

Significância Clínica e Mudança Confiável na Avaliação de Intervenções Psicológicas¹

Zilda Aparecida Pereira Del Prette²
Almir Del Prette
Universidade Federal de São Carlos

RESUMO - As evidências da efetividade de qualquer intervenção psicológica baseiam-se na demonstração de sua validade interna e externa, usualmente efetuadas por meio de estatística inferencial. Uma alternativa mais recentemente explorada, especialmente no âmbito da pesquisa em psicoterapia, é a demonstração da confiabilidade e significância clínica das mudanças. Dentre os métodos utilizados na produção desses indicadores, destaca-se o desenvolvido por Jacobson e Truax (1991), mais conhecido como “Método JT”, ainda com limitada divulgação em nosso meio. Este estudo apresenta, resumidamente, a racionalidade desse método e algumas questões metodológicas e práticas pertinentes ao uso desse método, exemplificando sua aplicabilidade no tratamento de dados de uma intervenção fictícia.

Palavras-chave: Método JT; validade interna; validade externa; psicoterapia; intervenção psicoeducacional.

Clinical Significance and Reliable Change in Evaluating Psychological Interventions

ABSTRACT - The evidences of the effectiveness of any psychological intervention requires the demonstration of its internal and external validity, what is usually made by inferential statistical methods. An alternative, more recently explored, especially in the realm of the psychotherapeutic processes, is based on indexes of reliability and clinical significance of the changes. Among the methods used to produce these indexes, it is highlighted that one developed by Jacobson and Truax (1991), known as “JT Method”, still rare in the Brazilian psychological research. This study briefly presents the rational and some methodological and practical issues concerned with this method, exemplifying its use in the data analysis of a fictitious intervention.

Keywords: JT Method; internal validity; external validity; psychotherapy; psychoeducational intervention.

A preocupação com a efetividade dos atendimentos psicológicos é pertinente tanto aos profissionais que atuam em diferentes campos, quanto à clientela atendida, especialmente considerando-se o contexto atual de reconhecimento (inclusive legal) dos direitos do consumidor e de consolidação de uma cultura de avaliação sobre a prestação de serviços. Esse contexto avaliativo, presente na cultura da maioria dos países desenvolvidos, cria o desafio de um progressivo aperfeiçoamento dos procedimentos e práticas que subsidiam a formação e a atuação dos profissionais e técnicos.

Na Psicologia, e particularmente no campo da psicoterapia (Estados Unidos, Canadá e Europa), pode-se identificar, nos últimos vinte anos, um investimento crescente em métodos e critérios para avaliar a efetividade das práticas clínicas e para identificar procedimentos efetivos (Levant, 2005). A preocupação com procedimentos sustentados por

evidências de efetividade tem também recebido atenção, em anos recentes, no contexto da Educação Especial (Brantlinger, Jumenez, Pugach & Richardson, 2005; Gersten, Fuchs, Coyne, Greenwood & Innocenti, 2005; Thompson, Diamond, McWilliam, Snyder & Snyder, 2005).

Ainda que, idealmente, a formação universitária deva habilitar o psicólogo como profissional de serviços e como pesquisador, os que se inserem no mercado de trabalho (funcionários de empresas ou autônomos) parecem preocupar-se primariamente com a solução de queixas ou problemas mais do que com a produção e divulgação de conhecimento. Por outro lado, o pesquisador interessado na aplicação da Psicologia comumente investe na produção de conhecimento sobre: (a) aperfeiçoamento de procedimentos efetivos; (b) explicitação de relações entre os resultados obtidos e as variáveis de intervenções; e (c) metodologia de avaliação das intervenções, o que inclui definição e refinamento dos critérios usados para aferir resultados.

Historicamente, à medida que a Psicologia se tornou uma disciplina aplicada, disseminou-se a preocupação com os resultados por ela obtidos. Por diversas razões, inclusive pela necessidade de melhorar a formação profissional, a preocupação com a avaliação das intervenções clínicas e psicoeducacionais, justifica-se a busca de alternativas de avaliação e aperfeiçoamento desses serviços. A questão da efetividade de qualquer intervenção (clínica, educacional, organizacional etc.) implica, necessariamente, reunir evidên-

1 Apoio do CNPq (Bolsa de Pesquisador para os autores). Uma versão inicial deste estudo foi apresentada, pela primeira autora, no II Congresso “Psicologia, Ciência e Profissão”, em mesa redonda intitulada *Clínicas-Escola de Psicologia: Questões relevantes*, coordenada pela Profa. Dra. Edwiges Silveiras. Os autores agradecem os componentes da mesa. O manuscrito foi lido pelas professoras Dra. Deisy das Graças de Souza e Dra. Ana Lucia de Oliveira Ulian, a quem os autores também agradecem as sugestões.

2 Endereço para correspondência: Universidade Federal de São Carlos, Departamento de Psicologia. Via Washington Luiz, Km 235. Caixa Postal 676. CEP 13565.905. Fax/Fone 16-33518361. E-mail: zdprette@ufscar.br. Web-page: <http://www.rihs.ufscar.br>.

cias empíricas pertinentes (Cozby, 2001/2003) sobre: (a) sua *validade interna*, ou seja, o quanto os resultados podem ser atribuídos aos procedimentos utilizados; e (b) sua *validade social ou externa*, em termos de impacto sobre o funcionamento adaptativo do cliente, generalização para outros ambientes ou para população mais ampla, aceitabilidade do tratamento, relação custo-benefício etc.

Na Psicologia Clínica, surgiram várias propostas para a análise da efetividade das intervenções, particularmente voltadas para a questão da significância clínica dos resultados obtidos³. Entre tais propostas, pode-se destacar a de Jacobson, Follette e Revenstorf (1984), retomada posteriormente por Jacobson e Truax (1991), que ficou conhecida como “Método JT”, de acordo com McGlinchey, Atkins e Jacobson (2002). Esse método articula a análise da significância clínica (mais voltada para a validade externa) com a verificação da confiabilidade das mudanças obtidas (mais relacionada à validade interna). Dada a reconhecida importância do “Método JT” (Currie, Wilson & Curran, 2002; Evans, Margison & Barkham, 1998; Middel & van Solderen, 2002) e sua potencial aplicabilidade a outros campos (como, por exemplo, a Educação), este ensaio visa apresentar, resumidamente, a racional, os principais conceitos e o procedimento geral de tratamento de dados preconizados por esse método.

Efetividade da Intervenção: Validade Interna e Externa

Conforme já referido, a validade interna de uma intervenção implica demonstrar que os resultados obtidos (variável dependente) podem ser atribuídos aos procedimentos e técnicas utilizados (variável independente), e não a outras variáveis ou condições (história, maturação, efeito da testagem ou instrumentação, viés de seleção etc.) que comprometem essa validade (Kazdin & Weiss, 2003).

Nos estudos realizados com sujeito único, a validade interna é geralmente testada e controlada por meio de técnicas de reversão (etapas de não tratamento intercaladas com etapas de tratamento) e procedimentos de linha de base múltipla (introdução do tratamento em etapas subsequentes com diferentes indivíduos ou sobre diferentes aspectos do problema). Quanto à validade externa, enquanto generalização para a população, assume-se que depende da replicação dos efeitos em novos casos.

Em delineamentos experimentais de grupo (Cozby, 2001/2003) envolvendo grupos de validação social ou de controle (que não passam pela intervenção), a validade interna é geralmente verificada usando-se técnicas estatísticas inferenciais, com base em medidas de tendência central (média, mediana) e de dispersão (desvio padrão, erro padrão) dos resultados do grupo. Por meio dessas técnicas estatísticas, avalia-se a probabilidade de ocorrência das diferenças entre a pré e a pós-intervenção, ou seja, se são suficientemente robustas para se descartar a hipótese de representarem meras oscilações atribuíveis a erro de medida e, portanto, aceitar que há mudanças, atribuíveis às condições de intervenção. Nesses delineamentos, a validade externa, principalmente em termos

de generalização, depende das características de amostragem dos grupos experimental e controle (ou seja, o quanto a amostra é representativa da população mais ampla) e também da replicação desses resultados em novas intervenções.

Nos estudos conduzidos fora do laboratório, principalmente em condições práticas de atendimento psicológico, além das dificuldades de se usar esses controles, tem-se ainda um conjunto de questionamentos pertinentes à validade interna e externa, por exemplo, os efetuados por Jacobson e Truax (1992). No caso dos delineamentos de grupo, um dos questionamentos refere-se à variabilidade dos resultados entre participantes de uma mesma intervenção. Do ponto de vista estatístico, essa variabilidade representa um problema, na medida em que pode comprometer a significância dos resultados. O mesmo raciocínio seria aplicável quando vários indicadores de mudança são avaliados e computados globalmente (escore geral, média), obscurecendo alterações importantes em indicadores específicos. No entanto, do ponto de vista clínico, essa variabilidade representa um dado importante em termos heurísticos: a identificação de participantes que apresentaram melhoras, piores ou manutenção do transtorno sugere análises sobre as condições que maximizam ou reduzem a efetividade dos procedimentos utilizados.

Um outro questionamento, mais central tanto no delineamento com sujeito único como no delineamento de grupo, refere-se à relevância das mudanças obtidas, mesmo quando estatisticamente significativas. O interesse, nesse caso, recai sobre o significado e a funcionalidade adaptativa das mudanças do indivíduo, o que não é necessariamente garantido pela significância estatística.

Ainda que reconhecendo a necessidade de indicadores quantitativos, Jacobson e Truax (1991) questionam as análises baseadas em resultados médios de grupo ou no “tamanho do efeito” (obtido na meta-análise de resultados de vários estudos), uma vez que permitem aferir a significância estatística das diferenças (entre pré e pós-intervenção, ou entre grupo alvo de intervenção e grupo “não tratado”), mas pouco informam a respeito: (a) do significado *clínico* de tais diferenças; (b) do significado dessas diferenças em termos de mudanças ou melhoras ocorridas em *participantes específicos* do grupo. Esse segundo aspecto aponta para a importância de se considerar resultados individuais, mesmo que o processo de intervenção tenha ocorrido em grupo, o que parece ser relevante para grande parte dos serviços de Psicologia. Essas questões estão na base dos dois conceitos-chave do Método JT tratado a seguir: o índice de mudança confiável e a significância clínica.

O desafio, para Jacobson e Truax (1992), era o de elaborar um método de identificação de mudanças clinicamente significativas que pudesse ser: (a) aplicável, em tese, a qualquer desordem clínica; (b) consistente com as expectativas dos profissionais quanto aos resultados da psicoterapia; e (c) útil na identificação dos clientes que se recuperaram, melhoraram e não melhoraram ou até pioraram após a intervenção.

Descrição do Método JT

O Método JT prevê uma análise comparativa entre escores pré e pós-intervenção com o objetivo de decidir se as diferenças entre eles representam *mudanças confiáveis* e se são *clínicamente relevantes*. Conforme Evans e cols. (1998), esse método

³ Uma análise de diferentes métodos, que não é objeto do presente estudo, pode ser vista em Wise (2004).

busca responder duas questões. A primeira é pertinente ao índice de mudança confiável: os ganhos do cliente foram além de uma mera oscilação (positiva ou negativa) decorrente de erro de medida? A segunda se refere à mudança clinicamente significativa: qual a condição final do cliente em relação aos escores de grupos não clínicos de referência?

A análise de dados por meio do Método JT implica, portanto, dois processos complementares: (a) cálculo da confiabilidade das alterações ocorridas entre a avaliação pré e a avaliação pós-intervenção, descrita em termos de um Índice de Mudança Confiável (IMC); e (b) análise do significado clínico dessas alterações. Com base em indicadores quantitativos, esses dois processos podem ser graficamente representados. A seguir, são apresentadas as bases conceituais e metodológicas do Método JT, com ilustrações gráficas de sua aplicabilidade no tratamento dos dados de uma intervenção hipotética.

Índice de mudança confiável

Pode-se afirmar que um procedimento é efetivo quando altera – para mais ou para menos – dimensões descritivas do desempenho do indivíduo em termos de frequência, intensidade, duração, topografia etc. As alterações dese-

jáveis são aquelas que sugerem a redução de problemas ou a maximização de medidas de ajustamento ou “saúde psicológica”. O Índice de Mudança Confiável (IMC) é um indicador de erro da medida obtida na avaliação do cliente, que vai ser comparado a uma distribuição teórica de erros do instrumento: se a mudança é verdadeira e não decorrente de erro de medida, a probabilidade associada ao IMC será tão pequena ($p < 0,05$) que a hipótese de erro será rejeitada. Para isso, os autores do Método JT desenvolveram uma fórmula específica, baseada na diferença entre pré e pós-teste dividida pelo erro padrão da diferença.

Para ilustrar graficamente a importância do IMC, considere-se, hipoteticamente, uma avaliação de ajustamento psicológico, como por exemplo, uma medida de enfrentamento de situações e interlocutores tidos como “difíceis”, cujos escores podem variar de zero a quatro. Os resultados de vários indivíduos (dados e nomes fictícios) são representados em um gráfico de dispersão das diferenças entre pré e pós-intervenção (pós *menos* pré), conforme mostrado na Figura 1.

Os pontos sobre a diagonal indicam escores iguais nas avaliações pré e pós-intervenção (diferença zero); os pontos acima da diagonal representam oscilações favoráveis, em

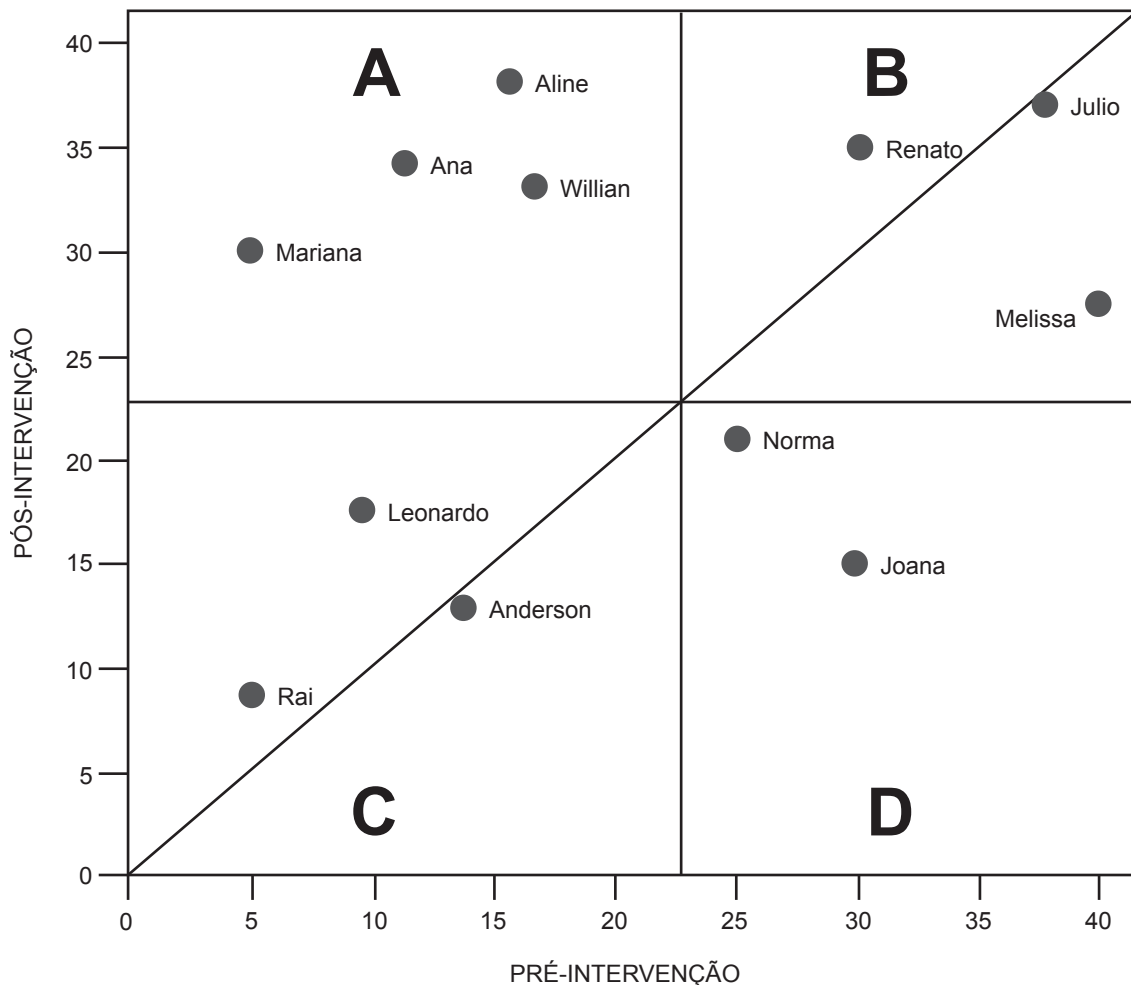


Figura 1. Dispersão das diferenças entre pós e pré-intervenção em uma medida hipotética de ajustamento psicológico de participante sob um atendimento.

termos de melhora no ajustamento psicológico (pós>pré), e os pontos abaixo, oscilações desfavoráveis ou piores nos indicadores de ajustamento (pós<pré). Na Figura 1, sete participantes se situaram acima da diagonal, indicando que foram beneficiados com a intervenção, enquanto cinco participantes ficaram abaixo, sugerindo efeito negativo da intervenção.

A Figura 1 apresenta, ainda, duas linhas, uma vertical e uma horizontal, que separam os indivíduos conforme os escores obtidos na pré e na pós-intervenção. A linha vertical separa o grupo de indivíduos que na *pré-intervenção* apresentaram escores baixos (população disfuncional, “clínica”, ou seja, com indicativo de problemas de ajustamento) daqueles que apresentaram escores altos (população funcional, “não clínica”, ou seja, com indicadores de bom ajustamento). A linha horizontal separa os indivíduos que na *pós-intervenção* apresentaram escores baixos daqueles que apresentaram escores altos.

O traçado das linhas horizontal e vertical delimita quatro quadrantes que representam as combinações entre oscilações positivas ou negativas em função da condição inicial e final dos participantes. O quadrante **A** (clínico - melhora) é constituído pelos participantes que necessitavam de intervenção (escores iniciais muito baixos) e apresentaram oscilações positivas, sugestivas de melhora (Marina, Ana, William e Aline). O quadrante **B** (não clínico – não clínico) inclui os participantes que já apresentavam inicialmente escores altos e que permaneceram altos (Renato e Julio), mesmo com oscilações negativas (Melissa). O quadrante **C** (clínico – clínico) representa os participantes com escores iniciais muito baixos e que permaneceram baixos mesmo que com oscilações (Leonardo, Raí e Anderson). E o quadrante **D** (não clínico – piora) é constituído pelos participantes com escores iniciais altos que apresentaram oscilações negativas ao final da intervenção.

Como se pode verificar na Figura 1, a extensão da “melhora” ou “piora” foi bastante variada entre os participantes da intervenção. Assim, uma análise mais refinada é necessária para se estabelecer de fato quais participantes apresentaram oscilações positivas ou negativas em amplitude suficiente para caracterizar melhoras ou pioras confiáveis. Essa análise mais refinada leva em conta os intervalos de confiança estabelecidos pela confiabilidade do instrumento de avaliação (linhas horizontal e vertical) e pela amplitude da diferença entre as duas avaliações (linha diagonal), relacionada à variabilidade natural dos fenômenos avaliados. No método JT, esses aspectos são computados em termos de um Índice de Mudança Confiável (IMC), que permite delimitar áreas de confiabilidade e de incerteza na representação gráfica dos resultados.

Em relação ao objeto da avaliação, as dimensões descritivas do comportamento (frequência, intensidade, duração, topografia etc.) apresentam uma variabilidade “natural”, imposta por condições internas e externas (Sidman, 1960/1976), que podem dificultar a análise da validade interna dos efeitos de uma intervenção. Parte dessa variabilidade está associada à confiabilidade dos instrumentos de avaliação psicológica (Anastasi & Urbina, 2000/1997). Um instrumento é tanto mais confiável quanto menor for seu erro na medida. No Método JT, o reconhecimento dessas fontes de variabilidade leva ao cálculo de um Índice de Mudança Confiável (IMC). Qualquer oscilação, positiva ou negativa, entre a avaliação pré e a pós-

intervenção, somente será considerada uma mudança confiável se for suficientemente robusta para superar a margem de incerteza associada a erros da medida ou variabilidade do objeto avaliado, situando-a dentro do intervalo de confiança para os resultados obtidos.

No caso da representação gráfica da Figura 1, intervalos de confiança devem ser delimitados tanto para a diferença entre a avaliação pré e pós-intervenção (linha diagonal) quanto para os pontos de corte dos escores iniciais (linha vertical) e dos finais (linha horizontal), pois cada um deles pode implicar em algum erro de medida (Evans & cols., 1998; Jacobson & Truax, 1992). As linhas que delimitam as áreas de confiabilidade e incerteza são traçadas por meio de fórmulas matemáticas baseadas nos indicadores de variabilidade dos resultados (desvio-padrão, erro padrão, confiabilidade do instrumento). A Figura 2 ilustra as três áreas de incerteza (mais claras) e as de confiança requeridas para a localização gráfica dos resultados que apresentam ou não apresentam um IMC.

No quadrante A, todos os participantes apresentaram resultados que os situaram fora das regiões de incerteza e dentro da área de confiabilidade (área mais escura), ou seja, todos apresentaram melhoras confiáveis. No quadrante B e no quadrante C, todos se situaram em alguma região de incerteza e, portanto, não se pode afirmar que melhoraram ou pioraram. No quadrante D, Joana representa um caso de piora confiável (área mais escura), mas Norma, não.

Em resumo, considerando a confiabilidade das oscilações positivas e negativas, indicativas de melhora e piora, respectivamente, a Figura 2 mostra que quatro participantes (Aline, Ana, William e Mariana) apresentaram melhoras confiáveis, que apóiam a efetividade ou validade interna da intervenção, e um participante (Joana) apresentou piora confiável, que não apóia essa possibilidade. Assim, como há casos de melhora e de piora, a questão da efetividade em termos de validade interna da intervenção ainda permaneceria indefinida e objeto de novas investigações. De todo modo, os resultados permitem levantar hipóteses sobre características associadas aos participantes que podem ter contribuído para os resultados obtidos ou sobre tipos de intervenção mais apropriados em função de características dos participantes.

Mudanças clinicamente significativas

Além de estatisticamente confiáveis, as melhoras associadas a programas de intervenção devem ocorrer em uma extensão que as caracterize como clinicamente relevantes. Como exemplo, uma pessoa com obesidade mórbida poderia apresentar uma redução de peso estatisticamente significativa e confiável, em relação ao peso inicial e, ainda assim, continuar com o diagnóstico desse transtorno, requerendo atendimento adicional.

Para se avaliar a significância clínica de mudanças obtidas em uma intervenção terapêutica, Kazdin (2003) identifica três indicadores: (a) a diferença entre as avaliações finais e as da pré-intervenção, considerando-se os indicadores de amostras normativas funcionais; (b) a avaliação subjetiva do cliente ou de seus significantes; e (c) o impacto social em medidas reconhecidamente relevantes de ajustamento psicológico, como internação, evasão, hospitalização, alta, morte etc. O

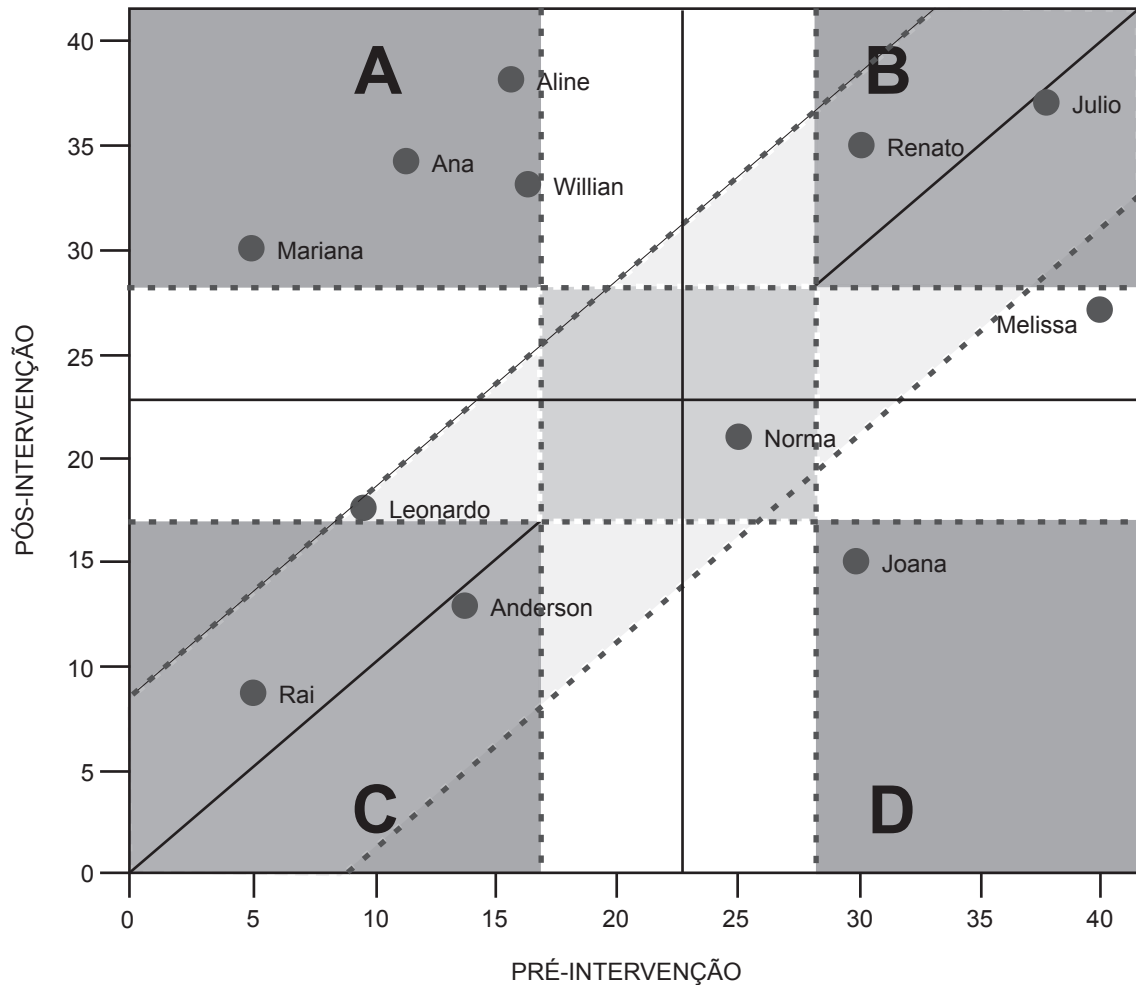


Figura 2. Intervalos de confiança para as diferenças entre as medidas obtidas na pré e na pós-intervenção e em torno da diagonal que representa as diferenças pós menos pré-intervenção (Quadrantes: A = clínico-melhora; B = não clínico-não clínico; C = clínico-clínico; D = não clínico-piora).

Método JT enfatiza principalmente o primeiro desses indicadores. Jacobson e cols. (1984) consideram clinicamente significantes aqueles indicadores de mudanças que, além de confiáveis, expressam o retorno do indivíduo ao funcionamento “normal”, o que pode ou não estar relacionado a um efeito estatisticamente significativo obtido na análise inferencial de resultados grupais. Descartada a hipótese de que as diferenças entre momentos ou entre grupos tenham ocorrido por acaso, ainda se coloca a questão da efetividade enquanto impacto sobre o funcionamento cotidiano do cliente.

Quando os resultados de uma intervenção se baseiam em critérios dicotômicos ou absolutos de sucesso-fracasso (vômito psicogênico, alucinações etc.), a avaliação dessa efetividade é relativamente simples. No entanto, ela se torna mais difícil quando se apóia em escores avaliados por meio de escalas, que situam o indivíduo em um ponto de uma distribuição funcional ou disfuncional, como é o caso das escalas de ansiedade, neuroticismo, depressão ou diferentes habilidades. Nesse sentido, uma mudança é clinicamente significativa quando o escore anterior do indivíduo, que o situava como membro de uma população disfuncional para determinada variável, altera-se, após a

intervenção, passando a situá-lo como membro de uma população funcional.

Considerando a disponibilidade de dados normativos a respeito da distribuição dos escores sobre um determinado aspecto do funcionamento psicológico avaliado, Jacobson e Truax (1991, 1992) sugerem três critérios para analisar se uma mudança pode ser considerada clinicamente significativa. Os três critérios implicam em diferentes “pontos de corte” na distribuição de escores da população funcional e/ou disfuncional, conforme descritos a seguir.

A aplicação desses critérios foi aqui considerada para indicadores de ajustamento positivo, para os quais a mudança desejável deve ocorrer na direção de aumento nos escores. No caso de indicadores de problemas (quanto menores mais situam o indivíduo dentro do grupo funcional), deve-se efetuar um raciocínio inverso em termos da direção da mudança desejável.

Critério A. Esse critério é particularmente útil quando não se dispõe de dados normativos para a população funcional (não clínica), podendo-se estimar média e desvio padrão com base nos dados pré-intervenção da própria amostra disfuncional sob tratamento. Nesse caso, uma mudança será

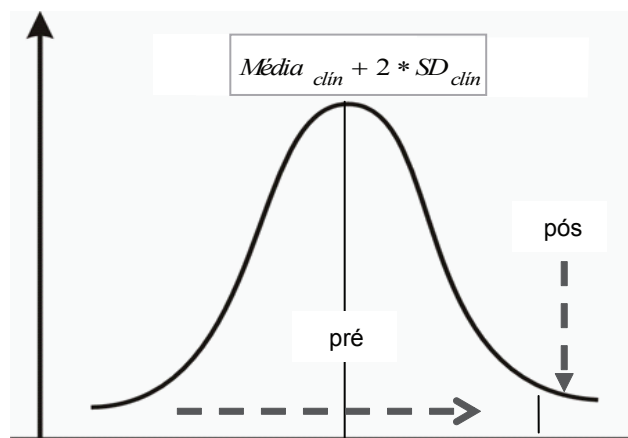


Figura 3. Direção de alteração representativa de mudança confiável quando se dispõe somente de dados da população disfuncional (Critério A).

considerada clinicamente relevante se a diferença entre a avaliação pré e a avaliação pós-intervenção for de, no mínimo, dois desvios padrão acima⁴ da média pré-intervenção, ou seja, *deve colocar o indivíduo fora* da distribuição de escores da população disfuncional. A Figura 3 ilustra esse critério, com a fórmula estatística de cálculo.

Critério B. Quando se dispõe de dados normativos sobre a distribuição dos escores da população funcional, uma mudança clinicamente relevante será considerada quando o escore pós-intervenção *desloca o indivíduo para dentro da distribuição da população funcional*. Em outras palavras, seus escores pós-intervenção devem se situar dentro do intervalo que começa no ponto de corte representado pela média *menos* dois desvios padrão dessa população. A ilustração é semelhante à anterior (Figura 3), porém no sentido inverso, ou seja, escores pré (fora da distribuição) e escores pós (dentro da distribuição).

Critério C. Quando se dispõe de dados normativos sobre a distribuição dos escores da população funcional e da disfuncional, uma mudança clinicamente relevante deve levar o indivíduo, após a intervenção, *simultaneamente para fora da distribuição disfuncional e para dentro da distribuição funcional*. Portanto, considerando as duas distribuições de escores, o escore final do indivíduo deverá situá-lo acima do ponto definido pela média *mais* dois desvios padrão da população disfuncional e acima também da média *menos* dois desvios padrão da população disfuncional.

A adesão ao Critério C coloca uma questão importante: o que ocorre quando os escores da população funcional e disfuncional, distribuídos ao longo de um contínuo (por exemplo, neuroticismo, estresse, condutas adaptativas), são muito separadas ou quando são parcialmente sobrepostas? A Figura 4 ilustra essas duas condições, juntamente com a fórmula de cálculo.

Comparando-se os três critérios, no caso de distribuições parcialmente sobrepostas, o critério A (apresentado na Figura 3) seria muito exigente pois, dada a sobreposição, não é necessário que o cliente saia totalmente da distribuição

disfuncional para se situar na funcional. O critério B seria muito indulgente, considerando que uma mudança muito pequena seria suficiente para situar o indivíduo na distribuição funcional. Nesse caso, o Critério C é o mais indicado.

Quando as duas distribuições são muito separadas, ocorre o contrário: o critério A seria muito indulgente, uma vez que sair da população disfuncional não significa retorno ao funcionamento adaptativo. Por outro lado, o critério B seria muito exigente, pois o indivíduo inicialmente classificado como “clínico” precisa de uma grande mudança para ser incluído na distribuição da população funcional. Ainda assim, os autores do Método JT defendem que o critério B deve ser utilizado por ser mais confiável.

Algumas limitações do Método JT

Mesmo diante da pertinência desses questionamentos, o Método JT parece representar uma alternativa importante, especialmente nos casos de intervenção em pequenos grupos, quando não se dispõe de grupos controle e, também, nos casos de intervenções individuais com replicações para um pequeno número de indivíduos (Margison & cols., 2000). No entanto, deve-se reconhecer que a sua potencialidade requer ainda estudos empíricos sobre diferentes conjuntos de dados e problemas. O investimento nesse tipo de pesquisa pode, conforme Middel e van Solderen (2002), contribuir para ampliar as informações de procedimentos baseados em evidência.

Considerações Finais

O Método JT foi originalmente desenvolvido para estabelecer a significância clínica de resultados da psicoterapia, com aplicabilidade, pelo menos teoricamente, a qualquer desordem clínica, de forma consistente com as expectativas dos profissionais (Jacobson & Truax, 1992). Nesse sentido, constitui uma importante ferramenta de tratamento de dados, tanto para o profissional de serviços, preocupado primariamente em resolver o problema de seu cliente e demonstrar a efetividade de sua atuação, como para o pesquisador, preocupado em desenvolver procedimentos e práticas progressivamente mais efetivos (Margison & cols., 2000).

Os resultados obtidos com o Método JT podem ser bastante diferentes dos produzidos pelos procedimentos da estatística inferencial, que compara indicadores descritivos do grupo pré *versus* pós ou funcional *versus* disfuncional (Evans & cols., 1998; Jacobson & Truax, 1992). Em outras palavras, para um mesmo conjunto de dados pré e pós-intervenção, pode-se encontrar diferenças estatisticamente significativas que não serão consideradas como mudanças confiáveis ou clinicamente relevantes e *vice-versa*. Ainda assim, Evans e cols. (1998) ressaltam que uma mudança confiável, mesmo que não clinicamente significativa (sem retorno ao funcionamento normal), pode fazer grande diferença na vida do indivíduo, especialmente no caso de intervenções com pacientes esquizofrênicos, autistas ou com outras formas de deficiência mental ou sensorial.

Não obstante a potencialidade do Método JT, os proponentes reconhecem algumas de suas limitações (Jacobson & Truax, 1992), tais como: (a) a falta de evidências sobre a

⁴ No caso de indicadores de problemas como, por exemplo, depressão, neuroticismo etc., o raciocínio é inverso: dois desvios-padrões abaixo.

$$\frac{(m\u00e9dia_{func} \times SD_{norm}) + (m\u00e9dia_{func} \times SD_{cl\u00edn})}{SD_{func} + SD_{cl\u00edn}}$$

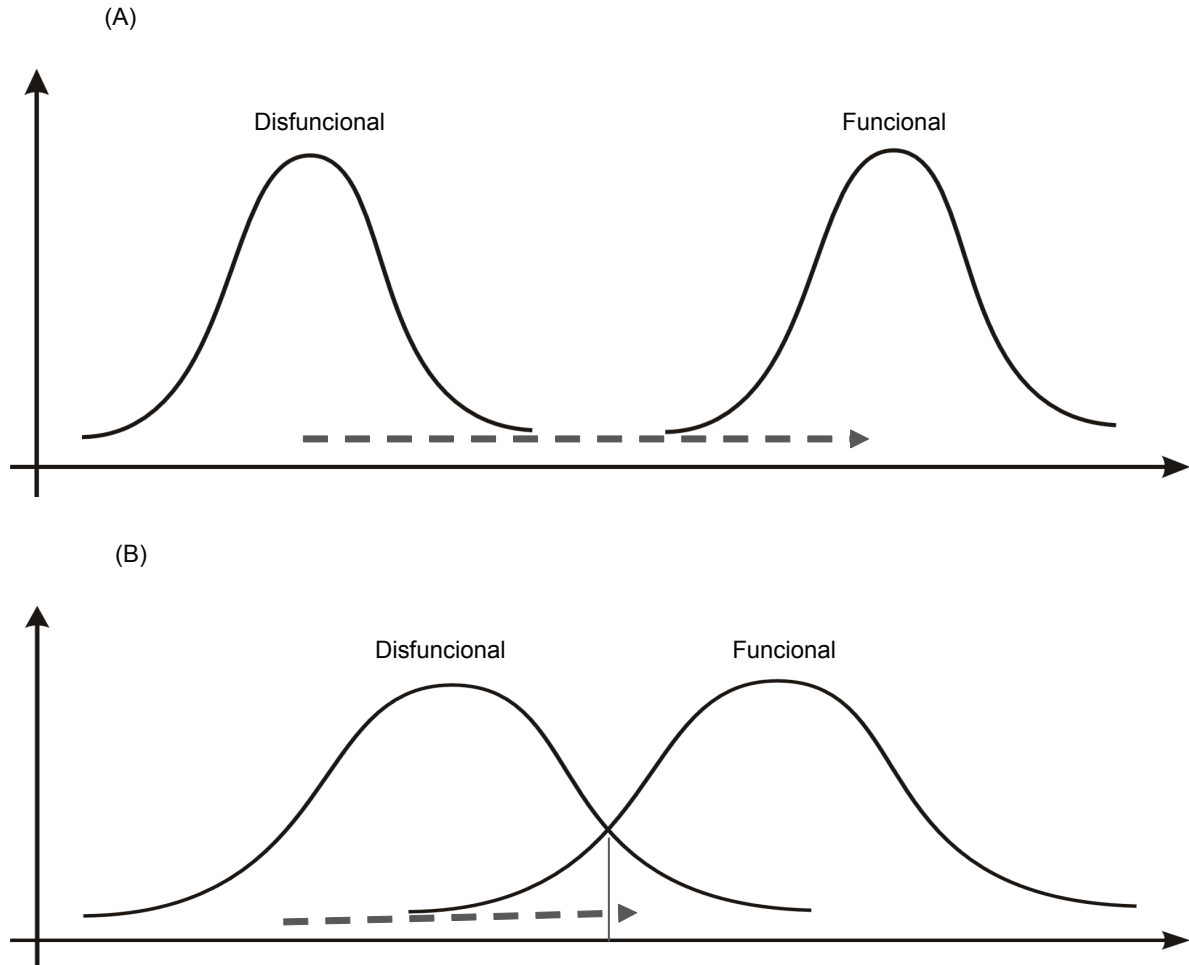


Figura 4. Ilustração do Crit\u00e9rio C, em que se disp\u00f5e da distribui\u00e7\u00e3o dos escores da popula\u00e7\u00e3o funcional e disfuncional: (A) quando as duas s\u00e3o afastadas; (B) quando as duas distribui\u00e7\u00f5es s\u00e3o sobrepostas.

robustez do m\u00e9todo quando os escores da popula\u00e7\u00e3o funcional e/ou disfuncional n\u00e3o apresentam distribui\u00e7\u00e3o normal, considerando que as evid\u00eancias dispon\u00edveis se basearam na suposi\u00e7\u00e3o de normalidade (ainda que a formula\u00e7\u00e3o do m\u00e9todo n\u00e3o dependa disso); (b) a n\u00e3o pertin\u00eancia da operacionaliza\u00e7\u00e3o da signific\u00e2ncia cl\u00ednica em termos de recupera\u00e7\u00e3o ou retorno ao funcionamento normal para todas as doen\u00e7as tratadas em psicoterapia como, por exemplo, esquizofrenia ou autismo, nos quais o padr\u00e3o de recupera\u00e7\u00e3o imposto pelo m\u00e9todo excederia as expectativas da maioria dos profissionais que trabalham nesse campo; e (c) a falta de familiaridade de muitos profissionais com o uso de indicadores de confiabilidade dos instrumentos e de dados normativos para popula\u00e7\u00f5es funcional e disfuncional, o que representa novos desafios \u00e0 dissemina\u00e7\u00e3o desse m\u00e9todo.

Outras fragilidades do M\u00e9todo JT s\u00e3o tamb\u00e9m apontadas. Massen (2004) discute e compara f\u00f3rmulas para o c\u00e1lculo do erro padr\u00e3o, indicando a necessidade de uma adapta\u00e7\u00e3o da f\u00f3rmula original utilizada no M\u00e9todo JT que d\u00ea conta da diferen\u00e7a entre a vari\u00e2ncia das avalia\u00e7\u00f5es inicial e final. Alguns estudos questionam a necessidade de maiores investiga\u00e7\u00f5es sobre a rela\u00e7\u00e3o entre os indicadores de mudan\u00e7a confi\u00e1vel e clinicamente relevante do M\u00e9todo JT *versus* os obtidos pelos m\u00e9todos inferenciais tradicionais, como signific\u00e2ncia estat\u00edstica das diferen\u00e7as entre grupos. Enquanto alguns autores consideram que esses indicadores podem ser contradit\u00f3rios (Sheldrick, Kendall & Heimberg, 2001), outros (e.g., Beutler & Moleiro, 2001) defendem que n\u00e3o devem ser vistos como equivalentes, mas como complementares. Neste caso, argumentam que nem sempre se disp\u00f5e dos dados

necessários para a análise da relevância clínica, mas que, quando viável, esta complementa a qualidade da interpretação estatística. Adicionalmente, Atkins, Bedics, McGlinchey e Beauchaine (2005) compararam resultados de intervenções sob diferentes métodos de análise, que deveriam produzir resultados diferentes (confiabilidade da medida, tamanho do efeito e correlação pré-pós), verificando que houve mais semelhanças do que diferenças. Esses autores concluem que muitos dos métodos de avaliação da efetividade de intervenções são comparáveis entre si quando baseados em níveis altos de confiabilidade, o que teoricamente reduz o caráter de novidade do Método JT para esses casos.

O interesse pelo Método JT está bastante documentado na literatura e é atribuído principalmente ao fato de combinar informação sobre o funcionamento pré e pós-intervenção com medidas normativas da população funcional e disfuncional. Ele foi indicado por Ferguson, Robinson e Splaine (2002) como procedimento referencial quando se usam medidas de auto-relato de saúde funcional como, por exemplo, o SF-36. Segundo Ogles, Lunnen e Banesteel (2001), a maioria dos estudos sobre algum tipo de significância clínica publicados no *Journal of Consulting and Clinical Psychology*, nos nove anos anteriores, usaram o Método JT ou uma variação dele.

Ainda que desenvolvido no contexto da clínica e dos serviços de psicoterapia, entende-se que o Método JT pode ser particularmente relevante também para a avaliação das intervenções conduzidas fora desse contexto, por exemplo, para intervenções educacionais. Nesse caso, certamente algumas especificidades precisariam ser consideradas, pois não se trata, via de regra, de “recuperar” a funcionalidade do indivíduo, mas de uma preocupação proativa ou preventiva, visando levá-lo a um funcionamento mais desejável do que o atual, estabelecido como objetivos da intervenção. Assim, paralelamente ao conceito de significância clínica, pode-se pensar que a melhora dos escores do indivíduo, mesmo dentro da distribuição funcional, além de indicar a consecução dos objetivos de tais intervenções (validade interna), poderia ser tomada como um indicador de “significância educacional” (validade social ou externa).

Essa possibilidade aponta para a necessidade e importância de maior investimento na avaliação empírica de estudos de intervenção nesse campo específico. No entanto, mesmo na pesquisa clínica, ainda são escassos os estudos baseados na significância clínica e na mudança confiável. Uma busca nos indexadores mais populares (*scielo, google* etc.) permite identificar estudos sobre significância clínica nas ciências biológicas e da saúde e referências a esse construto na Psicologia; porém, estudos empíricos, aplicando um método de tratamento de dados para aferir esses aspectos, não são usuais. Dentre os poucos estudos disponíveis em nosso meio, que usaram o método JT, pode-se citar o de Souza e Silveiras (2006), que analisaram os efeitos de um programa de intervenção, com grupo de mulheres em fase de menopausa, sobre indicadores de dificuldade de sono e depressão; o de Yoshida (2006), que avaliou os processos de psicoterapia dinâmica breve; e o de Pereira, Silveiras e Del Prette (no prelo), que examinaram os efeitos do tratamento da enurese sobre a redução de problemas de comportamento. Na Educação, mais especificamente na

Educação Especial, pode-se citar o estudo recente de Aguiar (2006), que efetuou a análise de mudança confiável para os resultados de um grupo de deficientes mentais adultos que participaram de um programa de promoção de habilidades sociais e comunicativas.

Uma dificuldade para a análise da mudança confiável e da significância clínica refere-se à falta de modelos e de instruções, mais apropriadas aos profissionais de Ciências Humanas, para efetuar o tratamento de dados. Visando suprir essa lacuna, o grupo de pesquisa dos autores do presente trabalho produziu um guia passo-a-passo, impresso, de instruções para a aplicação do Método JT em programa Excel, com geração de figuras ilustrativas dos resultados (Aguiar, Aguiar & Del Prette, no prelo).

Referências

- Aguiar, A. A. R. (2006). *Construção e avaliação de um programa multimodal de habilidades comunicativas para adultos com deficiência mental*. Tese de Doutorado, Universidade Federal de São Carlos, São Carlos, São Paulo.
- Aguiar, A. R., Aguiar, R. G. & Del Prette, Z. A. P. (no prelo). *Calculando a significância clínica e o índice de mudança confiável em pesquisa-intervenção*. São Carlos: EDUFSCar.
- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica* (7ª. ed.) (M. A. V. Veronese, Trad.). Porto Alegre: Artmed. (Trabalho original publicado em 1997)
- Atkins, D. C., Bedics, J. D., McGlinchey, J. B. & Beauchaine, T. P. (2005). Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology*, 73, 982–989.
- Beutler, L. E. & Moleiro, C. (2001). Clinical versus reliable and significant change. *Clinical Psychology: Science and Practice*, 8, 441-445
- Brantlinger, E., Jumenez, R. Pugach, M. & Richardson, V. (2005). Qualitative studies in special Education. *Council for Exceptional Children*, 71, 195-207.
- Cozby, P. C. (2003). *Métodos de pesquisa em ciências do comportamento* (P. I. Gomide & E. Otta, Trad.). São Paulo: Atlas. (Trabalho original publicado em 2001)
- Currie, S. R., Wilson, K. G. & Curran, D. (2002). Clinical significance and predictors of treatment response to cognitive-behavior therapy for insomnia secondary to chronic pain. *Journal of Behavioral Medicine*, 13, 135-152.
- Evans, C., Margison, F. & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health (EBMH Notebook)*, 1, 70-72.
- Ferguson, R. J., Robinson, A. B. & Splaine, M. (2002). Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. *Quality of Life Research*, 11, 509-516.
- Gersten, R., Fuchs, L. S., Coyne, M., Greenwood, C. & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Council for Exceptional Children*, 71, 149-164.
- Jacobson, N. S., Follette, W. C. & Revenstorf, D. (1984). Psychotherapy outcomes research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.

- Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.
- Jacobson, N. S. & Truax, P. (1992). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Em A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 521-538). Washington (DC): Copyright.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.
- Kazdin, A. E. & Weiss, J. R. (2003). Assessment and evaluation of interventions. Em A. E. Kazdin (Ed.), *Evidence-based psychotherapies for children and adolescents* (pp. 408-435). New York: Guilford Press.
- Levant, R. F. (2005). *Report of the 2005 Presidential Task Force on Evidence-Based Practice 1: American Psychological Association*. Recuperado em dezembro de 2006 de <http://www.apa.org/practice/ebpreport.pdf>.
- Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K. & Connel, J. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *British Journal of Psychiatry, 177*, 123-130.
- Massen, G. H. (2004). The standard error in the Jacobson and Truax Reliable Change Index: The classical approach to the assessment of reliable change. *Journal of the International Neuropsychological Society, 10*, 888-893.
- McGlinchey, J. M., Atkins, D. C. & Jacobson, N. S. (2002). Clinical significance methods: Which one to use and how useful they are? *Behavior Therapy, 33*, 529-550.
- Middel, B. & van Solderen, E. (2002). Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care, 2*, 1-22.
- Ogles, B. M., Lunnen, K. M. & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*, 421-446.
- Pereira, R. F., Silvaes, E. F. M. & Del Prette, Z. A. P. (no prelo). Problemas de comportamento em enuréticos: um exemplo de análise utilizando o índice de mudança confiável (IMC). *Arquivos Brasileiros de Psicologia*.
- Sheldrick, R. C., Kendall, P. C. & Heimberg, R. G. (2001). The clinical significance of treatments: A comparison of three treatments for conduct disorder children. *Clinical Psychology: Science and Practice, 8*, 418-429.
- Sidman, M. (1976). *Táticas da pesquisa científica: Avaliação dos dados experimentais na Psicologia* (M. E. Paiva, Trad.). São Paulo: Brasiliense. (Trabalho original publicado em 1960)
- Souza, C. L. & Silvaes, E. F. M. (2006). Grupos informativos sobre menopausa: Trabalhando com grupos de mulheres na clínica-escola. Em E. F. M. Silvaes (Org.), *Atendimento psicológico em clínicas-escola* (pp. 203-228). São Paulo: Alínea.
- Thompson, B., Diamond, K., McWilliam, R. & Snyder, P. & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Council for Exceptional Children, 71*, 181-194.
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future. <http://www.leaonline.com/loi/jpa>, 82, 50-59.
- Yoshida, E. M. P. (2006). Psicoterapia dinâmica breve: los procesos de cambio evaluados a través de la madurez de los mecanismos de defensa, los síntomas psicopatológicos y la estrategia terapéutica. *Libro de Resúmenes*. Simposio Internacional de Investigación en Psicoterapia, Montevideo, p. 34.

Recebido em 27.02.2007

Primeira decisão editorial em 18.03.2007

Versão final em 22.04.2007

Aceito em 30.07.2008 ■