

Além das diferenças inter-individuais: Abertura para uma psicometria “clínica”

Beyond inter-individual differences: The way to a “clinical” psychometrics

Além das diferenças inter-individuais: Abertura para uma psicometria “clínica”

Beyond inter-individual differences: The way to a “clinical” psychometrics

Roland Capel, Dr. psicologia, Prof. Remplaçant, Maître d’enseignement et de recherches
Université de Lausanne, Suíça

Raphaël Rossé, Psychologue, Assistant de recherche, Université de Lausanne, Suíça

Renzo Oswald, Sócio da Moityca, adaptador do L.A.B.E.L. para o Brasil, São Paulo

Endereço para correspondência:

Moityca Eficiência Empresarial Ltda

Av. Luis Romero Sanson 701

São Paulo, SP

CEP 04786-080

Tel. +11 5666-5661

Fax. +11 5666-0002

E-mail: roswald@moityca.com.br

Resumo

O método de construção de escalas de atitudes (interesses, personalidade, etc) é praticamente o mesmo utilizado na construção de medidas unidimensionais de aptidões. Essa similitude significa que as respostas aos inventários de atitudes são estabelecidas exatamente da mesma forma como se tratasse de respostas do tipo “verdadeiro / falso”. Se a utilização de métodos paramétricos se justificar na tabulação de respostas desse tipo, deveriam ser questionados quando se trata de escalas de atitudes. As respostas do tipo “gosto / não gosto” não possuem o mesmo valor e não deveriam ser tratadas com os mesmos métodos paramétricos. Da mesma forma a normalização sistemática das escalas de atitudes – prática diretamente importada dos construtos de aptidões – é imprópria quando aplicada a construtos multidimensionais da mensuração das atitudes. Esse artigo mostra dois exemplos do viés introduzido na avaliação das pessoas quando se normalizam “automaticamente” os elementos de um perfil de personalidade. Finalmente nosso intento é mostrar que além de uma avaliação diferencial das pessoas (estabelecendo uma comparação entre elas), seria necessário poder efetuar uma avaliação intra-pessoal, sem nenhum referencial a outras pessoas, em uma perspectiva de integração profissional e desenvolvimento pessoal.

Palavras Chave: escalas de atitudes, construto multidimensional, nomotética e idiografia

Abstract

The construction method of a scale of attitudes (interests, personality, etc.) is quiet the same already used for the construction of one-dimensional measures in the field of aptitudes. They are alike because the type of answers in attitude scales is treated as if they were made of “right/wrong”. If the use of parametrical method is justified for answers of this kind, it should no longer be the case when we talk about attitude scales. The answers “I like/I don’t like” are not the same and shouldn’t be treated with parametrical methods.

Moreover, the system of attitude scales is directly inherited from the field of aptitudes, and is definitely unsuitable to the multidimensional domain of attitudes' measure. This article shows two examples of the biases introduced in the evaluations of persons when one standardizes a profile of personality or any other kind of profile. Finally, the idea is to show that, next to a differential evaluation of persons (inter-evaluation), it would be necessary to give an intra-personal evaluation (without any comparisons with others), in a perspective of personal development.

Key words: Attitudes scales, multidimensional construct, idiographic and nomothetic.

1. Introdução: os testes sempre questionados e questionáveis...

Como toda prática que se inscreve no tempo, a utilização de testes e de medidas psicométricas conheceu períodos críticos e períodos fastos durante as últimas décadas. Neste início de século, a avaliação psicológica quantificada se beneficia de uma aureola científica de boa qualidade e parece se expandir com uma surpreendente facilidade, em ambientes de consulta a cada vez mais numerosos e diversos: orientação escolar e profissional, reclassificação profissional (*outplacement*), diagnósticos psiquiátricos, processos de seleção de pessoas, mapeamento das competências, etc. Como uma sombra fiel, a crítica dos testes segue naturalmente de perto essa evolução, mas a utilização das medições psicométricas entra ainda mais nas práticas e os opositores esgotam seus argumentos em um “reducionismo tecnocrático”. Atualmente o momento aparece mais propício à crítica construtiva e ao questionamento.

Cada construtor de instrumentos psicométricos se preocupa naturalmente com as críticas aos testes e ao controle da utilização deles, e particularmente com duas questões de atualidade: “Como reagir frente à utilização constante e generalizada de instrumentos a fim de conselho ou de seleção?” “Deve-se aproveitar desta renovação de popularidade, para desenvolver novos instrumentos sem nenhum controle e protegido da crítica pondo em causa sua legitimidade?” Esta problemática não é certamente bem original, mas permanece “sensível” e de atualidade e constitui indubitavelmente um desafio essencial da psicometria moderna. De uma certa forma o Conselho Federal de Psicologia trouxe uma resposta a essas perguntas editando a resolução 002/2003 que tanto irritou e provocou de incompreensão durante os anos passados.

A nossa intenção com esse artigo, é de abordar a questão de confiabilidade das escalas sob um novo ângulo, ou mesmo provocar ligeiramente o leitor, pois não trataremos nem da fidedignidade nem da validade. (cf.: Capel, 2003, capítulo 3).

Nossa interrogação pode ser formulada da seguinte forma: é justificável, tanto do ponto de vista ético como do ponto de vista técnico – estes dois aspectos sendo por outro lado intimamente ligados – utilizar os resultados das escalas dos testes para aconselhar, orientar ou selecionar as pessoas?

2. Aptidões e atitudes: mesmo tratamento?

Como preâmbulo a qualquer reflexão é bom lembrar que a utilização de testes ou provas parecidas, cujo objetivo é de selecionar pessoas, existe há muito tempo. Na Europa e na China, há séculos que se utilizam ferramentas de medição de competências e de conhecimentos escolares para selecionar os indivíduos de destaque, no intento de aproveitar as competências deles. Atualmente escolas e institutos de formação anexa, ainda pontuam o calendário de provas que se estendem em um período sempre mais comprido. No início do século XX a jovem ciência psicológica começa se preocupar com o fenômeno da inteligência, com o sucesso e as decepções que todos conhecemos.

Com Spearman e os primeiros passos da análise fatorial (Spearman 1904), a questão da medição das aptidões (*abilities of man*) ficou ainda mais presente na literatura (Spearman 1927): não se pretendia mensurar só o adquirido, mas o potencial de aprendizagem que foi sempre associado de forma mais ou menos íntima à inteligência, apesar da denominação do “fator g”, permitir desviar a atenção sobre a questão incômoda da definição dessa inteligência, sem realmente a resolver durante vários anos. A medição do fator g e as tentativas de defini-lo do ponto de vista psicológico, foi uma preocupação constante de inúmeros psicólogos, particularmente a partir da primeira guerra mundial (ver sobre esse assunto: Terman, 1916). Durante esse período, e mais particularmente a partir dos anos 1930, a técnica de construção das escalas de aptidões foi desviada de seu primeiro objetivo, não foi só utilizada para mensurar um desempenho, mas também atitudes e competências, sempre em auto-avaliação. No caso das escalas de aptidões, os itens determinam comportamentos, de fato, o que permite apreciar uma resposta certa, considerando todas as outras erradas. O sujeito que responde mostra assim se conhece a resposta ou se é capaz de executar uma determinada tarefa.

Essa década de 1930 pode ser considerada uma época chave na avaliação psicológica das atitudes: à maneira antiga de pensar em tipologias pouco a pouco substitui a exigência da mensuração.

Após as primeiras tentativas hesitantes dos questionários caracterológicos construídos com embasamento puramente teórico, foi percebido que as técnicas inauguradas por Spearman podiam contribuir, com soluções científicas, aos problemas criados em termos de exigências de precisão e de validade de uma escala. A avaliação dos caracteres se torna uma ciência, e a utilização de novas técnicas inspiradas nos modelos estatísticos tornam-se incontornáveis.

A técnica clássica de construção das escalas nasce conjuntamente com uma nova disciplina da psicologia: a psicometria, que vai impor suas regras de forma soberana praticamente a todos os construtos do comportamento humano (ver por exemplo Thurstone, um pai fundador dessa técnica: 1929, 1941, 1959)

A técnica clássica de construção das escalas já é bem estabelecida a partir dos anos 1930-1940 e apesar dos aperfeiçoamentos devido às contribuições de Cronbach (1965), Catell (1970), assim como as de outros nomes prestigiosos da disciplina, devemos reconhecer que não evoluiu praticamente desde então. Uma análise acurada dos manuais atuais dos testes mais recentes, mostra claramente que as escalas construídas nesse início de século XXI não se diferenciam, em suas concepções, das escalas elaboradas a mais de meio século.

Essa constatação nos permite afirmar que desde antes do início da segunda guerra mundial, a maioria das teorias e conceitos ainda hoje utilizados já estavam perfeitamente estabelecidos: a elaboração de escalas a partir da análise dos itens, a utilização do modelo normal e da teoria probabilística dos erros, a estimativa da fidedignidade baseada em uma teoria nomeada das “provas paralelas”, conceitos de validade múltiplos, etc. Na realidade nenhum desses “monstros sagrados” do método clássico foi questionado durante os últimos 70 anos. Se insistimos nesse aspecto é porque de nosso ponto de vista, esse conservantismo (para não dizer imobilismo ou falta de espírito de

inovação) surpreendente, é a fonte dos numerosos problemas que se enfrentam hoje, na utilização de testes em situações bem particulares, como para a seleção por exemplo, como vamos em seguida tentar demonstrar.

Em uma primeira fase, é necessário definir claramente a que correspondem os testes de atitudes que evocamos acima, e que também são chamados de “testes de comportamento típico”. Ao contrário dos testes de aptidões, o objetivo não é medir um nível de desempenho, de potencial de raciocínio ou de memória, mas de avaliar até qual nível alguns traços de personalidade ou de comportamento se manifestam em um indivíduo, *a partir de suas próprias respostas*. De fato, quando se pede para um sujeito responder a um questionário de atitudes (personalidade, valores, interesses, características profissionais, etc.) não se pede que ele responda “o melhor que ele pode”, mas simplesmente de manifestar como ele é, como se sente, acredita ou se sente atraído por um estímulo, geralmente um item, que seja uma palavra (adjetivo, verbo ou nome de profissão) ou um fragmento de comportamento (frase curta). Definidos esses termos, é importante relevar e entender que as escalas de atitudes são construídas exatamente da mesma forma, utilizando o mesmo modelo que as escalas dos testes de aptidão (Mitchel, 2003, p23): são depuradas a partir de uma análise de itens, calculadas somando algumas respostas cuidadosamente selecionadas, garantidas, do ponto de vista qualitativo, pela mesma teoria dos erros e, na quase totalidades dos casos são normalizadas (cf. por exemplo: Rust, J & Golombok, 1989). Essa maneira de operar quase automática (praticamente nunca foi questionada desde que foi proposta por Likert em 1932) da técnica de construção das escalas de aptidões transposta à construção das escalas de atitudes geram, de nosso ponto de vista, alguns problemas técnicos importantes. Paralelamente os profissionais de RH devem responder a essa pergunta crucial: é possível legitimar em todas as situações a utilização de escalas de atitudes construídas com técnicas originalmente destinadas a uma utilização completamente diferente, ou seja, o desenvolvimento de escalas destinadas a hierarquizar (selecionar) os sujeitos em função dos desempenhos?

3. Não confundir o que é paramétrico com o que não é!

No campo das aptidões é claro e perfeitamente natural esperar que um teste permita situar um indivíduo comparando-o com um outro, comparando-o com o desempenho “médio” observado em um grupo ou uma população que possui características homogêneas supostamente conhecidas. Por exemplo, quando se trata de avaliar a memória, a força do raciocínio, a faculdade de entender o sentido de algumas palavras ou frases, é normal emitir um julgamento do tipo: “O sujeito X se situa....na média, no quartil superior, ao percentil 74, etc”.

A avaliação das aptidões foi essencialmente desenvolvida no intento de diferenciar os indivíduos em uma hierarquia, e a técnica da escolha (puramente paramétrica) que permite alcançar esse objetivo é a normalização. É por isso que praticamente todos os manuais dos testes (aptidões ou atitudes) apresentam nos anexos ou apêndices toda uma série de gabaritos que permitem julgar o valor de um desempenho comparando-o com aquele de um grupo bem definido. Assim é possível, por exemplo, avaliar os resultados obtidos para uma criança de 12 anos a um teste de raciocínio comparando-os com aqueles de suas colegas da mesma idade, comparam os resultados de dois adultos, sem diferenciar os sexos, ou só comparando com resultados de mulheres ou homens, etc.

Se considerarmos essa técnica de um ponto de vista puramente teórico, é possível perceber que as medições de aptidões são construídas graças a técnicas puramente paramétricas (correlações, normalização, etc) e que essa técnica é perfeitamente justificada pela natureza do conceito a ser tratado que deriva diretamente das noções de “verdadeiro” ou “falso” que podem ser legitimamente consideradas como sendo “objetivas”. De fato, o nível de “veridicidade” ou de “falsidade”, conteúdo na correção de cada item de um teste de aptidão pode ser o objeto de um consenso: todos os itens “falsos” estão falsos de forma equivalente, isso por todos os juizes, como por todos os sujeitos que se submetem ao teste. Para os itens “verdadeiros” o raciocínio é o mesmo.

No campo das atitudes, qualquer seja o construto, o assunto é bem diferente: os itens não são considerados “verdadeiros” ou “falsos”, mas proporcionam a possibilidade de responder em função de um nível de acordo, ou de preferências, o que é longe de ser um critério “objetivo”. Cada sujeito

utiliza uma escala de respostas individual, como melhor a entende: uma avaliação máxima a um item por parte de um sujeito é dificilmente comparável com a resposta ao mesmo item de uma outra pessoa, etc. Devemos admitir e constatar que quando se trata de atitudes as respostas são puramente subjetivas. Isso significa que de um indivíduo a um outro, *e até para o conjunto das respostas de uma mesma pessoa*, (de um item característico de uma dimensão a um outro característico de uma outra dimensão) *as escalas de medição não podem ser decompostas em unidades fixas e determinadas*. De acordo com o que acaba de ser afirmado, o tratamento das respostas aos inventários de atitudes não deveria ser tratado de forma paramétrica, o que implica que a técnica de construção das escalas de aptidões não convém para construir escalas de atitudes. Particularmente a análise dos itens e a padronização são técnicas inapropriadas para tratar as respostas do tipo “gosto” / “não gosto”.

Como comentário a essa crítica devemos constatar que, a nosso conhecimento, a análise não paramétrica das respostas aos testes de atitudes não conheceu nenhum desenvolvimento interessante ou utilizável, sem dúvida por causa da complexidade do problema. É sem dúvida por isso que a receita simples e “sossegada”, que designamos com o termo genérico de “técnica clássica” que serviu para a construção de escalas de aptidões foi praticamente universalmente aplicada às atitudes, *aparentemente* sem muitos problemas. Nosso intento nesse artigo é de questionar essa *aparência* que consideramos enganosa. A linha de pesquisa do Instituto de Psicologia da Universidade de Lausanne na Suíça que fornece todo o apoio científico à Moityca visa dois objetivos: o primeiro é de confiabilizar uma técnica de construção de escalas de atitudes baseada em métodos não paramétricos, cujo objetivo é de comparar os dois sistemas de respostas “clássicas” e “não paramétricas”, para demonstrar que a segunda não fornece forçadamente os mesmos resultados que a primeira. Essa constatação seria o ponto de partida de um questionamento radical da qualidade de todas as escalas atualmente existentes. O segundo eixo de pesquisa visa achar uma alternativa às medidas normalizadas (tratadas de forma puramente paramétrica) que consistiria em medidas “absolutas”, o que permitiria não só comparações diferenciais como também, finalmente, análises de “casos únicos”, ou seja análises “intra-pessoais” (Cf. Gendre, 1985).

É bom constatar igualmente que essa exigência pode também ser aplicada as aptidões, a condição de adotar uma investigação multidimensional. Por exemplo: é possível se perguntar quais são as aptidões mais importantes de um indivíduo em uma perspectiva intra-sujeito e *não só em um*

quadro diferencial inter-sujeitos. Por exemplo: uma pessoa com um nível global considerado fraco, pode apresentar também um índice de memorização fraco, quando comparado a um grupo de pessoas da mesma idade. Mas não seria também interessante constatar que entre todas as suas aptidões, é a memória que funciona melhor? Responder a essa pergunta significa hierarquizar as aptidões em uma perspectiva interna que não tem mais nada a ver com uma triagem diferencial inter-indivíduos, mas baseada na análise das faculdades intra-pessoais. Essa análise seria sem dúvida muito útil para conhecer melhor o sujeito de forma personalizada e particularmente, focalizar o desenvolvimento dos pontos que ele considera “fortes”, *até quando se revelam fracos quando comparados à média de um grupo de referência*.

Para as aptidões, é possível alcançar esse objetivo teoricamente, construindo provas “emparelhadas” de mesmo comprimento selecionando cuidadosamente itens de dificuldade média para a população meta. Para as atitudes essa noção de “dificuldade média” não faz sentido, o que nos força a explorar outros caminhos, a inaugurar novas técnicas.

4. Uma nova maneira de mensurar.

A idéia de produzir medidas “absolutas” não é particularmente nova e corresponde parcialmente a definição do *nível de um traço latente* como utilizada na teoria de resposta ao item –TRI- (Lord & Novick, 1968). Essa teoria se apresenta como a única alternativa séria à teoria clássica por que diz respeito à medição das aptidões.

Ela permitiu o desenvolvimento de toda uma série de testes adaptativos que só estão normalizados (nesse contexto, calibrados) do ponto de vista dos itens, mas não do ponto de vista de uma medição final. Nosso intento não é abordar nesse trabalho os princípios dessa teoria cujos fundamentos estatísticos e matemáticos são bem mais complexos daqueles utilizados na teoria clássica. Nosso interesse é só de ressaltar a contribuição fundamental da TRI: a idéia de mensurar um traço de forma personalizada, na base de uma seleção de itens própria a cada indivíduo e sem nenhum referencial a alguma norma de um lado é pertinente e praticamente realizável. Todas as escalas desenvolvidas a

partir da TRI são exclusivamente escalas de aptidões, e o traço latente medido é sempre único e medido exclusivamente graças as respostas do tipo “verdadeiro”/ “falso”. Essa limitação não permitiu estender os benefícios dessa nova teoria dos testes adaptativos a medição das atitudes, apesar dessa possibilidade existir, pelo menos do ponto de vista teórico.

A produção de medidas absolutas nos construtos das atitudes tornou-se uma realidade a partir da década de 1970, graças aos trabalhos de Gendreau (cf. 1998, 2002 e a ser publicado) que permitiram a criação de uma nova técnica de construção de escalas, chamada de “método funcional” e destinada exclusivamente a medir traços de atitudes ou de comportamentos típicos. Sua originalidade consiste em várias propriedades inexistentes na técnica clássica.

Para cada construto de atitudes (personalidade, interesses, comportamento, etc.) o método funcional permite criar um espaço de medição multidimensional, o que autoriza a medição simultânea de vários traços.

Esse espaço de medição pode conter todos os itens existentes aparentados com o construto. Todos esses itens estão presentes na composição de todas as escalas. Nesse espaço cada item é representado por suas coordenadas denominadas “características do item”.

Cada série de respostas de um indivíduo é modelizada nesse espaço graças a uma equação de regressão linear múltipla cujos coeficientes são chamados de “vetor da estratégia de resposta do sujeito”. Cada sujeito se localiza nesse espaço de medição por suas coordenadas que são os elementos de seu vetor de estratégia.

Os itens representados nesse espaço de medição fundamental podem ser agrupados em clusters a partir das proximidades deles (escalas “funcionais”) como também a partir de algumas teorias existentes (escalas externas). Esses agrupamentos possuem um ponto médio ou centróide cujas coordenadas definem a “assinatura” da escala.

Nesse espaço de medição se encontram assim os itens, as escalas e as respostas dos indivíduos. As mensurações são calculadas projetando um vetor no outro por operação de produtos escalares. Assim as projeções do vetor “sujeito X” acima dos vetores “escalas” representam os escores brutos do indivíduo X nas escalas correspondentes.

Os escores do indivíduo não são o resultado de uma soma, mas um produto escalar que representa a projeção, no sentido vetorial, ou seja uma correlação no sentido estatístico.

Um dos principais postulados do método funcional é que os escores brutos estão em uma métrica absoluta, a das correlações que dispersa os resultados de -1 a +1. Nada impede depois de estandardizar (normalizar) essas correlações em um grupo de referência e de apresentá-los em um sistema de unidades cômodo, mas o fato é que o método funcional permite alcançar o objetivo definido no início, ou seja propor escalas de dois tipos: normalizadas e não.

5. Dois exemplos

Queremos mostrar agora dois exemplos de traços de personalidade medidos graças às escalas do Big Five apresentadas no L.A.B.E.L. (Gendre & Capel, 2003), já utilizado por numerosos psicólogos consultores no Brasil. Para entender melhor o interesse dos números mostrados nas tabelas, fornecemos a seguir algumas explicações:

Os escores brutos se encontram em uma métrica que vai de 1 (rejeição absoluta) até 5 (forte atratividade) a mesma que for proposta ao sujeito para fornecer suas respostas – escala de Likert.

Essa standardização “intra-sujeito” é linear e não altera em nada a hierarquia intra-pessoal. Tem mais; a métrica de 1 a 5 permite fixar claramente os limites para a interpretação dos escores brutos:

Menor que 2: rejeição absoluta do traço

Maior que 4: forte atratividade

Entre 3 e 4: atratividade moderada

Entre 2 e 3: rejeição parcial

A fim de diferenciar bem os escores brutos não standardizados dos escores normalizados, apresentamos esses últimos em escores G, de média 50 e desvio padrão 20, que são na realidade stanine a dois dígitos (por exemplo: escore G=47, equivale a stanine 5; escore G=5 equivale a stanine 1, etc). Os limites dos stanine se situam a 1 e a 9, enquanto os dos escores G se situam entre 0 e 100. A população de referência é composta de 382 pessoas (237 de sexo feminino, ou seja 62%) com idades de 19 a 63 anos. Do ponto de vista profissional 45% são psicólogos ou profissionais de RH, 12% administradores de empresas, 12 % engenheiros e o restante da amostra representa profissões diversas.

A análise da média de todas as respostas (como se a população fosse reduzida a um único indivíduo) fornece esses níveis (escores brutos funcionais nas escalas do Big Five do NEO-PI de Costa e McCrae, 1985)

Abertura as experiências (O) 3,3

Consciência (C) 4,6

Extroversão (E) 4,2

Afabilidade (A) 4,4

Estabilidade emocional (S) 4,5

Esses números mostram claramente que nossa população não é homogênea do ponto de vista de seus traços exprimidos nos conceitos OCEAS da teoria do Big Five. Se os traços C,E,A e S são bastante homogêneos, O se distancia nitidamente. Em consequência, é claro que essa disparidade vai influenciar os escores normalizados da seguinte forma: nossa população sendo constituída principalmente de pessoas com elevada consciência (C) e Estabilidade (S), as pessoas que apresentam os mesmos traços terão seus escores normalizados sub-avaliados e as pessoas mais curiosas (abertura as experiências – O) terão seus escores G super-avaliados. De forma geral a hierarquia dos traços revelada através dos escores brutos “absolutos” será sempre mais desorganizada quando a hierarquia intra-pessoal de um sujeito se afasta daquela da população.

5.1 Análise do «perfil» de uma coordenadora de formação de 25 anos. Cf. Tabela 1

Considerando a interpretação dos resultados obtidos com uma normalização clássica (escores entre parênteses) a pessoa aparece como um perfeito artista, pouco consciencioso, extrovertido, simpático e estável. Podemos formular a hipótese que não é ordenado e talvez lhe falte moralidade. Do ponto de vista INTRA, podemos anotar que se apresenta como uma pessoa extrovertida, simpática e estável, Abertura (O) recebe a nota absoluta muito elevada, a MAIS ELEVADA da amostra (daí o escore G = 100). A Consciência (C) é também muito elevada (no absoluto 4,6 em um máximo de 5), mas como todas as pessoas da amostra quiseram se mostrar extremamente conscienciosas ... ele recebe um escore baixo (39) quando normalizado!

Tabela 1

Escores absolutos e normalizados (entre parênteses) nas escalas do Big Five como primeiro exemplo

Abertura	4.9 – (100)
Consciência	4.6 – (39)
Extroversão	4.2 – (86)
Afabilidade	4.5 – (70)
Estabilidade	4.5 – (68)

5.2 Análise do «perfil» de uma coordenadora de treinamento de 30 anos. Cf, Tabela 2

Analisando os escores normalizados, estamos aqui, frente a uma pessoa que aparece como original e gentil, praticamente no mesmo nível normalizado (75 e 78), mas pouco organizada e instável. De fato, a nota absoluta menor é de 3,9, todas as notas são elevadas, mas como a amostra é pouco Aberta (O), a originalidade é SUPER estimada e a estabilidade (S) é SUB estimada pois a amostra é um modelo de estabilidade. Finalmente se trata de uma pessoa sociável que gosta de ser notada, não particularmente original com algumas fraquezas de equilíbrio, sem mais.

Tabela 2

Escores absolutos e normalizados (entre parênteses) nas escalas do Big Five como segundo exemplo (coordenadora de treinamento de 30 anos)

Abertura	4.2 – (78)
Consciência	4.3 – (30)
Extroversão	4.6 – (53)
Afabilidade	4,7 – (75)
Estabilidade	3.9 – (32)

6. Resumo e perspectivas

Nesse artigo quisemos abordar um assunto que praticamente nunca é discutido na literatura especializada, ou seja, qual é a pertinência de utilizar métodos paramétricos, diretamente herdados da técnica de construção das escalas de aptidões, para desenvolver as escalas de atitudes cuja natureza é completamente diferente.

A nosso conhecimento, nem o grande mestre anglo-saxão Cronbach (1990), nem Dicks (1994) em seus excelentes trabalhos ou Lavault (1997) tratam esse problema de forma explícita.

Ao que diz respeito ao problema específico da normalização, esperamos ter demonstrado que essa operação, muito apreciada pela psicologia “diferencial” e seus objetivos de seleção que presidiram seu evento, não é apropriada em múltiplos casos. Na realidade a utilização de testes que medem várias escalas simultaneamente, deve despertar imediatamente a atenção sobre a importância relativa dos traços, uns comparados com os outros em uma perspectiva INTRA-pessoal e não exclusivamente comparando com os resultados de uma população meta referenciada, que na maioria dos casos é mal conhecida, isso quando não é mal adaptada (gabaritos antigos, amostras pequenas, sujeitos mal amostrados, amostras diferentes da população meta, etc.)

A consequência desse fenômeno, portanto evidente, é que a hierarquia interna dos interesses (por exemplo) de uma população é socialmente, culturalmente e politicamente determinada. Isso implica que normalizando os resultados de um sujeito cuja hierarquia interna não corresponde a hierarquia da população referenciada (normalmente representativa da sociedade onde o sujeito vive) se cria um viés importante super-avaliando ou sub-avaliando alguns traços. Nessa perspectiva diferencial, esse viés de um novo tipo cria, em alguns casos, uma falsa imagem da importância relativa dos traços que pode induzir o psicólogo a identificar pontos fortes e fracos que são puros artefatos provocados por uma

técnica não adaptada ao primeiro objetivo (em princípio) de uma avaliação: ou seja, entender melhor um indivíduo.

Essa conclusão sugere que a avaliação psicológica não deveria se limitar exclusivamente ao nível diferencial (imposto por uma técnica inadequada) mas deveria contemplar também um nível “clínico” que consideraria a especificidade de cada indivíduo, sem compará-lo com outras pessoas.

Nossa reflexão revela com isso uma ambição que pode parecer irrealista: reconciliar as abordagens nomotética (quantitativa) e idiográfica (qualitativa), através do desenvolvimento de instrumentos de medição capazes de produzir não só, medidas normalizadas como também medidas “absolutas”, sem referência a nenhuma população. Isso permitiria abordar a avaliação das pessoas não mais exclusivamente no aspecto diferencial clássico, mas também na perspectiva de uma psicometria que aspiramos a chamar de “clínica”.

Já estamos na altura de poder responder a pergunta formulada na introdução. Um postulado da psicometria clássica preconiza que um teste acompanhado de um manual completo que explique a construção das escalas, o processo de normalização, a interpretação, fidedignidade e validades podem justificar sua utilização em praticamente todos os construtos.

Com nossa análise tentamos convencer o leitor que as “verdades” da psicometria clássica podem, em alguns casos, como na medição de aptidões, fornecer critérios suficientes de qualidade, mas sem dúvida, não em todas as situações de teste e particularmente nas situações de seleção de pessoas baseadas em auto-avaliações de atitudes normalizadas.

Bibliografia :

Capel R. (2003). *L'évaluation psychologique – Notes de cours, partie 1*. Actualités psychologiques, 16, 128 p.

Cattell, R. B., Eber, H. W. & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL : Institute for Personality and ability testing.

Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personal decisions*. Champaign, IL : University of Illinois Press.

Cronbach, L. H. (1990). *Essentials of psychological testing*. New York : Harper & Row.

Dickes, P., Tournois, J., Flieller, A. & Kop, J.-L. (1994). *La psychométrie*. Paris : P.U.F.

Gendre, F. (1985). Vers une orientation professionnelle individualisée : la méthode « synthèse ». *Revue de Psychologie Appliquée*, 35 (1), 19 - 34.

Gendre, F., Capel, R. (1999). L'évolution de l'outil informatique modifiera-t-elle la fonction du psychologue conseiller ? *Actualités psychologiques*. 4 , 99 - 112.

Gendre, F., Capel, R. & Monod, D. (2002). L.A.B.E.L. Un instrument d'évaluation. de la personnalité à visée universelle. *Psychologie & Psychométrie*, 28, 1/2.

Gendre, F., Capel, R. & Oswald, R. (2003). L.A.B.E.L. Manuel práctico. São Paulo. Moityca.

Gendre, F. (1988). L'orientation rationnelle unifiée synthétisée (O.R.U.S.) : un modèle d'aide à la décision professionnelle assistée par ordinateur. *Revue de psychologie appliquée*, 38 (1), 5 - 30.

Gendre, F. & Capel, R. (2001). *LIMET, Manuel pratique*. Institut de psychologie, Université de Lausanne.

Holland. J. L. (1983). *Making career choices: a theory of personality types and work environments*. Englewood Cliffs, Prentice Hall.

Holt R.R. (1978). *Methods in Clinical Psychology: Assessment, Prediction and Research* (2 volumes). New York: Plenum.

Lavault, D. & Grégoire, J. (1997). Introduction aux théories des tests en sciences humaines. Bruxelles : De Boeck & Larcier.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading (Mass.) : Addison Wesley.

Michell, J. (2003). The quantitative imperative: positivism, naïve realism and the place of qualitative methods in psychology. *Theory and Psychology*, 13, 1, 5-32.

Rust, J. & Golombok, S. (1989). *Modern psychometrics*. London : Routledge.

Spearman, C. (1904). « General intelligence » objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Spearman, C. (1927). *The abilities of man*. New York : Macmillan.

Terman, L. M. (1916). *The measurement of intelligence*. Boston : Houghton Mifflin.

Thurstone, L. L. & Chave, E. J. (1929). *The measurement of attitude*. Chicago : University of Chicago Press.

Thurstone, L. L. & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric monographs*, 2.

Thurstone, L. L. (1959). *The measurement of values*. Chicago : University of Chicago Press.