

TEORIA DA MENSURAÇÃO: INCONVENIENTE NECESSÁRIO?

Nicolino Trompieri Filho¹
José Anchieta Esmeraldo Barreto²

(...) mas há razões, se as procurarmos, encontramos-las sempre, razões para explicar qualquer coisa nunca faltarão, mesmo não sendo as certas (19:p.106).

RESUMO

O artigo trata de alguns aspectos da teoria da mensuração na área educacional e suas relações com a avaliação. Discute soluções teóricas apresentadas por Stevens e Torgerson, dando atenção aos limites e possibilidades das escalas de medida. Aborda também o modelo linear clássico de mensuração, discutindo os conceitos de medida verdadeira, medida observada e erro de medida. Finalmente, apresenta indicações para a construção, de forma adequada, de instrumentos de medida.

Palavras-chave: *avaliação – mensuração – instrumentos de medida*

ABSTRACT: *Measurement Theory: A necessary inconvenient?*

This article deals with some aspects of the theory of measurement as applied to education and especially its relationship with educational assessment. It discusses theoretical solutions presented by Stevens and Torgerson paying special attention to the limits

¹ Mestre em Avaliação pela UFC e professor adjunto do Departamento de Fundamentos da Educação da UFC.

² Ph.D em Educação pela Universidade de Tulane e professor titular da UFC. E-mail: anchietab@yahoo.com.br

and possibilities of measurement scales. It also discusses the classical linear model of measurement analyzing the concepts of real measure, observed measure as well as error in measure. Finally it offers suggestions for the construction in adequate form of instruments to measure securely.

Key-words: *evaluation – measurement – scales of measurement.*

Introdução

Um olhar sobre a trajetória da avaliação educacional mostra fases que devem ser analisadas, não somente nelas mesmas, mas também na relação com a função da medida.

De acordo com Penna Firme (15:p.5-12), nos últimos cem anos houve *quatro gerações* de avaliadores, que diferem umas das outras, principalmente nas concepções de avaliação que defendem. Não se pode falar, rigorosamente, em rupturas de paradigmas, embora seja possível registrar algumas diferenças operacionais e de ênfase nos diversos momentos da ação avaliativa.

Na primeira fase, avaliação era igual a medida. Tudo se resumia em medir. Por conta da transferência do paradigma positivista das Ciências Naturais para as Ciências Humanas, a medida apareceu como condição necessária e suficiente para avaliação. Dessa maneira, o importante era a construção de testes. Se havia um bom instrumento de medida, todas as dificuldades da avaliação estavam resolvidas. Este pensamento predominou com maior intensidade entre os anos 1920 e 1930.

Nos anos finais da década de 1930 e na de 1940, surge a *geração descritiva*. Uma única medida não era suficiente para se avaliar. Deviam-se considerar outras informações que possibilitassem melhor descrição daquilo que os objetivos determinavam. Além da ênfase na medida, havia a preocupação de que os dados fossem *obtidos em função do alcance dos objetivos* (15:p.7).

Foi a mudança do foco para os objetivos que provocou uma terceira onda no movimento de avaliação educacional. Rapidamente se verificou a dificuldade que apresentava definir objetivos claros e precisos. Portanto, era quase impossível deixar de reconhecer outro componente da avaliação: o juízo de valor, no que se refere ao seu mérito e relevância. Esta geração desenvolveu seus estudos a partir da década de 1950 até aos anos 1980.

A última etapa, que se iniciou nos primeiros anos da década de 1990 e ainda predomina hoje, é o da negociação. Parece, a um observador mais atento, que aqui se pretende uma fusão entre avaliação e desenvolvimento organizacional. É uma tarefa muito difícil de ser levada a cabo, principalmente pela deficiência de recursos humanos, tanto num setor quanto noutro.

Três observações se impõem: a) o percurso feito pelos representantes teóricos das quatro gerações mencionadas não tiveram seu correspondente nas atividades práticas, principalmente nas ações avaliativas dos professores em sala de aula e em provas seletivas, tais como concurso para cargos públicos, vestibulares e testes de avaliação de desempenho, como o Exame Nacional de Cursos. Nestas situações, avaliação e medida são normalmente iguais; b) apesar de serem conhecidas, pelo menos por alguns educadores, as exigências técnicas de um instrumento de medida, cada vez menos se observa em trabalhos científicos publicados, dissertações e teses a preocupação dos autores com a *precisão e validade* de suas medidas; c) qualquer que seja sua orientação filosófico-pedagógica, o resultado de uma avaliação será comprometido se não houver garantia de um mínimo de *precisão e validade*.

É claro que há razões para tal estado de coisas, mesmo não sendo as certas. Uma delas é a falta de embasamento teórico dos que eventualmente utilizam instrumentos de medida nos seus trabalhos. Os educadores, na sua grande maioria, não dispõem dos conhecimentos teóricos necessários uma vez que, há muito tempo, a disciplina *medidas educacionais*

deixou de ser ofertada nos cursos de formação de professores. É preciso ter sempre em mente o fato de que a condição necessária para uma avaliação eficiente reside na garantia de que os dados utilizados são dignos de confiança.

Estas considerações justificam a apresentação, neste trabalho, de maneira mais simples possível, de reflexões sobre a teoria da mensuração, mesmo reconhecendo ser ela, para muitos, apenas um *inconveniente*. Espera-se, entretanto, que os subsídios teóricos apresentados os convençam a acrescentar um adjetivo e transformá-la em *inconveniente necessário*. Concluindo, observações sobre construção de instrumento de medidas serão feitas.

Teoria da Mensuração

A teoria da mensuração opera utilizando conceitos que, na linguagem comum, dão margem a diferentes interpretações, daí a necessidade de defini-los. Estes conceitos são: objeto ou sistema, atributo, propriedade mensurável, magnitude e quantidade. De forma simples, pode-se defini-los da seguinte maneira:

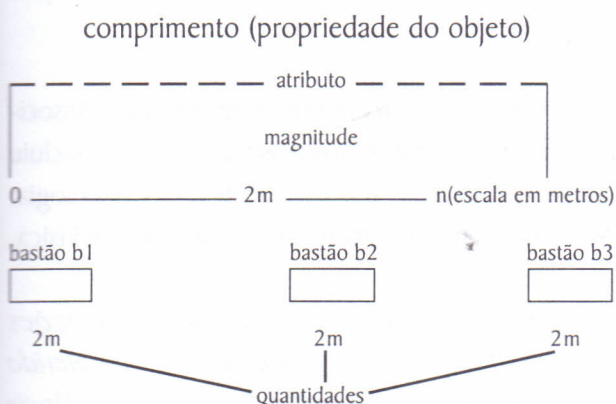
- *objeto ou sistema* – coisa que possui uma ou mais propriedades;
- *atributo* – qualquer propriedade mensurável do objeto;
- *propriedade mensurável* – propriedade do objeto capaz de apresentar graduações;
- *magnitude* – particularização de um atributo;
- *quantidade* – uma instância particular de uma magnitude singular.

Veja-se como, na prática, estes conceitos podem ser identificados. Torgerson (22:p.26) fornece como exemplo um *bastão de madeira*. Aqui o objeto é um elemento qualquer de um conjunto de artefatos de madeira que tem determinado comprimento; o atributo refere-se ao contínuo de comprimento que os objetos dessa classe podem apresentar, por exemplo, na escala métrica com valores no intervalo [0; n

] A magnitude é o ponto do contínuo que corresponde a um dado comprimento de bastão de madeira. Se, na classe constituída por bastões de madeira, existem bastões aos quais se pode associar uma mesma magnitude, neles o comprimento se apresenta com quantidades iguais. Russell explica a relação entre magnitude e quantidade ao assinalar:

quando uma magnitude pode ser particularizada por uma posição temporal, espacial ou espacial-temporal ou, quando sendo uma relação, pode ser particularizada levando em consideração um par de termos entre os quais se situa então, a magnitude, assim particularizada, é chamada uma quantidade (17:p.167).

Graficamente ter-se-ia:



A distinção entre os conceitos de magnitude e de quantidade é fundamental. A construção das escalas de mensuração se sustenta em postulados referentes a relações entre quantidades ou entre magnitudes. As regras de correspondência devem ser apropriadas para cada caso. No primeiro, a ênfase situa-se no sistema empírico, enquanto que, no segundo caso, está no sistema formal.

A teoria da mensuração desenvolveu-se fundamentada gnosiologicamente no idealismo platônico. Platão admite três classes de número:

- *números ideais* – são os números em si, pertencem ao mundo das idéias e deles procedem os demais números;

- *números matemáticos* – os conceitos, desprovidos de matéria e dotados de quantidade, e que se encontram na mente;
- *números sensíveis* – são os que se encontram unidos aos objetos do mundo material.

Assim, um atributo do objeto (propriedade mensurável) tem existência *a priori* na mente como magnitude que pode ter sua quantidade determinada. O objetivo da mensuração é fazer corresponder à propriedade, tal como se apresenta no objeto, uma magnitude que identifique com precisão a quantidade da propriedade no objeto. Como se espera demonstrar, todo o problema da medida reside no maior ou menor sucesso da associação *propriedade – magnitude – quantidade*.

Hölder, em 1901 (13:p.232-233), formulou condições para se levar a mensuração a cabo através dos axiomas de quantidade, que especificam as condições que as quantidades deveriam atender para ser usadas em equações matemáticas. Esses axiomas foram expressos em termos de quantidades reais de elementos específicos. São eles:

- a) axiomas referentes à identificação de ordem ou de igualdade:

a.1. Dadas duas quantidades quaisquer a e b, então $a > b$ ou $a < b$ ou $a = b$.

a.2. Dadas três quantidades quaisquer a, b e c, então, se $a > b$ e $b > c$, então $a > c$.

a.3. para cada elemento A com quantidade a, existe um elemento B, distinto de A, com quantidade B, tal que $a = b$.

a.4. Dadas três quantidades quaisquer a, b e b', se $a > b$ e $b = b'$, então $a > b'$.

a.5. Dadas duas quantidades quaisquer a e b, se $a = b$, então $b = a$.

a.6. Dada uma quantidade a qualquer, existe sempre uma quantidade b tal que $a > b$.

- b) Axiomas referentes à condição para que se possam somar quantidades para se obter nova quantidade.

b.1. Dadas duas quantidades quaisquer a e b, existe uma quantidade c tal que $c = a + b$.

b.2. Dadas duas quantidades quaisquer a e b , existe sempre uma quantidade a' tal que $a + b > a'$.

b.3. Dadas duas quantidades quaisquer a e b , há sempre duas outras quantidades a' e b' tais que $a + b = a' + b'$.

Dadas três quantidades quaisquer a , b e c , então

b.4. $a + b = b + a$

b.5. $(a + b) + c = a + (b + c)$

b.6. se $a < b$, existe um número inteiro n tal que $na > b$.

Campbell, em uma série de trabalhos publicados no período 1920 – 1938 (13: p.233), desenvolveu a primeira exposição completa dos fundamentos lógicos da mensuração. Campbell define mensuração como *a atribuição de numerais para representar propriedades de sistemas materiais, que não números, em virtude de leis que governam essas propriedades* (23:p.13). Considera a mensuração como o isomorfismo entre a idéia de magnitude e a quantidade da propriedade a ser medida; demonstrando-se o isomorfismo quando se prova que as quantidades atendem aos axiomas de Hölder. Mesmo na mensuração de propriedades físicas dos objetos isso se torna problemático. Se, por um lado, os axiomas de Hölder podem ser comprovados empiricamente para propriedades como massa, comprimento etc., podendo-se ordenar e somar quantidades, de outra parte, muitas outras propriedades físicas não atendem aos axiomas que se referem à soma de quantidades para se obter nova quantidade, como é o caso da densidade e da temperatura. Tome-se, por exemplo, a situação em que se tenham três magnitudes de água correspondentes às quantidades $a = 2$ litros, $b = 3$ litros e $c = 4$ litros. Neste caso, todos os axiomas de Hölder são atendidos. Apenas para ilustrar, o axioma b.5 se verifica, uma vez que $(2l + 3l) + 4l = 2l + (3l + 4l) = 9l$.

O mesmo não acontece com as quantidades de temperatura que não atendem às propriedades aditivas expressas nos axiomas enunciados no item b. Como exemplo, tomem-se duas magnitudes correspondentes

às quantidades $a = 30^{\circ}\text{C}$ e $b = 50^{\circ}\text{C}$. Empiricamente se verifica que, ao se somarem as duas magnitudes a quantidade c de temperatura não será igual a $a + b$. Se a e b representam a temperatura de dois volumes de água, juntando-se estes, a temperatura resultante não será a soma das duas temperaturas originais.

O problema se complica quando são consideradas propriedades psicológicas, educacionais ou sociais. Segundo Muñiz,

Esse tratamento da teoria da mensuração assume um ponto de vista muito restrito daquilo que se pôde considerar mensuração. Havia somente um tipo de escala tolerável; a quantidade tinha um conjunto de propriedades e estas eram estritamente não negociáveis. As magnitudes teriam que as cumprir ou não teriam que ser medidas... O sistema, claro está, não oferecia a possibilidade de incluir atributos psicológicos (13: p.234).

Em 1940, um comitê designado pela Associação Britânica para o Progresso da Ciência concluiu que a mensuração, em Educação e em Psicologia, não era da mesma espécie da mensuração em Física.

Stevens, referindo-se às declarações contraditórias no relatório final do referido comitê, conclui que talvez a concordância possa melhor ser alcançada se reconhecermos que a mensuração existe sob uma variedade de formas e que as escalas de mensuração situam-se em classes definidas. Essas classes são determinadas tanto pelas operações empíricas envolvidas no processo de medir, quanto pelas propriedades formais (matemáticas) das escalas (9:p.713).

A Solução de Stevens

Esta postura de Stevens o levou a propor uma solução para o impasse. Eliminando a restrição de

que as medidas teriam que obedecer necessariamente aos axiomas de quantidade, sugere, em 1946, uma inovação fundamental à teoria da mensuração. Enquanto Campbell trabalha com postulados referentes a relações entre quantidades, Stevens vai fundamentar sua proposta em postulados referentes a magnitudes, permitindo a obtenção de medidas que não seriam possíveis na visão de Campbell.

Definindo mensuração como a atribuição de numerais a objetos ou eventos de acordo com leis (20:p.4), e tomando os postulados de ordem de Huntington³ como capazes de garantir o isomorfismo entre idéias de magnitude e a de quantidade, distingue os diferentes tipos de atribuição (tipos de escala), classificando os níveis de medida de acordo com a regra de atribuição de numerais que torna invariante a forma da escala.

O problema então, é tornar explícitas: a) as várias regras para a atribuição de numerais; b) as propriedades matemáticas (ou estrutura do grupo) das escalas resultantes, e c) as operações estatísticas aplicáveis às mensurações feitas com cada tipo de escala... As escalas são possíveis em primeiro lugar, porque há um certo isomorfismo entre o que nós podemos fazer com os aspectos dos objetos e as propriedades das séries de numerais (20:p.4-5).

A eliminação da restrição de Campbell permitiu garantir a existência de quatro classes possíveis de atribuição de numerais, isto é, quatro tipos de escala: nominal, ordinal, intervalar e de razão.

Segundo Stevens, a escala de medida determina que tipo de técnica estatística é aplicável às medidas que a mesma produz.

³ Esses postulados são: sejam as magnitudes a, b e c .

a) se $a \neq b$, então, ou $a > b$ ou $a < b$

b) se $a < b$, então $a \neq b$

c) se $a < b$ e $b < c$, então $a < c$.

O quadro a seguir resume a classificação de Stevens:

QUADRO Nº 1
Classificação das Escalas de Medidas de Stevens, (19:p.5)

TIPO DE ESCALA	OPERAÇÕES EMPÍRICAS BÁSICAS (*)	ESTRUTURA MATEMÁTICA DO GRUPO (**)	ESTATÍSTICAS PERMITIDAS
1 Nominal	Determinação de igualdade	Grupo permutativo $x' = f(x)$, onde $f(x)$ significa qualquer substituição uma a uma	Número de casos, moda, coeficiente de contingência
2 Ordinal	Determinação de maior ou menor	Grupo isolônico $x' = f(x)$, onde $f(x)$ é qualquer função monótona crescente	Mediana, percentis, Correlação ordinal
3 Intervalar	Determinação de igualdade de intervalos	Grupo linear $x' = ax + b$ $a > 0$	Média aritmética, Desvio padrão, correlação, produto-momento
4 Proporcional ou de razão	Determinação de igualdade de proporções	Grupo de similaridade $x' = cx$ $c > 0$	Todas

(*) operações básicas para a criação de uma escala.

(**) transformações matemáticas que deixam invariante a forma da escala.

As restrições ao emprego de técnicas estatísticas conforme o tipo de escala levantou uma polêmica entre especialistas em psicometria. Argüía-se se era possível, com dados ordinais, o uso de técnicas estatísticas que requeriam dados intervalares. Em oposição à visão de Stevens, uma variedade de psicometristas,

(...) por exemplo, Lord (1953), Burke (1953), Anderson (1961), argüíram que McNemar (1962) e Hays (1963) argüíram que as estatísticas se aplicam a números e não a coisas e que as propriedades formais das escalas de medida, como tais, não teriam influência na escolha das estatísticas (2: p.17).

Os defensores da posição de Stevens ficaram conhecidos como teóricos da escola da mensuração fraca (weak measurement), em oposição aos teóricos da estatística robusta (strong statistics). O debate teórico em torno desta questão estendeu-se por toda a década de 1950 e ocorre até hoje. A posição de Burke (13:p.242) é reveladora do lado dos teóricos da estatística robusta. Demonstrando que as propriedades de uma escala de mensuração envolvem correspondências entre conjuntos de axi-

omas sobre objetos e números, com relações e operações apropriadas, e que os métodos estatísticos começam e terminam com números, deduz, dessas duas proposições, que as propriedades do conjunto de números de uma escala de medida não têm nenhum efeito sobre a escolha das técnicas estatísticas para representar e interpretar as medidas fornecidas pela escala.

Baker, Hardyck e Petrinovich, considerando que, para Stevens e seguidores, testes paramétricos de significância, tais como os testes *t* de Student e *F* de Snedecor, somente podem ser usados com medidas obtidas com escalas nas quais transformações lineares são admissíveis (escalas intervalares e de razão) vão buscar empiricamente resposta à pergunta se é possível *tornar decisões corretas sobre a natureza da realidade e negligenciar da escala de medida quando aplicamos testes estatísticos*.(2:p.19). O debate é deslocado para o campo empírico quando, até então, se desenvolvia somente no campo teórico.

Visando a responder empiricamente à indagação que, na verdade, se refere à possibilidade ou não do uso de testes estatísticos paramétricos, com dados obtidos com escalas ordinais, quando, na visão de Stevens, a utilização desses testes supõe, pelo menos, o nível intervalar de mensuração, esses autores, a partir de uma distribuição de escores intervalares, construíram, através de transformações não lineares, uma série de distribuições com dados ordinais e geraram, em computador, amostras aleatórias com esses dados transformados. Observaram que o uso de medidas ordinais não afeta seriamente os valores observados de *t* para os níveis de significância de 5% e 1%, quando comparados com os valores de *t* referentes a medidas intervalares correspondentes, e que as restrições de Stevens somente são válidas para as medidas da estatística descritiva.

Esta visão pode ser correta se consideram determinações específicas únicas de uma estatística no sentido descritivo... mas, é

*incorreta quando aplicada ao problema da inferência estatística... Os dados presentes indicam que estatísticas robustas como o teste *t* são mais que adequadas para serem usadas com medidas fracas e, com algumas reservas menores, as probabilidades estimadas a partir da distribuição *t* são pouco afetadas pelo tipo de escala de medida usada* (2:p.23).

Os autores sugerem que, no caso de distribuições assimétricas e amostras com tamanhos diferentes, deve-se usar somente teste bicaudal.

A Solução de Torgerson

Torgerson refuta a definição de mensuração de Stevens (atribuição de numerais a objetos ou eventos) assegurando que são as propriedades dos objetos que são passíveis de mensuração, *... a mensuração pertence a propriedades dos objetos, e não aos objetos. Então, um bastão de madeira não é mensurável em seu uso do termo, embora seu comprimento, peso, diâmetro e dureza possam sê-lo* (23:p.14). Para ele, o problema da mensuração refere-se ao processo e à racionalidade envolvidos na construção de uma escala e às propriedades que lhe podem ser atribuídas. Considerando o tipo de representação numérica da escala, Torgerson classifica as escalas de medida de acordo com as características dos números, capazes de garantir o isomorfismo entre as magnitudes e as quantidades.

Essas características são:

- a) de ordem – dadas duas quantidades quaisquer *a* e *b* e se *a* \neq *b*, então *a* > *b* ou *a* < *b*;
- b) de distância – dadas quatro quantidades quaisquer *a*, *b*, *c* e *d* e se *a* \neq *b* \neq *c* \neq *d*, então (*a* – *b*) > (*c* – *d*) ou (*a* – *b*) < (*c* – *d*) ou (*a* – *b*) = (*c* – *d*);
- c) de origem – a escala possui ou não origem única, indicada por zero.

Qualquer escala apresenta a característica

de ordem e pode, também, apresentar, ou uma das duas outras, ou ambas. Assim, a classificação de Torgerson é feita em função das características de distância e de origem, em uma escala.

- d) Dadas três quantidades quaisquer a, b, e c, se $a \# b \# c$, pode-se ou não determinar se $(b - a) > (c - b)$ ou $(b - a) = (c - b)$, ou $(b - a) < (c - b)$;
- e) ou o zero indica ausência do atributo (a escala possui origem natural), ou o zero é relativo, não indicando ausência do atributo (a escala possui origem não natural), ou não apresenta zero.

O quadro a seguir resume a classificação de Torgerson.

QUADRO Nº 2

Classificação de Torgerson das Escalas de Medida

Características de origem \ Características de distância	Não natural	Natural
Não determinável	Escala ordinal	Escala ordinal com origem natural
Determinável	Escala intervalar	Escala de razão

Torgerson não considera as escalas do tipo nominal da classificação de Stevens como escalas de medida, pois, de acordo com sua definição de mensuração, os valores fornecidos por uma escala nominal não são números que traduzem quantidades, os numerais utilizados são simplesmente símbolos indicativos de categorias e permitem somente a classificação de objetos em função de propriedades não mensuráveis. As escalas dos tipos ordinal, intervalar e de razão são as mesmas da classificação de Stevens. Quanto às escalas ordinais com origem natural, elas constituem um tipo de escala ordinal com um zero absoluto, indicando ausência no objeto da propriedade sob mensuração.

Na mensuração de atributos como atitudes, estética, preferência e calores, a origem natural ocorre dentro das séries e pode ser descrita com um ponto neutro onde todos os estímulos ou indivíduos em uma direção são favoráveis, agradáveis, apreciados ou desejados, conforme o caso, enquanto que todos aqueles no outro lado são desfavoráveis, desagradáveis, menosprezíveis ou indesejáveis... para esses atributos... os números atribuídos aos elementos em um lado da origem natural possuem sinal negativo e aqueles no outro lado têm sinal positivo (23:p.30).

Limites e Possibilidades das Escalas

Tanto na classificação de Stevens quanto na de Torgerson, os quatro tipos de escala restringem, em graus diferentes, a atribuição de números. A restrição à atribuição de números implica a quantidade de informação contida na medida obtida por um tipo de escala. Assim, a medida de uma propriedade obtida com uma escala ordinal possui menos informação do que a medida da mesma propriedade obtida por uma escala intervalar. Esta, por sua vez, contém menos informação do que a medida fornecida por uma escala de razão. As escalas de razão, por atenderem a todos os axiomas de Hölder, permitem o nível mais elevado de mensuração e, conseqüentemente, maior informação sobre a propriedade medida no objeto. Torgerson exemplifica com a mensuração do raio e da área de um conjunto de círculos, visando a expressar matematicamente a relação entre essas duas propriedades de círculo. Com as medidas dos raios e áreas tomadas com uma escala ordinal, a área é uma função monótona do raio. *Isto é, se a área ou o raio ou ambos são medidas com uma escala ordinal, nós podemos determinar somente que a relação entre os dois é monótona (23:p.21).* Com uma escala ordinal com origem natural, a área é uma função monótona do raio a partir

da origem; se o raio é igual a zero, a área é igual a zero. Com uma escala intervalar, a área é uma função quadrática do raio, e com uma escala de razão, a área é proporcional ao quadrado do raio. Nesse caso, é possível verificar a existência de uma constante de proporcionalidade: em todos os círculos, a razão entre a área e o quadrado do raio é igual ao número π .

Enquanto Campbell considerava as relações empíricas como depositárias de toda informação relevante (as quantidades deveriam atender aos axiomas de Hölder), Stevens e Torgerson deslocam a ênfase para as propriedades da escala: as magnitudes devem atender aos axiomas de ordem de Huntington, no caso de Stevens, ou às características da representação numérica capazes de garantir o isomorfismo entre as magnitudes e quantidades, em Torgerson.

Independentemente da ênfase estar no sistema empírico ou no formal, esses autores colocam, como condição necessária à mensuração, a relação axiomática entre os dois sistemas. Ocorre aí uma inversão de prioridade, pois a identificação da natureza das relações entre as magnitudes da escala e as quantidades (entidades empíricas) não é uma consequência da escolha de um tipo de escala para se realizar a mensuração de uma propriedade, e sim um pré-requisito.

Lord e Novick, ao desenvolverem a generalização do modelo clássico de mensuração, rompem com a concepção idealista platônica de *um procedimento para a atribuição de números (scores, medidas) a propriedades específicas de unidades experimentais de modo a caracterizar e preservar relações do domínio empírico* (11:p.17), fundamentando-se numa visão realista.

Os autores acrescentam à classificação de Stevens um quinto nível de mensuração, constituído pela escala absoluta que apresenta as características das escalas de razão e, adicionalmente, a unidade de mensuração é fixa. Qualquer operação de contagem produz medida absoluta. Ao atributo sob mensuração são associados números, de modo que todas as propriedades dos números refletem propriedades do atributo. Nesse caso, nenhuma transformação é

permissível, qualquer transformação destrói alguma propriedade da escala.

Em relação às escalas nominais, é feita uma restrição: somente as escalas dicotômicas produzem medida. Escalas nominais com três ou mais categorias são simplesmente escalas classificatórias, podendo, no entanto, ser dicotomizadas, tomando-se uma das categorias como referência.

O Modelo Linear Clássico de Mensuração

Verificadas as dificuldades teóricas para mensuração na área psicológica, educacional ou social, necessário se fez a busca de modelos que pudessem minimizá-las. Foi o que ocorreu com a utilização de testes na mensuração nestas áreas. Tornou-se necessária a busca de um modelo que permitisse, através do uso de técnicas estatísticas, estimar o erro da medida, uma vez que este não é observado diretamente.

O modelo linear clássico, desenvolvido por Spearman no início do século XX, tem por objetivo central a estimação do erro cometido quando se utiliza um teste para medir variáveis psicológicas, no entanto, o modelo pode ser estendido a situações de mensuração com os mais variados tipos de instrumento de medida.

O modelo linear clássico estabelece que a medida (X), atribuída por meio de um instrumento de medida a um sujeito em uma dada propriedade, é igual à soma de dois componentes: a medida verdadeira (X_v) do sujeito nessa propriedade (X_v é desconhecida) e o erro da medida (e) cometido no processo de mensuração⁴, isto é, $X = X_v + e$.

Na descrição do modelo acima, a expressão atribuída é um termo-chave para se compreender a fundamentação platônica do conceito de medida. O conceito de medida verdadeira era definido sintaticamente (i.e., matematicamente) e interpretado semanticamente (i.e., no mundo real) como a medida obtida através de um teste com um número infinito de itens.

⁴ São fontes do erro: o instrumento, o observador e o objeto sob mensuração.

Nesse teste ideal, os itens abarcariam, em seu conjunto, todos os estímulos passíveis de ensinar todas as respostas possíveis. A propriedade por ele medida poderia se apresentar, através da medida, em qualquer ponto de uma escala contínua.

O fato de a medida verdadeira de uma propriedade não poder ser observada diretamente cria um problema de ordem epistemológica, produzindo os três pontos de vista sobre o conceito de medida verdadeira encontrados na literatura sobre a importância prática e teórica do conceito. Segundo Loevinger (apud 11:p.27), a medida verdadeira, não sendo diretamente observável, tem seu conceito desprovido de importância prática. Para Thorndike, (apud 11:p.27), o fato de o conceito não ser diretamente observável reveste-o de um caráter místico e, portanto, não apresenta importância teórica. O terceiro ponto de vista é que o conceito é útil teoricamente e permite a obtenção de resultados práticos.

Esta não é uma posição metafísica, nós não patrocinamos uma teoria da mensuração que postule inúmeras proposições que são incapazes de verificação prática. Nós usamos a noção de score verdadeiro porque achamos que ela é útil teoricamente e produz implicações que podem ser verificadas na prática efetiva (11:p.27).

Em relação ao conceito de medida verdadeira, desenvolveram-se três concepções. A primeira, com fundamentação platônica, define a medida verdadeira a partir de uma situação específica de mensuração (m), na qual um sujeito (s) apresenta, na propriedade (p), que está sendo medida, única quantidade constante verdadeira (X_v) – a medida verdadeira de (s) em (p). Qualquer mensuração particular de (p) em (s) na mensuração (m) conduz a um valor (X) – a medida observada – que, em geral, difere da quantidade X_v em razão do erro de medida. O valor X é observável, mas o valor X_v não o é. Na Física clássica, essa concepção parece ser razoável. Se a situação de mensuração é especificada com precisão, há

especificação dos procedimentos de mensuração e da escala de medida utilizada, é correto falar na velocidade verdadeira da luz, no peso verdadeiro de um objeto etc. Por outro lado, nas Ciências Humanas, cujas teorias são elaboradas com base em construtos, a concepção platônica de medida verdadeira não é capaz de sustentar uma fundamentação axiomática satisfatória para essa teoria.

A segunda concepção a se desenvolver tem fundamentação empirista, com base na teoria de probabilidades, de Von Mises. Ainda é a mais freqüente na literatura sobre mensuração em Psicologia e em Educação. Von Mises, em reação ao subjetivismo e apriorismo da teoria clássica de probabilidades, elaborou uma estruturação da teoria análoga àquela das Ciências Naturais, dando ao conceito de probabilidade uma fundamentação empírica através do conceito de *coletivo*.

Coletivo constitui uma sucessão infinita de observações que realizam suas condições:

1) as freqüências relativas dos atributos particulares de elementos singulares do coletivo tendem para um limite definido; e 2) este limite não é afetado por seleções localizadas (princípio do azar). O valor limite da freqüência relativa de um atributo dado, admitido que é independente de qualquer seleção localizada, será chamado 'a probabilidade' deste atributo; dentro do coletivo dado, só é possível falar de probabilidade com respeito a um coletivo definido com propriedade (22:p.22).

Toranzos (22) apresenta as três objeções mais freqüentes na literatura à teoria de Von Mises: 1) a probabilidade como valor-limite para o número de provas tendendo para o infinito não fundamenta empiricamente o conceito, pois, um experimento com as n provas tendendo para o infinito não é realizável, uma vez que não é possível prolongar idealmente o experimento porque os processos regidos pelo acaso, se prolongados racionalmente, perdem o caráter aleatório, que é sua essência; 2) a segunda condição

do conceito de coletivo postula a aleatoriedade, introduzindo um elemento subjetivo que rompe, dessa forma, com o fundamento empirista, e 3) a supressão dos conceitos de probabilidade de um caso isolado e de probabilidade a priori restringe o cálculo de probabilidade; esses conceitos têm se revelado muito fecundos em aplicações.

A concepção empirista de medida verdadeira entende-a como valor-limite se, em uma dada situação (m) de mensuração, é feita, no sujeito (s), uma série sucessiva de mensurações da propriedade (p), a média das medidas X_i ($i = 1, 2, \dots, n$; $n \rightarrow \infty$) de (p) em (s) converge, com probabilidade 1, para um valor constante; esse valor constante é a medida verdadeira de (s) em (p).

A terceira posição, desenvolvida por Lord e Novich (1968) como base para a generalização da teoria clássica, fundamenta-se na teoria axiomática intuitiva de probabilidades, desenvolvida por Frechet, Cramer e outros (22:p.23).

A teoria axiomática intuitiva de probabilidades apresenta uma fundamentação axiomática própria da Matemática, contendo, ao lado dos axiomas formais, outros de caráter existencial. Segundo Toranzos (22:p.23), a adoção de suposições que, apesar de não se harmonizarem com a perfeição e rigidez axiomática, como, por exemplo, o conceito de aleatoriedade como idéia primitiva, permite fundamentar uma teoria capaz de aplicações à realidade. A teoria também parte, como a de Von Mises, da noção de frequência, diferindo, no entanto, por evitar a passagem ao limite com o número de mensurações tendendo ao infinito na situação de mensuração.

Toranzos reproduz a definição de probabilidade de Frechet.

Verifica-se que, numa determinada categoria de provas, quando se calculam as frequências de um acontecimento fortuito em diferentes grupos de provas, essas frequências diferem pouco, praticamente, quando cada um dos grupos se constitui de numerosas provas. As

frequências de um acontecimento fortuito E, nos grupos, cada um composto de numerosas provas, todas pertencentes a uma mesma categoria C, são valores experimentais de uma constante física, determinada pela natureza do acontecimento E, e pela categoria C. É uma constante física que chamaremos: a probabilidade do acontecimento E na categoria de provas C (22:p.23).

Em outras palavras, tomando-se uma sequência constituída por K experimentos aleatórios idênticos, cada um deles com um número grande de provas, observa-se que as frequências relativas de um evento, com ocorrência possível em cada prova, distribuem-se em torno de um valor constante; a esse valor denomina-se probabilidade do evento E.

Segundo Lord e Novick (11), os modelos sustentados por processos de mensuração que descrevem o comportamento de variáveis psicológicas e, por extensão, educacionais, são modelos probabilísticos em virtude do problema do erro da medida. Assim, esses modelos podem ser representados sintaticamente por:

$$x = \varphi(\theta_i) + e,$$

onde x é uma variável dependente observável, θ_i é um vetor de variáveis independentes, φ é uma função conhecida (não necessariamente linear) que relaciona x e θ e e, o erro, um conjunto de efeitos não associados a θ .

Tomando-se o modelo linear

$$X = X_v + e,$$

a medida verdadeira de X_v é definida como a esperança matemática de uma série de valores observados de uma variável. A série pode ser constituída por medidas da variável tomadas em mensurações repetidas em única vez em cada sujeito de uma amostra.

O modelo generalizado tem por base três postulados:

1) a medida verdadeira (X_v) é a esperança matemática de uma série de medidas empíricas: $X_v = E(x)$

2) Não existe correlação entre as medidas verdadeiras e seus respectivos erros de medida: $\rho(X_v; e) = 0$.

Não há razão plausível para se pensar que o tamanho do erro de medida esteja associado sistematicamente ao tamanho da medida verdadeira. O erro se distribui aleatoriamente.

3) Definindo-se instrumentos paralelos como dois ou mais instrumentos que apresentam a mesma variância dos erros de medida e as mesmas medidas verdadeiras (pode-se dizer, em linguagem comum, que dois ou mais instrumentos são paralelos, se medem a mesma propriedade, apesar de variarem na forma), postula-se que os erros de medida em um instrumento (J) não se correlaciona com seus erros de medida em um instrumento paralelo (K), $\rho(e_j; e_k) = 0$.

Se a mensuração é feita corretamente com os instrumentos paralelos, não existe razão para que as medidas covariem sistematicamente umas com as outras.

O modelo permite deduções imediatas de importância teórica e prática. Essas deduções estão formuladas para os valores paramétricos da população.

1) O erro de medida se define como a diferença entre a medida empírica e a medida verdadeira. $E = X - X_v$.

2) A esperança matemática dos erros de medida é zero. $E(x) = 0$.

3) A média das medidas empíricas (μ_x) é igual à média das medidas verdadeiras (μ_{xv}). $\mu_x = \mu_{xv}$

4) As medidas verdadeiras não covariam com os erros de medida. $Cov(X_v; e) = 0$

5) A covariância entre as medidas verdadeiras e as medidas empíricas é igual à variância das medidas verdadeiras. $Cov(X; X_v) = Var(X_v)$

6) A covariância entre as medidas empíricas de dois testes é igual à covariância entre as medidas verdadeiras. $Cov(X_1; X_2) = Cov(X_{v1}; X_{v2})$

7) A variância das medidas empíricas é igual à variância das medidas verdadeiras mais a variância do erro de medida. $Var(X) = Var(X_v) + Var(e)$

8) A correlação entre as medidas empíricas e os erros de medida é igual ao coeficiente entre o desvio padrão dos erros e o desvio padrão das medidas empíricas. $\rho(X; e) = \sigma_e / \sigma_x$

Para K testes paralelos:

9) as médias das medidas empíricas são iguais. $\mu_1 = \mu_2 = \dots = \mu_k$

10) as variâncias das medidas empíricas são iguais. $\sigma(X_1) = \sigma(X_2) = \dots = \sigma(X_k)$

11) as intercorrelações das medidas empíricas são iguais. $\rho(X_1; X_2) = \rho(X_1; X_3) = \dots = \rho(X_1; X_k)$

Santisteban acentua a importância atual do modelo linear:

A introdução deste modelo permitiu que se desenvolvesse uma das teorias mais prolíficas no campo da Psicometria... Ainda que posteriormente tenham sido desenvolvidas novas teorias, a teoria baseada no modelo de Spearman continua sendo influente em nosso tempo... (18:p.26).

A Construção de Instrumento de Medidas: indicações preliminares

As reflexões teóricas, até aqui apresentadas, tiveram duplo objetivo. Primeiro, mostrar que a

teoria da medida se fundamenta em bases mais sólidas do que muitos imaginam. A reação da corrente qualitativa ao *quantitativismo*, embora salutar quando chama a atenção para o exagero que se comete na utilização de resultados numéricos de testes, é prejudicial na medida em que, pelo menos no Brasil, responde pelo descaso com que vem sendo tratada a medida educacional. Praticamente o estudo teórico do problema não é mais feito na área educacional; migrou para institutos especializados como a Fundação Cesgranrio, a Fundação Carlos Chagas e para alguns Departamentos de Estatística das universidades.

O segundo intuito foi expor de maneira compreensível, espera-se, os elementos necessários para que professores se sintam mais à vontade na execução das tarefas de avaliação, presentes no seu dia-a-dia profissional. Com este mesmo desiderato, seguem estas considerações sobre a construção de testes.

De acordo com Thorndike (21:p.9-14), para a elaboração de um teste ou um conjunto de instrumento de medida, seja qual for o setor de conhecimento, é necessário sempre o cumprimento de três fases:

(1) identificação e definição da qualidade ou atributo que se quer medir, (2) determinação de um conjunto de operações através das quais o atributo pode se manifestar e ser percebido e (3) estabelecimento de procedimentos e definições que possibilitem traduzir as observações em afirmações quantitativas de grau ou magnitude.

A primeira fase diz respeito aos objetivos. Procedimentos de medida e sua relação com a definição de objetivos são fundamentais para a elaboração de testes. Em um de seus trabalhos, Anderson concluiu que

(...)os que trabalham em pesquisa educacional ainda não aprenderam como construir testes de desempenho que satisfaçam ao requisito fundamental exigido por qualquer sistema de

medida, isto é, uma definição clara e consistente das coisas que estão sendo medidas (1:p.145).

A observação de Anderson, infelizmente, continua atual, mesmo quase 30 anos depois de feita.

Bloom (4:p.289) e Thorndike (21:p.9-14) examinaram o problema dos objetivos de maneira mais específica. Primeiro estabeleceram a distinção entre objetivos educacionais e finalidades e metas da educação. As metas da educação são sempre de natureza ampla e dão pouca orientação para as atividades cotidianas do ensino e técnicas de avaliação.

Os objetivos educacionais, ao contrário, são específicos e fornecem a professores e avaliadores orientação tanto sobre o conteúdo que deve ser ensinado, como sobre a natureza do que precisa ser medido.

Quando se diz, por exemplo, que a finalidade da educação é promover o desenvolvimento integral do homem, ou que o ensino da Matemática visa ao treinamento do raciocínio lógico-dedutivo, pouca ou quase nenhuma orientação e informação, sobre que experiências e atividades devem ser incluídas no dia-a-dia da sala de aula, são fornecidas. Ainda mais, estas diretivas gerais em nada auxiliam na definição de técnicas e na elaboração de instrumentos de medida. Só os objetivos específicos de cada disciplina podem ajudar o professor a definir o conteúdo a ser ensinado e como tais assuntos devem ser tratados em sala de aula; também são muito úteis na hora de elaborar os testes necessários à avaliação, porque fornecem indicações mais precisas sobre as coisas que devem ser medidas.

Desde a publicação, em 1956, da hoje clássica *Taxionomia dos Objetivos Educacionais*, de Bloom, ganhou relevância a importância de se definir objetivos mais específicos.

Ao longo da segunda metade do século XX, muito se discutiu sobre como definir de forma mais clara, para efeito de medir, os objetivos colimados. Mager (12), Payne (14) e Popham (16) propuseram o uso do que denominaram de *objetivos comportamentais*. Como era de se esperar, e quase sempre

ocorre entre educadores, a discussão foi deslocada para a *forma* do problema e sua *essência* foi esquecida. Perdeu-se de vista o fato de que o problema em pauta não era decidir se os objetivos comportamentais eram preferíveis ou não a objetivos definidos de outra maneira, conforme observou Eisner (6). Tanto que Popham (16) concluía que, apesar de calorosa, a controvérsia se mostrou totalmente improdutiva.

Perdeu-se de vista a idéia de que o ponto crucial é: se os fins da educação não forem traduzidos em termos claros e significativos serão questionados, a construção e uso de testes, seja para fins preditivos ou de diagnóstico.

A definição de objetivos torna possível a determinação do conjunto de operações através das quais o atributo, que se deseja medir, pode ser percebido. Há uma interação das duas ações. A clarificação do atributo age como mecanismo seletor, entre as operações possíveis, daquelas que são mais relevantes e adequadas para revelar o atributo em estudo.

A esta altura, deve-se ser mais preciso na definição de medida. Para Russell a

(...) medida de magnitude é, em sentido geral, qualquer método através do qual se estabelece uma correspondência recíproca e única entre toda a magnitude de alguma natureza ou parte dela e todos ou alguns números inteiros, racionais, ou reais, conforme seja o caso (17:p.176).

O problema para quem elabora um teste é encontrar os procedimentos que lhe possibilitem estabelecer esta *correspondência única e recíproca*. A concepção teórica de medida de Russell envolve a definição de um isomorfismo entre a propriedade que está sendo medida e um sistema numérico apropriado. O essencial é a atribuição de números, de tal modo que uma função seja definida, tendo como *domínio* o conjunto formado por todas as possíveis magnitudes que estão sendo medidas, e *contradomínio*, um conjunto numérico adequado (*inteiro, racional ou real*) (23:p.13-40).

O instrumento de medida (teste, questionário, entrevista, escala ou outro qualquer), dentro desta perspectiva, será a regra de correspondência que tornará possível a realização concreta do isomorfismo pretendido.

É evidente que um isomorfismo de acordo com a definição matemática nunca é possível, portanto, é necessário que se aprenda a enfrentar uma situação de inexatidão. Por isto, como já explicitado, deve-se assumir a verdade segundo a qual o número produzido pelo teste tem dois componentes: a medida *verdadeira* e o *erro* de medida.

O termo erro engloba mudanças no contradomínio da função e ambigüidade da regra de correspondência. Desta forma, a preocupação maior de quem constrói um teste deve ser reduzir a imprecisão do instrumento de tal forma que pelo menos uma percentagem significativa da variância do conjunto de medidas seja explicada pelas diferenças que realmente existem no domínio da função (variância verdadeira), e somente uma percentagem mínima decorra da ambigüidade do instrumento ou mudança no contradomínio da função (erro de medida) (10:p.151-160).

Em termos práticos, o problema se reduz a determinar se o teste oferece informações que podem ser confiáveis e precisas. Tal informação é dada pelo coeficiente de precisão de um teste que possibilita calcular *a percentagem da variância obtida na distribuição dos escores que pode ser considerada variância verdadeira, isto é, variância não acarretada por falhas do instrumento de medida* (8:p.153-160).

Cronbach (5:p.151-193) tomou emprestado dos engenheiros de comunicação os conceitos de sinal e ruído para explicar a precisão, confiabilidade de um teste: o sinal é a informação que uma pessoa teria se houvesse recepção perfeita, e o ruído é qualquer sorte de interferência na transmissão.

Aplicados ao teste, sinal e barulho são relacionados negativamente: quanto maior o ruído, mais confuso e distorcido será o sinal, portanto, menor será a quantidade de informação verdadeiramente

recebida. O problema da precisão fica, então, reduzido a determinar quanto da informação dada pelo teste é sinal e quanto é ruído. Em outras palavras, que percentagem da variância total é variância verdadeira e qual a parte devida a falhas do teste.

A importância na precisão de um teste aumenta drasticamente quando seus resultados são utilizados como critérios de seleção (vestibulares e concursos em geral) ou para fins de avaliação de desempenho (Exame Nacional de Cursos e Exame Nacional do Ensino Médio – ENEM). Nestes casos, qualquer coeficiente de determinação abaixo de 0,8 já deve ser motivo de preocupação para os responsáveis pela escolha, aplicação e uso dos resultados do teste.

Outro aspecto técnico que deve ser levado em consideração na construção de instrumentos de medida é a sua validade. A validade diz respeito à capacidade de um instrumento qualquer, seja ele teste, questionário, escala etc, medir aquilo que pretende medir. Decorre deste entendimento que a validade de um teste está ligada especialmente a uma determinada situação, não sendo um conceito generalizado. *A validade é, portanto, não um critério geral mas um critério específico, no sentido de que um teste pode ser altamente válido quando usado em uma situação e altamente inválido quando usado em outra* (7:p.53-54). Desta forma, um instrumento pode ser válido para um grupo de pessoas, uma instituição, um programa e não o ser, se as pessoas, instituições ou programas forem outros. Este é um ponto a ser considerado cuidadosamente na análise de propostas de avaliação de caráter regional ou nacional. É o que parece estar ocorrendo com as avaliações do Sistema Nacional de Avaliação do Ensino Básico – SAEB.

Existe uma relação entre os dois indicadores técnicos do teste: a precisão é condição necessária para a validade, isto é, não se pode falar em validade de um teste, se este não for preciso. A analogia com ruído e sinal esclarece este fato. Suponha que alguém esteja tentando receber uma mensagem através de um rádio. Para que isto ocorra, é necessário que o sinal se sobreponha ao ruído. Se houver mais ruído do que sinal

nada ou quase nada será captado. Para que alguma coisa seja captada, é preciso que o sinal não seja apagado pelo ruído. Portanto, para que se possa medir o que se deseja medir, isto é, para que um instrumento seja válido, a condição necessária é que ele seja preciso. Entretanto, é bom lembrar que, em determinadas situações, o importante é captar o sinal, mesmo com a presença de algum ruído (3:p.47).

Conclusão

Das considerações teóricas feitas neste trabalho, podem ser destacados os seguintes pontos práticos.

- 1) O valor obtido numa medida não é absoluto. Isto significa que o professor, ao avaliar seu aluno, não deve adotar uma posição rígida em relação à nota resultante dos testes aplicados; deve sempre ter em mente o fato de que, em cada teste ou prova aplicada, existe um erro no resultado obtido, uma vez que, como visto, a medida verdadeira é igual à medida observada (nota ou escore do teste) mais o erro. É ridícula, portanto, a discussão entre professores e coordenação pedagógica sobre a concessão ou não de décimos de ponto a alunos, para que atinjam o mínimo necessário para aprovação.
- 2) A avaliação é um conceito mais amplo do que medida. Houve tempo em que os dois conceitos foram igualados. Atualmente, seja qual for a corrente teórica seguida, a avaliação é distinta dos procedimentos para aferição dos resultados de testes e provas. Avaliação é um julgamento de valor e quanto mais informações sobre o que se quer avaliar estiverem disponíveis, mais acurado será o julgamento feito.
- 3) A base da avaliação é a observação de um comportamento. O conceito de observa-

ção deve e pode ser ampliado para incluir as mais diversas formas de se obter informações. Além da utilização de instrumentos de medida (testes, escalas de atitude, opinião de juízes, provas elaboradas pelo professor etc), o professor pode e deve incluir informações obtidas diretamente por ele no convívio diário com os alunos (interesse pelo conteúdo ministrado, participação em sala de aula, manifestação de atitude crítica etc).

- 4) A aparente sofisticação matemática da teoria da mensuração pode sugerir aos menos avisados que a construção dos instrumentos de medida seja uma tarefa reservada para especialistas do ramo. Nada mais falso. Não há necessidade de um professor dominar e compreender, em profundidade, os fundamentos matemáticos da teoria da mensuração para tal. Basta que siga alguns procedimentos norteadores da construção de instrumentos de medida em educação e que, a cada prova ou teste elaborado, reflita sobre os resultados obtidos e os erros de elaboração cometidos apontados pelos próprios alunos. Os procedimentos mencionados podem ser encontrados em qualquer manual sobre o assunto. Em língua portuguesa, um dos melhores é o de Vianna (24), constante da bibliografia deste artigo.
- 5) É sabido que os professores sofrem solicitações das mais diversas ordens em seu exercício profissional e, por isso mesmo, tendem a relegar a um plano secundário sua função de avaliador. De fato, se é perguntado ao professor o que faz, ele responderá: *eu ensino*; dificilmente incluirá na resposta *avalio*. Daí a necessidade de que seja criado junto às coordenações pedagógicas um setor de apoio técnico para análise dos instrumentos utilizados, dos resultados obtidos e repasse das análises feitas com estes para o professor.

BIBLIOGRAFIA

- 1 ANDERSON, Richard C. How to construct achievement tests to assess comprehension. In: **Review of Educational Research**. 1972, 42(2), 145.
- 2 BAKER, B.O., Hardyck, C.D. e Petrinovich, L.F. Weak Measurement vs. Strong Statistics. Na empirical critique of S.S. Stevens proscriptions on statistics, in **Educational and Psychological Measurement**, 26, 291-309, (1967).
- 3 BARRETO, J. A. E. Avaliação: Mitos e Armadilhas In: **Ensaio – Avaliação e Políticas Públicas em Educação**. Rio de Janeiro, Fundação Cesgranrio, vol I, dezembro 1993.
- 4 BLOOM, Benjamin S. Testing cognitive ability and achievement. In GAGE, N.L. (Ed.) **Handbook of Research on Teaching**. Chicago, Rand McNally, 1963, 289.
- 5 CRONBACH, Lee J. Other Characteristics Defined in Tests In: **Essentials of Psychological Testing**. New York, Harper & Row Publishers, 1970, 151-193.
- 6 EISNER, E. W. **Educational Objectives: Help or Hindrance**. Trabalho apresentado na reunião anual da American Educational Research Association. Chicago, Fevereiro, 1966.
- 7 GEBERICH, J. Raymond et al. **Measurement and evaluation in the modern school**. New York, David Mackay Company, 1964, 53-54.
- 8 HOYT, Cyro J. Test Reliability Estimated by Analysis of Variance. In: **Psychometrika**. 1941, 6, 153-160.
- 9 KEEVES, John P. (ed.). **Educational Research, Methodology and Measurement: An International Handbook**. Cambridge, Pergamon, 1997.
- 10 KUDER, P. Frederic & RICHARDSON, Marion W. The Theory of the estimation of Test Reliability. In: **Psychometrika**. 1937, 2, 151-160.
- 11 LORD, Frederic M. e Novick Melvin R. **Statistical Theory of Mental Test Scores**. Massachusetts, Addison Wesley Publishing Company, 1968.

- 12 MAGER, R. F. *Preparing Instructional Objectives for Programmed Instruction*. São Francisco, Fearon Press, 1962.
- 13 MUÑIZ, José. *Teoría Clásica de los Testes*. Madrid: Ediciones Pirámide, 1991.
- 14 PAYNE, D.A. *The specification and measurement of learning outcomes*. Watham, Mass.: Blaisdel, 1968.
- 15 PENNA FIRME, Thereza. *Avaliação: Tendências e Tendenciosidades*. In: *Ensaio – Avaliação e Políticas Públicas em Educação*. Rio de Janeiro, Fundação Cesgranrio, vol I, dezembro, 1993.
- 16 POPHAM, W. James et al. *Instructional Objectives*. In: *Aera, Monograph Series on Curriculum Evaluation*. Nº 3, Chicago, Rand McNally, 1969.
- 17 RUSSEL, Bertrand. *Principles of Mathematics*. New York: Norton, 1938.
- 18 SANTISTEBAN, Carmen R. *Psicometria*. Madrid: Ediciones Norma, 1990.
- 19 SARAMAGO, José. *A Caverna*. São Paulo, Companhia das Letras, 2000.
- 20 STEVENS, S. S.– *On the Theory of Scales of Measurement*, in MEHRENS, William A e Ebel, Robert L. (ed). *Principles of Educational and Psychological Measurement*. Chicago: Rand McNally & Compang, 1967.
- 21 THORNDIKE, Robert L. & HAGEN, Elizabeth. *Measurement and Evaluation in psychology and education*. New York, John Wiley & Sons, Inc., 1969, 9-14, 35(a).
- 22 TORANZOS, Fausto I. *Estatística*. São Paulo: Ed. Mestre Jou, 1962.
- 23 TORGERSON, Warren S. *Theory and Methods of Scaling*. New York. John Wiley & Sons, 1958, 13-40.
- 24 VIANNA, Heraldo M. *Testes em Educação*. São Paulo: IBRASA, 1987.

Informação sobre os autores

Nicolino Trompieri Filho – professor adjunto da Faculdade de Educação da Universidade Federal do Ceará – UFC. Licenciado em Pedagogia pela UFRJ e mestre em avaliação pela UFC.

José Anchieta Esmeraldo Barreto – professor titular da Faculdade de Educação

da Universidade Federal do Ceará – UFC. Bacharel em Matemática pela Universidade Federal do Pará. PhD em Educação pela Universidade de Tulane – Nova Orleans – EUA.

Endereço para contato:

Rua Francisca Clotilde, 1543 – Rodolfo Teófilo

CEP – 60-431.070 – FortalezaCE.

Fone: (0xx85) 223-5124 e 9990-5124

Endereço Eletrônico: anchietab@yahoo.com.br