

Repensando o uso da escala Likert: tradição ou escolha técnica?

Rethinking the use of Likert scale: tradition or technical choice?

Rafael Lucian

Faculdade Boa Viagem | DeVry, Recife, PE, Brasil

RESUMO

O processo decisório dos executivos é diretamente condicionado à qualidade da informação que recebem que em abordagens empíricas é fruto da capacidade do instrumento de mensuração utilizado para respostas válidas e precisas sobre o público investigado. Nesse sentido, grande parte das informações obtidas por meio de pesquisas de opinião e disponibilizadas aos executivos estão limitadas à capacidade da escala Likert de captar o construto entendido. Com o objetivo de contribuir com os profissionais de Administração, este ensaio teórico se propõe a investigar se existe algum tipo de escala de mensuração de atitude que seja mais indicada para o uso em Administração do que o modelo tradicional de Likert de 5 pontos. Para tanto, foi feita uma profunda revisão da literatura e, ao final, uma nova escala de mensuração mais precisa e balanceada é proposta.

PALAVRAS-CHAVE: Mensuração de atitude, escala Likert, pesquisa de opinião.

ABSTRACT

The decision-making process of the executive is directly conditioned the quality of information they receive, which ultimately empirical approaches, is the result of the measurement instrument 's ability to use valid, accurate answers on the investigation public. In this sense, much of the information obtained by survey and made available to executives are limited Likert scale capabilities to capture the intendido construct. Therefore, in order to enhance a discussion that is of great practical and theoretical contribution to the administration, this theoretical essay propose to investigate whether there is some kind of an attitude scale of measurement that is most appropriate for use in management of the traditional model of Likert five points. To do so is made a literature review in depth and in the end, a new scale of measurement more accurate and balanced is proposed.

KEYWORDS: Attitude of measurement, Likert scale, opinion research.

Submissão: 26 outubro 2015

Aprovação: 21 janeiro 2016

Rafael Lucian

Doutor em Administração pela Universidade Federal de Pernambuco (UFPE). Coordenador do Núcleo de Apoio à Pesquisa na Faculdade Boa Viagem (FBV) | DeVry.

(CEP 51200-060 – Recife, PE, Brasil).

E-mail: rlucian@fbv.edu.br

Endereço: Campus Imbiribeira, Rua Jean Émile Favre, 422 - Imbiribeira - 51200-060, Recife, PE, Brasil.

1 INTRODUÇÃO

A atual sociedade possui diversos contornos de modernidade e de uso intensivo de tecnologia da informação e comunicação, porém há ainda um crescente espaço para preocupação com o indivíduo, sua natureza e seu comportamento (PETTER; STRAUB, 2012).

Este esforço para entender o comportamento humano, com incursões na Psicologia a fim de perceber-se o âmago do indivíduo e capturá-lo, em especial pela compreensão das atitudes, é uma tarefa colossal que tem sua base na Psicometria. Em Administração, autores como Lawrence e Lorsch (1967), Parasuraman et al. (1988), Venkatesh et al. (2003) e Mittelman (2014) fizeram uso de princípios psicométricos a serviço dos estudos organizacionais.

Atitude é uma característica das pessoas que faz referência ao conjunto de crenças sobre algo e sua resposta em relação a isso (THURSTONE, 1928). Mensurar atitude é importante pelo fato de que este conhecimento é útil na compreensão do comportamento das pessoas, no entendimento da forma como tomam decisões e no conhecimento do modo como se organizam em grupos.

As primeiras aproximações datam do século XIX, porém os estudos migraram e se intensificaram em Ciências Sociais entre as décadas de 1920 e 1930, quando surgiu o conceito de mensuração escalar da atitude, tema deste ensaio teórico.

Nos primórdios da academia, o atendimento dessa meta se dava por meio de testes psicológicos individualizados, sendo este um método que demandava muitos recursos e possuía baixa capacidade de atingir grandes grupos (THURSTONE, 1928).

Por outro lado, devido à grande aceitação que a pesquisa de atitude obteve na academia de Psicologia, houve um esforço dos estudiosos para desenvolver métodos que permitissem sua aplicação em grupos, entre estes, destacou-se a proposição de Likert (1932) que afirmou que a mensuração da atitude de grupos era possível por meio do uso de escala multi-item.

Essa nova forma de mensuração escalar, denominada mensuração multi-item, é um instrumento científico de observação e mensuração de fenômenos sociais idealizada com a finalidade de medir as atitudes por meio das opiniões de forma objetiva (LIKERT, 1932). Tal proposição obteve grande reconhecimento por parte de academia e sua aplicação se estende até os dias atuais. As escalas com esse intuito são também denominadas escalas de opinião.

Nessa perspectiva, embora a escala de Likert seja amplamente aceita e estável, existem questões sobre a mesma que até o momento ainda não foram resolvidas, como a forma de analisar os dados e as variáveis ordinais que admitem apenas testes não paramétricos, conforme afirmam Rasmussen et al. (1989), Jöreskog e Sörbom (1996) e Schriesheim e Castro (1996).

A limitação dos dados provenientes da escala de Likert, é de tal forma severa que alguns autores como Ockert (2005), por exemplo, afirma que os dados resultantes da escala em questão permitem apenas conjecturas relacionais de equivalência relativa, quais sejam igual, menor ou maior. Diz-se que, os dados oriundos de uma escala de Likert, permitem apenas ordenação dos indivíduos por meio da favorabilidade de sua atitude com relação a determinado objeto, mas não revelam uma base para se dizer quanto um indivíduo é mais favorável que outro, nem para medir a quantidade de mudança de certa experiência (SELLTIZ et al., 1975; SANCHES; MEIRELES; SORDI, 2011).

Entre as outras possíveis falhas existente na escala Likert está a questão da gradação da escala que, segundo Preston e Coleman (2000), ainda não foi resolvida mesmo após décadas de discussões. Enfatizando a presença desta falha, Vieira e Dalmoro (2008) afirmam que é muito comum em estudos no campo da Administração, pesquisas que apresentem em seu procedimento metodológico a

expressão que utilizou-se uma escala tipo Likert de n pontos, mas é raro o aparecimento de justificativas para a escolha do valor de n .

Sendo assim, diante da clara oportunidade de pesquisa, este ensaio teórico se propõe como objetivo geral a investigar se existe algum tipo de escala de mensuração de atitude que seja mais indicada para o uso em Administração do que o modelo tradicional de Likert de cinco pontos.

Para tanto, são apresentadas a seguir uma introdução à mensuração, uma revisão em profundidade da proposição original de Likert (1932), as principais críticas ao modelo multi-item de cinco pontos e, ao final, uma alternativa que reduz as críticas conhecidas em benefício de um instrumento com maior poder de mensuração.

2 MENSURAÇÃO POR ESCALAS: BREVE EXPLANAÇÃO

Em Arquitetura a função da escala é permitir que a planta baixa de determinado espaço possa ser feita e, mesmo sem os objetos presentes, haja capacidade de se conhecer a localização ideal de cada item.

Já na Medicina a aplicação das mensurações por escala possui um papel fundamental para os diagnósticos, pois suas informações básicas são por vezes mais precisas que a livre descrição dos pacientes (BRACHER, 2008). Um exemplo disso é a Escala Graduada de Dor Crônica (EGDC), que foi desenvolvida para proporcionar um método simples e breve para graduar a severidade da dor crônica ou recorrente (VON KORFF; MOORE, 2001). Neste caso, o construto é a dor e a mensuração se dá pela expressão do paciente em uma escala visual escalonada de zero a dez para cada uma das questões.

As escalas citadas pelo Inmetro (2007) configuram um tipo particular, conhecido como escalas paramétricas, justamente por caracterizar a absolutez dos números. Acontece que esse tipo de escala não existe em Psicologia nem em suas ramificações como a mensuração de atitude, em que até mesmo o teste de Quociente de Inteligência (QI) é fortemente criticado, por não atingir essa característica paramétrica (PATTO, 1997).

Sendo assim, um século após convenção do metro em 1799, Galton (1880) cravou o termo escala não paramétrica para as novas escalas que se dedicavam a mensurar o psicológico e não o mundo físico.

A principal característica das escalas não paramétricas é sua incapacidade de apresentar resultados absolutos, pois cada escala fornece um resultado que deve ser comparado sob condições muito rigorosas para que haja construção de conhecimento e esta é parte do desafio das métricas que se propõem a mensurar o psicológico, conhecidas como psicométricas.

A Psicometria é a mensuração de variáveis psicológicas com objetivo de medir e comparar indivíduos e grupos (MICHELL, 1997). Para atingir esta meta, Pasquali (2003) exalta a importância das escalas oriundas da Psicologia Social, como é o caso da escala de Likert.

Netemeyer, Bearden e Sharma (2003) definem a Psicometria como a mensuração de atributos psicológicos que não são diretamente observáveis, mas que podem ter sua presença evidenciada por um conjunto de comportamentos ou opiniões.

Em Administração, uma das formas mais tradicionais de se utilizar escalas é a Psicometria, pois é de interesse dos gestores observar e compreender o comportamento de pessoas e grupos por meio de sua atitude. Tal formato, entretanto, admite certos rigores para seu uso e análise, os quais só podem ser discutidos após uma breve revisão sobre os tipos de escalas e suas características.

2.1 TIPOS DE ESCALA: VISÃO SEMINAL

Os tipos de escalas foram originalmente propostos por Stevens (1946) e são: nominal, ordinal, intervalar e razão. Essa classificação foi proposta ao se observar que a classificação anterior, de intensiva e extensiva, já não era mais suficiente para explicar a diversidade de escalas que surgiam no campo da Psicologia. A intenção do autor, fundamental a essa seção, era criar um conjunto de regras de mensuração vinculado a cada tipo de escala, o que explicitaria as aplicações de cada uma.

A escala nominal, na qual não há relação numérica entre os graus, representa o uso mais descriterioso dos números, os quais servem apenas como títulos para os pontos da escala e poderiam ser substituídos, sem prejuízo, por letras, nomes ou imagens. A única estatística permitida para esse grupo é a relativa à frequência dos casos. Esse tipo é uma forma bastante simples de escala e naturalmente muitos não lhe atribuem a devida importância. Possui apenas uma única regra que é não designar números iguais para classes diferentes ou números diferentes para classes iguais. Um exemplo de escala nominal é uma lista de números de matrículas de alunos.

A escala ordinal que preserva a ordenação dos números, mas não de seus intervalos, surgiu da operacionalização do escalonamento gradual, tendo como característica preservar a ordem dos pontos de forma invariável. Esse tipo de escala não admite o cálculo da média e do desvio-padrão e seu uso revela apenas a posição relativa do objeto em uma escala gradual. Stevens (1946) justifica que o uso da média e do desvio-padrão são imprudentes, pois os intervalos da escala não são equivalentes em tamanho. É possível, contudo, realizar cálculos de percentis, além da contagem e do conhecimento de posição dos casos na escala. Um exemplo de escala ordinal é a atribuição de posições em um *ranking* de melhores bandas da década.

O uso da escala intervalar, na qual há conhecimento dos intervalos, mas não há definição do zero absoluto, é a forma mais usualmente requerida por pesquisas quantitativas. Quase todas as estatísticas são aplicáveis a esta forma de mensuração, porém o que lhe fornece certa limitação é a impossibilidade de se determinar o ponto zero verdadeiro, que é determinado por conveniência ou convenção. Exemplos desse tipo são as escalas de temperatura Celsius e Fahrenheit, que atribuem ambas um valor zero a algum ponto arbitrário que não são correspondentes.

As escalas de razão, que possuem ponto zero absoluto, são embasadas na propriedade de conversão entre escalas diferentes que medem o mesmo objeto, são as mais comumente encontradas na Física e são as únicas que possuem igualdade, ordenamento gradual e intervalos iguais. O zero absoluto está sempre presente na escala, mesmo quando abstrato. Acredita-se que a escala de razão tenha o real significado para o uso coloquial da palavra escala, pois, no dia a dia, todas as medidas (alimentos, litros, metros, velocidade) são exemplos deste tipo de escala, embora elas ainda possam ser subdivididas entre fundamentais e derivadas. As primeiras correspondem às escalas básicas do universo, como comprimento e peso, enquanto as derivadas são obtidas em função das fundamentais, como a divisão entre duas escalas (por exemplo, km/h), mas nem por isso perdem suas propriedades de escala de razão.

Um valioso alerta é que a escolha do tipo de escala é função da capacidade de mensuração utilizada e isso não implica necessariamente em uma relação de qualidade. A mensuração é tão acurada quanto a capacidade do instrumento em captar os fragmentos empíricos. Assim, qualquer escala, sensorial ou física, é sempre passível de tendenciosidade, baixa precisão e outras restrições.

Embora a proposição de Stevens (1946) seja posterior ao texto seminal de Likert (1932), sua classificação é válida para todas as escalas existentes, inclusive as psicométricas, sendo, portanto, linguagem padrão entre os livros e *softwares* estatísticos. Neste mister, é possível revisar em profundidade as ideias de Likert e suas propriedades.

3 PROPOSIÇÃO ORIGINAL DE LIKERT

A década de 1930 sofria a influência dos autores psicométricos de 1920 como Allport e Hartman (1925) e Thurstone (1928), porém o formato de escalas utilizado em tal época ainda era de grande discussão.

Para Likert (1932), a atitude não poderia ser mensurada a partir de uma única opinião, como pretendia Thurstone (1928), na verdade seria fruto de um conjunto de opiniões. Nessa ótica, o conjunto de atitudes que uma pessoa pode ter, varia de acordo com o conjunto de estímulos que ela pode receber, sendo assim, capturar a atitude de uma pessoa por meio de uma única opinião é impossível, segundo Thurstone (1928), já que não existe uma escala com parâmetros definidos, pois as possibilidades de atitudes são infinitas e marcar apenas um ponto no contínuo é impraticável.

Assim, a mensuração de atitude é, em sua essência, indireta. Ocorre a partir das opiniões, existindo naturalmente uma flutuação das respostas, que pode ser minimizada ao se considerar um conjunto de assertivas ao invés de apenas uma unidade como grau da escala, ou seja, considerar em verdade um *cluster* de opiniões internamente ligadas umas às outras, como um indicador reflexivo.

Em sua pesquisa sobre internacionalismo, racismo, economia, política e religião, Likert (1932) seguiu rigorosamente os passos de Thurstone (1928). As opiniões foram coletadas a partir de entrevistas com especialistas, declarações em jornais, revistas comerciais e alguns livros. As afirmativas foram editadas de forma a se encaixarem com os requisitos básicos de simplicidade, clareza e brevidade. Sem exceções, as questões apresentadas deveriam apresentar um julgamento de valores e não de fatos. Isso foi conseguido com o uso de afirmativas que remetiam a possibilidades futuras eliminando algum julgamento prévio das pessoas e, sobretudo, do ambiente. O teste foi realizado com 2000 respondentes, dos quais apenas 650 foram considerados válidos.

Diferenciando-se, pois, dos autores prévios, Likert (1932) propôs a mensuração por meio de uma escala multi-item. Como requisito de sua investigação, decidiu utilizar quatro tipos de escalas naquele estudo histórico, quais sejam:

- Escala de 3 pontos;
- Escala de múltipla escolha com cinco alternativas;
- Escala de 5 pontos com afirmativas e graus de concordância (a mais famosa delas);
- Escala de 5 pontos, exatamente igual à anterior, mas usando relato de notícias.

Como se tratava de um uso inovador, as formas de análise também eram igualmente inovadoras. Diante deste cenário seminal, Likert (1932) resolveu optar por duas abordagens. Em seu texto é possível notar sua consciência em relação a suas limitações estatísticas e talvez por isso certa insegurança ainda com a melhor abordagem de análise (embora óbvio, é necessário enfatizar que não fazia-se uso de nenhuma espécie de *software* estatístico à época e, para qualquer cálculo matemático, seria necessário dominar os algoritmos e suas propriedades).

As formas de análise sugeridas nesse estudo seminal, para uso em escalas, foram a soma dos pontos e a distância *sigma* (distância entre os pontos e a média). A média das distâncias representava a atitude do respondente naquela questão e como a atitude é definida por um conjunto de afirmativas, e não apenas uma, o que contaria era a média das médias para cada *cluster*. No caso da análise por soma de pontos, era considerado o valor da alternativa marcada pelo respondente em cada afirmativa e, ao final, a atitude seria revelada pelo somatório desses pontos e não pela média.

Algumas observações metodológicas adicionais foram registradas no artigo e deram base para a forte influência no mundo das pesquisas de mensuração de atitude: o resultado final da análise por somatório ou por média é similar e não parece interferir na interpretação final, sendo um critério de

escolha do pesquisador. O número de alternativas (3, 5 ou 7), segundo Likert (1932), não interfere no resultado.

Uma vez conhecida a origem da mais tradicional escala de mensuração de atitude em Administração, inicia-se o debate em profundidade sobre suas limitações. Como tais críticas são resultados de uma ampla investigação bibliográfica, ordenou-se didaticamente o texto, iniciando-se pelas críticas gerais (parte das quais feitas pelo próprio Likert), e afunilando-se até chegar ao ponto necessário para responder a pergunta de pesquisa deste ensaio teórico.

4 ANÁLISE E CRÍTICA ÀS DIVERSAS PROPOSIÇÕES DE ESCALAS DE MENSURAÇÃO

Uma das questões mais frequentes sobre a mensuração por escalas é de que não é possível, por meio de números, mensurar uma personalidade complexa. Isso é verdade, tanto para investigações por escalas como para qualquer outra forma de mensuração existente, já que nenhum método é capaz de extrair com fidedignidade a complexidade da personalidade humana.

De fato, mensurar é alocar um objeto em algum ponto de um contínuo abstrato (THURSTONE, 1928) e a mensuração de qualquer objeto descreve apenas o atributo (construto) mensurado. Esta é uma característica universal das mensurações. Assim, a mensuração de atitude apenas mensura, obviamente, a atitude. Em adição, só podem ser medidas as variáveis que podem ser classificadas em termos de ser maior ou menor que, com amplo perfil linear, ou seja, as variáveis devem ser relativizadas.

A proposição de Thurstone (1928) foi mensurar atitude por meio da conformidade do indivíduo com opiniões sob a limitação de não se ter certeza sobre a veracidade da resposta ao questionamento. Defendia-se este aspecto argumentando que o que está se mensurando é a atitude que a pessoa quer demonstrar e isso deverá estar em conformidade com seus atos, pois é mais provável que haja coincidência que contradição entre as respostas e as atitudes. Ou seja, se a pessoa possui tal atitude em seu íntimo, mas revela à sociedade um comportamento distinto, o interesse da pesquisa social é sobre sua atitude explícita e não sobre o que se passa em sua cabeça.

O que até hoje não foi esclarecido é se realmente as escalas existentes são capazes de mensurar algo pretendido, mesmo que seja um aspecto isolado. A escala mais usada em Administração é reconhecidamente de grande qualidade, sobretudo pela proposição original de se mensurar atitudes por dimensões multi-itens, porém, ainda existem pontos a ser melhorados. A principal falha da escala, reconhecida pelo próprio Likert, foi fazer uso de uma lógica intervalar que possui severas limitações.

De início, a mensuração de atitude por escalas foi construída sobre a estatística não paramétrica. A ideia discutida nesta seção é relativa a alguns dos pontos levantados por Galton (1880) que não foram e não são criteriosamente observados pela imensa maioria dos estudos contemporâneos em Administração.

A rigor, uma escala como a de Allport e Hartman (1925) ou de Likert (1932), não é de ordem métrica. Em qualquer grau, haverá sempre um espaço sem valores entre duas afirmações, seja numa escala de 5 ou 5.000 pontos e isso faz bastante diferença em relação à estatística disponível para análise, pois não há instrumento externo que possa ser utilizado como parâmetro; logo, qualquer comparação entre pessoas ou grupos por elas medidas só seria possível em termos relativos dentro da mesma amostra.

Como forma de amenizar esse problema assume-se, espelhando-se em Galton (1880), que todas as medidas por meio de escalas do tipo Likert seriam relativas dentro de uma escala arbitrária (ou não paramétrica) e, assim, algumas estatísticas serem mais favoráveis que outras ao escopo da mensuração.

Todavia, mesmo assim, não é possível eliminar essa limitação de relatividade das escalas de atitude na forma como são usadas atualmente, porém isto não as invalida. É possível e relevante utilizá-las para entender que, em determinado grupo, o sujeito A é mais propenso a uma dada ação que o sujeito B.

Em detalhe, se um indivíduo concorda (nível 4) com uma afirmação, é possível compará-lo com outro que assinalou o valor 1 e tirar conclusões relativas. No entanto, se a diferença for entre casas vizinhas, por exemplo, 4 e 5 (*concordo e concordo totalmente*), a certeza da diferença é reduzida e, a rigor, de certa forma, subjetiva.

Mas há, de fato, uma imprudência que deve ser destacada, visto que a escala é um instrumento que se utiliza para mensurar aspectos de pessoas e não a unidade abstrata que as compõem. Nos estudos de Thurstone (1931) já era possível observar essa preocupação.

Como exemplo atual desta imprudência teórico-prática, cita-se o entendimento usual consagrado no estudo de Parasuraman et al. (1988), no qual há crença de que a escala SERVQUAL mensura a qualidade do serviço, quando ela, de fato, está limitada a medir a atitude das pessoas em relação a isso. Na prática, as ações gerenciais deveriam objetivar práticas que melhorassem a percepção subjetiva de um serviço de qualidade ou em vez de realmente crer que haja parâmetro objetivos de qualidade nos serviços, sob pena de subjugar seu público.

Exibindo o paradoxo, o que se constata é que, em questões relativas, jamais a resposta de um grupo poderá ser comparada com a de outro se as condições ambientais e temporais (no caso, a qualidade do serviço) não se mantiverem estáticas, pois se a única referência de comparação, que é a unidade abstrata da escala, for alterada, nada mais é comparável entre grupos distintos com uso de escalas não paramétricas.

Pela teoria original de Allport e Hartman (1925) e Thurstone (1931), a escala SERVQUAL deveria ser um contínuo graduado com objetivo de posicionar as pessoas de acordo com sua atitude em relação ao tema qualidade do serviço e, como tal, mediria exclusivamente a atitude das pessoas e não a qualidade do serviço, como já mencionado.

A contribuição de Likert (1932), a seu tempo, atribuiu às escalas anteriores maior complexidade, permitindo que a atitude fosse mensurada por meio de um agrupamento de afirmações e arranjos estatísticos, porém, não alterou o princípio da escala. Suplementarmente, quando se incrementou o processo para a análise por média (ou somatório) da concordância/discordância com diversas afirmações, fez-se surgir um novo problema.

Talvez a principal contribuição de Likert (1932) tenha sido a unificação de dois vetores, antes mensurados de forma independentes, o sentido e a intensidade. Os autores pré-Likert identificavam, por meio de um instrumento, se a atitude do respondente era positiva ou negativa e, por outro instrumento, tentavam graduar a intensidade de tal construto. Ao propor seu modelo de 5 pontos, ambas as informações podem ser conhecidas com apenas uma aplicação.

Na época em que Likert propôs a sua escala, já se passara meio século da crítica que havia condenado o uso de média para análise de escalas não paramétricas. Contudo, seus sucessores não usaram a mediana, como sugerido, pois em uma escala de apenas 5 pontos haveria muitos respondentes com o mesmo resultado final e isso reduziria, em muito, a precisão das conclusões.

Recorda-se que a proposição de trocar a média pela mediana como medida central das escalas não paramétricas é imperativa e, diante do fato que em uma escala incompleta, em que só alguns valores são reais, qualquer ponto entre os valores (intervalos) não existe. Por exemplo, se 1 corresponde a discordo totalmente e 2 a discordo, 1,4 não possui significado algum. Há quem julgue que 1,4 é mais

próximo de 1 do que de 2, mas é conveniente lembrar que, em uma escala não métrica, os intervalos não são definidos e não precisam obedecer nenhum critério matemático conhecido, aqui incluso o conjunto dos números reais. Sendo assim, mesmo que 1,3 pareça mais próximo a 1 que a 2, isto é uma inferência sem base matemática, pois não se sabe se a distância entre 3 e 4 é a mesma que entre 4 e 5. Este era o risco alertado sobre o uso de escalas não paramétricas desde o século XIX.

Likert (1932), contudo, fazendo uso de uma estatística mais complexa, incorporou às análises estatísticas, técnicas descritivas como média, desvio-padrão e distância *sigma* dos pontos em relação à média, sugerindo inclusive a possibilidade de se utilizar o resultado das escalas como insumo para correlações.

O ponto central em questão é que, matematicamente, Likert (1932) iguala as respostas dos respondentes de acordo com um dos 5 pontos da escala. Ademais, é impossível afirmar que a concordância de uma pessoa em relação a alguma afirmação é a mesma de outro respondente. Mas essa interpretação inconsistente só se torna realmente perigosa quando é potencializada e o erro incorporado a cada resposta é somado e transformado em média com os erros médios. A rigor, sem que os intervalos sejam conhecidos e sabendo que os pontos são em algum grau imprecisos e incomparáveis intersubjetivamente, o cálculo da média é indevido, pelo fato de que não há como comprovar que os pesos estejam certos.

Sendo assim, pode-se entender que os cálculos estatísticos sugeridos originalmente na aplicação da escala Likert, hoje, após o avanço no estudo de escalas, são considerados impróprios, pois tanto para *sigma* quanto para a correlação, o problema da indefinição do intervalo existe e impede que os resultados sejam confiáveis, como já houvera sido alertado pelo próprio idealizador da escala.

De fato, em seu artigo seminal, Likert (1932) sugeriu que, ao invés de se calcular a média, os resultados fossem tratados por meio do somatório dos pontos. Isso, segundo o autor, seria mais coerente, embora ainda consignassem restrições para uso advindas de não conhecer o real valor dos intervalos. Aquele ícone do estudo de mensuração assumiu a limitação dos intervalos e afirmou que não desistiu de sua proposição, pois acreditava que, em algum ponto futuro, essa questão seria solucionada. Todavia, quase um século depois, não há ainda certeza sobre os valores dos intervalos, embora alguns autores tenham enveredado por esta busca.

Mager e Kluge (1987) propuseram que as duas dimensões de sentido e intensidade, unificadas por Likert (1932) fossem novamente separadas. Nessa proposta, o respondente primeiramente marcaria se sua atitude é positiva ou negativa em relação a algo e, em seguida, assinalaria em escala separada qual a intensidade de sua atitude, tal proposição se denomina escala de dois estágios.

Albaum (1997) afirmou que existem vantagens em se utilizar escalas do tipo Likert de dois estágios em detrimento daquelas escalas de um estágio. O autor afirma ainda que, ao se adotar escalas do tipo Likert de dois estágios, estaria se evitando que os respondentes confundissem as dimensões da atitude, obrigando, desta forma, que se posicionassem individualmente sobre cada uma delas. Embora tradicionalmente os metodólogos, como Albaum (1997), não adentrem à questão da análise de dados, suspeita-se que o resultado, na prática, não seja tão divergente assim, entre as escalas de um ou dois estágios, pois a principal limitação do estudo é a indeterminação do intervalo.

Autores como Kunin (1998) criticam também o uso de palavras nos rótulos dos pontos da escala. A interpretação do texto é uma barreira adicional à comunicação. Foi sugerido, então, que fossem eliminados os rótulos da escala e substituídos por símbolos, formando uma escala gráfica. As figuras escolhidas por Kunin (1998) foram rostos desenhados, nas quais a curvatura da boca representaria um nível maior ou menor de felicidade.

Embora conceitualmente haja uma contribuição, a escala de Kunin (1998) não altera o tipo de variável originado por ela, que é o mesmo da escala do tipo Likert, a qual, apesar de ser analisada no segmento da Administração como se fosse intervalar ou até mesmo de razão em casos mais extremos como, por exemplo, na modelagem de equações estruturais, não tem essa característica, pois não tem intervalos conhecidos. Assim, reforça-se a advertência do próprio criador da escala, qual seja, a questão intervalar como sua principal limitação.

Aliás, se o rigor de Stevens (1946) for aplicado, a escala de Likert (1932) e suas variações não podem, nem ao menos ser classificadas como ordinais, já que não é possível afirmar que dois pontos da escala não representam a mesma coisa, nem que duas coisas diferentes não sejam representadas pelo mesmo ponto. Essa limitação acontece pelo fato de que não é possível ter certeza de como os respondentes realmente interpretam as categorias da escala. Embora o bom senso presuma que não seja provável que o *concordo totalmente* de um indivíduo seja, na realidade, igual ao simples *concordo* de outro, isso não justifica, em hipótese alguma, que a escala em apreço seja classificada como intervalar.

Desta forma, a classificação ideal para a escala em estudo e suas variações é nominal, pois nela os números representam nomes (por exemplo, 3 representa *não concordo nem discordo*). Se assim for tratada e analisada, nenhuma crítica relativa à falta de rigor é aplicável (até mesmo porque a escala nominal é a de menor liberdade de análise). Caso seja tratada como ordinal, o pesquisador deverá apontar quais garantias sustentam essa classificação, mesmo em desacordo com Stevens (1946). Importante, porém, é ratificar que não há sustentação teórica no processo de construção desta escala que a justifique como de razão ou intervalar.

A característica pentatômica e, conseqüentemente, os intervalos da escala Likert (1932) são frutos da tentativa de se mensurar a intensidade da atitude. A ideia de que duas pessoas podem ter a mesma atitude, porém em intensidades diferentes, é aceita, embora a sua mensuração não pareça ser suficientemente precisa. De fato, não há como ter certeza de que as pessoas que marcam o mesmo ponto tenham a mesma intensidade de atitude: em verdade é muito mais plausível que não tenham e essa distância em casos extremos pode até mesmo inverter a ordem das categorias, sem que isso seja captado pela escala.

O número de categorias é arbitrário e, como mostrado no próprio estudo seminal, não há diferença significativa entre escalas de 3 ou 5 pontos. A cargo disso, o fato de alterar a característica da escala que poderia ser binária, mas é polinária, implica em uma redução das possibilidades de análise de dados. Estas, entretanto, não são as únicas críticas à escala proposta por Likert (1932), existem considerações mais severas.

4.1 CRÍTICAS À PROPOSIÇÃO ORIGINAL DE LIKERT

Embora a escala em si de Likert (1932) seja utilizada em diversas áreas, ela foi elaborada originalmente para o construto *atitude*. A escala foi teorizada considerando que a atitude não poderia ser captada por um único item (propondo então a escala multi-itens), teve desenvolvida uma forma de se mensurar simultaneamente o sentido e a intensidade desta atitude, e isso implicava na necessidade de uso da estatística para converter os itens em um construto único. É sobre esses dois pontos que a análise de dados deve ser planejada.

O método mais comum de conversão dos itens em um único construto é o cálculo da média, mas essa prática não é indicada para o tipo de escala em si de Likert, tanto pelas restrições aos cálculos não paramétricos de Galton (1880), quando pela classificação de variável nominal (categórica) de Stevens (1946), como já discutido.

A este ponto, já é possível observar a primeira divisão de tarefas existente nos estudos sobre mensuração. Enquanto a definição do construto e a elaboração dos itens é tarefa individualizada,

executada pelo pesquisador interessado em propor uma nova escala, as definições estruturais da escala são coletivas, com notáveis bases nos autores seminais. Desta forma, ao se adotar uma escala, as possibilidades estatísticas não necessitam de nova discussão, pois as definições já são conhecidas e as críticas já deveriam ter sido consideradas.

A discussão sobre análise de dados não faz parte do processo de construção de escalas, no entanto, o metodólogo tem a responsabilidade de, ao menos, esclarecer quais critérios a análise deve observar. Não se pode, em hipótese alguma, descartar a importância e a necessidade do conhecimento estatístico de análise dos dados, porém o seu papel na construção do instrumento deve ser definido.

Sendo assim, define-se a primeira limitação técnica da escala de Likert: a categorização da variável e o fato de os pontos da escala representarem categorias impede o uso de estatística multivariada (STEVENS, 1946; RASMUSSEN et al, 1989; JÖRESKOG; SÖRBOM, 1996; SCHRIESHEIM; CASTRO, 1996). Acrescente-se a isso mais três críticas que são alvo de discussão da Academia: o ponto neutro, os rótulos da escala e os elementos da gradação.

4.1.1 QUESTÃO DO PONTO NEUTRO

Com a popularização da escala de Likert, os debates sobre seus aspectos se intensificaram e um dos pontos mais explorados foi a importância do ponto neutro. Kamorita (1963), um dos principais teóricos desse aspecto, sugere que não é possível definir claramente um ponto neutro na escala Likert.

Essa inabilidade de se definir o ponto neutro em uma escala graduada, levou Guy e Norvell (1977) a realizarem um estudo empírico na tentativa de obter conclusões sobre a importância do ponto neutro em uma escala como de Likert. Os autores averiguaram que a presença do ponto neutro induz os respondentes a fugirem dos extremos e utilizarem mais os pontos médios em suas respostas, indicando que a escala sem ponto neutro pode ser confiável e precisa. Em adição, Garland (1991) interpretou que o ponto neutro serve como forma de o respondente anular a questão.

Anteriormente, em relação a pragmática do ponto neutro, Peabody (1962) afirmou que a presença ou ausência de uma categoria neutra é indiferente para a validação da escala, conclusão essa também compartilhada por Sjöberg e Nett (1968).

Assim, pode-se, então, levantar o questionamento sobre o papel do ponto neutro como instrumento de anulação da questão ou realmente uma atitude neutra. A atitude, entretanto, não possui faceta neutra, afinal desde o momento em que o sujeito toma conhecimento da existência do objeto em análise ele irá demonstrar alguma atitude positiva ou negativa, o que pode variar é o seu grau. Por exemplo, dizer que é indiferente à administração da empresa pode significar atitude positiva em relação aos gestores, pois isto é interpretável como um voto de confiança mesmo que sem apoio explícito.

Sendo assim, a verdadeira função do ponto neutro na escala Likert é anular a questão e não indicar uma suposta atitude completamente neutra. Entretanto, o rótulo *não discordo e nem concordo* não parece ser a melhor forma de dizer que tal ponto serve para anular a questão em caso de inaptidão para a resposta. Exemplificando: para uma pessoa que não tenha conhecimento sobre o plano de cargos e salários da empresa sua resposta para o item *Estou satisfeito com o plano de cargos e salários da empresa* deveria ser *não posso responder* ou *não estou apto a responder* ao invés de *não discordo e nem concordo*. No exemplo citado há uma clara diferença entre uma possível atitude neutra e o simples desconhecimento do objetivo mensurado.

4.1.2 USO DOS RÓTULOS

O formato dos rótulos das escalas é um tema recorrente em livros de metodologia de pesquisa, como Malhotra (2006) e Aaker et al. (2007), e é de fato um tema que tem recebido pouca atenção nos estudos de mensuração, portanto requerendo debate sobre o tema.

Desde o início têm sido utilizados palavras e números como rótulos, embora boa parte dos críticos questionem essa escolha e não tenha havido sucesso nas alternativas apresentadas por Boyd, Westfall e Stasch (1977) e Alwin (1997): o uso de rótulos gráficos. Reconhece-se, desde Reichmann (1964), que o estímulo visual é mais efetivo para capturar respostas corretas do que questionamentos diretos, pois as pessoas sentem-se mais confortáveis lendo figuras que interpretando sentenças ou marcando números. Há, também, o entendimento de que os rótulos gráficos são mais bem trabalhados para públicos-alvo específicos, como é o caso do uso de figuras para pesquisas com crianças (ALWIN, 1997).

Contudo, essa alternativa gráfica não é aceita como ideal por autores como Chrystal (2008), que afirmou que a linguagem visual promove expressão limitada de valor e a cultura específica do respondente pode afetar sua interpretação sobre o instrumento de coleta de dados. Diante deste impasse teórico, Derham (2011) realizou uma série de testes empíricos, observando o comportamento de três tipos de escalas do tipo Likert. A primeira utilizou apenas palavras nos rótulos, a segunda utilizou uma série numérica para indicar os graus da escala e a terceira foi completamente gráfica.

Os resultados do estudo podem ser vistos na Tabela 1, que versa de acordo com o estudo citado, o formato mais confortável ao respondente (e, portanto, mais indicado para o uso) seria palavras como rótulos. Este formato de escala apresentou melhor desempenho em seis dos sete atributos testados. À parte do formato dos rótulos, o número de graus na escala tem despertado grande interesse da academia e até hoje não há consenso sobre seu efeito na mensuração de atitude.

TABELA 1

Comparação entre os tipos de rótulos de escalas.

| ATRIBUTO / RÓTULO | PALAVRAS | NÚMEROS | GRÁFICOS |
|---|----------|---------|----------|
| Preferida pelos respondentes | 43% | 23% | 30% |
| Facilidade para responder | 40% | 21% | 29% |
| Rapidez para responder | 37% | 23% | 31% |
| Formato mais atraente | 35% | 18% | 39% |
| Melhor formato para expressar os sentimentos | 40% | 20% | 29% |
| Formato que traz as respostas mais verdadeiras | 45% | 23% | 23% |
| Formato que mais faz o respondente refletir antes da resposta | 27% | 25% | 36% |

Fonte: Derham (2011, p. 21).

O uso de palavras como rótulos destoa da proposição original de Likert em que foi utilizado números, porém entre as várias aplicações contemporâneas já é possível observar que, parte dos pesquisadores, substituiu os números por nomes.

4.1.3 ELEMENTOS DA GRADAÇÃO

As escalas do tipo Likert carregam dois componentes: direção e intensidade (que só pode ser aferida caso haja, no mínimo, dois pontos para cada sentido). Contudo, como já antevia Cronbach (1951), restariam dúvidas sobre a efetividade da mensuração sobre intensidade.

Por outro lado, Bendig (1954) e Komorita (1963) concluíram por investigações empíricas que a confiabilidade da escala independe do número de alternativas de resposta da escala. Na última obra citada, constatou-se até que escalas dicotômicas e politômicas tendem a ter o mesmo grau de confiabilidade quando comparadas, embora nada se possa afirmar sobre sua eficiência.

Analisando constatações como estas, Peabody (1962) evidenciou empiricamente que o comportamento de escalas de 2 a 6 pontos, para capturar informação de intensidade em escalas tipo Likert, pouco contribuíram para o resultado final da pesquisa.

Adicionalmente, a questão da eficiência da mensuração também foi observada por Jacoby e Matell (1971) ao testar escalas de 2 a 19 pontos em relação às suas confiabilidade e validade. Esses autores concluíram que, não há diferenças significativas entre elas e, mais taxativamente, afirmaram que o principal componente da escala de Likert (1932) é o direcional, restando à intensidade um papel secundário. O componente direcional foi o que mais representou importância para a determinação do poder total de mensuração da escala graduada nos testes empíricos efetuados.

Rodriguez (2005), por meio de meta-análise, concluiu que uma escala com três opções de resposta é suficiente. Este autor destaca que o efeito da diminuição do número de opções de escolha encolhe o teste, proporcionalmente, aumenta sua eficiência para grandes quantidades de respondentes e diminui a eficiência para pequenas quantidades de respondentes. Em complemento, denota que o tempo gasto na resposta do questionário é proporcional ao número total de alternativas e o uso de três itens na escala diminui o tempo na coleta de informação.

Wiswanathan, Sudman e Johnson (2004) também demonstram preocupação com a relação entre a escala e os testes estatísticos, visto que a definição do número de itens na escala vai afetar os testes estatísticos a serem realizados. De fato, o uso de uma escala com muitos pontos pode não prover uma base de dados válida para a realização de inferências estatísticas, visto que, de acordo com o tamanho da amostra, pode resultar em uma dispersão entre os respondentes, limitando o uso de alguns testes estatísticos.

Uma outra preocupação referente ao número de itens é relativa ao não balanceamento das escalas politômicas (NUNNALLY, 1978). Ao se propor uma escala de 5 pontos, os intervalos negativos tendem a ser maiores que os positivos e este comportamento de desbalanceamento independe do construto que se esteja mensurando (TOMAS; OLIVER, 1999).

Esse efeito de intensificação das respostas e distanciamento dos intervalos é potencializado se a afirmativa referente ao item mensurado também for escrita na forma negativa (MOTL; DISTEFANO, 2002) e isto é uma das explicações para Barnette (2000) ter identificado que as frases negativas possuem naturalmente pior desempenho nos testes de confiabilidade.

Em outras palavras, o componente negativo tende a ser supervalorizado e impede o correto balanceamento das escalas, seja pelo aumento do intervalo entre os itens negativos, seja pelo uso de afirmativas com palavras negativas.

Este comportamento assimétrico entre positivo e negativo é explicado por Rozin e Royzman (2001), pelo fato de as avaliações negativas serem mais fortes, intensas e rápidas que as positivas. Assim, o conjunto de várias percepções positivas contra apenas uma negativa pode resultar em atitude negativa, não respeitando a lógica aritmética. A sugestão de dicotomização da escala, lembrando Peabody (1962), é então justificada pela assimetria entre positividade e negatividade (ANDERSON, 1965).

Desta forma, para Cacioppo e Berntson (1994), as atitudes positiva e negativa são entidades independentes que podem coexistir, portanto, não se encaixam no conceito de escala.

Cacioppo, Gardner e Berntson (1999), então, assumem que as mensurações psicométricas são bivariadas com a existência bipolar do negativo e do positivo sem gradações intermediárias.

Baseado nesses mesmos critérios de assimetria entre positivo e negativo, e por meio de testes empíricos que mensuraram a intensidade de cada lado da escala, Alexandrov (2010) sugeriu que a escala do tipo Likert pentatônica não deve ser utilizada para mensuração de atitude.

5 CONSIDERAÇÕES FINAIS E PROPOSIÇÃO DE UMA ESCALA ALTERNATIVA

A fim de dar luz aos argumentos aqui apresentados como críticas a proposição original de Likert (1932), o Quadro 1 reúne resumidamente a contribuição dos autores já citados sobre o tema.

QUADRO 1

Resumo da argumentação crítica sobre a escala Likert.

| TÓPICO | ARGUMENTAÇÃO CRÍTICA | AUTORES |
|-----------------------|--|---|
| Elementos da gradação | A mensuração do sentido e da intensidade da atitude de forma separada confere maior precisão aos resultados. | Mager e Kluge (1987); Albaum (1997); Sanches, Meireles e Sordi (2011); Robertson (2012) |
| | Há um desbalanceamento natural nas escalas graduadas, e o lado negativo é o que possui intervalo com maior peso. | Rozin e Royzman (2001); Motl e DiStefano (2002); Alexandrov (2010) |
| | Não há definição de qual o número ideal de pontos em uma escala graduada de mensuração de atitude. | Preston e Coleman (2000); Vieira e Dalmoro (2008) |
| Ponto neutro | A presença ou ausência de ponto neutro é indiferente à validade da escala. | Peabody (1962); Sjoberg e Nett (1968) |
| | O ponto neutro deve ser incluído para que o respondente possa anular uma questão quando sentir necessidade | Garland (1991) |
| Rótulos | Escalas gráficas são mais facilmente entendidas pelos respondentes. | Kunin (1998) |
| | O uso de palavras nos rótulos é o formato que traz as respostas mais verdadeiras. | Derham (2011) |

Fonte: Compilação dos argumentos das obras citadas no quadro.

5.1 PROPOSIÇÃO DE UMA ESCALA

Respondendo a questão de pesquisa deste ensaio teórico, observou-se que a literatura indica ser possível propor uma escala de mensuração mais eficiente que a original pentatônica de Likert. Sendo assim optou-se por criar uma nova escala na tentativa de eliminar as limitações observadas na escala referida.

Existem três diferenças entre a escala proposta e a escala original de Likert (1932) que são o fato de não se utilizar números como rótulos para os pontos, o deslocamento do ponto neutro para fora da escala e a dicotomização da variável atitude.

Com relação a não utilizar números como rótulos seguiu-se a recomendação de Derham (2011), que sugere que se usem rótulos com palavras em vez de gráficos ou números. Este formato é mais confortável ao respondente e o torna mais propenso a entender o que está sendo questionado.

A segunda mudança efetiva foi o deslocamento do ponto neutro para fora da escala, a fim de não oferecer a ideia de *continuum*. Embora incluso e popularizado na escala de Likert, o ponto neutro não é item obrigatório em escalas graduadas de intensidade (PEABODY, 1962; SJOBERG; NETT, 1968; GUY; NORVELL, 1977), porém seu uso é recomendável, pois é a ele que o respondente recorre em caso de não querer ou não ter capacidade para responder a questão, sendo, portanto, uma forma legítima de anulação da alternativa.

Quando o ponto neutro se localiza no centro da escala, há uma tendência já reconhecida pela Academia de que as respostas sejam nele de fato. Como observaram Guy e Norvell (1977) e Garland, (1991) o ponto central neutro atrai as respostas pela sua localização central.

Assim, devido à conotação forte que a palavra anular pode representar aos respondentes, na escala aqui proposta optou-se por manter o rótulo de indiferente para o ponto neutro, com a expectativa de que fosse utilizado pelo respondente, no caso deste não possuir uma atitude qualquer em relação à afirmativa correspondente.

Esta escolha deveu-se à constatação de que, em verdade, as categorias de atitude positiva e negativa são independentes e o deslocamento do ponto neutro ressalta isso para o respondente, tal que seu uso deve ser para casos de anulação da resposta, quando não há atitude em relação à afirmação. Caso haja alguma atitude, independente de grau, o respondente é forçado a se posicionar.

Desta forma, na elaboração da escala aqui proposta, foi acatada a sugestão de Peabody (1962), Anderson (1965), Cacioppo e Berntson (1994), Cacioppo, Gardner e Berntson (1999) e Alexandrov (2010), de se utilizar das atitudes positivas e negativas como posições assimétricas que não devem ser mensuradas por grandezas escalares; sendo assim optou-se dicotomizar a variável.

As escalas dicotômicas são aceitas pela literatura e podem ser utilizadas sem prejuízo à confiabilidade ou efetividade da mensuração (BENDIG, 1954; KOMORITA, 1961; JACOBY; MATELL, 1971). Esse formato, entretanto, não permite que a intensidade seja mensurada, mas não se acredita que isso seja uma desvantagem, pois, como já confrontado antes, para parte dos teóricos, a mensuração da intensidade proposta por Likert (1932) é impraticável devido ao desbalanceamento entre a atitude positiva e negativa (PEABODY, 1962; ALEXANDROV, 2010). Enfatiza-se que o componente direcional que representa a essência da mensuração de atitude (JACOBY; MATELL, 1971) é preservado e, neste caso, potencializado.

5.2 IMPLICAÇÕES DA ESCALA PROPOSTA

A escala aqui proposta possui melhor balanceamento e, portanto, é mais precisa que a escala original de Likert (1932). Acredita-se que os futuros pesquisadores que adotarem esta escala para mensuração irão observar maior fidelidade nos resultados dos testes não paramétricos, inclusive em cálculos intersubjetivos e entre grupos.

O novo formato, dicotômico, permite ainda que cálculos avançados como a Teoria de Resposta ao Item (TRI) seja utilizada em Administração sem necessidade de adaptações e isso representa grande vantagem, uma vez que a estrutura matemática da TRI é muito complexa. Ainda sobre possibilidades de análise, o uso de escalas binárias se adequa perfeitamente ao uso de correlações e testes de hipóteses não paramétricas e até mesmo cálculos métricos como Regressão Logística, dependendo das demais variáveis envolvidas.

Menos notório, mas igualmente importante é o fato da nova escala permitir que o respondente conclua o questionário mais rapidamente, pois, uma escala mais simples e visualmente mais enxuta cansará menos o indivíduo que tenderá a ter mais concentração até o final dos itens.

Espera-se que tal contribuição seja adotada e posta à prova pela Academia de Administração para que, paulatinamente, com o esforço conjunto, as mensurações sejam cada vez mais precisas e as informações fruto destas, mais valiosas.

6 REFERÊNCIAS

AAKER, D. A.; KUMAR, V.; DAY, G. S.; LAWLEY, M.; STEWART, D. *Marketing research*. Sidney: John Wiley, 2007.

ALBAUM, G. The Likert Scale Revised: An alternative version. *Journal of the Market Research Society*, v. 39, n. 2, p. 1-12, 1997.

ALEXANDROV, A. Characteristics of Single-Item Measures in Likert Scale Format. *The Electronic Journal of Business Research Methods*, v. 8 n. 1, p. 1-12, 2010.

ALLPORT, F. H.; HARTMAN, D. A. The Measurement and Motivation of Atypical Opinion in a Certain Group. *The American Political Review*, v. 19, n. 4, p.735-760, 1925.

ALWIN, Duane F. Feeling Thermometers Versus 7-Point Scales Which are Better? *Sociological Methods & Research*, v. 25, n. 3, p. 318-340, 1997.

ANDERSON, N. H. Averaging versus Adding as a Stimulus-Combination Rule in Impression Formation. *Journal of Experimental Psychology*, v. 70, n. 4, p. 394-400, 1965.

BARNETTE, J. J. Effects of Stem and Likert Response Option Reversals on Survey Internal Consistency: If You feel the Need, There is a Better Alternative to Using Those Negatively Worded Stems. *Educational and Psychological Measurement*, v. 60, n. 3, p. 361-370, 2000.

BENDIG, A. W. A Factor Analysis of Student Ratings of Psychology Instructors on the Purdue Scale. *Journal of Educational Psychology*, v. 45, n. 7, p. 385, 1954.

BOYD, H. W; WESTFALL, R; STASCH, S. F. *Marketing research*. Text and cases. 4. ed. Illinois: Richard D Irwin Inc., 1977.

BRACHER, E. S. B. *Adaptação e validação da versão em português da escala graduada de dor crônica para o contexto cultural brasileiro*. São Paulo, 2008. Tese (doutorado). Faculdade de Medicina da Universidade de São Paulo.

CACIOPPO, J. T.; BERNTSON, G. G. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, v. 115, n. 3, p. 401-423, 1994.

CACIOPPO, J. T; GARDNER, W. L; BERNTSON, G. G. The affect system has parallel and integrative processing components: Form Follows Function. *Journal of Personality and Social Psychology*. v. 76, n. 5, p. 839-855, 1999.

CRONBACH, Lee J. Coefficient alpha and the internal structure of tests. *Psychometrika*, v. 16, n. 3, p. 297-334, 1951.

DERHAM, P. A. J. Using preferred understood or effective scales. How scale presentations effect online survey data collection. *Australasian Journal of Marketing & Social Research*, v. 19, n. 2, 2011.

GALTON, F. Statistics of Mental Imagery. *Mind*, v. 5, n. 19, p. 301-318, 1880.

GARLAND, R. The mid-point on a rating scale: is it desirable? *Marketing Bulletin*, v. 2, p. 66-70, 1991.

GUY, R. F; NORVELL, M. The neutral point on a Likert scale. *The Journal of Psychology*, v. 95, p.199-204, 1977.

INMETRO. *Sistema internacional de unidades – SI*. 8. ed. Rio de Janeiro: 2007.

JACOBY, J.; MATELL, M. Three-point Likert scales are good enough. *Journal of Marketing Research*, v. 8, p. 8, 1971.

JÖRESKOG, K. G.; SÖRBOM, D. *Preliis2: user's reference guide*. *Scientific Software International*. Chicago: 1996.

KOMORITA, S. S. Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, v. 61, p.327-334, 1963.

KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology*, v. 51, n. 4, p. 823-824, 1998.

- LAWRENCE, P. R.; LORSCH, J. W. *Organization and Environment: Managing differentiation and integration*. Boston: Harvard University, 1967.
- LIKERT, R. A Technique for the Measurement of Attitudes. *Archives of Psychology*, v. 140, p. 1-55, 1932.
- MAGER, J.; KLUGE, A. One-stage versus two-stage attitude measurement: separating the dimensions of direction and intensity. In: *Academy of Marketing Science World Marketing Congress*. Barcelona: 1987.
- MALHOTRA, N. K. *Pesquisa de marketing: uma orientação aplicada*. 4. ed. Porto Alegre: Bookman, 2006.
- MICHELL, J. *Quantitative science and the definition of measurement in psychology*. New York: Brit J Psychol, 1997.
- MITTELMAN, M.; ANDRADE E. B.; CHATTOPADHYAY, A.; BRENDL, C. M. The Offer Framing Effect: Choosing Single versus Bundled Offerings Affects Variety Seeking. *Journal of Consumer Research*, v. 41, n. 4, p. 953-964, Dec. 2014. ISSN: 00935301.
- MOTL, R. W; DISTEFANO, C. Longitudinal Invariance of Self-Esteem and Method Effects Associated with Negatively Worded Items. *Structural Equation Modeling: A multidisciplinary Journal*, v. 9, n. 4, p. 562-578, 2002.
- NETEMEYER, R. G; BEARDEN, W. O; SHARMA, S. *Scaling Procedures*. Thousand Oaks: Sage Publications, 2003.
- NUNNALLY, J. C. *Psychometric theory*. 2. ed. New York: McGraw-Hill, 1978.
- OCKERT, D. Substantive Scale Verification: A Likert Scale Analysis and Critique of university Student Pedagogical Activity Preferences. *JALT Hokkaido Journal*, v. 9, p. 48-64, 2005.
- PARASURAMAN, A; ZEITHAML, V.; BERRY, L. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, v. 64, n. 1, p.12-40, 1988.
- PASQUALI, L. A medida psicométrica. In: Pasquali L. *Psicometria: teoria dos testes na psicologia e na educação*. 2 ed. Petrópolis: Editora Vozes, 2003.
- PATTO, M. H. S. Para uma Crítica da Razão Psicométrica. *Psicologia USP*, v. 8, n. 1, p. 47-62, 1997.
- PEABODY, D. Two components in bipolar scales: Direction and extremeness. *Psychologia Revista*, v. 69, p. 65-73, 1962.
- PETTER, S.; RAI, A.; STRAUB, D. The Critical Importance of Construct Measurement Specification: A Response to Aguirre-Urreta and Marakas. *MIS Quarterly*. 36, 1, 147-156, Mar. 2012. ISSN: 02767783.
- PRESTON, C. C.; COLEMAN, A. M. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, n. 104, p. 1-15, 2000.
- RASMUSSEN, J. L. et al. Univariate and Multivariate Groups by Trials Analysis Under Violation of Variance-Covariance and Normality Assumptions. *Multivariate Behavioral Research*, 24, 1, 93, Jan. 1989. ISSN: 00273171.

REICHMANN, William John. *Use and abuse of statistics*, 1964.

ROBERTSON, J. Likert-type Scales, Statistical Methods, and Effect Sizes. *Communications of the ACM*, v. 55, n. 5, 2012.

RODRIGUEZ, M. C. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*. Summer, 2005.

ROZIN, P; ROYZMAN, E. B. Negativity Bias, Negativity Dominance, and Contagion. *Personality Social Psychology Review*, v. 5, n. 4, p. 296-320, 2001.

SANCHES, C.; MEIRELES, M.; SORDI, J. O. de. Análise qualitativa por meio da lógica paraconsciente: Método de interpretação e síntese de informação obtida por escalas Likert. In: *Anais do III Encontro de Ensino e Pesquisa em Administração e Contabilidade*. João Pessoa: 2011.

SCHRIESHEIM, C.; CASTRO, S. Referent effects in the magnitude estimation scaling of frequency expressions for response anchor sets: an empirical investigation. *Educational and Psychological Measurement*, v. 56, p. 557-569, 1996.

SELLTIZ, C.; JAHODA, M.; DEUTSCH, M.; COOK, S. W. *Métodos de pesquisa nas relações sociais*. São Paulo: EPU, 1975.

SJOBORG, G; NETT, R. *A methodology for social research*. New York: Harper & Row, 1968.

STEVENS, S. S. On the Theory of Scales of Measurement. *Science*, v. 103, n. 2684, p. 677-680, 1946.

THURSTONE, L. L. Attitudes Can Be Measured. *American Journal of Sociology*, v. 33, n. 4, p. 529-554, 1928.

TOMAS, J. M; OLIVER, A. Rosenberg's Self-Esteem Scale: Two Factors or Method Effects. *Structural Equation Modeling*, v. 6, n. 1, p. 84-98, 1999.

VENKATESH, V. et al. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, v. 27, n. 3, p. 425-478, 2003.

VIEIRA, K. M.; DALMORO, M. Dilemas na construção de escalas tipo Likert: o número de itens e a disposição influenciam nos resultados? *Anais... XXXII ENANPAD*. Rio de Janeiro: 2008.

VON KORFF, Michael; MOORE, James C. Stepped care for back pain: activating approaches for primary care. *Annals of Internal Medicine*, v. 134, n. 9, Part 2, p. 911-917, 2001.

WISWANATHAN, M.; SUDMAN, S.; JOHNSON, M. Maximum versus meaningful discrimination in scale response: implications for validity of measurement of consumer perception about products. *Journal of Business Research*, v.12, n. 57, p. 08-24, 2004.