida do que a cristalizada, o caso é que os instrumentos, que no passado foram tão criticados por apontar as diferenças das pessoas, são os mesmos que, atualmente, apontam para um aumento das capacidades intelectuais entre as pessoas. Trata-se, sem dúvida, de uma boa notícia.

Por outro lado, os testes podem auxiliar na verificação do efeito de políticas governamentais para redução das desigualdades sociais entre etnias, como a política americana de “ação afirmativa”. Recentemente, Nyborg e Jensen (2001) apresentaram dados sobre o fator *g* (extraídos de 19 testes cognitivos), renda econômica e status ocupacional, coletados entre 1985 e 1986, de veteranos das forças armadas americanas nos últimos 30 anos. A comparação entre pessoas negras e brancas de igual *g* revela que as primeiras apresentam maior renda econômica e posição ocupacional do que as segundas desde que o *g* se encontre acima do percentil médio.

Crítica 2: “Houve prática fraudulenta na produção psicométrica”

As crenças políticas dos pesquisadores, em qualquer área de estudo, podem, afinal de contas, ser apoiadas ou não pelos resultados de pesquisas que possuem um rigor científico. No entanto, um estudo que descreva o rigor científico utilizado não garante a probidade dos dados. Em outras palavras, não podemos ser ingênuos em acreditar que todo trabalho científico é verdadeiro ou correto. As razões variam: há estudos que escondem falhas metodológicas e há outros que não percebem a inadequação metodológica. Os primeiros

⁽¹⁾ Um fator “*g*” pode ser extraído como o primeiro fator principal não rotado de uma análise de componentes principais. O famoso estudo de Detterman & Daniel (1989) sobre a hipótese de diferenciação do fator *g* foi realizado utilizando os dados da padronização americana do WISC_R.

mostram que seus responsáveis adotaram uma atitude incorreta e não ética, enquanto os segundos mostram que seus responsáveis ainda não adquiriram maturidade científica. Contudo, a atitude mais condenável que um pesquisador pode adotar é a fraude científica. No caso das diferenças individuais, o episódio mais citado é o de Cyril Burt, um psicólogo escolar inglês de meados do século passado. Os dados mostrados por Burt, com efeito, sugerem a prática fraudulenta em alguns estudos relacionados às bases biológicas da inteligência. O descobridor foi o psicólogo social Leon Kamin.

Kamin (Eysenck & Kamin, 1982) descreveu a estranheza que provocou nele a irregularidade dos valores das correlações apresentados por Burt em estudos de gêmeos criados em separado. Segundo Burt, a correlação positiva entre QIs de gêmeos seria uma evidência de que a inteligência é fundamentalmente influenciada pela genética. Em 1943, Burt teria comunicado que a correlação entre QIs de gêmeos idênticos era de 0,77. Em 1955, quando a amostra aumentou de 15 para 21 pares, o valor da correlação era de 0,771. Em 1958, a amostra teria aumentado para 30 pares e o valor manteve-se em 0,771. No final de 1958, aumentou a amostra para 42 pares e a correlação se modificou muito pouco (0,778). Em 1966, quando a amostra era de 53 pares, a correlação era a mesma de 1955, isto é, de 0,771. Há, pois pouca dúvida de que os estudos relatados por Burt, no mínimo, eram irregulares e suspeitos. Kamin descreveu outros estudos efetuados em gêmeos criados juntos e separados e tentou mostrar diversas falhas metodológicas.

A suspeita de fraude por parte de Burt (que até então era considerado um dos melhores pesquisadores da inteligência humana) reforçou a posição da ciência radical e provavelmente obscureceu os avanços das pesquisas na área das diferenças individuais durante alguns anos.

Uma das melhores armas da ciência para mostrar equívocos (ex. falhas metodológicas) ou comportamentos suspeitos (ex. fraude) é a replicabilidade dos estudos. E surpreendentemente um deles, iniciado em 1979, apresentou um valor correlacional próximo do apregoado por Burt. Trata-se do estudo de Minnesota (Bouchard, et al., 1990). Foram reunidos mais de 100 pares e trios de gêmeos, monozigóticos e dizigóticos, criados juntos e separados. Os gêmeos eram procedentes da Inglaterra, Estados Unidos, Austrália, Canadá, China, Nova Zelândia, Suécia e Alemanha. Os participantes foram submetidos a mais de 50 horas de avaliação médica e psicológica. No que diz respeito à avaliação psicológica, foram aplicados quatro questionários de personalidade, três inventários de interesses e duas baterias de capacidade cognitiva. Os resultados do estudo de Minnesota apresentaram um valor correlacional de 0,70 entre QIs de gêmeos monozigóticos adultos. Bouchard et al. (1990) concluíram que seus resultados estão bastante próximos de outros, nos quais se obteve valores oscilando entre 0,64 e 0,74.

Embora ainda seja comprometedora a irregularidade dos dados de Burt, a semelhança entre os valores supostamente obtidos por aquele e os de Minnesota causa surpresa no meio acadêmico. No entender de vários autores (Colom, 2000; Fletcher, 1990; Rushton, 1994), o "*Burt Affair*" ainda não foi bem esclarecido.

Crítica 3: "Os testes não medem aquilo que pretendem medir"

É a expressão mais ouvida daqueles que se posicionam contrários aos testes psicológicos. Trata-se da questão da validade dos instrumentos, isto é, como podemos saber, por exemplo, se os testes de inteligência medem, de fato, a capacidade cognitiva de um indivíduo e em que medida eles o fazem? Para responder a essa questão cabe remontar à psicometria, particularmente, às técnicas da análise fatorial (validade do construto) e às

teorias sobre o construto investigado, aspectos esses freqüentemente não considerados ou desconhecidos pelos críticos dos testes psicológicos (Colom, 1998).

A análise fatorial, como outras técnicas, tem sido aperfeiçoada com o passar dos anos, permitindo o desenvolvimento de formas de análise cada vez mais poderosas. Tecnicamente, a análise fatorial se propõe verificar todas as relações possíveis entre as variáveis de uma matriz de dados e as reduz a um conjunto de dimensões chamadas “fatores” (Hair, Anderson, Tatham & Black, 1995). Assim, por exemplo, se o objetivo é descobrir as variáveis subjacentes ao sucesso de venda de um determinado produto, é muito provável que se apontem diversas delas, tais como a imagem do fabricante, o preço do produto, a quantidade, a embalagem, a tradição da marca no mercado, a durabilidade do produto, seu cheiro e a cor, as informações contidas no produto, etc. Uma análise fatorial poderá reduzir todas essas variáveis a três dimensões que podem receber o nome de qualidade (durabilidade, cheiro, textura, cor), custo (preço, quantidade do produto) e publicidade (imagem do fabricante, embalagem do produto, tradição do produto, informação do produto). Essas dimensões podem ser chamadas de “fatores” e o peso de cada uma na venda do produto determinará a atuação do fabricante (ex. reforçar a qualidade e manter a publicidade ou manter o preço e a qualidade e aumentar a publicidade).

No caso do construto chamado “inteligência” aplica-se também a análise fatorial. E, antes de aplicá-la, deve-se ter alguma clareza a respeito do construto que se quer avaliar. No caso da inteligência, esta pode ser entendida como um construto que designa uma condição abstrata do ser humano. Entre os acadêmicos existe uma diversidade de definições, dependendo da orientação teórica adotada, embora se observe certo consenso de que o ato de resolver problemas seja uma das características principais da

inteligência (Flores-Mendoza & Nascimento, 2001). Na população leiga, praticamente todos os indivíduos reconhecem, desde o menos instruído até o mais culto, que as pessoas se comportam de maneira diferenciada ao resolver problemas ou tomar decisões. Algumas pessoas são rápidas e eficientes e são chamadas de pessoas muito inteligentes; outras não são tão rápidas e nem tão eficientes e são chamadas de pessoas pouco inteligentes. O criador de um teste parte de uma concepção do que venha a ser inteligência, a fim de elaborar itens cuja solução represente a manifestação desse construto. Saber calcular, possuir informações sobre o ambiente em que reside, resolver problemas do cotidiano, realizar analogias, possuir um vocabulário extenso são todos indícios de comportamento inteligente. Quais desses comportamentos representam melhor a inteligência? Talvez nunca se chegue a um acordo. Sabemos que as atividades humanas, quaisquer que sejam, implicam sempre, em maior ou menor grau, o uso da inteligência. E se essa inteligência se mantém relativamente estável num indivíduo, enquanto resposta a determinadas situações, então a inteligência é um traço do comportamento humano que pode ser medido. A existência diferenciada nos seres humanos desse traço de comportamento constitui a base das diferenças individuais. O teste psicológico delineado para avaliar a inteligência deverá, portanto, apresentar itens que demandem o uso de habilidades cognitivas que representem o traço de inteligência, embora como diz Cronbach (1998) “*Nenhum construtor de testes pode elaborar um teste com todas as qualidades desejáveis*”(pág. 172).

Se as pontuações em testes de inteligência acompanham, em algum grau, o comportamento fisiológico e social de um indivíduo, então, que inteligência é essa que é medida pelos testes? Qual é a natureza e a estrutura da inteligência?

Responder a essas questões implica em ir além da construção de testes. Saber que um teste apresenta resultados estáveis no decorrer do tempo (precisão), que se correlaciona com um critério externo (validade de critério) e que se utilizou uma amostra representativa da população para a elaboração das normas não é suficiente para explicar o construto que o teste pretende medir; para tanto, é necessário aprofundar as discussões no terreno teórico científico e também analisar as interações entre os testes e a vida quotidiana. Realizar essa análise significa o que os psicometristas chamam de estabelecimento da validade de construto, ou seja, investigar se o teste constitui uma representação legítima do construto (Pasquali, 1997) e implica utilizar técnicas de análise fatorial e métodos interdisciplinares de investigação experimental que possam contribuir para a elucidação do construto.

A análise fatorial teve seu início no final do século XIX, quando o inglês Charles Spearman elaborou o método tetrádico. Segundo esse método, quando há correlação significativa entre dois ou mais testes, essa correlação é devida a um fator comum (ou fator “*g*”). A parte que não se correlaciona corresponderia a fatores específicos a cada teste. Portanto, a análise fatorial isola ambos os fatores (fator “*g*” e fatores específicos). Os fatores extraídos, então, são expressões estatísticas e não pressupõem entidades concretas. Mas, Spearman (1955) afirmou que o fator “*g*”, em última instância, reflete uma “energia” mental, uma capacidade geral dos indivíduos para resolver testes de natureza diversa. Posteriormente, tanto o fator “*g*” como os fatores específicos foram reunidos no “modelo dos dois fatores”.² Outros métodos fatoriais, além do tetrádico, foram surgindo na literatura psicométrica e em todos eles obteve-se um *g* significativo, desde que os estudos empregassem amostras amplas e

representativas da população, assim como também testes variados que demandassem do sujeito distintas atividades cognitivas (Jensen & Weng, 1994; Jensen, 1998a).

O que a prática psicométrica tem mostrado? As estimativas dos fatores têm mostrado, como dito anteriormente, um fator geral permeando o desempenho em todas as tarefas intelectuais; daí a importância do fator “*g*” proposto por Spearman. Outros psicometristas têm exigido a análise minuciosa dos numerosos componentes e, portanto, dos diversos fatores que seriam responsáveis pelas diferenças individuais. Dessa forma, desde o trabalho pioneiro do psicólogo inglês Spearman até pouco tempo atrás, os psicometristas se dividiram; uns afirmando a importância de um fator geral para explicar as diferenças individuais na inteligência (Eysenck, 1982; Jensen, 1987; Burt, 1955; Gustafsson, 1988; R. Thorndike, 1994) e outros afirmando a existência de numerosos fatores como importantes componentes estruturais da inteligência (E. Thorndike, 1921; Thurstone, 1921; Guilford, 1954, 1988).

Tais arranjos fatoriais dividiram-se em dois grupos: o grupo dos modelos hierárquicos (que pressupõem dimensões mais gerais na parte superior da estrutura da inteligência e dimensões de menor generalidade, frequentemente as mais numerosas, que se encontram na parte inferior da estrutura) e o grupo dos modelos não-hierárquicos (que pressupõem diversos componentes que compõem a estrutura da inteligência desconsiderando qualquer relação entre as dimensões). Até o momento, o modelo hierárquico é o que mais tem contribuído para a prática da avaliação psicológica, dado que é mais parcimonioso estimar a capacidade de um indivíduo a partir de testes de ampla generalidade do que utilizar numerosos testes específicos (Colom, 1998; Juan-Espinosa, 1997).

² Posteriormente, Spearman aceitaria o fato de que entre o fator “*g*” e o fator “*e*” haveria um certo número de fatores de grupo, mas que considerava de menor importância se comparado à importância do “*g*” (Thorndike, 1994).

Os modelos hierárquicos ganharam força também graças ao trabalho de John Carroll (1993), considerado um marco no campo das diferenças individuais. Trata-se de uma árdua investigação em que o autor reanalisou aproximadamente 460 estudos fatoriais realizados em diversos países num período de 40 anos, a maioria nos Estados Unidos e que envolveram mais de 130.000 sujeitos. A conclusão dessa reanálise foi que os fatores podem ser agrupados, de acordo com seu grau de generalidade, em três dimensões ou estratos. O primeiro estrato corresponde aos fatores lingüísticos, raciocínio, memória e aprendizagem, percepção visual, recepção auditiva, produção de idéias, velocidade cognitiva, capacidade de conhecimento e rendimento e capacidades motoras. O segundo estrato corresponde a fatores que caracterizam a inteligência fluida, inteligência cristalizada, memória geral, percepção visual e auditiva, capacidade de recuperação da informação (memória a longo prazo) e velocidade cognitiva. O terceiro estrato corresponde ao fator *g* relacionado à velocidade de processamento de informação e capacidade de memória de trabalho. Os fatores dos estratos superiores apenas resumem os fatores dos estratos inferiores. Assim, o fator *g* do terceiro estrato resume os fatores do segundo e estes últimos resumem os fatores do primeiro estrato. Dado que o estudo de Carroll representa as respostas dos sujeitos a situações definidas (testes) em diferentes épocas e em diversos ambientes culturais, a teoria dos três estratos representa, atualmente, o modelo psicométrico mais bem aceito na comunidade acadêmica; nele se confirma a superioridade do fator *g* como a dimensão mais geral subjacente a toda atividade cognitiva (Juan-Espinosa, 1997).

A obra "*The g Factor: The science of mental ability*", do professor americano Arthur Jensen (1998a), constitui a compilação de seus numerosos estudos efetuados durante quase 50 anos sobre a validade do fator *g*. O profícuo trabalho de Jensen mereceu uma edição especial da revista *Intelligence*, na

qual Detterman (1998) afirma: "...*belamente escrito eu predigo que o livro será o fundamento das pesquisas nas próximas décadas.*" (pág. 176). A obra, em resumo, mostra o fator *g*, a partir de estudos das áreas da genética, da neurologia e da psicologia da conduta humana, como a dimensão mais geral na análise hierárquica e que parece estar associado, inclusive, a variáveis não-psicométricas como tempo de reação. Dessa forma, o fator "*g*" representaria uma variável que ligaria a psicologia à biologia; uma tese compartilhada nos últimos anos de vida por R. L. Thorndike (1994).

Para elucidação da natureza e estrutura da inteligência e sua medição, a validade de um teste de inteligência é verificada utilizando-se do princípio de correlação. No caso da validade de critério, correlaciona-se o resultado obtido com o instrumento (teste) e um critério (ex. rendimento escolar). A relação entre dois fenômenos, no caso os resultados do teste e o desempenho em outra atividade cognitiva, expressa-se mediante o coeficiente de correlação *r*. Uma correlação pode ser positiva ou negativa: se positiva, os dois fenômenos assumem valores na mesma direção; se negativa, cada fenômeno assume um valor em sentido inverso ao do outro. Se o coeficiente de correlação for próximo de 1,00, a relação é quase perfeita (ex. 0,90 ou -0,90) e pode-se prever o valor de um fenômeno a partir do valor do outro. Inversamente, se o coeficiente de correlação for próximo de zero a relação é imperfeita (ex. 0,09 ou -0,09) e, portanto, o valor de um fenômeno não prediz o valor do outro.

O valor da correlação é denominado "coeficiente de validade". É comum que as correlações entre um teste de inteligência e os critérios externos não sejam muito altas, posto que diversos fatores concorrem e afetam as correlações (ex. amadurecimento do sujeito, variação no sistema de atribuição de notas escolares, experiências que afetam o sujeito - como as familiares e sociais -, a margem de erro na precisão do instrumento, etc.). Portanto, no campo da medição da inteligência

é improvável uma correlação perfeita entre o instrumento e o critério. Segundo Cronbach (1998), é pouco freqüente, inclusive, que o índice de validade alcance um valor acima de 0,60. Esperar, portanto, que um teste prediga perfeitamente o nível de inteligência de um indivíduo ou afirmar que um teste de inteligência nada prediz da conduta inteligente de um indivíduo são posturas, no mínimo, insensatas.

Encontrar correlatos tanto fisiológicos quanto sociais do desempenho em testes de inteligência não significa encontrar uma relação de causa e efeito entre eles. As correlações apenas indicam em que medida dois fenômenos se relacionam. Dependendo do índice de r obtido, a predição de uma variável em função da outra é possível. Para efeitos práticos, é útil a um psicólogo organizacional ou escolar saber em que medida se pode prever o desempenho de uma pessoa num cargo de chefia ou no sistema de ensino regular a partir do desempenho em testes de inteligência. Prever a partir de outros critérios pode não ser uma conduta profissional acertada. Segundo Juan-Espinosa (1997), a predição do rendimento laboral a partir de testes de personalidade gira ao redor de 0,24; a partir de entrevistas, 0,14 e a partir de inventários de interesses, 0,10 (os testes de inteligência conseguiriam uma predição em torno de 0,52).

Enquanto os avanços psicométricos não são do domínio de muitos dos que praticam a avaliação psicológica, é comum ouvir dizer que os testes de inteligência não são úteis para prever o rendimento na vida quotidiana (ex. na escola ou no trabalho) dado que seus itens freqüentemente parecem alheios à vida quotidiana, isto é, devido a seu viés interno. Aqui se deve fazer uma distinção entre o uso de testes importados sem nenhuma adaptação ao novo meio cultural e o uso de testes adaptados, validados e normatizados para a população que se deseja testar. Enquanto a primeira situação reflete quase a metade dos testes utilizados no Brasil, a crítica, no sentido acima, estará sempre correta; no entanto, as

críticas à segunda situação não se sustentam, pois trata-se de um dos princípios de construção de testes. O item que compõe um teste não deve necessariamente representar fielmente a rotina do dia-a-dia (tampouco se afirma que o item deve ser o mais alheio possível à experiência do indivíduo), posto que o conteúdo dos testes importa pouco. A chave está na complexidade de processamento que o teste solicita do sujeito. Por exemplo, quando se extrai o fator g de um conjunto de testes cognitivos, observa-se que sujeitos mais inteligentes obtêm altas pontuações em provas como dígitos ou vocabulário do WAIS do que sujeitos menos inteligentes. Há, portanto, uma relação entre a memória (dígitos) e o fator g , assim como entre fluência verbal (vocabulário) e o fator g . Mas esse resultado *não* indica que o fator g equivale a “guardar números na cabeça” ou “saber o significado das palavras”. O fator g não se relaciona ao conteúdo específico dos problemas dos testes ou com suas características superficiais. Isto constitui o *princípio da indiferença do indicador*, teorema cunhado por Spearman (1955) ao observar as correlações entre as distintas tarefas que compunham a bateria de Binet e o alcance acadêmico. O fator g relaciona-se com o que há de comum no desempenho de diversas tarefas e que exige dos indivíduos determinado nível de cognição. Os tratamentos fatoriais conseguem detectar esse aspecto comum e os psicometristas o chamam de *g* psicométrico ou inteligência geral. O fator g está relacionado à complexidade da atividade cognitiva exigida pelos diversos problemas, isto é, com a capacidade de estabelecer relações entre os elementos, entre os conceitos abstratos, raciocinar, analisar, discriminar entre informações relevantes e aquelas irrelevantes e inferir conclusões a partir dos elementos de informação (Colom, 1998). Nesse sentido, Jensen (1998a) enfatiza que ***g não é*** o processo cognitivo ou a configuração dos circuitos neuronais com que opera o cérebro; ***g seria*** a capacidade para realizar operações cognitivas, uma noção semelhante ao conceito de CPU ou

de potência e eficiência de um computador. E o que é mais interessante, como bem destacam Colom e Andrés-Pueyo (1999), *g* não exclui as diversas aptidões que compõem a inteligência. Tanto é assim que o modelo de Carroll (1993) mostra que a inteligência enquanto construto científico seria composta de aproximadamente 60 aptidões. O fator *g* é a dimensão compartilhada por essas aptidões.

Um século de investigações com testes de inteligência tem mostrado resultados interessantes:

- a) A correlação entre as pontuações de testes de inteligência e o rendimento escolar gira em torno de 0,50 (Neisser et al., 1996).
 - b) Há correlação significativa entre o fator *g* e rendimento laboral. Os coeficientes giram em torno de 0,50 a 0,90, dependendo do tipo de emprego (Hunter & Hunter, 1984; Jensen, 1998a; Ree & Earle, 1992; Ree, Earles & Teachout, 1994).
 - c) Há significativas correlações negativas, porém pequenas, entre QI e problemas sociais como delinqüência, alcoolismo, autoritarismo (Brand, 1987), taxa de acidentes automobilísticos graves (O'Toole & Stankov, 1992).
 - d) Não há diferenças significativas entre os sexos com respeito ao desempenho em testes de inteligência geral (Colom et al., 1999; Aluja-Fabregat et al., 2000; Andrés-Pueyo, 1997).
 - e) Há correlação significativa entre o desempenho em testes de inteligência e variáveis psicobiológicas como tempo de reação e tempo de inspeção. Alguns pesquisadores situam a relação entre -0,40 e -0,50 (Bates & Eysenck, 1993; Chaiken & Young, 1993; Kranzler & Jensen, 1986; Deary, 1994, 1996), mas
- outros consideram que a relação não deverá ultrapassar a barreira de -0,30 a -0,40 (Hunt, 1980).
- f) Há correlação significativa, porém pequena, entre QI e tamanho cerebral (medição intracraniana por neuroimagem). Os coeficientes giram em torno de 0,35 (Willerman et al., 1994) a 0,40 (Andreasen et al., 1993).

Na abertura do *Tenth Meeting of International Society of the Study of Individual Differences - ISSID*, realizado em Edimburgo (2001), foi descrito um dos estudos mais consistentes sobre a validade dos testes de inteligência. Trata-se da retomada dos estudos de Godfrey Thomson (que coordenou a testagem, em 1932 e 1947, de todas as crianças da Escócia nascidas entre 1921 e 1936), pelo pesquisador escocês Ian Deary. Os primeiros resultados da avaliação de pessoas sobreviventes da testagem de 1932, isto é, da retestagem de pessoas ($n=101$) depois de 66 anos da primeira testagem utilizando-se o mesmo instrumento, mostram, entre outras coisas, a estabilidade das diferenças individuais nas habilidades mentais na ausência de enfermidades ou doenças crônicas ($r= 0,63$). As implicações de tais resultados para outras áreas do conhecimento humano, especialmente a geriatria, são extremamente positivas, dado que é possível estudar, com certo grau de precisão, a extensão e a época do declínio de funções cognitivas no ser humano (veja-se também Deary, Whalley, Lemmon & Crawford, 2000).

Assim sendo, conclui-se que a inteligência é uma variável psicológica estável e, portanto, passível de ser medida. Ela é também ubíqua, presente em maior ou menor grau em todas as atividades humanas e o instrumento que a mede, o teste psicométrico, pode ser considerado uma das mais importantes contribuições da ciência psicológica.

Crítica 4: “Interessa mais a Inteligência Emocional do que a Inteligência Acadêmica”

Uma dimensão importante do comportamento inteligente é a capacidade de gerenciar as emoções de forma a permitir às pessoas a sociabilidade e a aquisição de um padrão de vida adequado. A psicologia tem chamado essa capacidade de “inteligência intra e interpessoal” (Gardner, 1994) de “inteligência prática” (Sternberg, 1985, 2000) e, desde 1990, de “inteligência emocional” por Mayer e Salovey (1997).

A popularização do termo “inteligência emocional” veio com Goleman, um psicólogo formado pela Universidade de Harvard. Goleman (1995) enfatiza que o sucesso das pessoas na vida cotidiana depende mais do uso inteligente das emoções do que da inteligência acadêmica tal como é medida pelos testes de QI. Nesse sentido, o autor relata diversos casos de pessoas que não tiveram sucesso na vida apesar dos seus altos níveis intelectuais. O “analfabetismo” emocional seria o responsável pelo fracasso social dessas pessoas.

Os resultados dos testes de inteligência, como visto anteriormente, correlacionam-se fortemente com o desempenho acadêmico, laboral e, em menor grau, com o desempenho social e afetivo. Isto porque seus itens visam avaliar a capacidade subjacente à resolução de problemas lógico-matemáticos, verbais, espaciais, psicomotores e outros. Os testes de inteligência não foram criados para avaliar a capacidade de reconhecer os próprios sentimentos e os dos outros; tampouco avaliam a capacidade de controlar e gerenciar as emoções, embora se saiba há muito tempo que tais capacidades podem auxiliar ou dificultar o desempenho acadêmico e laboral. A “negligência” da emoção nos testes de inteligência pode dever-se ao fato de que os casos em que a influência de emoções negativas altera significativamente o desempenho acadêmico e laboral, a ponto de

arruinar a qualidade de vida de uma pessoa, constituem uma pequena parcela e não é, portanto, a regra. As evidências psicométricas ainda continuam apoiando a tese de que quanto mais inteligente for uma pessoa maior será a sua capacidade em gerenciar seus sentimentos, suas emoções e seu funcionamento adaptativo. Talvez, as capacidades relacionadas à administração das emoções possam ser mais bem entendidas na esfera da personalidade e das habilidades sociais (Flores-Mendoza & Nascimento, 2002).

Um estudo que reforça a suspeita da ligação da inteligência emocional a aspectos da personalidade constitui o trabalho de Ciarrochi, Chan e Caputi (2000). Esses autores aplicaram em 134 universitários australianos uma bateria de testes composta por medidas de QI, personalidade, satisfação de vida, qualidade de relacionamento e a Escala Multifatorial de Avaliação da Inteligência Emocional (MEIS). Os estudantes também foram induzidos a experimentar diversos humores para observar se aqueles que tivessem altas pontuações no MEIS apresentavam melhor capacidade em gerenciar seus humores e discriminavam melhor seus julgamentos sociais para prevenir humores inadequados. Os resultados mostraram que a inteligência emocional não estava relacionada ao QI ($r = 0,05$) e sim a aspectos da personalidade como empatia ($r = 0,43$) ou aspectos motivacionais como satisfação na vida ($r = 0,28$). Surpreendentemente, enquanto o QI se relacionou tanto com o gerenciamento quanto com a prevenção de estados humorais, a partir de julgamentos adequados, a inteligência emocional esteve relacionada apenas com o gerenciamento humoral.

Mais tarde, uma análise realizada por Ciarrochi et al. (2001) sobre as diversas escalas e questionários de inteligência emocional apontou que as medidas de inteligência emocional baseadas no desempenho estariam mais relacionadas a medidas tradicionais de QI do que a medidas de auto-relato. Estas

últimas estariam mais relacionadas com as dimensões da personalidade. Contudo, um estudo recente de Roberts, Zeidner e Matthews (2001) foi realizado para investigar as propriedades psicométricas e correlatos psicológicos do MEIS (uma medida baseada em desempenho). Participaram 704 sujeitos que, além do MEIS, completaram o Inventário de Auto-Descrição do Traço-IADT (uma medida do modelo dos Cinco Fatores de Personalidade), e o ASVAB (*Armed Services Vocational Aptitude Battery*), uma medida de inteligência. Os resultados mostraram que, embora o MEIS tenha se correlacionado moderadamente com o ASVAB e levemente com o IADT, encontraram-se índices pobres de precisão de alguns subtestes do MEIS, assim como a pontuação dada por peritos foi bastante diferente da que foi dada pelos não peritos. Tais resultados levantam sérias dúvidas sobre a estabilidade e validade do construto. Em outro documento Zeidner, Matthews e Roberts (2001) analisam as condições necessárias para que um construto se torne cientificamente válido; entre elas, estão: a necessidade de demonstrar um número de habilidades primárias emocionais, que reunidas definam um construto de ordem superior; apresentar evidências genéticas, neuropsicológicas, cognitivas e de desenvolvimento; identificar os componentes fluido e cristalizado. Sem essas condições a inteligência emocional ainda constitui uma miragem.

CONSIDERAÇÕES FINAIS

A revisão dos pontos mais polêmicos relacionados aos testes psicológicos, especificamente daqueles relacionados à medição da inteligência humana, teve como objetivo auxiliar a corrigir algumas atitudes erradas a respeito dos testes. A principal delas é a crença bastante generalizada no país de que os testes não servem para medir a conduta humana. Como já se argumentou, são significativas as correlações entre o

desempenho intelectual e as atividades fisiológicas, acadêmicas e sociais. Embora ainda não haja um consenso sobre o que venha a ser inteligência, pode-se afirmar que os testes medem aquela capacidade utilizada na cultura ocidental para resolver e enfrentar desafios quotidianos (por exemplo, escola, emprego). No entanto, deve-se ressaltar que testes mal adaptados e/ou construídos com inadequadas propriedades psicométricas, como é o caso da maioria dos testes nacionais, prejudicam qualquer avaliação psicológica. A segunda crença, que provavelmente fornece base à primeira, é a de que os testes inferiorizam ainda mais alguns segmentos étnico-sociais. Os testes não podem ser responsabilizados pelas diferenças observadas entre as pessoas, eles apenas as retratam. Infelizmente, a interpretação errônea dos resultados obtidos nos testes e a suspeita de invenção de dados por parte de alguns psicometristas do início do século passado, parece ter provocado uma resistência aos avanços ocorridos na esfera das diferenças individuais.

E, ultimamente, um interesse popular tem surgido sobre a chamada inteligência emocional. Trata-se de um construto para o qual ainda não há evidência científica da sua validade porque os instrumentos que tentam medi-lo ainda apresentam problemas psicométricos. Para que a inteligência emocional conquiste o mesmo status que a inteligência geral (ou fator *g*) é necessário replicar e delinear numerosos estudos por um bom período de tempo a fim de verificar a solidez e abrangência do referido constructo.

Em relação à inteligência propriamente dita, embora ainda persistam diversas perguntas, conforme mostra o relatório da APA (Neisser et al., 1996), é inegável que a medição psicológica tem melhorado com o passar dos anos e tem auxiliado a compreender melhor a cognição humana. Ignorar tais avanços e afirmar que os testes somente servem aos interesses da classe dominante significa voltar

às discussões já superadas das décadas de 60 e 70. O principal desafio agora não é saber se existem diferenças entre as pessoas ou grupos, o desafio maior é saber, por um lado, a que se devem essas diferenças e, por outro lado, como ajudar as pessoas desfavorecidas. O novo milênio exige que cientistas e pesquisadores encontrem respostas a esses desafios.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALUJA-FABREGAT, A., Colom, R., Abad, F. & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as g among young adolescents. *Personality and Individual Differences*, 28, 813-820.
- ANDREASEN, N.C., Flaum, M., Swayze, V., O'leary, D.S., Alliger, R., Cohen, G., Ehrhardt, J. & Yuh, W. T. C. (1993). Intelligence and brain structure in normal individuals. *American Journal of Psychiatry*, 150, 130-134.
- ANDRES-PUEYO, A. (1997). *Manual de psicología diferencial*. Barcelona: McGraw-Hill.
- BATES, T.C. & Eysenck, H. J. (1993). Intelligence, inspection time, and decision time. *Intelligence*, 17, 523-531.
- BOUCHARD, T., Lykeen, D., Mc Gue, M. Segal, N. & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science*, 250, 223-228.
- BRAND, C. (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and Controversy* (pp.251-265). New York: Falmer.
- BURT, C. (1955). The evidence for the concept of intelligence. *British Journal of Educational Psychology*, 25, 159-177.
- CARROLL, J. B. (1993): *Human cognitive abilities. A survey of factor analytic studies*. Cambridge, Cambridge Univ. Press.
- Chaiken, S. R. & Young, R. K. (1993). Inspection time and intelligence: Attempts to eliminate the apparent movement strategy. *American Journal of Psychology*, 106, 191-210.
- CIARROCHI, J. V., Chan, A. Y. C. & Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality and Individual Differences*, 28, 539-561.
- CIARROCHI, J. V., Chan, A. Y. C., Caputi, P. & Roberts, R. D. (2001). Measuring emotional intelligence. In: J. Ciarrochi, J. Forgas, & J.D. Mayer (Eds.). *Emotional intelligence in every life: A scientific inquiry*. (pp. 25-45). Philadelphia, PA: Psychology Press.
- COLOM, R. (1995). *Tests, inteligencia y personalidad*. Madrid: Pirámide.
- COLOM, R. (1998). *Psicología de las diferencias individuales. Teoría y práctica*. Madrid: Pirámide.
- COLOM, R. (2000). Algunos <<Mitos>> de la Psicología: Entre la Ciencia y la Política. *Psicothema*, 12, 1, 1-26.
- COLOM, R., Abad, F.J., Garcia, L. F. & Juan-Espinosa, M. (no prelo). Education, Wechsler's Full Scale IQ, and g.
- COLOM, R., Andrés-Pueyo, A. & Juan-Espinosa, M. (1998). Generational IQ gains: Spanish data. *Personality and Individual Differences*, 25, 927-935.
- COLOM, R. & Andrés-Pueyo, A. (1999). El estudio de la inteligencia humana: recapitulación ante el Cambio de Milenio. *Psicothema*, 11, 3, 453-476.
- COLOM, R., Quiroga, M. A. & Juan-Espinosa, M. (1999). Are cognitive sex differences disappearing?. Evidence from Spanish populations. *Personality and Individual Differences*, 27, 1189-1195.
- CRONBACH, L. J. (1998). *Fundamentos de los tests psicológicos*. Madrid: Biblioteca Nueva.
- DEARY, I. J. (1994). Sensory discrimination and intelligence: Postmortem or

- resurrection? *American Journal of Psychology*, 107, 95-115.
- DEARY, I. J. (1996). Intelligence and inspection time. *American Psychologist*, 51, 6, 599-608.
- DEARY, I.J., Whalley, L.J., Lemmon, H. & Crawford, J.R. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, 28, 1, 49-55.
- DEFTERMAN, D. K . (1998). Kings of men: introduction to a special Issue. *Intelligence*. 26, 3, 175-180.
- DEFTERMAN, D. K. & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low-IQ groups. *Intelligence*, 15, 247-250.
- EYSENCK, H. J. (1982). *A model for intelligence*. New York: Springer-Verlag.
- EYSENCK, H. J. & Kamin, L. (1982). *O grande debate sobre a inteligência*. Trad. Maria Julieta A. A. C. Penteado. Brasília: Editora UNB.
- FLETCHER, R. (1990). *The Cyril Burt scandal: Case for the defense*. New York: MacMillan.
- FLORES-MENDOZA, C. E. & Nascimento, E. (2001). *Inteligência: o melhor construto investigado em Psicologia* (manuscrito submetido a publicação).
- FLORES-MENDOZA, C.E. & Nascimento, E. (2002). *Inteligência emocional: Um construto científico?* (manuscrito submetido a publicação).
- FLYNN, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.
- FLYNN, J.R. (1998). IQ gains over time: Toward finding the causes. In: Ulric Neisser (ed.). *The rising curve: Long-term gains in IQ and related measures*. (pp.25-66). Washington: APA.
- GARDNER, H. (1994). *Estruturas da mente. A teoria das inteligências múltiplas*. Trad. Sandra Costa. Porto Alegre: Artes Médicas.
- GOLEMAN, D. (1995). *Inteligência emocional*. Trad. Marcos Santarrita. Rio de Janeiro: Objetiva (original publicado em 1995).
- GOULD, S. J. (1996). *The mismeasure of man*. 2. ed. New York: Norton.
- GUILFORD, J. P. (1954). *Psychometric methods*. NY : McGraw-Hill.
- GUILFORD, J. P. (1988). Some changes in the structure of intellect model. *Educational and Psychological Measurement*, 48, 1-4.
- GUSTAFSSON, J. (1988). Hierarchical models of individual differences in cognitives abilities. In: R. J. Sternberg (ed.). *Advances in the Psychology of Human Intelligence*, vol. 4. New Jersey : LEA.
- HAIR, J. F., Anderson, R. E, Tatham, R. L. & Black, W. C. (1995). *Multivariate data analysis*. New Jersey: Prentice Hall.
- HERRNSTEIN, R. J. & Murray, Ch. (1994). *The bell curve - intelligence and class structure in American life*. New York: A Free Press Paperbacks Books.
- HUNT, H. (1980). Intelligence as an Information-Processing Concept. *British Journal of Psychology*, 71, 4, 449-474.
- HUNTER, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- JENSEN, A. R. (1987). Psychometric "g" as a focus of concerted research effort. *Intelligence*, 11 3, 193-198.
- JENSEN, A. R. (1998a). Jensen on "Jensenism". *Intelligence*, 26, 3, 181 - 208.
- JENSEN, A. R. (1998b). *The g factor*. New York: Praeger.
- JENSEN, A. R. & Weng, L. (1994): What is a good g?. *Intelligence*, 18, 231-258.
- JUAN-ESPINOSA, M. (1997). *Geografía de la inteligencia humana*. Madrid: Pirámide.
- KAMIN, L. (1974). *Ciencia y política del cociente intelectual*. Madrid: Pirámide.

- KRANZLER, J. H. & Jensen, A. R. (1989). Inspection time and intelligence: A Meta-Analysis. *Intelligence*, 13, 4, 329-347.
- MAYER, J. D. & Salovey, P. (1997). What is emotional intelligence?. In P. Salovey & D. Sluyter (Eds.) *Emotional Development and Emotional Intelligence: Implications for educators* (pp. 3-31). New York: Basic Books.
- MAYER, J. D., Caruso, D. R. & Salovey, P. (2000). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27, 4, 267-298.
- NEISSER, U. (1998). *The rising curve - long term gains in IQ and related measures*. Washington, DC: APA.
- NEISSER, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., Sternberg, R. & Urbina, S. (1996): Intelligence: knowns and unknowns. *American Psychologist*, 51, 2, 77-101.
- NYBORG, H. & Jensen, A.P. (2001). Occupation and income related to psychometric g. *Intelligence*, 29, 45-55.
- O'TOOLE, B. J. & Stankov, L. (1992). Ultimate validity of psychological tests. *Personality and Individual Differences*, 13, 699-716.
- PASQUALI, L. (1997). *Psicometria: teoria e aplicações*. Brasília: Editora UnB.
- REE, M. J. & Earle, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1, 86-89.
- REE, M. J., Earles, J. A. & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79, 518-524.
- ROAZZI, A., Spinillo, A., & Almeida, L. S. (1991). Definição e avaliação da inteligência: Limites e perspectivas. In: L. S. Almeida. *Cognição e Aprendizagem Escolar*. Lisboa: APPORT.
- ROBERTS, D. R., Zeidner, M. & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some New Data and Conclusions. *Emotion*, 1, 196-231.
- ROSENTHAL, R. & Jacobson, L. (1966). Teachers' expectancies: determinants of pupils' IQ gains. *Psychological Report*, 19, 115-118.
- ROSENTHAL, R. & Jacobson, L. (1968). *Pygmalion in the classroom: teacher expectation and pupils' intellectual development*. New York: Holt, Rhinehart & Winston.
- RUSHTON, J. P. (1994). The equalitarian dogma revisited. *Intelligence*, 19, 263-280.
- RUSHTON, J. P. (1998). The "Jensen effect" and the "Spearman-Jensen Hypothesis" of black-white IQ differences. *Intelligence*, 26, 3, 217-225.
- SPEARMAN, C. H. (1955). *Las habilidades del hombre*. Trad. Jaime Bernstein. Bs. As.: Paidós (original publicado em 1927).
- SPITZ, H. H. (1999). Beleaguered Pygmalion: A history of the controversy over claims that teacher expectancy raises intelligence. *Intelligence*, 27, 3, 199-234.
- STERNBERG, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- STERNBERG, R. J. (1992). *As capacidades intelectuais humanas: uma abordagem em processamento de informação*. Trad. Dayse Batista. Porto Alegre: Artes Médicas (original publicado em....).
- THORNDIKE, E. P. (1921). Contribution to intelligence and its measurement. *Journal of Educational Psychology*, 12, 124-127.
- THORNDIKE, R. L. (1994). "G". *Intelligence*, 19, 2, 145-155.
- THURSTONE, L. L. (1921). Contribution to intelligence and its measurement. *Journal of Educational Psychology*, 12, 201-207.

- WECHSLER, D. (1971). Intelligence: definition, theory and the IQ. In: Cancro R. (ed.) *Intelligence: genetic and environmental influences*. New York: Grune Straton.
- WILLERMAN, L., Shultz, R., Rutledge, A. N. & Bigler, E. D. (1992). Hemisphere size asymmetry predicts relative verbal and nonverbal intelligence differently in the sexes: Na MRI study of structure-function relations. *Intelligence*, 16, 315-328.
- WILSON, E. (1995). Ciencia e ideología. *Academic Questions*, vol. 8, 06-10. Trad. Andrés-Pueyo e R. Colom (Comps): *Ciencia y política de la inteligencia en la sociedad moderna*. Madrid: Biblioteca Nueva, 1998.
- ZEIDNER, M., Matthews, G. & Roberts, R. D. (2001). Slow down, you move too fast: emotional intelligence remains an "elusive" intelligence. *Emotion*.1, 265-275.