

# MEDIDAS IPSATIVAS NA AVALIAÇÃO PSICOLÓGICA

Giselle Müller–Roger Welter - Universidade São Francisco  
Cláudio Garcia Capitão<sup>1</sup> - Universidade São Francisco

---

## RESUMO

O objetivo do presente artigo é fomentar a discussão sobre os instrumentos de avaliação psicológica no formato de escolha forçada multidimensional, ou medida ipsativa. A medida ipsativa é obtida por meio da ordenação de itens ou quando o respondente deve escolher os itens que “mais se parecem comigo” e itens “que menos se parecem comigo” em um conjunto de itens igualmente socialmente desejáveis. Esse formato contribui para a resistência ao falseamento das respostas ao dificultar a atribuição de um valor maior a si mesmo em cada construto contido em cada item. O artigo discute a possibilidade de se submeter dados ipsativos a critérios de validação psicométricos sob a ótica da Teoria Clássica dos Itens. As limitações dos dados ipsativos devem ser levadas em consideração ao se determinar as propriedades internas dos instrumentos, pois o valor das medidas ipsativas supera suas limitações psicométricas quando há risco de falseamento de respostas.

*Palavras-chave:* Avaliação psicológica; Escolha forçada; Medida ipsativa; Falseamento de resposta; Desejabilidade social.

## IPSATIVE MEASURES IN PSYCHOLOGICAL EVALUATION

### ABSTRACT

This article aims to promote the discussion on the multidimensional forced choice instruments, or ipsative measure, in psychological evaluation. Measure can be ipsative through the properties of the item response format such as rank-ordered scales or through forced choice responses in which the respondent must choose items that are “most like me” or “less like me” from a set of items with equal social desirability. The forced choice format was designed to control response distortions due to social desirability, by means of attributing greater value to oneself to each latent construct, represented by the items contained in each item set, with equivalent social desirability value. The article discusses psychometric limitations of ipsative measures considering the Classic Test Theory, which must be taken into account at the determination of internal properties of the instruments, since its value overcomes its limitations, when prevention of item response distortion is needed.

*Keywords:* Psychological evaluation; Forced choice; Ipsative measure; Response distortion; Social desirability.

---

## INTRODUÇÃO

O objetivo do presente artigo é fomentar a discussão sobre os instrumentos de avaliação psicológica no formato de escolha forçada. Os instrumentos de avaliação psicológica que visam apreender traços de personalidade caracterizam-se por inventários de autodescrição, medidas de interesses e atitudes, técnicas projetivas e *assessment*, ou testes de situação, como as técnicas de dinâmica de grupo (Anastasi, 2003). Em geral, para se obter o escore das diferentes dimensões avaliadas nos inventários autodescritivos adota-se um método normativo, baseado na múltipla escolha, como é o caso da escala Likert, na qual o respondente deve atribuir o grau de concordância ou discordância para cada item do inventário, adotando valores, por exemplo, de 1 a 5. Os testes de múltipla escolha permitem a realização de comparação interpessoal e intrapessoal e expressam

a qualidade e a intensidade dos sentimentos em

relação aos objetos ou pessoas, por meio de adjetivos e da intensidade dos advérbios (Anastasi, 2003; Dilchert & cols., 2006; Heggstad & cols., 2006a e b; Meade, 2004; Stark, Cherneyshenko & Drasgow, 2005; Sydell & Snell, 2003).

Enquanto as técnicas projetivas têm o seu uso predominante no contexto clínico, jurídico e em pesquisa, os demais instrumentos têm encontrado ampla aplicação no contexto organizacional, em especial no âmbito da seleção de pessoal, o que traz em seu bojo uma série de reflexões quanto à atitude do respondente frente à situação de avaliação. Os inventários autodescritivos são especialmente sujeitos à simulação de respostas tidas como desejáveis ou socialmente valorizadas, pois o respondente tem interesse em criar uma impressão favorável, enquanto candidato a uma vaga (Anastasi, 2003; Dilchert & cols., 2006; Heggstad & cols., 2006a e b; Meade, 2004; Stark, Cherneyshenko & Drasgow, 2005; Sydell & Snell, 2003). Em outras situações, o testando pode buscar

---

<sup>1</sup> Contato:

E-mail: gcgcapitao@uol.com.br

um parecer que ateste uma perturbação psicológica, dependendo do seu interesse, como é o caso de criminosos acusados, recrutas ou funcionários que buscam o afastamento do posto de trabalho (Anastasi, 2003).

Baron (1996), Stark, Cherneyshenko & Drasgow (2005) também chamam atenção para a possibilidade de os respondentes darem respostas diferentes do escore verdadeiro, consciente ou inconscientemente, ao responder a escalas Likert, com desvios de tendência central, evitando as categorias de respostas extremas (respostas aquiescentes com a afirmativa apresentada de desejabilidade social ou refletindo baixa auto-estima, passando imagem mais negativa de si). No entanto, a tendência de escolher respostas tidas como socialmente convenientes não indica, necessariamente, um falseamento deliberado, mas pode refletir falta de autoconhecimento, resistência em aceitar as próprias limitações ou uma autopercepção distorcida. Em contrapartida, autodescrições desfavoráveis podem ter como objetivo evocar reações de simpatia e a prestação de ajuda. Sydell e Snell (2003) julgam importante considerar que os indivíduos que não respondem a um determinado conjunto de itens da mesma maneira em diferentes situações, questionando a validade do conceito de escore verdadeiro na avaliação da personalidade, pois à medida que a motivação em relação à situação de avaliação aumenta, diminui a validade de critério da medida. Os autores destacam a necessidade de se compreender o contexto organizacional para que se possa descobrir uma relação válida e significativa entre o preditor e o critério.

Diversos estudos comprovaram o êxito com que é possível fraudar inventários de resposta (Anastasi, 2003; Dilchert & cols., 2006; Heggstad & cols, 2006a e b), o que, certamente, contribuiu amplamente para o descrédito atribuído aos testes psicológicos nas décadas de 70 e 80 (Braidó, 2002; Heggstad, 2006). No entanto, o interesse por medidas da personalidade ressurgiu nos últimos anos a partir de estudos que mostraram que os escores nos testes de personalidade conseguem prever com segurança o desempenho no contexto educacional e ocupacional (Stark, Cherneyshenko & Drasgow, 2005).

A pesquisa sobre a desejabilidade social ganha importância ao se lidar com questões práticas enfrentadas pelas organizações que usam inventários de personalidade para melhorar o processo de tomada de decisão. Como são provas

autodescritivas, os inventários de personalidade podem ser fortemente influenciados pela desejabilidade social. Muitas organizações consideram essa possibilidade inevitável, o que diminui significativamente a instrumentalidade das medidas de personalidade enquanto ferramentas úteis no processo de tomada de decisão (Ellingson & Heggstad, 2003).

Considerando a retomada do interesse pelos testes de personalidade e a relativa facilidade com que os respondentes identificam as respostas corretas, ou esperadas, nas escalas autodescritivas, torna-se necessário desenvolver formas alternativas de construção de itens de testes de personalidade à prova de falseamento das respostas. Quando há forte motivação por parte do testando em apresentar um bom desempenho, ele pode distorcer propositadamente suas respostas para causar uma boa impressão (Sydell & Snell, 2003).

Pesquisas sobre a desejabilidade social geraram muitas alternativas para diminuir o impacto do falseamento da resposta, como avisos sobre as conseqüências associadas à distorção das respostas e que estas podem ser detectadas, construção de medidas no formato de escolha forçada ou instrumentos contendo itens sutis e a aplicação de corretivos de desejabilidade social e reteste (Ellingson & Heggstad, 2006).

A detecção do falseamento das respostas pode ser feita por meio da construção de itens relativamente “sutis” ou socialmente neutros e pela inclusão de chaves de verificação para detectar fraudes nos grupos de resposta (Anastasi, 2003). Como exemplo de cuidado na construção dos itens, Del Prete e Del Prete (2003), no IHS – Inventário de Habilidades Sociais, procuraram minimizar o efeito da desejabilidade social no instrumento, redigindo 15 dos 38 itens que o compõem com fraseado negativo, fazendo que a reação sugerida seja indicadora de falta de habilidade. Nesses 15 itens, ao contrário dos demais, as freqüências maiores são interpretadas como indicadoras de déficit, enquanto as menores são indicadoras de recursos em habilidade social. Esse procedimento requer a inversão do valor atribuído ao item na computação dos escores. No IFP – Inventário Fatorial de Personalidade (Pasquali, 1997) foram adotadas duas escalas de controle, uma para avaliar a consistência das respostas (Validade) e outra para avaliar a Desejabilidade. Escores elevados em Validade invalidam o resultado do teste, pois as respostas são consideradas inverídicas, enquanto que escores elevados na escala Desejabilidade

Social indicam que o testando procurou apresentar-se de acordo com as expectativas externas, ao invés de expressar as próprias necessidades, o que faz com que o resultado obtido deva ser analisado com cautela.

Cattell (Russell & Karol, 2004) introduziu no 16PF três escalas de controle: AI – Administração da Imagem, NF – Não Frequência e AQ- Aquiescência. Percentis iguais ou acima de 95 nessas escalas apontam para a necessidade de questionar os resultados do teste, anular o protocolo ou, eventualmente, fazer uma retestagem, dependendo do contexto da avaliação. No entanto, Ellingson & Heggstad (2003) consideram que, apesar das instruções contidas nos manuais sobre escalas de controle e corretivas, pouca informação consta sobre o procedimento a ser adotados pelas organizações ao se depararem com a suspeita de falseamento das respostas. Em geral, elas optam pelo reteste, embora não haja pesquisas sobre a confiabilidade desses resultados para avaliar as implicações ao permitir que um indivíduo refaça uma medida de personalidade.

Chan (2003) menciona que vários métodos têm sido empregados com o objetivo de minimizar os efeitos do falseamento das respostas em inventários autodescritivos, visando aumentar a validade do teste, destacando dois em especial. O primeiro é o controle do falseamento por meio da construção cuidadosa dos itens que compõem o experimento, questionário ou inventário, citando a orientação de Skinner no sentido de minimizar o efeito da deseabilidade social mediante questões simples e objetivas, o que aumenta a confiabilidade do teste e diminui a deseabilidade das respostas. O segundo é o controle estatístico, excluindo o fator principal e separando os fatores com deseabilidade social.

Uma outra possibilidade é a prevenção do falseamento das respostas por meio da técnica de escolha forçada, denominada medida ipsativa, amplamente utilizada nos testes em uso no contexto organizacional, como o Quati (Zacharias, 2000), MBTI (Bents & Blank, 2001; Costa, 2006), DISC (Harazim, 2006; Hicon, 1999; Price, 2006; SuccessInsights, 2006), VECA e Web Veca (Alba Consultores, 2006), LABEL (Moytica Desenvolvimento e Consultoria, 2005), entre outros. Nesse caso, o respondente deve escolher entre dois ou mais itens, adotando como critério de escolha aqueles que mais se parecem com ele. Em alguns casos o respondente deve fazer uma escolha positiva, indicando os itens preferidos em meio a

um conjunto. Em outros casos deve fazer uma escolha positiva e uma negativa, indicando os itens que mais e menos se parecem com ele. Uma terceira forma de escolha forçada é colocar os itens numa ordem de preferência.

Heggstad e cols. (2006b) apontam o crescente otimismo em relação à possibilidade de diminuir a capacidade de testandos responderem de maneira socialmente desejável por meio da escolha forçada multidimensional, pois reduzem a inflação dos escores e apresentam maior validade de critério em comparação à escala Likert. Por outro lado, consideram que esse tipo de medida apresenta sérias limitações, pois não fornecem dados normativos, como a escala Likert, necessários para a realização de comparações interindividuais, presentes no contexto da seleção de pessoal. A escolha forçada multidimensional consiste de duas ou mais afirmações, cada uma representando um traço latente diferente, as quais o respondente deve escolher, ordenar ou indicar sua preferência entre as afirmações. Contribui para sua resistência ao falseamento das respostas o formato dos itens, pois ele dificulta a atribuição de um valor maior a si mesmo no construto contido em cada item, criando itens equivalentes em termos de deseabilidade social. Em não havendo itens mais ou menos favoráveis nos conjuntos de itens, a possibilidade de responder com base na deseabilidade social é consideravelmente reduzida.

Baron (1996) considera que as distorções decorrentes da deseabilidade social são menos frequentes nas escalas interdependentes das medidas ipsativas, pois nelas os respondentes são forçados a atribuir um valor diferente aos itens, sem a possibilidade de concordar com todos eles. Argumenta a favor dos testes de escolha forçada, pois, no momento em que o indivíduo tem que escolher entre dois itens igualmente desejáveis, está simulando uma situação semelhante ao processo de escolha real que o indivíduo tem que fazer na vida, na qual é forçado a fazer uma escolha, como entre dois objetos de compra, com quem casar, que emprego aceitar, que curso fazer e assim por diante. De modo alternativo pode ser uma escolha entre diferentes tipos de comportamento que são mais característicos do indivíduo (Heggstad & cols., 2006b, Karpaschof & Elkaer, 2000; Meade, 2004;). Além disso, as escolhas forçadas resultam numa maior diferenciação dos escores no respondente, pois não é possível lhes atribuir um valor igual. Baron (1996) coloca, ainda, que tanto as medidas normativas como as ipsativas estão sujeitas a

distorções de diferentes tipos, não sendo possível determinar a extensão e a frequência dos mesmos nos dois modos.

Enquanto os dados normativos atendem aos critérios psicométricos, podendo ser submetidos às análises estatísticas mais frequentemente empregadas, como a análise fatorial, por exemplo, diversos autores informam sobre as dificuldades existentes ao se submeter dados ipsativos a análises estatísticas que pressupõem uma distribuição normal das respostas (Baron, 1996; Bartram, 1996; Bartram & Hambleton, 2006; Chan. & Cheung, 2002; Hammond & Barrett, 1996, Karpaschof & Elkaer, 2000; Kayes, 2006; McCloy, Heggstad & Reeve, 2005; Meade, 2004). Embora a ipsatividade represente uma questão importante na psicometria, há carência de artigos sobre ela no Brasil, não tendo sido encontrado nenhuma referência usando as palavras chave ipsatividade e medida ipsativa na Biblioteca Virtual de Saúde - BVS-Psi (2006). Já no Scientific Electronic Library Online – Scielo (2006), empregando as palavras-chave ipsativo, ipsatividade, medida ipsativa e escolha forçada foram encontrados apenas três artigos, nos quais o conceito é mencionado superficialmente. Esse termo também aparece em dois artigos no âmbito da arquitetura e da educação física. No entanto, o termo ipsative aparece em 111 artigos indexados nos últimos 10 anos na base de dados PsycINFO (2006). Enquanto os termos ipsativo e ipsatividade não geraram nenhuma referência no site de busca Google (2006), o termo correspondente em inglês ipsativity gerou 279 achados.

Do ponto de vista matemático, os dados são considerados ipsativos quando a soma de um determinado conjunto de respostas sempre resulta no mesmo total e é igual para todos os respondentes (Dorsch, 1992; Meade, 2004). Embora muitos fatores possam contribuir para a criação de dados ipsativo, na prática o termo ipsativo é usado de uma maneira genérica como sinônimo de “dados interdependentes”. A interdependência dos itens resulta do processo cognitivo envolvido na tomada de decisão entre os itens do conjunto, que é afetada pelo nível latente dos construtos avaliados do respondente. Suas expectativas positivas e negativas em relação ao meio-ambiente, sua percepção de desejabilidade e suas hipóteses sobre as consequências da escolha de cada item podem afetar a consistência das respostas (Baron, 1996; Meade, 2004).

#### *Limites psicométricos das medidas ipsativas*

As interdependências presentes nas escalas ipsativas e nos escores ipsativos observados podem alterar as propriedades psicométricas da escala, pois o item selecionado não depende só do nível de latência do traço que está sendo medido pelo item, mas também do conjunto de itens do qual faz parte e das suas propriedades. Isso faz com que cada escore observado seja contaminado pelos demais escores do conjunto de itens. O processo de decisão envolve processos cognitivos complexos, que podem resultar no aumento do índice de erro aleatório, na dificuldade de simular consistentemente ou na tendência de afirmar o primeiro item do conjunto (Baron, 1996; Karpaschof & Elkaer, 2000; Meade, 2004). Os efeitos da interdependência se tornam maiores quando o instrumento é constituído de poucas escalas e há elevada correlação entre os construtos medidos. Meade (2004) considera que, com aproximadamente 30 escalas, já é possível obter-se parâmetros psicométricos passíveis de interpretação. Baron (1996) julga que a adoção de um número consideravelmente maior de escalas permite que estas possam se assemelhar aos dados normativos. Com mais escalas, para a maioria dos indivíduos, os perfis normativos e ipsativos podem se assemelhar e ser interpretados da mesma maneira, com coeficientes de validade similares, embora os dados ipsativos apresentem a vantagem de controlar o viés inerente às respostas do tipo Likert, descrito anteriormente.

A problemática ipsativa está pautada na violação de um dos principais pressupostos da teoria clássica dos testes que é a independência da variância do erro, pois a correlação negativa que se aproxima de zero aumenta à medida que aumenta o número de escalas do teste. Além disso, não atende ao critério paramétrico padronizado, pois o nível de medida não é ordinal, e nem escalar.

Segundo Baron (1996), os argumentos a favor da medida ipsativa partem do princípio de que é lícito tratar a maioria dos métodos de medida em psicologia como uma medida escalar, desde que a informação obtida se mostre útil. Para ela, os argumentos contrários rejeitam a realização de análises padronizadas em medidas ipsativas, com base em medidas individuais ou por razões puramente teóricas, desconsiderando a possibilidade de se obter resultados passíveis de interpretação. Entretanto, julga necessário reconhecer que a interdependência das escalas nas medidas ipsativas cria dificuldades maiores do que as encontradas nas escalas Likert, o que traz

implicações importantes para a análise estatística e na análise de escores ipsativos.

Segundo Baron (1996) o escore ipsativo pode ser entendido como um escore normativo controlado, equivalente para o escore total em todas as escalas. Como não há medida de escore total em todas as escalas, ele mostra quais traços são mais fortes e mais fracos no indivíduo. É possível observar o formato do perfil, mas não os valores absolutos das diferentes escalas, permitindo comparar as diferenças intra-individuais, mas não as inter-individuais. Muitas vezes a normalização dos dados ipsativos pode não ser apropriada, pois impossibilita comparações intra-individuais. A distribuição dos escores ipsativos pode ser afetada pelo número de escalas do teste e pela correlação entre as escalas. Quando todas as escalas se correlacionam positivamente, é mais provável que o indivíduo apresente um perfil com valores predominantemente positivos ou negativos. Assim, Baron recomenda que se considere a expectativa de viés nas respostas normativas, a relativa ameaça à validade devido aos diferentes tipos de distorção e os possíveis meios de controle ao se escolher o tipo de medida.

Segundo McCloy, Heggstad & Reeve (2005), o fato de os itens competirem entre si tem implicações nas intercorrelações dos itens e, conseqüentemente, na análise de validade e da fidedignidade. Devido à baixa intercorrelação dos itens, as subescalas também tendem a apresentar baixa consistência interna. No entanto, julgam importante considerar que a baixa consistência interna, estimativa de fidedignidade, e as baixas correlações com outras medidas são decorrentes do procedimento de obtenção do escore como, por exemplo, por meio da escolha forçada.

A consistência interna é um índice que informa o quão bem os itens de uma escala medem o mesmo construto e está associada à fidedignidade do instrumento, sendo especialmente utilizada na área de personalidade. É útil para caracterizar a área do comportamento ou o traço do teste medido pelo teste, pois tem alguma significação para a sua validade de construto. A fórmula alfa é uma das muitas análises que podem ser empregadas para atestar a fidedignidade de medidas psicológicas ou educacionais, tendo sido inicialmente concebida para ser aplicada na comparação dos resultados obtidos por pessoas em duas condições diferentes. Valores elevados são obtidos quando os itens apresentam correlação elevada, indicando que os itens medem o mesmo construto, enquanto valores

baixos indicam que os itens não estão medindo um construto claramente ou que estão medindo mais de um construto, os quais estão correlacionados (Cronbach, 2006; Hammond & Barrett, 1996; Price, 2006).

Clark e Watson (1995) questionam a utilidade do índice alfa para se medir a fidedignidade de dados ipsativos. Consideram que valores elevados de alfa depõem contra a validade do construto ao invés de corroborá-lo, pois quando a escala consiste de poucos itens fortemente correlacionados, há uma atenuação da medida do construto em função da redundância dos itens. Paradoxalmente, consideram que uma escala fornece muito mais informações, contribuindo para uma melhor medida do construto, quando contém itens mais diferenciados, ou seja, que apresentam uma intercorrelação moderada, pois a maximização da consistência interna faz com que a escala se torne muito fechada em relação ao construto que pretende medir. Clark e Watson sintetizam sua opinião colocando que para que se possa garantir a unidimensionalidade, praticamente todas as correlações inter-itens, que são uma medida direta da consistência interna, deveriam ser moderadas quanto à magnitude e se agrupar ao redor do valor médio.

No caso de testes ipsativos, os valores de consistência interna são atenuados devido ao fato de a maioria das correlações inter-itens serem negativas, fazendo com que escores elevados em uma determinada dimensão provoquem escores baixos nas outras dimensões (Kays, 2006). Escalas com mais itens sofrem maior atenuação do valor de consistência interna, pois há maior competição entre os itens de uma determinada escala. A fidedignidade por meio da consistência interna dos testes com características ipsativas é em geral baixa, com valores medianos ao redor de 0,20, o que pode, contribuir, também, para a redução dos índices de correlação com outros instrumentos (Greer & Dunlap, 1997; McCloy, Heggstad & Reeve, 2005). Para Meade (2004) a estimativa da consistência interna, a partir de fórmulas derivadas da Teoria Clássica dos Testes, não é compatível com dados ipsativos.

Cronbach (2006), ao refletir sobre o coeficiente alfa e os procedimentos que o sucederam, considera que a fórmula alfa é apenas uma das diversas análises que podem ser utilizadas para medir a fidedignidade das medidas psicológicas e educacionais, além da correlação com outros instrumentos, re-teste e medida do erro

padrão. Uma outra questão que se coloca é a ausência de padrões claros sobre qual seria o nível de fidedignidade considerado aceitável. Embora o valor mínimo recomendado seja 0,80 e 0,90 para pesquisa básica e aplicada, respectivamente, o fato de pesquisadores contemporâneos considerarem adequados valores alfa em torno de 0,60 e 0,70 em dados ipsativos adequados não é incomum (Clark & Watson, 1995; Hammond & Barrett, 1996; Meade, 2004). Em função das controvérsias existentes em torno da medida de precisão, muitos psicometristas deixaram de adotar índices de fidedignidade, ou consistência interna, como forma de estabelecer a homogeneidade de uma escala ipsativa, sendo que explicitamente recomendam a adoção de índices de estabilidade como o teste-reteste em detrimento do cálculo dos valores de alfa nesses casos (Hammond & Barrett, 1996).

Embora, as evidências obtidas nos estudos realizados com dados ipsativos relatados por Baron (1996) indicam que há pouca diferença no tocante à fidedignidade entre as escalas ipsativas e normativas, o mesmo não acontece na análise fatorial. Os estudos realizados sobre o efeito da análise fatorial sobre dados ipsativos relatados por ela, mostraram ser possível obter o mesmo número de fatores que os obtidos com medidas normativas, quando se emprega mais de 400 itens. No entanto, como dificilmente se aplicam questionários com mais de 30 itens, os dados ipsativos apresentam poucas chances de contribuir para a compreensão da estrutura dos construtos subjacentes. Baron (1996) atribui maior importância à validade externa das medidas ipsativas, enquanto preditora do comportamento.

Karpatschhof e Elkjaer (2000), em seu relato de pesquisa, discutem os estudos realizados por Baron (1996) com objetivo de verificar se o uso de dados ipsativos na comparação interpessoal é totalmente ilegítimo, como afirmam os opositores, ou se há um argumento racional favorável à comparação interpessoal ipsativa. Fornecem suporte psicométrico para os adeptos dos testes ipsativos mediante evidências de validade obtidas tanto por meio da abordagem teórica e como da investigação empírica. Seus estudos foram realizados considerando três perspectivas de análise: a perspectiva formal da matemática (teoria da probabilidade), a perspectiva da simulação e a perspectiva empírica. Os resultados obtidos mostraram que, apesar da expectativa contrária, os testes ipsativos são úteis para avaliar diferenças individuais.

A TRI (Teoria de Resposta ao Item) constitui uma alternativa bastante promissora na construção de escalas ipsativas, nas quais há pareamento de itens de preferência multidimensionais, ou dicotômicos. Ao ter que escolher a opção que mais se parece com ele, quando o respondente se depara com um par de afirmações relativas à personalidade (estímulo), A e B, faz uma avaliação de cada estímulo separadamente para tomar uma decisão quanto ao endosso ou não dos mesmos. Caso o respondente deseje endossar ou rejeitar ambos os estímulos, terá que reavaliá-los para, então, apresentar sua preferência. A TRI permite o estabelecimento de diferentes pesos aos itens que compõem o banco de itens em relação à desejabilidade social, possibilitando o pareamento adequado deles. Já a FRI (Função de Resposta ao Item) representa a probabilidade de o respondente preferir o estímulo A em relação ao estímulo B, considerando os parâmetros do estímulo e os valores dos traços latentes (McCloy, Heggstad & Reeve, 2005; Stark, Cherneyshenko & Drasgow, 2005).

A identificação de diferentes graus de desejabilidade dos itens permite que itens semelhantes sejam pareados quanto ao grau e dimensão que representam, bem como parear itens de diferentes dimensões que compartilham o mesmo grau de desejabilidade, como apontado por Heggstad e cols. (2006b). A escala ipsativa construída dessa maneira deveria conter tanto itens unidimensionais, como itens multidimensionais, porém sempre pareando itens com grau de desejabilidade semelhante. Para que se possa identificar o traço latente e realizar comparações interindividuais, o que não é possível com as escalas ipsativas tradicionais, é importante que haja ao menos um pequeno número de pareamentos unidimensionais. Um outro aspecto a ser considerado diz respeito à possibilidade de se estimar parâmetros para os diferentes estímulos (itens) de cada dimensão separadamente, por meio da calibração das respostas aos itens unidimensionais, o que permite que esse modelo seja denominado multi-unidimensional (Stark, Cherneyshenko & Drasgow, 2005).

#### CONSIDERAÇÕES FINAIS

Embora haja discordância entre a posição dos diferentes autores quanto à possibilidade de se submeter dados ipsativos a critérios de validação psicométricos, há convergência entre eles no que

diz respeito à necessidade de se dispor de métodos adequados para a construção de itens de testes de personalidade que não possam ser facilmente falseados. Fora do âmbito da Teoria Clássica dos Testes, aumenta o interesse pela construção de testes ipsativos, ou de escolha forçada, o que traz consigo dificuldades psicométricas que precisam ser superadas. É nesse cenário que a TRI representa um avanço significativo, cuja aplicação deve ser estendida a novos contextos da avaliação da personalidade. A controvérsia na interpretação apropriada das propriedades psicométricas dos instrumentos de avaliação da personalidade no formato de escolha forçada é, portanto, bastante atual, constituindo objeto de debate sobre suas implicações teóricas e empíricas. Os autores pesquisados concordam que as limitações dos dados ipsativos devem ser levadas em consideração ao se buscar determinar as propriedades internas dos instrumentos de avaliação psicológica, sugerindo que o valor das medidas ipsativas supera suas limitações psicométricas quando há risco de falseamento de respostas. Há, portanto, necessidade de ampliação de pesquisas psicométricas com o objetivo de explorar a acuidade dos procedimentos multi-unidimensionais, com duas ou mais dimensões. Com o objetivo de examinar diretamente a questão da adequação de dados ipsativos para a avaliação de diferenças interindividuais, poderão ser conduzidas pesquisas, comparando a precisão da medida, empregando os mesmos itens administrados no formato Likert e no formato escolha forçada multidimensional.

#### REFERÊNCIAS

- Alba Consultoria (2007). *Web-Veca*. Disponível em [http://www.albaconsultoria.com.br/produtos\\_veca.htm](http://www.albaconsultoria.com.br/produtos_veca.htm). Acessado em 21 de fevereiro de 2007.
- Baron, H. (1996). Strengths and Limitations of Ipsative Measurement. *Journal of Occupational and Organizational Psychology*, 69, 49-56. Public domain. Disponível em <http://www.psychology.org.nz/industrial/Baron>. Acessado em 10 de outubro de 2005.
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, 69, p. 25-39.
- Bartram, D. & Hambleton, R. K. (2006). *Computer-Based Testing and the Internet*. West Sussex: John Wiley & Sons, Ltd
- Bentler, P. & Chan, W. (1996). Covariance Structure Analysis of Partially Additive Ipsative Data Using Restricted Maximum Likelihood Estimation. *Journal article by Multivariate Behavioral Research*, Vol. 31, 1996. Disponível em <http://www.questia.com/PM.qst?a=o&se=gglsc&d=96264734>. Acessado em 6 de outubro de 2006.
- Bents, R. & Blank, R. (2001). *Der MBTI: die 16 Grundmuster unseres Verhaltens nach C. G. Jung; eine dynamische Persönlichkeitstypologie*. München: Claudius.
- Braido, A. N. G. (2002). *Validação preliminar do inventário de adjetivos de personalidade (PACL) em universitários*. Dissertação de mestrado da Universidade São Francisco. Itatiba.
- Costa, C. (2006). *Análise de tipos psicológicos*. Disponível em [http://www.gestaoerh.com.br/site/visitante/artigos/rese\\_003.php](http://www.gestaoerh.com.br/site/visitante/artigos/rese_003.php). Acessado em 5 de outubro de 2006.
- Chan, W. & Cheung, M. W. L. (2002). Reducing Uniform Bias with Ipsative Measurement in Multiple-Group Confirmatory Factor Analysis em *Structural Equation Modeling*, 9 (1) p. 55-77.
- Chan, W. (2003) Analyzing ipsative data in psychological research. *Behaviormetrika*, 30, 1, 99-121. Disponível em <http://www.psychology.org.nz/industrial>. Acessado em 12 de maio de 2006.
- Cronbach, L. J. (2006). *My Current Thoughts on Coefficient Alpha and Successor Procedures*. Disponível em [http://www.stanford.edu/dept/SUSE/SEAL/Reports\\_Papers/My%20Current%20Thoughts\\_Submit.doc](http://www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/My%20Current%20Thoughts_Submit.doc). Acessado em 01/12/06.
- Del Prete, Z.A. P. & Del Prete, A. (2005). *IHS – Inventário de Habilidades Sociais*. 3ª edição. São Paulo: Casa do Psicólogo.
- Dilchert, S., Ones, D. S., Visweswaran, C. & Deller, J. Response distortion in personality measurement: born to deceive, yet capable of providing valid assessments? *Psychology Science*, Volume 48, 2006 (3), P. 209-225. Disponível em [http://www.psychology-science.com/3-2006/ps\\_3\\_2006\\_209-225.pdf](http://www.psychology-science.com/3-2006/ps_3_2006_209-225.pdf). Acessado em 6 de outubro de 2006.
- Ellingson, J. E. & Heggstad, E. D. (2003). The viability of Retesting and Warnings for Controlling Faking Tendencies. *Practical*

- considerations for Implementing Personality Testing in Organizations. Symposium conducted at the annual meeting of the society for Industrial and Organizational Psychology, Orlando, FL.
- Hammnod, S. & Barrett, P. (1996). The Psychometric and Practical Implications of the use of Ipsative, forced-choice formal, Quotionaires. *The British Psychological Society: Occupational Psychology Conference. Book of Proceedings*, p. 135-144.
- Harazim, P. (2006). *Alavancando o capital humano*. Disponível em <http://www.hicontreinamento.com.br/artigos.asp>. Acessado em 5 de outubro de 2006.
- HICON Treinamento (1999). *Gerenciando comportamentos*. Human Perspectives Inc. Exclusive Distributor – Brazil. © 1994 Carlson Learning Company & the Gary Little Company Inc.
- Heggstad, E. D. (2006). Summary Abstract. *Practical considerations for Implementing Personality Testing in Organizations*. Symposium conducted at the annual meeting of the society for Industrial and Organizational Psychology, Orlando, FL.
- Heggstad, E. D., Morrison, M., Reeve, C.L. & McCloy, R.A. (2006a) Forced-choice assessments of personality for selection: evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91, 1, 9-24.
- Heggstad, E. D., Morrison, M., Reeve, C.L. & McCloy, R.A. (2006b) *Do Mutidimensional Forced-Choice Measures Provide Normative Data? Practical considerations for Implementing Personality Testing in Organizations*. Symposium conducted at the annual meeting of the society for Industrial and Organizational Psychology, Orlando, FL.
- Karpatschof, B. & Elkær, H.K. (2000). *Yet the bumblebee Flies – The reliability of Ipsative Scores – Examined by Empirical Data and a Simulation Study*. Research report. Department os Psychology. University of Copenhagen. Disponível em <http://www.psy.ku.dk/forskning/Udgivelser/Forskrap/forskrap1.pdf>. Acessado em 10 de outubro de 2005.
- Kayes, D. C. (2005). Internal Validity and Reliability of Kolb’s Learning Style Inventory Version 3 (1999). *Journal of Business and Psychology*, 29 (2), 249-257.
- McCloy, R., Heggstad, E. D. & Reeve, C. L. (2005). A Silk Purse From the Sow's Ear: Retrieving Normative Information From Multidimensional Forced-Choice Items. *Organizational Research Methods*. 2005; 8: pp. 222-248. <http://orm.sagepub.com/cgi/content/abstract/8/2/222> Acessado em 6 de outubro de 2006.
- Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77, p. 531-552.
- Moytica Desenvolvimento e Consultoria (2005). *L.A.B.E.L.* Disponível em [www.moityca.com.br/](http://www.moityca.com.br/). Acessado em 20 de abril de 2005.
- Pasquali, L., Azevedo, A.M. & Ghesti, I. (1997). *IFP - Inventário Fatorial de Personalidade: manual técnico e de avaliação*. São Paulo: Casa do Psicólogo.
- Price, L.R. (2006). *DISC instrument Validation Study – Technical Report*. The Institute for Motivational Living, Inc. Disponível em [http://www.discinsights.com/cyber/images/web\\_IML\\_DISC\\_Validation\\_Study\\_2006.pdf](http://www.discinsights.com/cyber/images/web_IML_DISC_Validation_Study_2006.pdf). Acessado em 14/03/07.
- Russell, M. T. & Karol, D. L. (2002). *16PF, quinta edição: manual*. Atualizado e adaptado pelo Departamento de Estudos do CEPA. Rio de Janeiro: CEPA.
- Stark, S., Cherneyshenko, O. S. & Drasgow, F. (2005) An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, 29, 3, 184-203
- SuccessInsights (2006). *SuccessTools*. Disponível em <http://www.successtools.com.br>. Acessado em 6 de outubro de 2006.
- Sydell, E. J. & Snell, A. F (2003). The influence of Test-Taker Characteristics and Perceptions On the Validity of Personality Measures. *Practical considerations for Implementing Personality Testing in Organizations*. Symposium conducted at the annual meeting of the society for Industrial and Organizational Psychology, Orlando, FL.
- Zacharias, J. J. M. (2000). *QUATI: Questionário de Avaliação Tipológica (versão II)*. São Paulo: Vetor.

**SOBRE OS AUTORES:**

*Giselle Müller-Roger Welter*: Psicóloga, Especialista em Psicologia Escolar e Educacional, Mestre em Psicologia pela Universidade São Francisco. Orientadora Profissional, consultora organizacional e *coach*. Dedicou-se à tradução e adaptação de testes e livros de psicologia.

*Claudio Garcia Capitão*: Psicólogo, Especialista em Psicologia Clínica e Hospitalar, Pós-doutorado em Psicologia Clínica PUC-SP, docente dos cursos de graduação e de pós-graduação *stricto sensu* em Psicologia da Universidade São Francisco.