OBJETIVOS DA AVALIAÇÃO DA FIDEDIGNIDADE EM ESTUDOS OBSERVACIONAIS

Cecília Guarnieri Batista
Pontifícia Universidade Católica de Campinas

RESUMO - Na literatura da análise aplicada do comportamento, verifica-se que a avaliação da fidedignidade é feita através do cálculo do acordo entre observadores. Já na literatura etológica, constata-se maior diversificação de técnicas de avaliação de fidedignidade, incluindo-se a consideração de que a replicação da identificação das mesmas categorias por outros autores se constitui em indicação de que o observador foi fidedigno. Procura-se identificaras razões pelas quais existe essa diferença de enfoque à questão. Conclui-se sugerindo que a avaliação da fidedignidade deve estar subordinada aos objetivos e características de cada projeto de pesquisa, retendo-se o significado mais amplo do termo "fidedignidade", entendido como exatidão e replicabilidade.

GOALS OF RELIABILITY EVALUATION IN OBSERVATIONAL STUDIES

ABSTRACT - In the applied behavior analysis literature, reliability evaluation is done through the measurement of interobserver agreement. In the ethological literature, on the other hand, several techniques of reliability evaluation are used, including the replication of the identification of the same categories of behavior by other researchers. Reasons for this difference are identified. As a conclusion, it is suggested that reliability evaluation must be subordinated to the goals and characteristics of each research project, while keeping general meaning of reliability as accuracy and replicability.

As últimas décadas de estudo do comportamento têm assistido a um crescente interesse por estudos observacionais, por influência de determinadas abordagens teóricas, como a etologia (Blurton Jones, 1972; Hinde, 1966) e a análise do comportamento aplicada (Bijou, Peterson e Ault, 1968; Ramp e Semb, 1975). Recentemente são realizados estudos observacionais que decorrem direta ou indiretamente dessas abordagens e que abordam diferentes temas, como, por exemplo, o apego e as reações da criança à separação (Ferreira, 1984), a comunicação não-verbal (Scherer e Ekman, 1982) e as relações entre professor e aluno em sala de aula (Marturano, 1978, 1979; Marturano, Bertoldo e Camelo, 1982).

Paralelamente ao desenvolvimento desses estudos, tem crescido a preocupação com a aferição da fidedignidade do sistema de observação e, especificamente, do observador humano como parte desse sistema (Johnson e Bolstad,

1973; Caro, Roper, Young e Dank, 1979). No presente artigo, pretende-se examinar a forma como a fidedignidade tem sido considerada em Psicologia e, mais especificamente, nos estudos observacionais, e estabelecer algumas conclusões a partir desse exame.

FIDEDIGNIDADE EM PSICOLOGIA

A definição de fidedignidade em dicionários de Psicologia varia de acordo com as diferentes situações em que o conceito é aplicado, como se pode ver a seguir:

Com relação a relatos verbais, fidedignidade é entendida como "o grau de exatidão de um dado relato (de um evento, fenômeno, etc.) ou o grau de confiança no testemunho de um dado indivíduo" (Warren, 1934).

Com relação aos testes psicológicos, é considerada como "uma das qualidades de uma medida. A fidedignidade de um teste é expressa com o auxílio dos coeficientes de constância, de equivalência e de homogeneidade". Constância, por sua vez, é entendida como "correlação entre duas séries de medidas de um mesmo atributo ou caráter, efetuadas sobre um mesmo grupo, qualquer que seja o tempo decorrido entre as duas observações". Equivalência refere-se à "correlação entre duas formas de um teste, que se tomam como equivalentes e se aplicam com intervalo reduzido". E homogeneidade é definida como a "correlação entre as duas partes de um teste (questões pares e ímpares)". (Piéron, 1966, original de 1951).

Para a pesquisa experimental, fidedignidade refere-se ao "grau em que os resultados são consistentes na repetição do experimento" (Wolman, 1973).

Com relação a aparelhagens eletrônicas ou a sistemas homem-máquina, entende-se fidedignidade como a "probabilidade de que estes funcionem sem defeito, sob certas condições e durante um período determinado" (Sillamy, 1 980),

Verifica-se, assim, que ao longo das diferentes acepções que o termo assume, destacam-se as seguintes características: exatidão, constância de propriedades ao longo de um período, repetição de um resultado obedecidas as mesmas prescrições. Sabe-se, também, que a técnica específica a serutilizada para a avaliação da fidedignidade varia de acordo com a área de conhecimento em que o conceito é aplicado: fidedignidade de um teste psicológico é aferida de modo diferente do que fidedignidade dos resultados de uma pesquisa experimental.

FIDEDIGNIDADE DO OBSERVADOR NA LITERATURA DA ANÁLISE DO COMPORTAMENTO APLICADA

A questão da fidedignidade do observador tem sido extensivamente tratada pelos autores que atuam em análise do comportamento aplicada. Essa área de pesquisa originou-se da análise experimental do comportamento, que trabalha com uma ou poucas respostas que produzem um efeito ambiental claro, efeito esse registrado eletromecanicamente, em laboratório. A análise aplicada do comportamento, por sua vez, lida com respostas escolhidas em função de sua relevância social, ao invés de sua importância para a teoria (Baer, Wolf e Risley, 1968). Isso leva o analista a atuarem ambientes complexos como salas de aula, instituições, residências, etc, e a anotar as respostas relevantes para o problema estudado, respostas estas que raramente permitem um registro eletromecénico. Lança-se mão, então, do observador humano, que registra as respostas selecionadas de

acordo com instruções predeterminadas. Da necessidade de se aferir a precisão desse observador, menos confiável do que os equipamentos eletromecânicos, é que provavelmente surgiu a importância atribuída à estimativa da fidedignidade do observador.

O objetivo básico da avaliação da fidedignidade tem sido descrito na literatura da análise do comportamento aplicada como o de estabelecer a credibilidade dos dados de observação (Hawkins e Dotson, 1975; Kazdin, 1977; Kratochwill e Wetzel, 1977). Essa credibilidade vai permitir que se confie nas demonstrações de relações entre escores de comportamento e variáveis externas, nas demonstrações de mudanças no comportamento, em termos de taxa de resposta ou duração, ao longo de diferentes fases experimentais (Johnson e Bolstad, 1973; Hopkins e Herman, 1977). O acordo entre observadores que registram o comportamento simultânea e independentemente é considerado como indicação da fidedignidade desses observadores (Kazdin, 1977; Yelton, 1979). Os pesquisadores assumiram que o acordo em ocorrências e/ou não ocorrências de comportamentos-alvo sugere que os observadores estão respondendo aos mesmos eventos. Assim, o acordo entre observadores é uma indicação da replicabilidade dos dados de observação (Wildman e Erickson, 1977).

Entretanto, ao longo do uso de índices de acordo entre observadores em estudos dessa natureza, foram surgindo elementos para indicar que esses índices estão sujeitos à influência de vários fatores (Johnson e Bolstad, 1973; Kazdin, 1977). Além disso, foi sendo aprofundada a discussão sobre a relação entre acordo entre observadores e definição do comportamento,

Em um texto sobre características de uma boa definição, distribuído para alunos universitários, Michael (sem data) afirma que o teste final da qualidade de uma definição *é* o cálculo de fidedignidade. Em seguida a essa afirmação, ele faz a ressalva de que tem havido motivos que impedem o estabelecimento dessa relação direta na maioria das pesquisas realizadas e menciona algumas, entre as quais se inclui o uso do mesmo conjunto de definições no treino e na coleta de dados, o que pode levar os observadores a serem controlados por uma definição implícita (de consenso) e não pela definição explícita. Ele afirma que, nessas condições, o observador aprendeu a emitir o mesmo comportamento de registrar que seu treinador, mas que o grau em que isso corresponde à definição verbal relatada é completamente desconhecido. Assim, a definição **explícita** (relatada na pesquisa) não foi avaliada; apenas a definição **implícita** (não expressa verbalmente) foi avaliada. Ele afirma que isso não seria problema exceto porduas limitações: a) esse tipo de definição provavelmente é mais suscetível de modificação não intencional (como, por exemplo, o desejo do experimentador de obter certos resultados); b) é mais difícil a replicação do método de mensuração por cientistas de outros laboratórios, uma vez que a definição real é desconhecida. Ele considera, a seguir, que mesmo quando o observador que faz o teste de fidedignidade não foi pré-treinado nas definições de interesse, ainda podem haver problemas, pois: a) esse segundo observador pode ter uma definição de comportamento implícita semelhante à do primeiro observador, e, b) a definição pode ser pobre e produzir alto acordo devido à topografia que facilita a discriminação de outros comportamentos: desse modo, ela continua sendo pouco adequada para replicação do estudo.

Uma discussão semelhante é encontrada em Hawkins e Dotson (1 975) sobre acordo entre observadores em registro de intervalo. Esses autores mostram que, quando se usa para cálculo da porcentagem de acordo entre observadores uma fórmula em que se computam tanto os acordos na ocorrência como os acordos na

não-ocorrência do comportamento, esse índice é altamente afetado pela taxa (ou duração) do comportamento, e não serve a nenhuma das seguintes funções: 1)como índice de quão precisa, clara, objetiva e completa é uma definição; 2) como ínice de quão competente é o observador; 3) como índice da confiabilidade no efeito experimental relatado.

A partir dessa colocação de Hawkins e Dotson (1975), a discussão de fidedignidade em análise do comportamento aplicada passou a girar em torno de modos de se evitar índices altos devidos ao acaso. Para isso, foram feitas sugestões no sentido de: a) calcular separadamente índices de acordo para ocorrência e para não ocorrência do comportamento (Bijou, Peterson e Ault, 1968; Hawkins e Dotson, 1 975); b) estabelecer meios de identificara proporção de acordo devida ao acaso (Johnson e Bolstad, 1973; Hartmann, 1977; Hopkins e Hermann, 1977; Birkimer e Brown, 1979, a, b; Yelton, 1979), muitos deles envolvendo uma certa sofisticação estatística. A reação de autores com tradição na área não tem sido muito favorável a esse segundo tipo de proposição. Ao comentar o artigo de Hartmann(1977), por exemplo, Baer(1977) afirma que o que interessa ao analista do comportamento é responder à questão: "Será que quaisquer dois observadores, usando o mesmo código de comportamentos, veriam os mesmos comportamentos do mesmo modo ao mesmo tempo?" questão essa que se refere à avaliação da homogeneidade dos observadores que atuam num mesmo projeto. Ele considera que essa resposta pode ser obtida através do cálculo separado das porcentagens de acordo sobre ocorrência e não ocorrência do comportamento registrado em intervalos e que as proposições sobre outros modos de calcularo acordo, inclusive levando em conta os valores obtidos ao acaso, são dispensáveis e até prejudiciais à área da análise do comportamento aplicada. Curiosamente, as colocações de Michael (sem data) não são retomadas nos trabalhos referentes a índices de acordo entre observadores.

Levando-se em conta as considerações acima, verifica-se que a questão da fidedignidade do observador é tratada na análise do comportamento aplicada privilegiando-se o cálculo do acordo entre observadores como indicador de fidedignidade. Busca-se identificar e controlar os fatores de erro que afetam esses índices, sempre na direção de preservar essa forma de aferir fidedignidade. Isso está ligado ao interesse central da análise do comportamento aplicada, que é o de demonstrar o efeito de variáveis experimentais sobre a freqüência (ou duração) do comportamento, produzindo mudanças de magnitude consideradas socialmente significativas. A exatidão da contagem é, portanto, crucial, para que não sejam confundidos os efeitos dessas variáveis com os erros de mensuração.

FIDEDIGNIDADE DO OBSERVADOR NA LITERATURA ETOLÓGICA

O livro organizado por Blurton Jones (1972) com uma coletânea de artigos sobre comportamento infantil e maternal marca o início da pesquisa etológica sistemática do comportamento humano. Em vários artigos, a questão da fidedignidade do observador é abordada, no sentido de enfatizar a replicabilidade das observações. Vários autores afirmam que um teste adequado da realidade das categorias identificadas é o fato de que outros pesquisadores independentes estarão também identificando categorias semelhantes (Blurton Jones, 1972; Leach, 1972; Richards e Bernal, 1972). Nota-se que, aqui, a preocupação básica é com a fidedignidade na identificação de categorias replicáveis, definidas em termos de unidades observáveis, contrapondo-se aos estudos de psicologia social

e do desenvolvimento da época, que faziam uso de categorias amplas de comportamento, e que exigiam dos pesquisadores certo grau de interpretação para decidir em que categoria enquadrar os eventos observados.

Uma outra ênfase notada nos comentários de alguns autores, no livro já citado, refere-se à desconfiança no índice de acordo entre observadores como indicador de fidedignidade. Argumenta-se que um grupo de observadores, depois de um treino suficiente, provavelmente começa a compartilhar entre si de interpretações quanto à identificação de categorias, mas que este grupo pode divergir de um outro grupo utilizando o mesmo sistema de categorias. Isso é visto como um obstáculo para a comunicação com leitores posteriores e, portanto, para a replicação do estudo (Blurton Jones, 1972; Smith e Connolly, 1972).

Nesse sentido, uma das soluções propostas e a do calculo do acordo entre observadores em que um deles não está familiarizado com as categorias do estudo; foi o que fizeram Smith e Connolly (1972) quando utilizaram um observador com treino de observação de crianças que estudou apenas as definições escritas dos comportamentos referentes àquele estudo específico e efetuou o cálculo de acordo com o observador principal.

A ênfase dos etólogos, então, é na replicabilidade das categorias selecionadas para estudo e na busca de padrões de exatidão do registro que evitem o mero consenso entre observadores, eventualmente desviando-se das próprias definições escritas.

Um texto mais recente na revista "Behaviour" (Caro, Roper, Young e Dank, 1979) discute a questão da fidedignidade entre observadores em estudos etológicos. Os autores apresentam índices para cálculo do acordo entre observadores e arrolam fatores que afetam o índice, baseando-se, em parte, na literatura sobre fidedignidade produzida pela análise do comportamento aplicada. Eles consideram que os métodos de concordância podem ser úteis, especialmente no que se refere à ocorrência do comportamento a cada intervalo, no treino de observadores ou no exame detalhado de padrões de comportamento.Eles fazem uma analogia entre o observador e um instrumento utilizado na observação científica: ambos funcionam diferentemente sob diferentes condições. Assim, os fatores que afetam o modo como os observadores registram o comportamento são análogos às diferentes condições que afetam o funcionamento de um instrumento científico. Os autores comentam, no final do artigo, que a fidedignidade entre observadores não é uma condição suficiente para garantir a validade de um estudo. Eles consideram que é possível obter uma alta fidedignidade entre observadores e um resultado estatisticamente significante a partir de uma amostra viesada; assim, o único teste da validade de um estudo é sua replicação. Eles afirmam, ainda, que as diferenças entre observadores não são a única fonte de erro, constituindo-se a avaliação dessas diferenças em uma estimativa moderada do erro total.

Assim, nesse texto, verifica-se uma valorização maior do acordo entre observadores como indicação de fidedignidade do que nos textos do livro de Blurton Jones, ao mesmo tempo que se mantém a ênfase na importância da replicação dos resultados. Isso está ligado aos objetivos da pesquisa etológica, de detectar e descrever padrões de comportamento para responder a perguntas em um ou mais dos seguintes níveis de análise: causal, ontogenética, funcional e filogenética (Blurton Jones, 1972). Assim, a detecção dos mesmos padrões por outros pesquisadores é uma indicação muito significativa de que se está selecionando categorias adequadas para análise. Por outro lado, reconhece-se a utilidade relativa do cálculo do acordo entre observadores, como uma das formas

de se aferir a precisão do registro de um observador ao longo de um estudo. Dessa forma, a questão da fidedignidade do observadora tratada na etologia de um modo mais diversificado do que na análise do comportamento aplicada: um observador fidedigno tanto pode ser aquele que obtém resultados replicáveis como aquele que obtém um alto índice de acordo com outro observador, preferindo-se para esse teste um observador não treinado naquele conjunto específico de categorias.

Assim, duas vertentes de pesquisa, com perguntas diferentes, enfocam diferentemente a questão da avaliação da fidedignidade do observador.

EM BUSCA DE UMA SOLUÇÃO

A avaliação da fidedignidade do observador em diferentes áreas de pesquisa apresenta pontos comuns e pontos divergentes, relacionados aos objetivos específicos da investigação em cada área.

Um ponto levantado por autores de ambas as áreas revisadas refere-se à relação entre teste do acordo entre observadores e o teste da qualidade de um sistema de categorização do comportamento. Um grupo de autores que trabalham na área de comunicação não-verbal chegou a uma solução interessante (Frey e Pool, mencionados em Fisch, Frey e Hirsbrunner, 1983). Eles desenvolveram um sistema de notação de posições de cada parte do corpo, a serem codificadas a partir da observação da gravação da sessão em video-teipe; em seguida, calcularam índices de acordo intra e intercodificadores e obtiveram valores acima de 90%. No entanto, esses autores consideraram que o índice de fidedignidade alto não garante a qualidade de um esquema de notação e que a avaliação dessa qualidade depende do grau de resolução do esquema de notação, ou seja, o grau em que o mesmo está apto a descrever fidedignamente as diferenças comportamentais existentes. Para tanto, procuraram reconstruir 40 posições randomicamente selecionadas com base nos dados codificados. Para esse objetivo, tomou-se uma pessoa que atuou como modelo e cada parte de seu corpo foi colocada nas posições correspondentes à codificação dos sujeitos originais. Fotos tiradas desse modelo foram então codificadas e comparadas aos códigos obtidos a partir das posições originais. Os resultados indicaram acordo acima de 98%, dos 2400 itens de dados obtidos para cada conjunto de fotos.

Nesse estudo, verificaram-se alguns pontos a destacar. Em primeiro lugar, foi feita uma distinção entre acordo na codificação e teste da qualidade do sistema de categorização. Além disso, o teste de acordo na codificação foi efetuado tanto entre codificadores como intracodificadores. Essa é uma nova possibilidade a ser explorada a partir do uso mais extensivo do equipamento de video-teipe em estudos observacionais.

De fato, uma possível solução para a avaliação da fidedignidade do observador é o uso exclusivo da aferição do acordo intra-observador, em duas codificações separadas por um intervalo de tempo do mesmo teipe. Isso foi feito por Marturano (1978), em um estudo descritivo do comportamento de crianças de pré-escola em sala de aula. Nesse trabalho, o estudo de fidedignidade abrangeu um teste do catálogo de comportamentos, através do acordo entre observadores na utilização do mesmo, e uma avaliação da estabilidade do pesquisador na utilização desse catálogo para a codificação do comportamento observado através do video-teipe. Essa avaliação de estabilidade foi realizada levando-se em conta que todas as codificações das sessões foram realizadas por um mesmo pesquisador. Calculou-se, então, a proporção de variação na codificação da mesma fita, a

intervalos de 48 horas e de seis meses da codificação original, obtendo-se valores considerados satisfatórios. Desse modo, apresenta-se uma possibilidade de o pesquisador atuar sozinho na análise dos dados e aferir sua precisão no registro de seu catálogo de comportamentos. Note-se que o pesquisador só utilizou o cálculo do acordo entre observadores para aferir inicialmente os itens do catálogo.

A guisa de conclusão, a partir de todas as considerações feitas, sugere-se que a avaliação da fidedignidade do observador seja pensada pelo pesquisador de modo a estar relacionada especificamente a seu objetivo de estudo e às características específicas de seu projeto de trabalho. Algumas sugestões são apresentadas a seguir;

a) Estudos preliminares, de caracterização de um fenômeno, podem se beneficiar mais de discussões informais entre pesquisadores do que de testes formais de fidedignidade. Uma vez formalizados os critérios para delimitação de unidades e estabelecido um sistema de categorias, deve-se passar à seleção de uma técnica adequada de avaliação da fidedignidade do observador.

b) Estudos em que um observador é o principal codificador do comportamento podem ter avaliações de fidedignidade através de várias técnicas, entre as quais pode-se propor a avaliação do acordo entre observadores, em que o observador convidado é um pesquisador experimentado que efetua alguns testes de acordo, tendo tido contato com o sistema de categorias utilizadas nesse estudo apenas por escrito. Outra possibilidade, quando se conta com gravações esporádicas ou sistemáticas em video-teipes, é a avaliação da estabilidade desse observador ao longo do tempo, na codificação do mesmo teipe.

c) Projetos de pesquisa que envolvam grupos de observadores para a coleta de dados terão necessariamente de contar com um sistema de aferição da precisão de cada um desses observadores. Na elaboração desse sistema de aferição, dever-se-á levar em conta a literatura que aponta para diferentes fatores que afetam os índices de acordo entre observadores. Esse sistema de aferição não indicará necessariamente a adequação do sistema de categorias utilizado. Entretanto, a análise diferenciada dos índices de acordo para diferentes categorias poderá fornecer elementos para a revisão do sistema ou para o treino diferenciado de algumas categorias em que os observadores apresentem maior dificuldade.

d) O teste do sistema de categorias, conforme foi discutido, pode receber contribuições dos dados obtidos a partir do teste de acordo entre observadores mas não deve se confundir com este, uma vez que é preciso levar em conta, também, o risco de observadores com alto índice de acordo estarem reagindo a uma definição implícita da categoria, que não coincide necessariamente com a definição explícita. A solução proposta por Fisch et al. (1983) parece muito boa: a partir do registro codificado, alguém reproduz o comportamento e este é codificado de acordo com o mesmo sistema. Comparam-se, a seguir, as duas codificações. Talvez não seja tão fácil representar certos tipos de comportamentos, da mesma forma que posturas, como foi feito no estudo citado. Entretanto, talvez seja útil tentar a execução dessa e de outras propostas com o objetivo explícito de teste do sistema de categorias.

Verifica-se, então, que a questão da fidedignidade do observador pode ser trabalhada de vários modos, e que essa discussão tem sido implementada nos últimos anos. O importante é reter o significado mais amplo do termo fidedignidade e estabelecer formas compatíveis com os objetivos de cada trabalho para lidar com a questão da avaliação da fidedignidade do observador.

REFERÊNCIAS

BAER, D. M. (1977). Reviewer's comment: just because it's reliable doesn't mean that you can use it. *Journal of Applied Behavior Analysis, 10,* 117-119.

BAER, D. M., WOLF, M. M., & RISLEY, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91-97.

BIJOU, S. W., PETERSON, R. F., & AULT, M. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis, 1,* 175-191.

BIRKIMER, J. C., & BROWN, J. H. (1979). A graphical judgmental aid which summarizes obtained and chance reliability data and helps assess the believability of experimental effects. *Journal of Applied Behavior Analysis, 12,* 523-533(a).

BIRKIMER, J. C, & BROWN, J.H. (1979). Back to basics: percentage agreement measures are adequate, but there are easier ways. *Journal of Applied Behavior Analysis, 12,* 535-543(b).

BLURTON JONES, N. (1972). Characteristics of ethological studies of human behaviour. Em N. Blurton-Jones. (Ed.). *Ethological studies of child behaviour.* Cambridge: Cambridge University Press.

CARO, T. M-, ROPER, R., YOUNG, M., & DANK, G. R. (1979). Inter-observer reliability. *Behaviour.* LXIX, 3-4, 303-315.

FERREIRA, M. C R. (1984). O apego e as reações da criança à separação da mãe: uma revisão bibliográfica. *Cadernos de Pesquisa, 48,* 3-19.

FISCH, H., FREY, S., & HIRSBRUNNER, H. (1983). Analyzing nonverbal behavior in depression, *Journal of Abnormal Psychology, 92* (3), 307-318.

HARTMANN, D. P. (1977). Considerations in the choice of inter-observer reliability estimates.*Journal of Applied Behavior Analysis, 10,* 103-116.

HAWKINS, R.P.,&D0TS0N,V. A. (1975). Reliability scores that delude: an Alice in Wonderland trip through the misleading characteristics of inter-observer agreement scores in interval recording. Em E. Ramp, & G. Semb (Ed.). *Behavior analysis: areas of research and application,* Englewood Cliffs, N. Jersey: Prentice-Hall.

HINDE, R. A. (1966). *Animal behaviour.* N. York: McGraw-Hill.

HOPKINS, B. L,& HERMANN, J. A. (1977). Evaluating inter-observer reliability of interval data. *Journal of Applied Behavior Analysis, 10,* 121 -126.

JOHNSON**,** S. M.,& BOLSTAD, 0. D. (1973). Methodological issues in naturalistic observation: some problems and solutions for field research. Em L. A.

Hamerlynck, L. C. Handy, & E. J. Mash (Eds.). *Behavior change: methodology, concepts, and practice.* Champaign, Illinois: Research Press.

KAZDIN, A. E. (1977). Artifact, bias and complexity of assessment: the ABCs of reliability. *Journal of Applied Behavior Analysis, 10,* 141-1 50.

KRATOCHWILL, T. R., & WETZEL, R. J. (1977). Observer agreement, credibility, and judgment: some considerations in presenting observer agreement data. *Journal of Applied Behavior Analysis, 10,* 133-139.

LEACH, G. M. (1977). A comparison of the social behaviour of some normal and problem children. Em N. Blurton-Jones (Ed.). *Ethological Studies of Child Behaviour.* Cambridge: Cambridge University Press.

MARTURANO, E. M. (1978). Um método para a observação e análise do comportamento da criança em sala de aula. *Psicologia, 4* (2), 37-73.

MARTURANO, E. M. (1979). Características do comportamento no jardim da infância: I-repertório básico. *Psicologia, 5* (1), 69-89.

MARTURANO, E. M., BERTOLDO, A. A., & CAMELO, A. L. P. (1982). Estudo descritivo do intercâmbio verbal em sala de aula através da análise de contingência - uma contribuição metodológica. *Psicologia,* 5(3), 19-36.

MICHAEL J. (sem data). Characteristics of a good definition. Apostila mimeografada para distribuição interna na American University.

PIERON, H. (1966). *Dicionário de Psicologia.* Tradução de Cullinan, D. B., do original francês de 1951. Rio de Janeiro: Globo.

RAMP, E., & SEMB, G. (Eds.). (1975). *Behavior analysis: areas of research and application.* Englewood Cliffs, N. Jersey: Prentice-Hall.

RICHARDS, M. P. M., & BERNAL, J. F. (1972). An observational study of mother-infant interaction. Em N. Blurton-Jones (Ed.). *Ethological Studies of Child Behaviour.* Cambridge: Cambridge University Press.

SCHERER, K. R., & EKMAN, P. (1982). *Handbook of methods in nonverbal behavior research.* Cambridge: Cambridge University Press.

SILLAMY, N. (1 980). *Dictionnaire encyclopedique de Psychologie.* Paris: Bordas.

SMITH, P. K.,& CONNOLLY, K. (1972). Patterns of play and social interaction in pre-school children. Em N. Blurton-Jones (Ed.). *Ethological Studies of Child Behaviour.* Cambridge: Cambridge University Press.

WARREN, H. C (1934). *Dictionary of Psychology.* Boston: Houghton Mifflin.

WILDMAN, B. G., & ERICKSON, M. T. (1977). Methodological problems in behavioral observation. Em J. D. Cone, & R. P. Hawkins (Eds.). *Behavioral assessment: New directions in clinical psychology.* N. York: Brunner, Mazel.

WOLMAN, B. B. (1973). *Dictionary of behavioral science.* N. York: Van Nostrand Reinhold.

YELTON, A. R. (1979). Reliability in the context of the experiment. *Journal of Applied Behavior Analysis, 12,* 565-569.

Artigo recebido em setembro de 1985.