METHOD OF EQUAL-APPEARING INTERVALS

Authored by

mohammad looti

October 13, 2025

METHOD OF EQUAL-APPEARING INTERVALS

Primary Disciplinary Field(s): Psychometrics, **Scaling Theory**, Social Psychology

1. Core Definition

The Method of Equal-Appearing Intervals, often referred to simply as the **Thurstone Scale**, is a rigorous psychometric scaling technique designed to construct interval scales for the measurement of attitudes, opinions, or subjective perceptions. This method operates on the fundamental assumption that specialized judges or raters can accurately perceive and sort stimulus statements such that the resulting intervals between the scale points are perceived as subjectively equal. The technique ensures that the values obtained using the final scale assume that the psychological distance between consecutive intervals is uniform and consistent, thereby moving measurement beyond mere ordinal ranking into the domain of **interval data**. This characteristic is crucial because interval data allows for arithmetic operations (such as calculating means and standard deviations), which are impossible or inappropriate for purely ordinal measures. The core objective is to locate specific stimuli--such as statements about an attitude object--at specific, equidistant points along an underlying psychological continuum, ensuring that stimuli used in the experiment are, by design, equidistant on the internal scale.

Unlike simpler scaling methods that rely on the respondent to define both their attitude and the item weight simultaneously, the Method of Equal-Appearing Intervals separates these two tasks. The weight or scale value of each item is determined externally by a panel of judges, independent of the final respondents who will use the scale. This method transforms qualitative judgments (whether a statement expresses a strong or weak opinion) into quantifiable scale values, typically ranging from 1 (least favorable) to 11 (most favorable), or similar ranges. By utilizing specific statistical procedures, the method effectively filters out ambiguous or poorly defined statements, resulting in a highly reliable and objective scale intended to map a single latent trait or attitude structure.

2. Historical Development and Origin

The Method of Equal-Appearing Intervals was pioneered in the late 1920s by the American psychologist and psychometrician **Louis Leon Thurstone**. Thurstone's work represented a groundbreaking effort to introduce true measurement precision into the nascent field of social science, which had previously been constrained by crude, largely ordinal instruments. Before Thurstone, social measurements were typically limited to simply ranking preferences, failing to quantify the magnitude of the difference between those ranks. Thurstone sought to apply principles derived from physical science measurement to the measurement of psychological phenomena, arguing that if attitudes existed on a continuum, they must be measurable using objective intervals.

Thurstone formalized the approach in his 1929 work, *The Measurement of Attitude*. His development of this method was deeply influenced by the concepts of **psychophysics**, particularly the work of Fechner and Weber, which dealt with the relationship between physical stimuli and psychological sensation. Thurstone adapted these principles, specifically his formulation of the **Law of Comparative Judgment**, to address social stimuli like attitude statements. The Law of Comparative Judgment provided the mathematical framework necessary to transform paired comparisons or rating assignments into scale values, laying the foundation for modern psychometric scaling. The method was rapidly adopted for measuring politically sensitive topics, such as attitudes toward the church, war, and various social groups, demonstrating that subjective human experience could be systematically quantified.

3. Underlying Assumptions and Mathematical Basis

The success of the Method of Equal-Appearing Intervals hinges on two critical underlying assumptions. The first is the assumption of **unidimensionality**, meaning that all statements included in the instrument measure only a single underlying psychological continuum (e.g., attitude toward environmental protection). If the statements tap into multiple distinct attitudes, the resulting scale values become meaningless. The second, and more specific, assumption is that the panel of judges is capable of ignoring their personal opinions or attitudes toward the concept being measured and act purely as objective measurers of the statement's position on the favorable/unfavorable continuum. They must be able to assign statements to categories based solely on the intensity of the feeling expressed by the statement, rather than agreeing or disagreeing with the statement itself.

Mathematically, the method relies on statistical analysis of the distribution of the judges' assignments for each item. When judges sort statements into categories (e.g., 1 to 11), the cumulative proportions of assignments for each statement across the categories are plotted. The scale value for an item is typically determined by the median of the distribution of assignments made by all judges. This median represents the point on the continuum above which 50% of the judges placed the statement. Furthermore, the spread or variability of the assignments, usually quantified by the **interquartile range (IQR)**, serves as a crucial measure of the item's ambiguity. A small IQR indicates high agreement among judges regarding the item's position, confirming its clarity and suitability for inclusion in the final scale, whereas a large IQR signifies high ambiguity and requires the item's rejection.

This mathematical rigor, particularly the use of the median and IQR, differentiates Thurstone scaling from later methods. It aims to eliminate measurement error inherent in vague statements, ensuring that the selected stimuli are stable markers along the psychological scale. The objective scale value assigned to each statement is considered fixed and external to the subsequent respondents, who merely indicate the statements with which they agree.

4. The Scaling Process: Detailed Steps

The construction of a scale using the Method of Equal-Appearing Intervals is a multi-stage, labor-intensive process requiring meticulous preparation and analysis.

The initial stage involves **Statement Generation**. A large pool of statements (often 100 to 200) concerning the attitude object is collected or written. These statements must cover the entire continuum, ranging from extremely favorable expressions to extremely unfavorable expressions, and must be concise, unambiguous in their content, and relevant to the attitude being measured.

The second stage involves the selection and instruction of the **Panel of Judges**. A substantial number of judges (ideally 50 to 300) are recruited. They are explicitly instructed to ignore their personal feelings about the topic and to sort the statements purely according to the intensity of the attitude expressed by the statement. They are given a set number of categories (e.g., 9 or 11) that are visually and verbally defined as representing equidistant intervals along the favorable/unfavorable continuum.

The third stage is the **Sorting Task and Data Collection**. Each judge independently places every statement into one of the designated categories. This generates a frequency distribution for every statement across all categories. For example, Statement A might be placed in Category 1 (Extremely Unfavorable) by 10% of judges, Category 2 by 15%, Category 3 by 40%, and so on.

The fourth stage is **Calculation and Item Analysis**. For each statement, the scale value (S) is calculated, typically as the median category assignment across all judges. Simultaneously, the measure of ambiguity, the IQR, is calculated. Statements with high ambiguity (large IQR) are rejected because they do not reliably occupy a single position on the continuum.

The final stage involves **Scale Construction**. The researcher selects a final set of 20 to 30 statements that are highly non-ambiguous (low IQR) and whose scale values (S) are distributed approximately equally across the entire continuum. The final scale consists only of these selected statements and their predetermined scale values.

5. Key Characteristics and Features

Judge-Centered Scaling: The scale values for the items are derived entirely from the opinions of the initial panel of judges, not the final survey respondents. This separation is the defining feature of the Thurstone method, ensuring item objectivity.

Interval Measurement: It is designed specifically to yield a measurement scale with assumed equal intervals between points, allowing researchers to claim that the psychological difference between a score of 3 and 4 is the same as the difference between 8 and 9.

Ambiguity Control: The method incorporates a direct, statistical mechanism (the IQR) for

measuring and eliminating items that are interpreted differently by the judges, resulting in a cleaner and less noisy final instrument.

Use of Median Scale Values: The standard score assigned to a statement is the median of the categories assigned by the judges, reflecting the central tendency of their interpretation.

6. Comparison to Other Scaling Techniques

While foundational, the Method of Equal-Appearing Intervals is often contrasted with two other major scaling methods: the **Likert Scale** (Method of Summated Ratings) and the **Guttman Scale** (Scalogram Analysis). Understanding these differences clarifies the unique position of the Thurstone technique in psychometrics.

The most significant difference lies between Thurstone and Likert scaling. Likert scaling is far simpler and involves respondents indicating their level of agreement (e.g., Strongly Agree to Strongly Disagree) with a statement. The respondent's final score is the sum or average of their responses, and the item weights are determined implicitly by the response patterns of the final sample, not by external judges. Crucially, the Likert scale generates data that is technically ordinal, though often treated as quasi-interval; the Thurstone scale is explicitly designed to produce **interval data**. Furthermore, the Likert scale does not inherently possess a mechanism to eliminate ambiguous items in the same rigorous, pre-testing manner as the Thurstone method.

In contrast, the Guttman scale focuses on **cumulative scaling**, assuming that agreement with a strongly worded statement implies agreement with all less strongly worded statements (e.g., if one agrees to "I would die for my country," they must agree to "I like my country"). This implies a perfect, deterministic ordering of items. The Thurstone method, while requiring unidimensionality, does not assume this perfect, hierarchical structure, relying instead on the perceived equality of psychological intervals.

7. Limitations and Practical Criticisms

Despite its theoretical elegance and objective design, the Method of Equal-Appearing Intervals faces significant practical and conceptual criticisms, which have limited its widespread use compared to the Likert method.

The primary limitation is its **intensive labor requirement**. The need to recruit, instruct, and analyze the sorting data from a large panel of specialized judges makes the Thurstone method expensive and time-consuming. Preparing 200 statements, gathering 100 judges, and performing the subsequent statistical item analysis represents a substantially higher methodological hurdle than simply generating 20 statements for a Likert scale.

A key conceptual criticism centers on the assumption that judges can remain objective. Critics

argue that judges' personal attitudes inevitably influence their perception of the intensity of a statement, potentially biasing the final scale values. Research has suggested that the scale values determined by favorable judges and unfavorable judges may sometimes differ, violating the assumption of total objectivity. However, Thurstone argued that even if bias exists, it is often minimal and is generally canceled out by the large number of judges involved.

Furthermore, the Thurstone scale is less versatile than the Likert scale. Because the final scale relies on agreement with a select few statements that serve as objective markers, researchers cannot easily modify or update items without re-running the entire judge-sorting and scaling process. The inherent difficulty in construction means that subsequent researchers often prefer methods that allow for rapid iteration and adaptation.

8. Significance and Impact

The Method of Equal-Appearing Intervals remains profoundly significant not necessarily for its current widespread use, but for its **foundational role** in establishing psychometrics as a rigorous scientific discipline. Thurstone provided the first widely accepted mathematical framework for translating subjective psychological variables into objective interval-level data. This innovation legitimized the measurement of abstract social constructs like attitude, values, and perception.

Thurstone's work solidified the concept of **item analysis**--the systematic process of evaluating and selecting items based on statistical criteria (like the IQR) before they are administered to the target population. This principle has been incorporated into virtually all subsequent scaling methodologies, including modern item response theory (IRT) models. Although researchers today rarely construct a pure Thurstone scale due to logistical constraints, its core logic--the belief that psychological distances can be quantified and that item weights must be objectively determined--is the bedrock upon which all sophisticated attitude and personality measurements rest. It demonstrated conclusively that rigorous measurement of the "internal scale" was achievable.

Further Reading

Louis Leon Thurstone (Wikipedia entry)
Scaling (Social Sciences Wikipedia entry)
Psychometrics (Wikipedia entry)