

Criação de Escalas Quantitativas para Estudos de Eventos Mentais: Uma Revisão da Lei do Julgamento Comparativo

Carlo Martins Gaddi^{1*, 1, 2}

Orcid.org/0000-0003-2389-7386

Marcelo Fernandes da Costa^{1, 2}

Orcid.org/0000-0002-3944-8457

¹Laboratório de Psicofisiologia Sensorial, Departamento de Psicologia Experimental, Instituto de Psicologia, Universidade de São Paulo, SP, Brasil

²Núcleo de Neurociências e Comportamento e Neurociências Aplicada, Universidade de São Paulo, SP, Brasil

Resumo

Louis Leon Thurstone (1887-1955), interessado em quantificar atitudes, crenças, preferências e opiniões, criou uma equação conhecida como Lei do Julgamento Comparativo. O modelo de escala psicológica assim proposto foi o de considerar, como parte integrante da medida, a flutuação dos julgamentos ao longo das sucessivas exposições, uma vez que não se sabe seu grau de variação. Sua equação se tornou uma poderosa ferramenta para criação de escalas intervalares para quantificar os mais diversos atributos psicológicos. Embora o método não seja novidade, diversos estudos sociais e comportamentais ainda são normalmente estudados por abordagens qualitativas, através de escalas ordinais e nominais, e que não possuem poder estatístico para a realização de uma investigação verdadeiramente quantitativa.

O artigo pretende revisar a Lei do Julgamento Comparativo e como aplicá-lo em diferentes procedimentos experimentais, como o método de comparação aos pares e ordenamento de posições, trazendo facilidades procedimentais para estudar maiores grupos populacionais e em condições de poucos recursos experimentais, como ocorre em muitas condições de estudos sociais. A expectativa é de que essa revisão seja útil como fonte de referência para facilitar e encorajar a criação de escalas quantitativas nas mais diversas áreas do comportamento e experiência humana.

Palavras-chave: Lei do Julgamento Comparativo, procedimento de ordenamento, escala intervalar, medidas quantitativas, psicofísica.

* Correspondência: Universidade de São Paulo, Instituto de Psicologia, Depto. Psicologia Experimental, Av. Prof. Mello Moraes, 1721, 05508-030, Butantã, São Paulo - SP, Brasil. Tel.: +55 11 30911918, gaddimartins@gmail.com
Financiamento: CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) para Carlo Martins Gaddi (88882.376512/2019-01).

Creation of Quantitative Scales for Studies of Mental Events: A Review of the Law of Comparative Judgment

Abstract

Louis Leon Thurstone (1887-1955) was a pioneer in developing scales for non-metric stimuli, creating an experimental model centered on the subject without the need for physical units of comparison. This model, known as the Law of Comparative Judgment, relies on probabilistic measures rather than discrete ones, making it a powerful tool for quantifying psychological attributes through interval scales. Thurstone proposed that sensory organs operate probabilistically, influenced by an inherent noise within the organism, which can be either internally or externally induced. This noise impacts the hypothetical psychological continuum, causing the observer's response to be not absolute but based on information received at the moment. His model functions analogously to a statistical sample. The Law of Comparative Judgment is structured around three fundamental premises: (i) each stimulus initiates a discriminative process represented on a psychological continuum; (ii) momentary fluctuations, both internal and among different observers, can cause a stimulus to be perceived variably, characterized by a normal distribution; (iii) the mean (μ) and standard deviation (σ) of this distribution represent the scale value and its discriminative dispersion, respectively. One of the most widely adopted experimental methods derived from Thurstone's model is the pairwise comparison protocol. In this protocol, the observer's task is to discriminate between two stimuli presented simultaneously. In each trial, the observer must judge which of the two stimuli is superior in a specified dimension. The process involves comparing each possible pair of stimuli, with each pair being repeated a specific number of times. To organize and analyze the results, a matrix is constructed to capture the differences between all stimuli in the pairwise comparison. The final value assigned to each stimulus is determined by the average difference between that stimulus and all others in the set. The comparative judgment equation provides a method for transforming these differences from all pairwise comparisons into scalar values, whether metric or non-metric. This protocol not only offers a systematic and rigorous approach for measuring psychological attributes but also allows for a detailed and quantitative analysis of the observer's perceptions and preferences. Faced with the challenge of conducting pairwise comparisons under various experimental conditions, Thurstone developed a statistical approach to transform data from a simple ranking task into a binary matrix of comparisons. This method allows for the application of the comparative judgment equation, overcoming the challenges of traditional psychophysical methods related to the duration and complexity of experimental sessions. The ranking method, by simplifying data collection, has proven highly effective in obtaining interval scales in various psychological and social measurements, particularly in contexts of limited experimental control. This efficiency suggests that the method could be adapted for large-scale studies and outside laboratory environments, enhancing research in applied fields of psychology and related areas.

Keywords: Law of Comparative Judgment, ranking order procedure, interval scale; quantitative measure, psychophysics.

Creación de Escalas Cuantitativas para Estudios de Eventos Mentales: Una Revisión de la Ley del Juicio Comparado

Resumen

Louis Leon Thurstone (1887-1955), interesado en cuantificar actitudes, creencias, preferencias y opiniones, creó una ecuación conocida como la Ley del Juicio Comparativo. El modelo de escala

psicológica consideraba la fluctuación de juicios a lo largo de exposiciones sucesivas como parte integral de la medición, dado que su grado de variación es desconocido. Su ecuación se convirtió en una herramienta poderosa para crear escalas intervalares para cuantificar diversos atributos psicológicos. Aunque el método no es nuevo, muchos estudios sociales y conductuales todavía se abordan típicamente utilizando métodos cualitativos, a través de escalas ordinales y nominales, careciendo de poder estadístico para una investigación verdaderamente cuantitativa. El artículo tiene como objetivo revisar la Ley del Juicio Comparativo y cómo aplicarla en diferentes procedimientos experimentales, como el método de comparaciones pareadas y ordenamiento, proporcionando facilidades procedimentales para estudiar grupos de población más grandes y en condiciones de recursos experimentales limitados, como a menudo ocurre en muchos estudios sociales. La expectativa es que esta revisión sea útil como fuente de referencia para facilitar y fomentar la creación de escalas cuantitativas en diversas áreas del comportamiento humano y la experiencia.

Palabras-clave: Derecho de Juicio Comparado, Procedimiento de Clasificación, escala de intervalo, medidas cuantitativas, psicofísica.

A psicologia começou a estabelecer as bases para uma teoria da quantificação dos eventos privados da mente com Gustav Theodor Fechner (1801-1887), separando-se da filosofia para ganhar força e autonomia, aproximando-se do estatuto de uma ciência à medida que conseguiu estabelecer seus próprios métodos relacionando, de forma quantitativa, os eventos do mundo físico com a mente, uma vez que a consciência foi interpretada como uma manifestação do cérebro (Fechner, 1860; Stevens & Marks, 2017).

O objetivo de Fechner era determinar a relação existente entre um estímulo (sua unidade física mensurável) com a sua sensação (valor psicológico). Devido a impossibilidade de medir a sensação diretamente, criou métodos indiretos, perguntando ao observador se o estímulo estava presente ou ausente, se ele era mais ou menos intenso, sempre em relação a um estímulo de referência. Chamou de limiar absoluto o mínimo de energia física necessária para que um estímulo fosse percebido, e limiar diferencial o mínimo de diferença de energia física entre dois estímulos para que eles fossem percebidos como diferentes. Esses limiares foram sua unidade de medida para a construção das escalas de sensibilidade, e é conhecida como unidade de diferença apenas perceptível (Baird & Noma, 1978; Fechner, 1860; Gescheider, 1988; Stevens & Marks, 2017).

Fechner desenvolveu três métodos experimentais para medir os limiares da percepção humana através da sua unidade de diferença apenas perceptível (Fechner, 1860). Esses métodos ficaram conhecidos como métodos psicofísicos clássicos, e forneceram as bases para o desenvolvimento de uma psicologia experimental baseada em evidências, chamada de psicofísica, que é usada até hoje para medir diversas modalidades sensoriais. O desenvolvimento da psicofísica vem ajudando em diversos problemas de várias áreas do conhecimento, como biologia, fisiologia e ciências sensoriais (Baird & Noma, 1978; Dzhafarov & Colonius, 2011; Stevens & Marks, 2017).

No entanto, havia eventos relacionados à mente que ainda eram difíceis de quantificar devido seu alto grau de subjetividade. O próprio Fechner já havia concebido a ideia de que um experimento psicofísico poderia também requerer que um observador fizesse julgamentos em dimensões psicológicas que não possuíssem um correlato físico (estímulos não-métricos). Por exemplo, os observadores poderiam julgar a preferência por diferentes cores, ou o prestígio de diferentes profissões, gravidade de diferentes condições clínicas, preferências por políticos, níveis de inteligência (Baird & Noma, 1978; Stevens & Marks, 2017). Esses estudos vinham ganhando força na psicologia experimental na virada do século XX.

Foi Louis Leon Thurstone (1887-1955), matemático e psicólogo, um dos pioneiros no desenvolvimento da psicometria e análise fatorial, que elaborou escalas para estímulos não-métricos (Thurstone, 1927a, 1927b, 1928, 1929, 1931a, 1931b). Thurstone considerou, como parte integrante da medida, a flutuação dos julgamentos ao longo das sucessivas exposições, uma vez que não se sabe seu grau de variação (Gulliksen, 1946; Maydeu-Olivares, 2004; Maydeu-Olivares & Böckenholt, 2005; Pérez-Ortiz & Mantiuk, 2017). Sua medida experimental é puramente psicológica, centrado no julgamento discriminativo de dois eventos mentais e não necessitando de uma unidade física de comparação. Desta forma, as distâncias discriminativas passaram a ser baseada em medidas probabilísticas, e não mais em medidas discretas. Seu modelo ficou conhecido como Lei do Julgamento Comparativo (Thurstone, 1927a), e é uma poderosa ferramenta para a criação de escalas intervalares para quantificar os mais diversos atributos psicológicos (Arons et al., 2012; Costa et al., 2021; Costa et al., 2014; Franceschini & Maisano, 2020; Stadthagen-González et al., 2018)

Thurstone concretizou a concepção de Fechner, criando um método para escalonar itens tais como qualidade da escrita, seriedade das ofensas criminais e preferência por nacionalidades (Thurstone, 1928, 1929, 1931a). Nenhum desses conceitos tem correlatos físicos ou contrastam com estímulos, tais como intensidade de brilho e som, comprimento de linhas, diferenças colorimétricas. Seu interesse em escalonar atitudes complexas levou a uma reformulação da psicofísica para incorporar a manipulação de um espectro muito mais amplo de medidas do que apenas estímulos sensoriais (Stevens, 1966).

Esse trabalho tem como objetivo revisar a Lei do Julgamento Comparativo (Thurstone, 1927a), e como aplicá-la experimentalmente para estudar eventos subjetivos normalmente abordados de forma qualitativa. Também será revisado o método do ordenamento (Thurstone, 1931b), que traz facilidades procedimentais

ao permitir aplicar as equações do julgamento comparativo em um simples procedimento de ordenamento de posições, o que é vantajoso para estudos populacionais e em condições de poucos recursos experimentais, como ocorre em muitas condições em experimentos sociais.

Embora o modelo proposto não seja novo, muitos estudos nas ciências humanas limitam-se às escalas nominais e ordinais para as análises. A expectativa é de que essa revisão possa ser usada para encorajar a criação de escalas psicofísicas intervalares para as mais diversas áreas do comportamento e experiência humana.

A primeira seção trata de uma revisão a respeito das operações matemáticas permitidas em cada nível hierárquico de escala: nominal, ordinal, intervalar e de razão (Stevens, 1946, 1960), e que devem ser seguidas com rigor. A segunda seção revisa a Lei do Julgamento Comparativo (Thurstone, 1927a), e como ela pode ser aplicada em diferentes procedimentos experimentais, tais como comparação aos pares (Thurstone, 1927b) e ordenamento de posições (Thurstone, 1931b). A terceira seção ilustra didaticamente etapa por etapa da criação de uma escala intervalar a partir de um procedimento de ordenamento de posições.

Hierarquia das Escalas

Para a teoria clássica de medidas, medir consiste em estabelecer a razão de uma grandeza com uma outra grandeza de uma mesma espécie, a qual foi escolhida como unidade de medida. Reconhece-se se uma espécie de grandeza pode ser medida se for possível definir a igualdade e a soma dessas duas grandezas. Entretanto, para que isso seja possível, é necessário que se tenha uma unidade constante e um zero absoluto dessa grandeza (Ferris, 2004).

Stanley Smith Stevens (1906-1973), importante psicólogo americano que trabalhou extensivamente com psicofísica experimental, apresentou três argumentos para abandonar a teoria clássica de medidas (Stevens, 1946, 1960, 1966). Primeiro, é difícil de ter uma

unidade de medida rigorosamente constante. Segundo, o desenvolvimento das ciências evidencia a utilidade de certos dados numéricos que não correspondem a estas exigências. Terceiro, as próprias escalas psicofísicas mostraram a possibilidade de medidas quantitativas para diversas modalidades subjetivas.

Para Stevens, medir é atribuir números a objetos e eventos de acordo com determinadas

regras (Stevens, 1946, 1960, 1966). Isso o levou a criação de uma hierarquia das escalas (Tabela 1), listando quatro diferentes tipos: nominal, ordinal, intervalar e de razão. Cada nível possui propriedades específicas que determinam como seus dados podem ser analisados. As propriedades avaliadas são identidade, magnitude, intervalos iguais e valor mínimo de zero, e a lei básica que governa a hierarquia das escalas é o princípio da invariância.

Tabela 1
Hierarquia das Escalas

Escalas	Propriedades	Estrutura Matemática do Grupo	Estatística permitida
Nominal	Nomeação, classificação	Permutação de grupo; $x' = f(x)$; $f(x)$ significa substituição de um-a-um	Número de casos; Moda
Ordinal	Ordem, intervalos não-iguais	Grupo isotônico $x' = f(x)$; $f(x)$ significa função de aumento monotônico	Mediana; Quartis; Percentis
Intervalar	Intervalos iguais, zero arbitrário	Grupo linear geral; $x' = ax + b$	Média; Desvio-Padrão; Correlação de produto de momento
Razão	Intervalos iguais, zero natural	Grupo de similaridade; $x' = ax$	Coefficiente de Variação

Nota. Os tipos de escala e as regras para cada uma estão descritas na primeira coluna. As condições necessárias para criar as escalas são listadas na segunda coluna. A terceira coluna mostra a transformação matemática que deixa a escala invariante. A quarta coluna mostra as operações estatísticas permitidas para cada escala. Somente as escalas intervalares e de razão permitem estatísticas descritivas, sendo verdadeiramente quantitativas (adaptado de Stevens, 1960).

As escalas nominais apresentam uma forma mais livre para atribuir números aos eventos medidos. Os únicos procedimentos estatísticos permitidos para essas escalas são número de casos e moda. As escalas ordinais surgem quando os eventos observáveis apresentam variações crescentes ou decrescentes. Um exemplo são as escalas de dureza de minerais, que embora apresentem isotomia, não possibilitam saber a real distância entre os pontos na escala, o que significa que a ordem do numeral 4 não possui o dobro da dimensão atribuída para o numeral 2. Escalas ordinais são muito comuns na psicolo-

gia, através da escala Likert, e usada para medir nível de instrução e status socioeconômico. As escalas nominais e ordinárias são monotônicas em grupos, categorias ou ordens, o que significa que nenhuma informação é possível sobre as distâncias entre elas.

Nas escalas intervalares, define-se um zero de forma arbitrária, e recupera-se as distâncias dos pontos medidos, possibilitando saber numericamente quanto um ponto na escala é diferente do outro. Um exemplo é a escala de temperatura Celsius. São muito presentes na psicologia, em escalas de inteligência e atitude, que são dimen-

sões mentais que podem ser bem representadas por essas escalas. Seus pontos se baseiam nas diferenças, e tal condição permite o uso de estatística paramétrica, médias, desvio padrão e coeficiente de correlação.

As escalas de razão devem preencher todos os critérios anteriores, e ainda possuir um zero absoluto (uma dimensão física existente, mas não alcançável empiricamente). O zero absoluto permite multiplicação e divisão nas escalas, além de todas as operações estatísticas dos níveis anteriores. Um exemplo de escala de razão é a escala Kelvin, onde o zero da escala é o mínimo de temperatura fisicamente possível, representando a ausência total de movimento das moléculas. É a escala mais desejável para a ciência, por possuir maior robustez matemática. Diversas modalidades sensoriais são bem descritas nas escalas de razão, como percepção de saturação cromática e intensidade sonora (Stevens, 1966).

A compreensão da hierarquia das escalas de medição é crucial em pesquisas quantitativas e análise estatística, determinando as técnicas matemáticas adequadas e a interpretação dos resultados. Embora as escalas intervalares ofereçam mais poder estatístico, as escalas ordinais são mais viáveis em certos contextos, exigindo a habilidade de adaptar o nível de análise. A estatística paramétrica é usada com variáveis que seguem uma distribuição conhecida, geralmente associada a escalas intervalares, enquanto a estatística não paramétrica é empregada em dados com pressupostos não atendidos, como escalas ordinais. A escolha entre essas abordagens é essencial para garantir análises precisas e interpretações adequadas dos resultados (Anderson, 1961; Kemp & Grace, 2021).

Lei do Julgamento Comparativo

Thurstone assumiu que os órgãos sensoriais se comportam de forma probabilística devido um ruído inerente, seja causado por eventos externos, variabilidade do estímulo, estado neural do observador naquele instante, ou até induzido deliberadamente pelo experimentador.

Esse ruído tem um efeito no contínuo psicológico hipotético, e qualquer que seja a sua fonte, seu efeito sensorial não poderá ser distinguido do efeito causado pelo estímulo. Esse modelo é análogo a uma amostra estatística, onde a resposta não é tida como absoluta, mas sim que o observador prefere ou não tomar determinada decisão, com base na informação recebida naquele momento específico (Maydeu-Olivares, 2004; Pérez-Ortiz & Mantiuk, 2017; Thurstone, 1927a).

A Lei do Julgamento Comparativo tem como base as seguintes premissas:

- (i) Cada estímulo dá origem a um processo discriminativo, que tem algum valor no contínuo psicológico;
- (ii) Devido a flutuações momentâneas (que podem ser consideradas como flutuações internas que ocorrem para o mesmo observador, ou entre diferentes observadores), o valor de um estímulo pode ser julgado como maior ou menor em diferentes repetições. A distribuição dessa flutuação poderá ser caracterizada por uma distribuição normal;
- (iii) A média (μ) e o desvio padrão (σ) da distribuição associada ao estímulo são, respectivamente, os valores da escala e sua dispersão discriminativa;

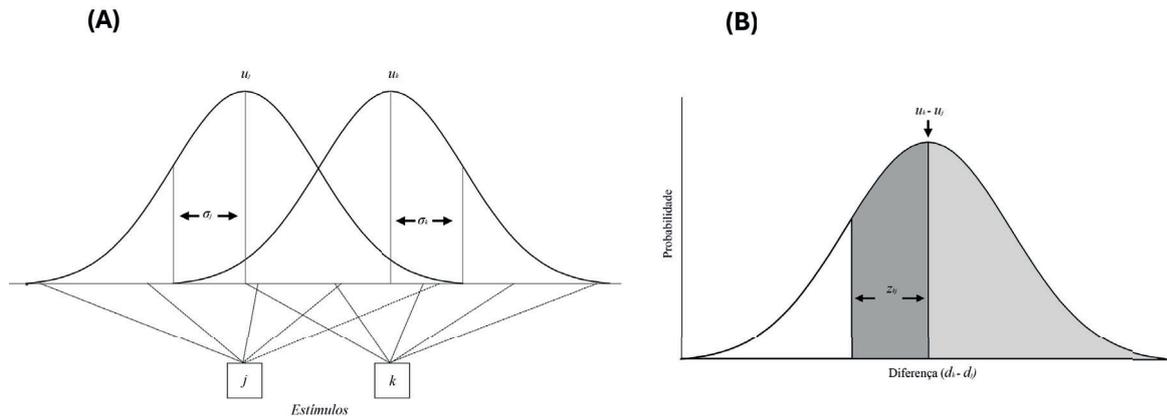
Ao apresentar juntos dois estímulos, j e k , cada um irá gerar uma dispersão discriminativa normal, com uma variabilidade expressa pela seguinte expressão:

$$\sigma_{jk}^2 = \sigma_j^2 + \sigma_k^2 - 2\rho_{jk} \cdot \sigma_j \cdot \sigma_k$$

onde ρ_{jk} é o coeficiente de correlação (quando os processos discriminativos não forem correlacionados, $\rho_{jk} = 0$).

A Figura 1 ilustra as distribuições discriminativas originadas para os estímulos j e k , dispostos em uma escala psicológica com suas respectivas dispersões normais com médias μ_j e μ_k , e desvios padrão σ_j e σ_k (Figura 1 A). A distribuição das diferenças também será normal (Figura 1 B), uma vez que a diferença entre as médias de duas distribuições é igual à média de suas diferenças.

Figura 1
Dispersão Discriminativa para os Estímulos j e k



Notas. (A). Dispersão discriminativa para os estímulos j e k . As médias das distribuições são u_j e u_k com desvios padrão $\sigma_j = \sigma_k$. A intensidade aumenta para a direita ao longo do eixo x (contínuo psicológico – métrico ou não-métrico). (B). Distribuição normal das diferenças entre os processos discriminativos ($d_j - d_k$).

A linha vertical da distribuição das diferenças (Figura 1 B) representa o ponto zero, e a área sob a curva à direita dessa linha representa a proporção de vezes que a diferença entre os dois processos discriminativos ($d_j - d_k$) foi positiva e, conseqüentemente, o observador julgou que o estímulo k é maior que o j . A área sob a curva que está à esquerda do ponto zero representa a proporção de vezes que essa diferença foi negativa, ou seja, a proporção de vezes que o observador julgou que j é maior que k . Portanto, deve-se calcular a média da distribuição das diferenças, uma vez que essa média representa a distância psicológica que existe entre os dois estímulos.

Se os dois estímulos forem apresentados um número suficiente de vezes, é possível calcular a proporção de vezes em que k é julgado maior que j .

$$z_{kj} = \frac{(u_k - u_j)}{\sigma_{kj}}$$

Assumindo que σ_{kj} é igual para todos os pares de estímulos, ele pode ser arranjado como igual a 1. Logo,

$$u_j - u_k = z_{jk}$$

onde z_{jk} é a proporção de vezes que o observador julga que o estímulo j é maior que o estímulo k ,

em unidades de desvio padrão (distribuição-z), e u_k e u_j são os valores escalares para os estímulos k e j . Substituindo na equação 2 os parâmetros da distribuição de diferenças u_d e σ_d , teremos:

$$u_j - u_k = z_{jk} \sqrt{\sigma_j^2 + \sigma_k^2 - 2\rho_{jk} \sigma_j \sigma_k}$$

que é a forma mais geral da Lei do Julgamentos Comparativo.

Para poder resolver a equação 4, é necessário simplificá-la, aceitando várias suposições. Thurstone (1927a) derivou cinco casos diferentes da lei.

O caso I é a própria lei enunciada na equação 4, que em sua forma completa, requer a aplicação dos estímulos a um elevado número de vezes ao mesmo observador. O caso II é idêntico ao anterior, exceto que se supõe que a lei se aplica também para n observadores que façam apenas uma única comparação. Embora esses dois casos não sejam aplicáveis na prática, seu valor como um modelo teórico é fundamental.

O caso III assume que não existe correlação entre as sensações ($\rho_{jk} = 0$). Essa suposição é especialmente plausível quando os estímulos não são semelhantes. Segundo essa suposição, a lei pode ser expressa como:

$$d_j - d_k = z_{jk} \sqrt{\sigma_j^2 + \sigma_k^2}$$

a qual tem uma clara aplicação prática e é solucionável, já que se podem estimar σ_j e σ_k , além de conhecer z_{jk} experimentalmente.

O caso IV necessita de outra suposição; que os desvios das distribuições dos diferentes processos discriminativos sejam iguais ($\sigma_j = \sigma_k = \sigma_n = \dots = \sigma_l$). Aceitando-se essa suposição, a equação anterior pode ser expressa dessa forma:

$$d_j - d_k = z_{jk}\sqrt{2\sigma^2}$$

sendo σ , o desvio comum a todas as distribuições dos processos discriminativos. Essa fórmula não acrescenta nada na prática, sendo preferido o Caso III.

Por fim, o caso V é o mais conhecido e usado devido a sua fácil aplicação e formulação intuitiva (Maydeu-Olivares, 2004; Pérez-Ortiz & Mantiuk, 2017; Thurstone, 1927a). Se assumirmos que $\rho_{jk} = 0$ e que $\sigma_j^2 = \sigma_k^2$, a lei do julgamento comparativo pode ser expressa assim:

$$d_j - d_k = z_{jk}\sigma_j\sqrt{2}$$

e, se tomarmos como unidade da escala $\sigma_j\sqrt{2}$, então, $d_j - d_k = z_{jk}$.

Comparação aos Pares

Um dos protocolos experimentais mais utilizados que se deriva do modelo de Thurstone é conhecido como comparação aos pares (Gulliksen, 1946; Maydeu-Olivares, 2004; Pérez-Ortiz & Mantiuk, 2017; Thurstone, 1927b). A tarefa do observador na comparação aos pares é discriminar entre dois estímulos. Em cada tentativa, se apresentam dois estímulos, e o observador deve julgar qual deles é o maior ao longo dessa dimensão (o mais brilhante, o mais intenso, o mais pesado, o mais agradável, o mais estético). O observador é solicitado a comparar todos os possíveis pares de estímulos, por exemplo, dado os estímulos E_i, E_j, E_k e E_r , um observador compara pares de estímulos $E_i - E_j, E_i - E_k, E_i - E_r, E_j - E_k$ e $E_k - E_r$. Cada possível comparação se repete um número k de vezes, de maneira que o número total de tentativas

será igual ao número de pares de estímulos multiplicado por k .

Os estímulos de cada par podem ser apresentados em duas posições diferentes ao mesmo tempo, ou na mesma posição em diferentes momentos. Em geral, se n é o número de estímulos a escalonar, e o número de pares será $n(n-1)/2$. Assim, o número total de tentativas será $kn(n-1)/2$. Uma matriz pode ser construída de modo a fornecer as diferenças entre os valores na escala para todos os pares de estímulos apresentados. O valor final assinalado a cada estímulo será a média das distâncias entre aquele estímulo contra todos os outros. A equação do julgamento comparativo fornece um procedimento para converter as proporções observadas de todas as comparações aos pares em valores escalares, seja métrico ou não-métrico.

Existem certos cuidados que o experimentador deve considerar ao criar um experimento de comparação aos pares, visando diminuir os vieses de julgamento e histerese sensorial. Seguem alguns:

Os diferentes pares devem ser apresentados sempre ao acaso, nunca um atrás do outro, formados por um estímulo comum ($E_i E_j, E_i E_k, E_i E_r, E_i E_m$).

O experimento deve ser balanceado, e a posição de um mesmo estímulo dentro dos diferentes pares também deverá ser aleatória; por exemplo, o par $E_i E_j$ deve ser apresentado o mesmo número de vezes que o par $E_j E_i$.

Existem diferentes formas de aplicar o método:

- (i) Pode-se solicitar a um observador que julgue k vezes cada um dos pares possíveis.
- (ii) Pode-se solicitar a n observadores que julguem cada par uma única vez.
- (iii) Pode-se solicitar a diferentes grupos de observadores que julguem cada um dos pares k vezes.

A escolha do modo específico de aplicação dependerá de vários fatores, mas principalmente de dois: (a) o tempo disponível dos observado-

res (b) o objetivo preciso que o experimentador pretenda alcançar. Por exemplo, se os observadores têm pouco tempo, o experimentador deverá optar pela forma II. Por outro lado, se o experimentador estiver interessado nas diferenças individuais, então ele deverá empregar a forma I. Se o experimentador quiser obter uma escala psicológica média representativa da população, a forma mais apropriada será a III.

Há uma certa semelhança entre o método de comparação aos pares e o método clássico dos estímulos constantes, utilizado para determinar o limiar diferencial. O procedimento no método dos estímulos constantes consiste em comparar dois estímulos: um de referência, tomado como estímulo-padrão, e que permanece fixo durante toda a sessão, e outro variável. O procedimento de comparação aos pares pode descrever-se como um procedimento de estímulos constantes mais geral, em que cada um dos estímulos é utilizado como sendo o estímulo de referência (Baird & Noma, 1978; Gulliksen, 1946; Stevens & Marks, 2017).

Procedimento de Ordenamento

A dificuldade de realizar procedimentos de comparação aos pares em diversas condições experimentais levou Thurstone a elaborar uma metodologia estatística para transformar dados de um simples procedimento de ordenamento de posições em uma matriz binária de comparação que permitia aplicar sua equação do julgamento comparativo a partir de um certo número de observadores (Maydeu-Olivares & Böckenholt, 2005; Thurstone 1931b).

Em uma tarefa de ordenamento de posições, onde se deve ordenar n estímulos pela sua intensidade, todas apresentadas de uma só vez, o observador está, na verdade, realizando julgamentos comparativos entre as opções disponíveis. Implicitamente, ele está realizando $n(n-1)/2$ comparações mentais. A diferença para o método de comparação aos pares é de que esses padrões de comparações realizados serão sempre transitivos (Maydeu-Olivares & Böckenholt, 2005). Assim, o total de ordenamentos possíveis

para n em um experimento de ordenamento de posições é $n!$. Já para um experimento de comparação aos pares, esse será apenas o total de padrões transitivos possíveis. O número de padrões intransitivos possíveis será de $2^n - n!$. Por exemplo, considere três estímulos, $\{i, j, k\}$, em um experimento de comparação aos pares. Suponha que um observador escolha $i > j, i > k, j > k$. Esse padrão binário de escolhas é transitivo e a sua pintura preferida é i , a segunda é j e a menos preferida é k . Suponha agora que em outro momento do experimento, o mesmo observador escolha $i > j, i > k, k > j$. Esse padrão é intransitivo, já que não é possível ordenar as intensidades com base em suas escolhas binárias.

Thurstone propôs transformar a ordem absoluta do ordenamento de posições em padrões binários de comparação aos pares, permitindo aplicar sua equação do julgamento comparativo (Thurstone, 1931b). Com a ordem das posições dos estímulos de cada observador, é possível calcular as proporções em que estímulo A foi escolhido na primeira colocação (p_{a1}) e a proporção em que o estímulo B foi escolhida na primeira colocação (p_{b1}). Calcula-se a soma das probabilidades para o estímulo A ser escolhido em cada uma das outras posições do rank, $p_{a2} + p_{a3} + p_{a4} \dots = p_{a>1}$. Isso deverá ser realizado para todos os estímulos, de forma que $p_{a1} \cdot p_{b>1}$ represente a probabilidade de cada observador, aleatoriamente, escolher o estímulo A na primeira posição, e o estímulo B nas posições remanescentes. Esse procedimento deverá ser realizado para todos os estímulos em todas as posições. O produto geral será a probabilidade do estímulo A ser mais intenso (P_e) no rank K e o estímulo B ser mais intenso no rank $>K$:

$$p_{ak} \cdot p_{b>k} = p_e$$

Somando o total dos ordenamentos dos estímulos A e B , obtém-se a probabilidade de que o estímulo B seja mais intenso no rank $>A$:

$$\sum (p_{ak} \cdot p_{b>k}) = V$$

Assume-se que os dois estímulos são pelo menos ligeiramente diferentes no valor

psicológico (V), de modo que $B > A$, espera-se que $p_{a>b} > 0.5$.

Com isso, é possível medir a probabilidade de todas os estímulos em uma dada série serem percebidos na posição $>A$. A descrição geral da qual cada par de estímulo é colocada no rank K maior do que os outros estímulos e pode ser formulada como:

$$p_{b>a} = \sum(p_{ak} \cdot p_{b>k}) + \frac{\sum(p_{ak} \cdot p_{bk})}{2}$$

Essa etapa permite construir a matriz de dominância idêntica à de um experimento de comparação aos pares, tornando possível aplicar a equação do julgamento comparativo (Thurstone, 1927a, 1931b).

O ordenamento de posições é uma medida de saliência dimensional, no sentido de que estímulos mais salientes/evidentes/prioritários são externalizados primeiramente em comparação a estímulos de menor valor subjetivo (Alwin & Krosnick, 1985; Maydeu-Olivares & Böckenholt, 2005; Thurstone, 1931b). Uma vez que a escala é produzida diretamente dos julga-

mentos discriminativos realizados (Thurstone, 1927a), a cada julgamento de diferença estamos realizando um afastamento dimensional, adicionando uma unidade de diferença subjetiva. Quando julgamos não haver diferença, estamos realizando uma redução dimensional, subtraindo uma unidade mental.

Construção da Escala Intervalar

Suponha que se queira medir a saliência de um conceito psicológico, por exemplo, nome de cor. Uma tarefa bastante comum é o método de listagem livre (Costa et al., 2021), onde os observadores devem escrever os nomes das cores que vierem à mente. Esse método não necessita do uso de estímulo físico, trazendo uma abordagem mais cognitiva do conceito psicológico da cor. O exemplo a seguir ilustra valores hipotéticos desse tipo de experimento, com um total de 110 observadores.

O experimentador deve anotar em uma tabela a lista dos nomes das cores dos observadores (Tabela 2 A) e transformar a ordem absoluta dos valores em frequências (Tabela 2 B), dividindo pelo total de observadores.

Tabela 2

(A) Ordem absoluta; (B) Frequência

(A) Ordem	Vermelho	Verde	Azul	Amarelo
1	16	13	56	18
2	25	20	21	26
3	17	30	24	18
4	18	24	9	18
(B) Ordem	Vermelho	Verde	Azul	Amarelo
1	0.15	0.12	0.51	0.16
2	0.23	0.18	0.19	0.24
3	0.15	0.27	0.22	0.16
4	0.16	0.22	0.08	0.16

O segundo passo (Tabela 3) consiste em somar todas as frequências das posições de cada nome de cor no rank (k), excluindo a frequência da posição anterior. Somando os valores de vermelho das posições 2, 3 e 4 (0.23, 0.15, 0.16,

ver Tabela 2), obtém-se o resultado da primeira posição do *vermelho*> $k1$ (0.53, ver Tabela 3). Somando os valores de vermelho das posições 3 e 4 (0.15 e 0.16), obtém-se o resultado da segunda posição do *vermelho*> $k1$ (0.32), e assim sucessivamente (equação 8).

Tabela 3
Frequência por Ordem Cumulativa

Ordem	Vermelho $k1$	Vermelho > $k1$	Verde $k1$	Verde > $k1$	Azul $k1$	Azul > $k1$	Amarelo $k1$	Amarelo > $k1$
1	0.15	0.55	0.12	0.67	0.51	0.49	0.16	0.56
2	0.23	0.32	0.18	0.49	0.19	0.30	0.24	0.33
3	0.15	0.16	0.27	0.22	0.22	0.08	0.16	0.16

No terceiro passo (Tabela 4), deve-se construir tabelas contendo todas as possíveis comparações aos pares para calcular as probabilidades de escolher determinado nome de cor contra todos os outros. O total de possíveis comparações binárias para quatro cores é 12. O primeiro valor (0.24, 4), *verde*>*vermelho*, é obtido através da soma das multiplicações entre

cada valor da coluna do *vermelho* $k1$ contra a coluna do *verde*> $k1$ (Tabela 3). O segundo valor (0.05), *verde*>*vermelho* (Tabela 4) é obtido através da metade da soma das multiplicações entre a coluna do *vermelho* $k1$ contra a coluna do *verde* $k1$ (Tabela 3). O terceiro valor é obtido através da soma desses dois valores (equações 9 e 10).

Tabela 4
Cálculo da Dominância Individual

Verde>Vermelho	Vermelho>Verde	Azul>Amarelo	Amarelo>Azul
0.24	0.17	0.16	0.39
0.05	0.05	0.08	0.08
0.29	0.22	0.25	0.47
Azul>Vermelho	Vermelho>Azul	Amarelo>Verde	Verde>Amarelo
0.15	0.37	0.17	0.26
0.08	0.08	0.05	0.05
0.23	0.45	0.22	0.32
Verde>Azul	Azul>Verde	Amarelo>Vermelho	Vermelho>Amarelo
0.48	0.13	0.18	0.19
0.08	0.08	0.05	0.05
0.56	0.21	0.23	0.24

A matriz de dominância (Tabela 5) é obtida através do terceiro valor das colunas da Tabela 4 (em negrito). É necessário converter os valores das frequências da matriz de dominância em distribuição-z, de modo valores $>50\%$ sejam mapeados positivamente e $<50\%$, negativamente, cruzando o valor de 1 unidade na escala com a

probabilidade de 0,84, 2 unidades com 97,7 e assim por diante (Gulliksen, 1946).

Se o experimentador optar por realizar um procedimento de comparação aos pares, é possível obter a matriz de dominância diretamente da organização dos dados dos observadores, não sendo necessário nenhuma das etapas acima.

Tabela 5

(A) Matriz de Dominância; (B) Correspondência em Distribuição-z

(A)	Vermelho	Verde	Azul	Amarelo
Vermelho		0.22	0.45	0.24
Verde	0.29		0.56	0.32
Azul	0.23	0.21		0.25
Amarelo	0.23	0.22	0.47	
(B)	Vermelho	Verde	Azul	Amarelo
Vermelho		-0.78	-0.13	-0.70
Verde	-0.54		0.15	-0.48
Azul	-0.75	-0.80		-0.68
Amarelo	-0.73	-0.76	-0.08	
Média	-0.67	-0.78	-0.02	-0.62
Ordem	3	4	1	2

Para aplicar a equação do julgamento comparativo, deve-se organizar os dados de forma que o nome de cor com a menor média seja colocado lado a lado com a segunda menor, e assim sucessivamente (Tabela 6). Realiza-se a subtração entre a primeira linha com a segunda linha de cada coluna, anotando o valor ao lado (em módulo, uma vez que estamos trabalhando com distâncias).

Para aplicar o caso V (equação 7), soma-se os valores obtidos das subtrações para então calcular a raiz quadrada de dois multiplicada pelo produto da soma obtida, dividido pela quantidade das diferenças calculadas entre os nomes das cores (que no exemplo é de 2).

Tabela 6
Lei do Julgamento Comparativo

	Azul	Amarelo	Amarelo	Vermelho	Vermelho	Verde		
Vermelho	-0.13	-0.70	0.57	-0.70			-0.78	
Verde	0.15	-0.48	0.63	-0.48	-0.54	0.06	-0.54	
Azul		-0.68		-0.68	-0.75	0.06	-0.75	-0.80 0.05
Amarelo	-0.08			-0.73		-0.73	-0.76	0.03
Soma			1.21		0.12			0.08
Caso V			0.85		0.09			0.06
Escala final	Verde	Vermelho	Amarelo	Azul				
	0.00	0.06	0.15	1.00				

A última etapa consiste em escolher um zero de referência para a ancoragem da escala. Por conveniência, escolhemos como zero o nome da cor que obteve o menor valor (verde). A partir disso, soma-se o valor obtido do caso V (Tabela 6), com o valor do nome da cor da posição anterior.

É importante lembrar que para o cumprimento de uma escala intervalar, deve-se atender os critérios necessários para poder realizar uma mudança da escala, isso é, mudar a origem e a unidade, através de uma transformação linear do tipo $x' = ax + b$ cujos valores de a e de b sejam conhecidos, uma vez que a origem da escala não é fixa, e o valor nulo não representa a ausência do atributo sendo medido. Essa escala incorpora todas as propriedades das escalas ordinal e nominal, e ainda especifica uma correspondência 1-1 entre os elementos do domínio observável e o conjunto dos números reais, permitindo assim que a distância entre os pontos medidos tenha

significado lógico (Gescheider, 1988; Stevens, 1966).

Consistência da Escala

Como a escala do Thurstone permite apenas uma estimativa dos dados reais, é importante verificar o nível de confiabilidade da escala. Essa consistência deve ser calculada usando o produto das diferenças entre a matriz de dominância em distribuição-z (Tabela 5 B), que representa os dados do experimento, contra a matriz de dominância final, ou seja, construída a partir dos valores da escala final (Tabela 7 A). Para isso, deve-se calcular a diferença de todos os possíveis pares dos nomes das cores usando os valores escalares finais (escala final, Tabela 6). O produto da diferença das duas matrizes (também em módulo, pois representam distâncias), revela (em distribuição-z) o quanto a escala desviou em relação aos dados originais do experimento (Thurstone, 1928, 1929).

Tabela 7
Consistência Interna da Escala

(A) Matriz de dominância escalar final	Vermelho	Verde	Azul	Amarelo
Vermelho		-0.06	0.94	0.09
Verde	0.06		1.00	0.15
Azul	-0.94	-1.00		-0.85
Amarelo	-0.09	-0.15	0.85	

(B) Tabela 5B - Tabela 7A	Vermelho	Verde	Azul	Amarelo
Vermelho	0.00	0.72	1.07	0.79
Verde	0.60	0.00	0.84	0.63
Azul	0.19	0.20	0.00	0.17
Amarelo	0.64	0.61	0.93	0.00

(C) Discrepância média	0.46			
------------------------	------	--	--	--

Nota. (A) matriz de dominância calculada a partir dos valores finais da escala; (B) produto da diferença entre a matriz de dominância dos dados do experimento (Tabela 5 B) com a matriz de dominância escalar final; (C) média do produto da diferença entre as duas matrizes de dominância, experimental e final, revelando a discrepância final da escala.

No exemplo, o valor do desvio foi de 0.46, quase metade do maior valor na escala final. É preciso ficar atento com valores experimentais muito baixos (<0.025) ou muito altos (>0.975) obtidos na matriz de dominância (Tabela 5 A). Os estímulos que obtiverem esses valores devem ser removidos da análise, dada a impossibilidade de assumir linearidade nas extremidades da distribuição normal, o que faz com que eles comprometam a consistência da escala. De forma geral, a consistência tende a aumentar conforme cresce o número de observadores no experimento.

Discussão

Fechner foi um dos pioneiros na psicologia experimental quantitativa por meio da sua psicofísica, que ainda é relevante clinicamente para a medição de diversas modalidades sensoriais (Baird & Noma, 1978; Dzhanfarov & Colonius, 2011; Fechner, 1860; Gescheider, 1988; Stevens & Marks, 2017). Posteriormente,

Stevens realizou uma série de experimentos usando técnicas indisponíveis na época de Fechner, como os métodos de estimativa de magnitude, o que levou ao desenvolvimento de escalas psicofísicas diretas para escalonar estímulos sensoriais muito mais complexos. Seus métodos possibilitaram a criação de medidas mais apuradas, usadas até hoje clinicamente para diferentes escalas sensoriais (Stevens, 1946, 1960, 1966), como na saturação cromática ou percepção sonora. Seu interesse em matemática e física o levou a elaborar níveis hierárquicos para a criação de escalas: nominal, ordinal, intervalar e de razão, onde cada nível possui propriedades específicas que determinam como seus dados podem ser analisados (Anderson, 1961; Ferris, 2004; Kemp & Grace, 2021; Stevens, 1960; Stevens & Marks, 2017). Sua teoria de medição, aceita em toda a ciência, o fez ser conhecido como pai da psicofísica moderna.

Porém, foi Thurstone que descreveu o processo mental subjacente aos julgamentos subjetivos e comportamento humano por meio de

um modelo probabilístico, conhecido como Lei do Julgamento Comparativo (Thurstone, 1927a). Esse modelo permitiu a aplicação de estudos quantitativos não apenas a estímulos físicos comparativos, como comprimento de linha e diferenças de brilho, mas também a eventos puramente mentais, como escalas de inteligência, preferências estéticas e opiniões políticas (Thurstone, 1928, 1929, 1931a). Essa expansão do uso de escalas quantitativas para eventos comportamentais abriu novas possibilidades de pesquisa em áreas como opinião pública, atitudes políticas e raciais, preferências, gerando avanços significativos em disciplinas como sociologia, comunicação, psicologia industrial e marketing. Essas escalas têm sido aplicadas com sucesso em estudos sobre percepção de qualidade de saúde em pacientes com diversas condições clínicas (Arons et al., 2012), nomeação e categorização de cores em crianças e adultos (Costa et al., 2021), escolhas de produtos para fabricação industrial (Franceschini & Maisano, 2020), atitude sexual em adultos brasileiros (Costa et al., 2014), alternância de código linguístico na comunicação (Stadthagen-González et al., 2018). O sucesso desses estudos explicita a versatilidade das escalas de Thurstone nos mais variados estudos do comportamento humano.

Thurstone também conseguiu superar um dos principais desafios dos métodos psicofísicos clássicos, que era a duração prolongada das sessões experimentais devido à necessidade de repetições extensas dos estímulos e ao uso de equipamentos sofisticados. Ele desenvolveu um método alternativo que, através de uma simples tarefa de ordenamento, conseguiu obter escalas estatísticas intervalares, usando como base as proporções derivadas do ordenamento em comparações entre os possíveis pares de estímulos (Thurstone, 1931b). O método de ordenamento vem se demonstrando ser bastante eficiente para diversas medidas psicológicas e sociais (Alwin & Krosnick, 1985), especialmente em situações de pouco controle experimental (Yu et al., 2023). Isso sugere que o uso desse método pode viabilizar estudos em grande escala e fora de ambientes laboratoriais.

Embora a psicologia tenha passado então a priorizar a criação de escalas intervalares ou de razão para quantificar as manifestações da mente (Kemp & Grace, 2021), é crucial distinguir entre medidas físicas e medidas comportamentais. Enquanto uma medida física pode ser observada diretamente (como altura e peso) e repetida infinitamente para resultados consistentes, uma medida comportamental é construída a partir de conceitos latentes e não observáveis, usando indicadores diretos. Fatores externos podem influenciar a precisão da medida comportamental, introduzindo um componente de erro que deve ser considerado.

Ainda assim, a característica evidente para usar a metodologia proposta está na associação de uma tarefa simples, compreensível e executável por observadores de diferentes idades e níveis socioeconômicos, fácil de ser criado e parametrizado pelo experimentador, mas que possui robustez psicofísica, permitindo que os valores absolutos sejam convertidos em valores escalares de intervalos iguais. As escalas intervalares possibilitam uma investigação com o uso de estatística paramétrica, incluindo média, desvio padrão e coeficiente de correlação (Anderson, 1961), trazendo diversas facilidades para uma abordagem quantitativa abrangente dos eventos e manifestações subjetivas da mente e do comportamento humano.

Considerações Finais

O objetivo principal do escalonamento psicofísico é a quantificação da experiência perceptual, através do entendimento das relações entre o mundo físico, e a representação interna do indivíduo. O desafio para unir diferentes atributos psicológicos envolvidos é gigantesco, mas através de uma metodologia simples, e muito substancial, é possível ao menos investigar essas relações em escalas intervalares, possibilitando estudar diversos eventos psicológicos de forma mais confiável e objetiva, e não apenas em laboratório e com testes que necessitem equipamentos sofisticados, mas em grandes populações e

em seus ambientes específicos, independentemente da idade ou nível educacional.

Espera-se que essa revisão traga o embasamento necessário para que modelos experimentais possam ser criados e aplicados para estudar as mais diversas áreas do comportamento e experiência humana, das quais ainda existem poucos estudos quantitativos.

Contribuição dos autores

Todos os autores participaram de todo processo, desde a elaboração da ideia; esboço; levantamento bibliográfico; elaboração do texto. Todos os autores aprovaram a versão final do manuscrito.

Conflitos de interesse

Os autores declaram não haver conflito de interesses relacionado à publicação deste manuscrito.

Referências

- Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly*, 49(4), 535–552. <https://doi.org/10.1086/268949>
- Anderson, N. H. (1961). Scales and statistics: Parametric and nonparametric. *Psychological Bulletin*, 58(4), 305–316. <https://doi.org/10.1037/h0042576>
- Arons, A. M., Krabbe, P. F., Schölzel-Dorenbos, C. J., van der Wilt, G. J., & Olde Rikkert, M. G. (2012). Thurstone scaling revealed systematic health-state valuation differences between patients with dementia and proxies. *Journal of Clinical Epidemiology*, 65(8), 897–905. <https://doi.org/10.1016/j.jclinepi.2012.01.018>
- Baird, J. C., & Noma, E. (1978). *Fundamentals of scaling and psychophysics*. John Wiley.
- Costa, M. F., Gaddi, C. M., Gonsalez, V. M., & de Paula, F. V. (2021). Psychophysical scaling method for measurement of colors concept in children and adults. *Methods in Psychology*, 5. <https://doi.org/10.1016/j.metip.2021.100077>
- Costa, M. F., Nomura, T. N., & Ribeiro, G. D. C. (2014). Psychophysical scaling of sexual attitude in Brazilian adults. In C. Pracana (Ed.), *Proceedings Inpact 2014: International Psychological Applications Conference and Trends* (pp. 322-324). World Institute for Advanced Research and Science.
- Dzhafarov, E. N., & Colonius, H. (2011). The Fechnerian idea. *The American Journal of Psychology*, 124(2), 127–140. <https://doi.org/10.5406/amerjpsyc.124.2.0127>
- Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.
- Ferris, T. L. (2004). A new definition of measurement. *Measurement*, 36(1), 101-109. <https://doi.org/10.1016/j.measurement.2004.03.001>
- Franceschini, F., & Maisano, D. (2020). Adapting Thurstone's Law of Comparative Judgment to fuse preference orderings in manufacturing applications. *Journal of Intelligent Manufacturing*, 31(2), 387–402. <https://doi.org/10.1007/s10845-018-1452-5>
- Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology*, 39(1), 169-200. <https://doi.org/10.1146/annurev.ps.39.020188.001125>
- Gulliksen, H. (1946). Paired comparisons and the logic of measurement. *Psychological Review*, 53(4), 199. <https://doi.org/10.1037/h0061673>
- Kemp, S., & Grace, R. C. (2021). Using ordinal scales in psychology. *Methods in Psychology*, 5, 100054. <https://doi.org/10.1016/j.metip.2021.100054>
- Maydeu-Olivares, A. (2004). Thurstone's Case V model: A structural equation modeling perspective. In *Recent developments on structural equation models* (pp. 41-67). Springer.
- Maydeu-Olivares, A., & Böckenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods*, 10(3), 285–304. <https://doi.org/10.1037/1082-989X.10.3.285>
- Pérez-Ortiz, M., & Mantiuk, R. K. (2017). *A practical guide and software for analysing pairwise comparison experiments* [Online forum post]. arXiv Stat.AP [stat.AP]. <https://arxiv.org/abs/1712.03686>
- Stadthagen-González, H., López, L., Parafita Couto, M. C., & Párraga, C. A. (2018). Using two-alternative forced choice tasks and Thurstone's

- law of comparative judgments for code-switching research. *Linguistic Approaches to Bilingualism*, 8(1), 67-97. <https://doi.org/10.25417/uic.19346456.v1>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680. <https://doi.org/10.1126/science.103.2684.677>
- Stevens, S. S. (1960). The psychophysics of sensory function. *American Scientist*, 48(2), 226-253. <https://www.jstor.org/stable/27827540>
- Stevens, S. S. (1966). A Metric for the Social Consensus. *Science*, 151(3710), 530-541. <http://www.jstor.org/stable/1717034>
- Stevens, S. S., & Marks, L. E. (2017). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. Routledge.
- Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review*, 34(4), 273-286. <https://doi.org/10.1037/H0070288>
- Thurstone, L. L. (1927b). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology*, 21(4), 384-400. <https://doi.org/10.1037/h0065439>
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554. <https://doi.org/10.1086/214483>
- Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological Review*, 36(3), 222-241. <https://doi.org/10.1037/h0070922>
- Thurstone, L. L. (1931a). The measurement of change in social attitude. *Journal of Social Psychology*, 2(2), 230-235. <https://doi.org/10.1080/00224545.1931.9918969>
- Thurstone, L. L. (1931b). Rank order as a psychophysical method. *Journal of Experimental Psychology*, 14. <https://doi.org/10.1037/h0070025>
- Yu, L., Yun, C., Xia, G., Westland, S., Li, Z., & Cheung, V. (2023). Analysis of research strategies to determine individual color preference: N-alternative forced choice, rank-order, rating and paired comparison. *Color Research & Application*, 48(2), 222-229. <https://doi.org/10.1002/col.22836>

Recebido: 13/09/2023
1ª revisão: 05/04/2024
Aceite final: 19/04/2024



O(s) autor(es), 20234 Acesso aberto. Este artigo está distribuído nos termos da Licença Internacional Creative Commons Atribuição 4.0 (<http://creativecommons.org/licenses/by/4.0/>), que permite o uso, distribuição e reprodução sem restrições em qualquer meio, desde que você dê crédito apropriado ao(s) autor(es) original(ais) e à fonte, fornecer um link para a licença Creative Commons e indicar se as alterações foram feitas.