A new generation of scaling methods: Level-anchored ratio scaling

Gunnar Borg and Elisabet Borg∗

RESUMO

Apresentam-se aqui princípios, métodos e experiências relativos ao escalonamento de razão com níveis de ancoragem verbais. Entre os grandes princípios figuram: os métodos de “escalonamento de razão” de Stevens; o Modelo de Variação segundo Borg; as determinações da extensão dinâmica subjectiva; a semântica quantitativa para a seleção e colocação de âncoras verbais; o evitamento dos efeitos terminais; as validações psicofisiológicas; uma âncora especial, a “estrela fixa”, decisiva para a intersubjectividade e a comunicação bilateral. A escala final é uma escala de razão com níveis de ancoragem, que abrange uma gama natural de intensidades e evidencia congruência entre os números e as âncoras verbais. No desenvolvimento de escalas psicométricas, esta é a última geração de métodos para o escalonamento de sensações, percepções, experiências e emoções. Resultados empíricos que apoiam esta escala têm sido recolhidos em estudos laboratoriais e de campo. A escala possui um vasto leque de aplicações nos campos do diagnóstico clínico, treino e reabilitação, desportos e ergonomia.


1. INTRODUCTION

Verbally level-anchored ratio scaling, also called CR scaling (meaning category (C) - ratio (R) - scaling) combines the advantages of “ratio scaling” for determining stimulus-response (S-R-) functions, with “category scaling” for direct levels of intensity. The two now mainly used CR scales are the CR10 scale, and the CR100 scale. They have great advantages towards the “ratio scaling methods” developed by S. S. Stevens, because they permit, not only determinations of relative S-R-functions, but also “absolute” levels of intensity.
The measuring sense.

Humans have a very good ability to judge perceived magnitudes. We can “measure” intensities of all kinds of attributes: sensations, experiences and emotions; both by ”scaling” relative growth functions, and by determining intensity levels that have a direct and natural (“absolute”) meaning to us. To accomplish this there has to be an operating ”measuring sense,” a judgmental system consisting of many ”measuring sticks” and ”gauges” integrated in a network, a multidimensional evaluative cognitive space.

The physical stimulus and its intensity variation may be well defined, and exact relations and levels determined in comparison to an arbitrarily set unit. This means that in natural science relations may be invariant and ”absolute”, while on the other hand the levels are arbitrary without ”absolute” meaning, unless regarded in relation to a specific class of objects or events. As Quine (1987) states: ”There is no place in science for bigness, because of this lack of boundary; but there is a place for the relation of bignerness. Here we see the familiar and widely applicable rectification of vagueness: disclaim the vague positive and cleave to the precise comparative”.

In our human world, the frequency of ”distal” and ”proximal” stimuli form mainly positively skewed distributions. For most attributes, weak stimuli are very common and very strong ones very unusual. We do not need to discriminate especially well between the extremely strong intensities, but between weak stimuli. This is in line with the Weber-fraction and its formulation by Ekman (1959) for perceptual intensities. It is then important to note that the different adjectives and adverbs that are used to identify perceptual intensity levels seem to be spread over the range according to this frequency distribution, roughly separated by a certain factor (see Figure 1).

![Frequency Distribution Graph]

**Figure 1** — Two possible frequency distributions (among many) of stimuli in the human world, and how the different adjectives and adverbs spread over the range according to this.
In a special “absolute” sense, there may not be any boundaries in science, as Quine states. (With some exceptions, as, e.g., zero temperature at 0° Kelvin or -273° C, or the speed of light). Most objects and events in our human world have, however, their boundaries, since they — by nature or definition — belong to specific classes with their limits. The biological systems set the boundaries for human experiences. Our sensory organs function from extremely low, “noise” levels and/or threshold intensities, to maximum. This limits the natural variation of the stimulation that can be perceived. The sensory reactions are to a large extent genetically determined and level anchored, and evaluated depending upon its position in the possible intensity range (Borg, 1961, 1962, 1990). Our perceptions, thus, depend very much upon the sensory organs. They constitute the basis for judgments of relations as well as levels. Our personal experiences and use of language also play an important role in the evaluation. We learn to use adequate verbal expressions for different intensities, in about the same way as we learn to name objects and events. The different linguistic terms become, in a very direct way, connected with different intensities of an “absolute” character. This seems to be true for all normal persons in normal situations. A perception is, however, not a static and isolated event, but a dynamic process affected and changed depending upon the figure-ground relation, the context, the level of adaptation, etc. With many linguistic terms we can, in a very refined and adequate way, describe perceptual intensities and communicate with other people in most “normal” situations. The demand on intersubjectivity can thus be met in a proper way for people with the same language and/or similar cultural background. The verbal expressions for perceptions denote fairly stable intensities thanks to the inner frame of reference founded on basic physiological processes and experiences from the earliest days of life and throughout the years. This notion is rather common and has been expressed among others by Borg (1962, 1964), Witte (1966), and Zoekes and Sarris (1983).

Responses reflecting perceptual intensities may also be given in numbers. There is, however, no direct connection between perceptual intensities and numbers as there is for verbal expressions. To claim that a certain juice is “15” in sourness or that a perceived exertion is “40” would be meaningless. The numbers have to be anchored to a certain chosen unit. For numerosness the numbers, however, get a certain “absolute” character. We learn to count the number of things in a similar way as we learn to name an object. To communicate intensities we sometimes also use other methods, such as showing with our arms and hands how big something is. We use analogies like giving a distance on a line, a sector of an area, or showing with “finger span” the intensity. Variation in colors from very light to very dark, or the saturation of colors, are also used. From other sensory modalities we use analogies in the form of the variation of pitch or heat or force of handgrip, and, maybe most common of all, conceived magnitudes, such as degrees of heaviness. The latter is very common in the literature. One good example of this is Dayden’s quote: “His heels to heavy, and his head to light”.

The natural connection between perceptual intensity and verbal expressions does not hinder us from using the terms in different ways depending upon the class of objects. As a complement to the fundamental (inner) determined intensity we also have "class dependent" identifications. We learn to identify a certain class of
objects and specify the range and frequency-relations for that class. A lemon that we call "not especially sour" may still be more sour than an orange that we judge to be "rather sour", and a "warm beer" can be colder than a "cold cup of coffee". Several examples showing forms of "semantic interaction" are expressed by Marks. In an empirical study subjects judged the intensity of the experience represented by the expression "the sun whispers" as stronger than "the moon roars" (Marks, 1982).

The fairly stable world that we are living in has formed us and made us adapt to a certain kind, range, and frequency of stimuli. Since our sensory organs are also fairly stable, our lifelong experiences give us many stable memories. To meet the demand on intersubjectivity of relations we have to assume that the sensory organs are of a similar character for different people. We know that sensory deficits and other deviations exist in some people, and that experiences differ, but, as a main assumption we can assume that most people have similar impressions with regard to relations between events: an intersubjective isomorphy. Then we may also postulate a high degree of intersubjectivity in perception of intensity levels. This is given for the same reason as above for relations, but also because we can never reach outside our own limited range. An intensity has to be judged in relation to its position in each individual's range. A very "right-leading" assumption is then that this total perceptual range is about the same for each individual, and also that the perception of a maximal intensity, e.g. in the form of a certain maximal perceived exertion and heaviness, is of similar intensity, and can be used as a main point of reference, "a fixed star", for interindividual comparisons, (Borg, 1962, 1992). Without this assumption of similarity — giving it an epistemic priority — communication would be chaotic. With this assumption, existing individual differences can be studied as deviations from what is similar.

History and background

Psychophysical "ratio" scaling (Stevens, 1957, 1975) was a great breakthrough. Before these methods were developed, scaling mainly contained a hypothetical S-R-function proposed by Fechner. Stevens' methods made it possible to describe relative S-R-functions mathematically with power functions (see Stevens, 1975). These methods were developed in analogy with physical scaling methods, using an arbitrary and "meaningless" unit that all other measures were related to. The choice of a relative unit did not permit any valid direct comparisons of "absolute" levels of intensity. The senior author pondered over this problem very much (Borg, 1961, 1962, 1998).

Stevens' methods have been exposed to a lot of criticism (see e.g. Falmagne, 1985; Poulton, 1989; Lockhead, 1992). They are, however, still the best for general descriptions of perceptual growth functions. The very strict mathematical demands that define a true ratio scale have, however, never been met. Most of these functions have been described by power functions. Stevens preferred a function without any extra constants. However, as shown by Ekman's work with taste perception (Ekman, 1959), and Borg's with perceived exertion (Borg, 1962), one extra constant is needed because of a certain perceptual "noise", which exists in the
A new generation of scaling methods: level-anchored ratio scaling

absence of distal stimulation. In at least one special case — perception of exertion for walking — two constants are needed (Borg, 1973a; see Borg, 1998). The general power law can thus be expressed (Borg, 1961, 1962).

\[ R = a + c (S - b)^n \]  

where \( S \) is the stimulus and \( R \) is the response given according to the chosen psychophysical method. The exponent, \( n \), describes the curvature of the function (the slope in log-log) and \( c \) is a measure constant (sometimes somewhat misleadingly called the “intercept”). The two extra constants, \( a \) and \( b \), refer to a basic perceptual noise or the absolute threshold. This power law can be applied, not only to perceptual, but also to physiological functions (Borg, 1962; Mountcastle, Poggio and Werner, 1963).

Even though Stevens was primarily concerned with studying group-functions, and paid very little interest in individual functions (Stevens, 1971), a small group of researchers, e.g. Hellman and Zwischlocki (1963), Zwischlocki and Goodman (1980), Collins and Gescheider (1989), have claimed that free magnitude estimation (ME) (with a special instruction) may meet with the “absolute” (A) demands necessary to produce reliable “level estimates” (for a review, see Gescheider, 1988, 1997). The concept of AME has, however, been questioned by several psychophysicists, e.g. Borg (1982), Marks (1988), Ward (1987, 1992). For a review, see Ellermeier, Westphal and Heidenfelder (1991).

In most psychological studies, simple scales of an ordinal or interval character are used. The proportions of right or wrong responses, e.g. solution frequencies, are used for scale construction. This gives only an arbitrary and no absolute zero, and therefore no ratio scale. Some fundamental concepts in such psychometric analyses are reliability, validity, and applicability. When an analysis is performed on data obtained from a ratio scaling method, the validity in a differential meaning is often found to be less than if a category-rating method has been used (see, e.g., Borg 1962, 1972; and Marks, Borg, and Ljunggren, 1983). This advantage with the rating scale seems to depend upon the fundamental (“absolute” and “intersubjective”) level sense of verbal expressions.

Category-ratio scales

A good review of the earlier literature on rating scales was given by Guilford (1936, 1954). He did not, as Stevens did, neglect the value of well-anchored ratings. He emphasized the importance of good “descriptive terms” and “landmarks,” and pointed out that “rating methods have overshadowed the applications of all the other evaluative methods based upon personal judgement” (p. 263, 1936). This was before the ratio scaling methods became popular. Stevens argued that no restrictions should be imposed on the subjects. He condemned category scaling and hoped it would fall into disuse. The rating methods never disappeared in spite of the efforts to make people stop using them. In his article from 1971, Stevens wrote: “Nevertheless, category scaling is occasionally used in psychophysical experiments, despite the demonstrated inefficacy of the procedure. Stevens and Galanter (1957) examined
some 70 different category scales on a dozen different perceptual continua and found little to justify the use of category scaling in quantitative studies. Their hope was that category scaling would thereafter fall into disuse. They hoped in vain.” (Stevens, 1971). It is thus evident that Stevens did not understand the value of well-anchored rating scales, and meant that the ongoing use of them mainly depended upon bad knowledge and ignorance. However, the reason for the continued use of these methods was, instead, their manifest usefulness for direct level determinations and interindividual comparisons (see, e.g., Borg, 1962, Ellermeier et. al., 1991).

The RPE-scale (the scale for Ratings of Perceived Exertion, Borg 1970, 1985) was, by special arrangements of verbal anchors, constructed to grow linearly with work load and heart rate for exercise on a bicycle ergometer. The construction of the RPE-scale is a unique example of scale construction, when the differential aspect is most important, and the demands on “semi-interval” properties (according to aerobic demands) are essential (see below, Figure 3). The senior author (G. Borg), who probably was the first to propose a “solution”, has strongly emphasized the need of a level-anchored ratio scale. In the end of the 1950s Borg started to use Stevens’ ratio scaling methods, and also simple, symmetrical rating scales, for studies of perceived exertion during heavy ergometer work. He also did studies in “quantitative semantics”, using ratio scaling to determine “interpretation” and “preciseness” of some verbal expressions for exertion. He found that the Swedish word “ansträngande” (hard; or in one first translation “laborious”) was used for a perception about three times stronger than when “lätt” (light) was used. “Rather hard” was less precise (higher coefficient of variation) than “very hard”. Borg then concluded that: “different quantitative linguistic expressions can be fitted into a subjective ratio scale so that exact mathematical relations between them are obtained” (p.55, Borg,1962). The interest from colleagues in psychophysics for this problem was, however, at this time very weak — to say the least. Because of this, it was not until 1973 that Borg presented the first “category-ratio” scale, consisting of numbers from 0, 1 to 20 (Borg, 1973b, 1977). See also below, Figure 3. — A similar approach to scaling has been utilized by Gracely for scaling pain (Gracely, Dubner, McGrath, and Heft, 1978), and later also by Green for scaling taste perception (Green, Shaffer, and Gilmore, 1993).

In verbally level-anchored ratio scales there are several main principles that have to be considered. In this article we will first present these main principles and then present methods developed accordingly. The value of the new generation of psychophysical scaling methods thus developed will then be supported by some empirical examples.

2. BASIC PRINCIPLES AND FOUNDATIONS BEHIND CR-SCALING

There are several fundamental principles that need to be considered. Some of these are common to all psychometric methods, including psychophysical scaling (e.g., definitions, instructions, etc.), and others are more specific to category-ratio scaling. This section will focus on the latter.
Definitions

The first principle is the necessity of good definitions concerning validity. Sometimes this problem is neglected. The observer may not be sure what to observe, e.g., if it mainly is the local, nociceptive pain or the suffering. For simple sensory perceptions, such as brightness, sourness, and loudness this is not a major problem. However, as soon as emotion, preference, and motivation are involved the importance of a good definition is more evident. Such is the case, for instance, if subjects are asked to estimate the degree of fear or anxiety. Also, for some rather simple perceptions such as noise, there may be an interpersonal disagreement depending upon the kind of noise, how annoying it is, or in what context it is presented. Most often, it is good to start with a constitutional definition that can be obtained directly from "common sense", dictionaries, and interviews. These definitions are further clarified by contextual definitions, in which the concepts and variables are defined in a relevant context with regard to the target situation. The methods developed will then lead to good operational definitions.

Target group

A good scaling method should be possible to use by most people. This means that about 90% of an ordinary population should be able to use it. "Most people" should not only mean people within a limited cultural setting, but people all over the world. It should be possible to use the basic scale construction, even if there may be difficulties with translations. The main principles of the scale should be universal, since they should be based on features common to all sensory functioning. The scale should preferably also be possible to use by people with special deviations and sensory deficits.

Subjects for scale construction

In the above paragraph, it was stated about the target group that a scale should be constructed for "most people". That the scale is intended for a big population does not mean, however, that it should be constructed with a sample of people drawn from this population. It has been shown by Borg and Borg (1992) that the age, education, and mental capacity of subjects influences the form of the S-R-function. Commonly a lower exponent is found for less educated people. It is our definite opinion that the scale should be constructed according to basic principles and results obtained from competent observers (in the meaning of, e.g., Quine and Stevens), i.e., people who have especially good possibilities to perceive and observe, to use mathematical concepts and verbal anchors correctly. When a good scale is then constructed, it is important to have an instruction and training method so that every one can use one and the same scale in the same way, giving a high degree of interpersonal agreement.

Norms

Since the scale is constructed using a selected group of "competent obser-
vers,” the obtained data are not directly representative of the target group. A sample of the target group has to be selected and trained to use the new scale correctly. After that, data can be obtained for norms.

Choice of scale type

In an absolute mathematical sense no rating method can yield the perfect properties of an interval scale (as a thermometer), or of a ratio scale (as a meter scale). However, it is sometimes quite evident that the magnitude of sensory perception does not grow linearly with stimulus intensity, because of physiological transductions, that in a meaningful way inhibit or magnify the stimulation. This may be related to the need for discrimination at certain levels of intensity, or for protection of the sensory system, or the human as a whole. Examples of this deviation from linearity is found for loudness (e.g., the dB-scale), and for perceived speed in car-driving (a positively accelerating S-R-function). The “semi-ratio scale” that can be obtained with a good “ratio scaling” method has, therefore, a great advantage over the ordinal rating scale. We can find an analogy to this, e.g., from cooking, and the use of “ratio values” in spite of the crude measures of number of potatoes or “spoons” and “cups” that are used in a recipe, or similarly from medical prescriptions. That we are using a “ratio scaling method” does not mean that the responses obtained with the method perfectly fits a mathematical ratio scale. We are always dealing with human responses and conceptions of numbers and not with abstract mathematical concepts. Responses obtained with a kind of “semi-ratio scale”, i.e., a scale, reflecting most peoples experiences, can then be handled statistically and mathematically as belonging to a ratio scale, adding to the degrees of explanatory power with regard to intersubjectivity.

Psychophysical scaling reveals a general and fundamental response pattern of the human sensory system. As stated by S. S. Stevens (1971) it is not probable that great individual differences should exist concerning these functions except for special cases, as in sensory deficits. According to him a “ratio scaling” method must not impose any restrictions on the subject, if responses are to be assessed on a ratio scale. This may have been a necessary condition in the first scale development, and the determinations of relative S-R-functions. After that knowledge had been obtained it was no longer a necessary condition. The knowledge should, however, be utilized as a foundation for further scale development. Of utmost importance then is that the scale is constructed so that there is a congruence in meaning between the anchors and the numbers.

The Range Model

Stevens was not interested in trying to scale “absolute” intensities or individual differences, nor was Ekman in Sweden. Ekman pointed out that we cannot say that a certain physical intensity is of the same subjective magnitude for different subjects, because we do not have an interindividually valid unit. This statement is difficult to deny, and may also concern intermodal comparisons (misleading figures
A new generation of scaling methods: level-anchored ratio scaling

are often seen in handbooks!). However, clinical applications show us the great need for reliable and valid scaling methods for natural ("absolute") identifications of intensity levels.

There is another statement by Quine, that is of interest in this concern: "The requirement of intersubjectivity is what makes science objective" (Quine, 1990). This statement emphasizes the importance of intersubjectivity, i.e., interpersonal agreements to make individual comparisons possible. A main requisite for intersubjectivity is that human beings perceive the world of objects and events in about the same way. We can never state, however, that we have perfect intersubjectivity. We can never perceive anything in exactly the same way as another person without being that person. We can, however, assume a good similarity. This idea may have an epistemic priority. People must perceive things in a similar way in order to understand each other. If they did not, good communication would be impossible. This may also imply that similarity has an epistemic priority over dissimilarity (Borg, 1992). People's differences can then only be understood as deviations from something that is similar.

It was, thus, a great challenge to try to come up with a theoretical model for interindividual comparisons. One day during the autumn of 1960, a hundred years after Fechner's "aha experience", the senior author also had an "aha experience". It was the following: Why not set as subjectively equal the perceptual intensities at individual maxima, i.e., at the highest tolerable exercise intensity for each individual according to physiological or performance measurements. The idea of "sameness" seemed so fundamental that immediately an experiment was done to test the idea. It was then found that predictions made according to the model worked very well (Borg, 1961, 1962). According to the model — the Range Model — the total subjective range from zero or an extremely weak intensity, to a maximal intensity, should be subjectively about the same for most individuals, in spite of the fact that the physical or performance range might vary very much. A perceptual intensity is, thus, evaluated according to its position in the range. The assumption of interpersonal similarity of the magnitude for a certain maximal intensity is especially important and can then be used as an interindividually valid point of reference, also in many intermodal comparisons.

The size of the subjective dynamic range

The physical dynamic range, i.e., the S-variation from an absolute threshold (S₀) to a maximal level (S), varies to a great extent between modalities, and to some degree also across subjects. The corresponding subjective dynamic range may also vary somewhat, but can, as a first order of approximation, be assumed to be the same, both across modalities and individuals (see Figure 2). In Borg's first presentation of the Range Model and its application on perceived exertion, "sameness" across subjects was assumed, but the size of the range was not estimated, since that was not a necessary condition for interindividual comparisons. It is, however, necessary for determinations of "absolute" levels and functions. According to Te Hagtsoonian (1971), the ranges across modalities are the same. This is not, however, a necessary assumption and it can be tested empirically (see Borg, 1972;
1998). From knowledge about the physical and subjective ranges, together with knowledge of the exponent, the measure constant of the psychophysical power function can also be estimated.

![Graph showing S-R functions for three theoretical modalities A, B, and C, with different physical dynamic ranges from the threshold \(S_0\) to a maximal level \(S_n\), and the exponents, \(n = \{0.5; 1.0; 2.0\}\).](image)

**Figure 2** — S-R functions for three theoretical modalities A, B, and C, with different physical dynamic ranges from the threshold \(S_0\) to a maximal level \(S_n\), and the exponents, \(n = \{0.5; 1.0; 2.0\}\).

From psychophysical studies and our own experiences, it seems probable that the total perceptual range is greater than 1:20 but typically not greater than 1:100. According to Teghtsoonian and Teghtsoonian, what they term the Range of Acceptable Stimulus Intensity (RASIN) — “the broadest range ... found acceptable and reasonably comfortable” — is 1.53 log units (Teghtsoonian and Teghtsoonian, 1997), i.e. 1:34. This may often function as a good correlate to the total range. For most sensory perceptions, it is possible to make about six doublings or halvings from “Min” to “Max”. This gives possibilities for a very correct discrimination of seven classes of intensities. However, it does not mean that with seven categories we only need the numbers 1 - 7 to describe the variation in magnitude. Instead we need many more numbers (N), according to the equation \(N = a^{-1}\), where \(a\) is 2 and \(c\) is the number of categories. In the case above with six doublings, we thus get a subjective dynamic range of 1:64. This is in accordance with a magic exponent of 6±1.

**Quantitative semantics**

With psychophysical methods the metric properties of verbal expressions (or other level labels, e.g., pictures or other symbols) may be determined on a “ratio level”. Such important semantic concepts as “interpretation” and “preciseness” may be defined operationally both with regard to quality and quantity. With regard to
quantity, "interpretation" is then, the intensity behind an expression (e.g., the mean), and "preciseness" the relative dispersion (e.g., the coefficient of variation, showing how much people agree). This new way of semantic studies with the use of ratio scaling methods, was first introduced and utilized by Borg on perceived exertion (Borg, 1962, 1964). In experiments by Borg and Hosman (1970), and Hosman and Borg (1970), the metric properties of several verbal expressions were determined. The subjective intensities of frequently used adjectives and adverbs in descriptions of subjective symptoms were also studied by Borg and Lindblad (1976). The finding was that adjectives and adverbs may function as multiplicative constants, and numerical relations between words can be used to identify intensity levels that are congruent with the numbers. Iterative trials are then needed to find the more exact positions.

**Congruence between numbers and anchors**

The bases for determinations of accurate S-R functions should be founded on the ratio scaling methods (Stevens, 1957, 1975), e.g., magnitude estimation. The special relation between a "symmetrical" category scale (rating scale) and a ratio scale (Stevens and Galanter, 1957) makes it possible to change the position of the verbal anchors of the category method so that the scales grow linearly to each other (see Figure 3). Using this method for scale construction will result in a "CR-scale" that gives approximately the same growth function as is found with magnitude estimation.

**Rating scale**

<table>
<thead>
<tr>
<th>Minimal</th>
<th>Weak</th>
<th>Moderate</th>
<th>Strong</th>
<th>Maximal</th>
</tr>
</thead>
</table>

**Converted scale**

- Linear
- "Semi-ratio"

**Stimulus intensity, S**

*Figure 3 — Knowledge of the relation between a verbally anchored category (rating) scale and a ratio scale can give the positions of the anchors on a ratio scale. An interval scale, such as the RPE-scale, may be constructed in a similar way.*

In an article a few years ago (Borg and Borg, 1987) a method to convert a category scale to a ratio scale was proposed and tested in some experiments. The transformation equation (regarding the category scale as a semi-log scale) worked
fairly well and may therefore be of special interest when constructing a CR scale. The main idea is that the distribution of the perceptual intensities roughly follows a “natural”, positively skewed distribution (log-normal) and that the verbal anchors should differ with a certain factor reflecting a geometric series. If scaling could be this simple, starting on a scale value of 1.5 for the first expression (to be able to use 100 as the upper reference point and in order not to get too big a range), and then using a factor of 2 and 7 expressions, would roughly result in scale values of: 1.5; 3; 6; 12; 25; 50; 100; with the verbal expressions from “Minimal” to “Maximal” being: “Very weak, Weak, Moderate, Strong, and Very Strong (Moderate at 12). Unfortunately, scaling does not seem to be that simple.

The general principle that the distances between the verbal anchors should increase according to a certain factor (F) may be formulated:

\[ F = \sqrt[2]{P} \]  

(2)

where \( K \) is the total number of verbal categories except those at zero and the absolute maximum (outside the fixed numbers), \( P \) is the perceptual range of a given numerical size. The scale value \( (v) \) for each expression, using the factor \( (F) \) described above, is then:

\[ v = (\sqrt[2]{P})^k \]  

(3)

where \( k \) is the number of the specific expression.

For a starting value other than \( v \) for the first expression, \( k=1 \), (e.g., if 1.5 is chosen as suggested above) the scale values (on the R-scale) has to be adjusted accordingly, e.g. as:

\[ v' = a \cdot (\sqrt[2]{\frac{v}{a}})^{k-1} \]  

(4)

where \( a \) is the lowest, starting value. This may be arbitrarily chosen, or selected to get numbers that are familiar to most people, and easy to use. The first verbally anchored number (except 0) may, e.g., be 0.5; 1.0; 2.0; or as in the example above 1.5.

**Avoiding end effects**

In physiology it is known that a response is sometimes higher than expected or thought possible (“supermaximal”). A subject may sometimes want to give a response that is higher than he or she has used before (or anticipated to ever use), corresponding to a higher perception (e.g., in pain) than ever previously experienced, a response outside the given range. A problem with many scales, e.g., the VAS, is a tendency for truncation of responses, especially at the top of the scale (a ceiling effect). To avoid this, one can leave the scale “open” at the top (and also at the bottom) and, e.g., mark the absolute maximum with a dot, permitting a subject to use any, reasonable, higher number. What is reasonable must then be handled carefully by the test leader. If, e.g., “10” is anchored as “Maximal” (see the CR10 scale), a value higher than 12 is seldom reasonable for perceived exertion, unless the activity causes special aches and pain. An extreme pain is always the most intense experience and may in this example be rated higher, e.g., 15.
Direct method

A great advantage with a category-ratio method is that it is a "direct" method, in the meaning that it can give data without any need for special "calibration" beforehand or afterwards. When the stimulus gives a direct impression that is easy to identify, the method should enable the person to respond directly. It should also be easy to directly understand and interpret the response.

Psychophysiological foundations

An important consideration comes from psychophysiology. A neuro-psychophysiological study on human beings in the area of taste perception was conducted by Borg, Diamant, Ström, and Zotterman (1967). This showed remarkably high correlation between psychophysical judgements and neural responses in taste perception. This may be regarded as a validation of the ratio scaling method. From the area of perceived exertion come further psychophysiological foundations. It is easy to find valid and easily measured physiological variables, e.g., heart rate (HR) and blood lactate ([La]), that can be used as criteria when identifying "absolute" levels of intensity and individual differences (See Borg, 1998). When support for the extra constants (a and b) in the psychophysical power law can be found in the underlying physiology there is good reason to believe that they may also be found in the psychophysical function. This may also be the case in persons with sensory deviations.

Iterative trials

When a scale has been constructed from the above principles, there is still a need for it to be tested and further adjusted by iterative trials. To construct a perfect scale that fits all modalities, individuals and situations are, of course, impossible. There will always be some "conflict" between the intensity that the numbers will give and that which is given by the verbal anchors for a special modality in a certain situation. For instance the "absolute" position of a "maximal" loudness is not the same as that of a "maximal" sweetness (see Borg, 1992). The relative position of different expressions within the range may differ somewhat between modalities. Still, it should be possible to construct a scale that is the best for most "situations," and in which the above mentioned "conflict" is minimized.

One specific anchor, a "fixed star"

It is important that we have many different linguistic terms, numbers and analogies interconnected in a big network that forms our judgmental system. This does not mean, however, that in psychophysical scaling all different expressions can be used as "units". They help us as "landmarks" and "anchors" to make direct judgments, but we also need a certain expression at a certain level as a main point of reference. This point does not have to be a "maximal intensity", but it should be such an intensity and anchor that most people agree upon its position (e.g., not
“Strongest imaginable”). If we do not use such a “fixed star” in scaling we will get many different reference points that all have a kind of direct connection to our perceptions and memories, and will form a judgmental scale that is somewhat floating around. Therefore, we need one main ”constant” level that we can use as a “fixed” standard and main reference point in order to compare and ”calibrate” the other expressions against.

The visual design

The scale values and their verbal expressions may be presented in many different ways. If they are spread out linearly with equal distances between the numbers the lowest expressions will visually be very close and the highest expressions very far from each other. If the expressions are placed logarithmically, they will be parted by the same visual distance. A good design may be a compromise between these two extremes. This can be accomplished by increasing the equal distances between the expressions, when placed on a log-scale (log v) according to a power function, where the measure constant will give the size of the scale and the exponent the increase of the distances between the expressions. Sometimes it might also be tempting to incorporate some background cues (e.g., colors) to denote the different levels of the scale. This can, however, not be done without first checking the congruence between the chosen cues and the numbers, as well as the verbal labels.

The form of the S-R-function

All demands should result in an S-R-function, that as well as possible reflects the “true” sensory functioning “behind” the responses. There are several criteria for these: internal psychophysical, psychophysiological, performance, etc. In this concern we will not discuss this matter any further, as it is a subject of its own.

Two-way communication

A good scaling method should not only be possible to use as an exploratory instrument, to have a person estimate his/her symptoms. It should also be possible to use for prescriptions, so that a patient can monitor just the right intensity of exercise or medication by producing the desired intensity. This possibility is not easy to get to function with the popular VAS method, but is much easier with the CR10 scale.

Some common psychometric demands

The demands on reliability are similar to those for most psychometric methods, except for one important difference, viz. the kind of S-R-functions, e.g., the exponents. The error variation that depends upon individual differences in rating behavior should be minimized. In psychophysical scaling, validity deals with the extent to which the responses obtained with a scale are linearly related to the fundamental internal psychological process of interest (see Gescheider, 1997). High validity is
promoted by a good definition, good instruction, and administration (see Borg, 1998). Sometimes one can also have the advantage of using physiological validity criteria, as, e.g., in the areas of pain and exertion.

3. THE CR10 SCALE

The CR10 scale is an improvement of the first CR scale, presented above (Borg, 1973b; see Borg, 1977). It consists of a variation in numbers from 0 to 10, with the first anchor at 0.5, and with possibilities to give ratings with decimals and also ratings below 0.5 and above 10 (to avoid end effects). The range from 0.5 to 10 is a rather small intensity range, but for practical purposes it is often good enough. In many studies, the validity of the CR-10 scale has been shown. (Borg, Ljunggren, and Marks, 1985; Neely, Ljunggren, Sylvén, and Borg, 1992; Noble, Borg, Jacobs, Ceci, and Kaiser, 1983; see also Borg and Ottosson, 1986). It is a general scale for measuring intensities of most kinds of sensory perceptions, experiences, and feelings. Its primary use has, however, been in clinical diagnosis of aches and pain; in determination of perceived exertion including breathlessness and fatigue in connection with work tests, training, and rehabilitation. The main fields of application are found in medicine, ergonomy and human factors, and in sports (see Borg, 1998).

```
0  Nothing at all       "No I"
0.3
0.5  Extremely weak      Just noticeable
0.7
1  Very weak
1.5
2  Weak                  Light
2.5
3  Moderate
4
5  Strong                Heavy
6
7  Very strong
8
9
10  Extremely strong     "Strongest I"
11

* Absolute maximum      Highest possible
```

Figure 4 — The Borg CR10 scale (© Gunnar Borg, 1982, 1998).
To increase the number range and to make it easier to discriminate at low intensity levels and to avoid a floor effect, other category-ratio scales have been constructed by Borg. One of the scales (the CR12) had a number range 1.7 times greater than that of the CR-10 scale and the other (CR20), a 2.9 times greater range. The numbers of the latter scales were spaced logarithmically, whereas the verbal labels were spaced nearly linearly. The validity of the scales was tested against some physiological criteria (Marks et al., 1983; Borg, Hassmén, and Lagerström, 1987). The psychophysical functions obtained with these scales were good (see also Sebald, 1990), but they were judged to be more difficult to use, for ordinary people, than the CR10 scale, which now has become a rather commonly used scale.

4. THE CR100 SCALE

The CR10 scale is a rather rough scale mainly for practical use. Sometimes, however, there is a need for a more fine-graded scale with a number range that comes closer to what is obtained with ME, or in better congruence with the size of the subjective dynamic range. Previous studies comparing CR scaling with free ME show two important things: First, that CR scaling functions as well as ME in generating general psychophysical functions (as shown by, e.g., the exponents). Second, while CR scaling gives comparable levels of intensity with good possibilities for interindividual comparisons, “free ME” does not (see, e.g., Borg, Ljunggren, and Marks, 1985; Marks, Borg, and Ljunggren, 1983; Neely, Ljunggren, Sylvén, and Borg, 1992). For an example of this, see the results obtained with loudness and blackness presented in Figures 5 and 6.

![Graphs showing loudness and CR functions](image)

**Figure 5** — Loudness functions for two different experiments. To the left: ME-functions showing large differences in levels. To the right: CR-functions showing good agreements in levels.
A new generation of scaling methods: level-anchored ratio scaling

Figure 6 — Blackness functions for four different experiments. To the left: ME-functions showing large differences in levels. To the right: CR-functions showing good agreements in levels.

A CR scale constructed by the principles in section 2, and based on Equation 3, gave exponents that were slightly high compared to ME, and not altogether linear in log-log. It was thus assumed that the distances between the verbal labels do not follow such a simple increase, but that the factor decreases. This does not have to be in conflict with the principle of equal response ratios, since this principle does not have to apply to the scaling method itself, but to the responses obtained with the method. Such a decrease could be obtained by progressively diminishing the factor F in Equation 2 so that the scale value \( v \) for each expression could be described by:

\[
v = f^k (1 - p)^{\frac{1-k}{2}}
\]  

(5)

where \( f \) is the starting value for the first expression (or the factor for the first level), \( p \) is the proportional decrease of the multiplicative factor, and \( k \) is the number of the specific expression. The value for \( p \) (\( p = 0.0825 \)) was obtained from the Equation 4 with \( f = 2.5 \), \( v = 100 \) and the total number of verbal expressions, \( K = 7 \). The resulting CR100 scale gave rather high exponents compared to ME (Borg and Borg, 1994). The subjects also found the scale a little difficult to use as none of the verbal labels were placed at whole numbers on the scale. This is a drawback when the scale is used for a two-way communication and intensity prescriptions. The scale was further adjusted by moving the verbal labels to the closest rounded numbers. A "minimal" and a "maximal" level were also estimated and added to the scale.
A number range of about 1:60 may be suitable with regard to the total range of perceptual magnitudes from "minimum" to "maximal". However, for most people it is not a common number range that is easy to deal with. A scale variation of 1:100 with one (1) as a lowest unit or starting point, as in percentage estimates, is much more natural. We encounter, however, a difficulty when we want to combine such a simple number range and make it congruent to the actual possible perceptual range, since that is often somewhat smaller. After many trials and tests the solution was to choose one and a half (1.5) as a starting value for "Minimum" giving a suitable number range of 1:67. "Minimum" on the scale approximates the absolute threshold, and "Maximal-Max X" is anchored in a previously experienced strongest intensity. As the absolute maximum can be at a level above this, and in order to avoid a ceiling effect, the "Absolute maximum" was placed with a dot outside the given numerical scale. To avoid a floor effect it should also be possible to go below 1.5 (see Figure 7). "Maximal-Max X", is the main point of reference and all other scale values are set in comparison to this. The responses are thus experienced in centigrade of this "Max" value as a "unit". The scale may thus also be called the "centi-Max" scale and the values cM (centi-Max) values. In conformity with this the CR10 scale may be called a "deci-Max" scale and the values dM (deci-Max) values. This might be preferred to the abbreviations "CR" (for Category-Ratio) because of the risk of confusion: CR must not be taken as an abbreviation for Category Rating. Since we have many names for the things we love, we have sometimes also called a CR-scale a "LIME-scale", meaning a "Level-anchored Intersubjective ME-scale".

Figure 7 — To the left of the scale there are some triangles "hidden". They should remain so (belong to an older version of the scale).
5. WHEN USING A CATEGORY RATIO SCALE

Even if the category-ratio scaling method is a "direct" method, this does not mean one can neglect proper instructions. It enhances the reliability if the instruction and administration proposed by Borg (1998) is followed as closely as possible. By using a training material (e.g., the blackness material, Borg and Borg, 1991) reliability is most probably further strengthened. The test leader must be certain that the individual understands the instructions by testing the subjects with some example questions about the perceptual intensity of certain things, such as the blackness of velvet (approx. "9"); the sourness of a lemon (approx. "6"); the sweetness of a ripe banana (approx. "3.5"); or the loudness of normal conversation (approx. "2.5"). They must also avoid using questions that might interfere with the purpose of the test. It is very important that the test leader prepares himself well and arranges the situation so the person to be tested feels comfortable without any disturbing elements. A good rapport with the person should be established, and all necessary information about the test and the purpose of it is given well in advance, permitting questions to be asked.

There is sometimes a need to use a special reference modality, to make possible valid interindividual (and intermodal) comparisons, e.g., when studying perception of taste or somatic symptoms. The criteria for the choice of such a reference modality are that it should be a simple, well-defined modality, easy to define and use, stable over time, with good interpersonal agreement, and it should also be valid to use in the form of imagined magnitudes, such as conceptions. Perceived exertion and heaviness is one such modality, since it meets the above requirements, and since a voluntary "maximal" intensity is not harmful to healthy people. Most persons have also experienced an extremely strong exertion that gives a well-defined "maximum" (or "peak value") and fits a natural subjective range.

The special properties of the CR scales are a great advantage when working on obtained responses and interpreting results. The construction makes it possible to use parametric statistics (descriptive and inferential, e.g., ANOVA) including determinations of mathematical S-R-functions.

6. SOME SPECIAL EXPERIMENTAL STUDIES

There is much support for the superiority of CR-scaling over other scaling methods. In studies of the fundamental principles presented above experiments have been performed for validation. References to these studies are found under each subheading. Above these we will also present the results of a few more specific experiments elucidating the following issues: the advantage of level anchoring; the importance of where verbal expressions are placed on the scale and; the impact of the visual design, and a summary of a comparison between AME and CR100 in tests of perceived exertion.

To be able to describe as correctly as possible the form of the S-R-function is of fundamental importance in order to increase the knowledge about sensory functioning. It is also important in practical applications, when we want to predict a
perceptual intensity from physiological functions, or when we want to predict physiological responses and performances from perceptions. For determinations of S-R-functions it is necessary to have a reliable and valid scaling methodology, not only a simple ordinal scaling method, but preferably a kind of ratio scaling method, permitting results to be described on a “semi-ratio” scale. In several previous studies, it has been shown that ratings of perceived exertion according to the RPE scale give possibilities for very good predictions of maximal heart rates from the sub-maximal relation between heart rate and RPE values. The RPE scale may then be regarded to be an interval scale (see Borg, 1998). Using scales with good metric properties gives possibilities to make many valuable estimations and predictions, above what can be obtained from simple, ordinal scales.

An experiment on the advantage of level anchored ratio scaling

A simple form of anchoring is the use of a conception of a certain intensity as a reference point. In a study by Borg (1972) the conception of maximal perceived exertion was used as such a reference, when rating perceived exertion for work on a bicycle ergometer.

A group of twenty-eight, 20-year old students were used as subjects. All had to start on a low exercise intensity, 12.5 W, and worked for 4 minutes. The workload was increased stepwise to 50, 100, 150 W etc., with 50 W increase every fourth minute, until the subjects could not go on exercising, i.e. to a voluntary maximum. Heart rate was recorded towards the end of each workload and subjects rated the perceived exertion according to the RPE scale and according to percentage ratings, i.e., in percentage of the conceived maximal intensity as 100. High correlations were obtained between the two rating methods, and also with heart rates, and there were good possibilities to predict maximal performances from sub-maximal responses.

In this article, we have focused on the percentage ratings, to reanalyze data and see what the possibilities are to predict maximal performances from very low percentage ratings. The group of subjects were divided into three subgroups, one with a low working capacity (group A), one with a normal working capacity (group B), and one with a high working capacity (group C). These groups differed with about 50 W, with regard to final working capacity according to their maximal performances. From high ratings, it is not difficult to predict a maximum performance, but from ratings below 50% the difficulties are greater. To be able to do any valid predictions we have to know the “true” S-R-function. From previous studies of this kind of work (see Borg, 1962, 1998), the exponent is about 1.5-1.6. This exponent is rather sensitive for the choice of the basic noise value (a-value). An exponent of 1.5 is very reasonable for this kind of work. If we do not know what the S-R-function will be we can thus use this previously determined function.

A low workload was chosen, 100 W, that all subjects could finish without problems, and that definitely was below 50% of the estimated maximal exertion. In Figure 8 we can see the results from these groups and predictions to 100% perceived exertion according to the exponent 1.5, and also according to an exponent of 1.0. The latter was chosen as an example of an S-R relation that may be chosen,
if no previous knowledge is accessible, or if no valid S-R-function can be obtained because of the bad metric qualities of the scaling procedure. In Table 1 the resulting predictions are shown in comparison to the obtained working capacity (estimated as the mean of the last level that the subjects could work on and the level that they could not fully manage).

![Graph](image)

**Figure 8** — Predictions of maximal working capacity for three groups of subjects (A, B, and C) from percentage ratings (of a conceived maximal perceived exertion) at an exercise level below 50% (100 W), with an exponent of 1.5 (full lines), and an exponent of 1.0 (dashed lines).

**Table 1** — Three groups of subjects with different working capacity: low (group A); normal (group B); and high (group C), their percentage ratings (R) at 100 W, predicted working capacity (Pred. Wmax) at 100% perceived exertion with two different exponents (n = 1.0 and n = 1.5), and their obtained working capacity (Ob. Wmax).

<table>
<thead>
<tr>
<th>Group</th>
<th>R(100 W)</th>
<th>Pred. $W_{\text{max}}$ n = 1.0</th>
<th>Pred. $W_{\text{max}}$ n = 1.5</th>
<th>Ob. $W_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>46.5</td>
<td>215</td>
<td>167</td>
<td>175</td>
</tr>
<tr>
<td>B</td>
<td>32.5</td>
<td>398</td>
<td>212</td>
<td>225</td>
</tr>
<tr>
<td>C</td>
<td>24.0</td>
<td>417</td>
<td>259</td>
<td>275</td>
</tr>
</tbody>
</table>

As can be seen from the figure and the table, the method works very well and very valid predictions were possible to make when the exponent of 1.5 was used. This is good evidence of the validity of the scaling methodology.

**A new experiment on the importance of the placement of the verbal expressions**

The shape of the S-R-function (i.e., the exponent) is not only affected by the range of numbers of the scale but also by the placements of the verbal expressions.
The purpose of this experiment was to show that the spacing of the verbal expressions on a numerical scale will affect the exponent of a psychophysical function. Starting from Borg's CR10 scale (Scale 1), the verbal expressions from "Extremely weak" to "Very strong" were moved so that three new scales were obtained: Scale 2 with equal spacing between the expressions (as on a traditional category rating scale); Scale 3 with the verbal expressions clustered towards the top of the scale; and Scale 4 with them clustered at the bottom of the scale. The expression "Extremely strong" was kept at the number "10" on all scales to serve as a special reference point.

If the numerical values for all expressions are plotted with CR10 as the independent variable the relations between scales shown in Figure 9 is found. A hypothesis of this experiment is that results obtained with the scales should reflect these relations. As can be understood from the figure, the second and third scale can be expected to give lower exponents (and b-values in the power function) when compared to the CR10, and the fourth scale can be expected to give an exponent that is higher than that for the CR10 scale (and an a-value).

![Figure 9](image)

**Figure 9** — Theoretical relationship between the four scales, with CR10 as the reference.

In order to study the relationship between these scales it was important to find an "intensive" modality, easy to use, and with "absolute" perceptual magnitude (i.e., not an "extensive" one). A modality that can meet with these conditions is perceived blackness (Borg and Borg, 1991). Ten subjects, without previous experience of psychophysical scaling, rated the degree of blackness. The stimulus levels used were $S = \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$ percent blackness according to the Swedish Natural Color System (NCS) (Hård, 1966). The test material consisted
of 5 x 5 cm squares of the stimulus pasted to A4 sheets of white paper and were presented in a randomized order. The psychophysical method chosen was the CR10 scale and three manipulated versions of the scale presented above.

The psychophysical function obtained for the CR10 scale (based on geometric means, and calculated by the method of least squares), shown in Table 2, with n = 0.94, is in agreement with what was previously found by Borg and Borg (1991). There was also a need for a small a-value, giving an exponent of n = 1.04. An a-value for blackness has previously been observed with ME as well as with the CR10 scale by Neely and Borg (1995).

The second scale gave an exponent of n = 0.70, without any need for an a- or b-value. This is in agreement with the results for blackness obtained by Neely and Borg (1995) who, for an ordinary category rating scale with 7 expressions, obtained an exponent of n = 0.66 (without an a-value) and n = 0.71 with a small a-value of 0.2. The third scale gave an exponent of n = 0.65, and with a need for a b-value of b = 3.8 giving the exponent, n = 0.55. As a contrast to this, the fourth scale, without extra constants, gave an exponent of n = 0.98. However, there was a need for a rather high a-value of a = 0.60 giving an exponent of n = 1.60.

**Table 2 — Obtained psychophysical functions for the four differently spaced scales**

<table>
<thead>
<tr>
<th>Scale No.</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(CR10)</td>
<td>0.23</td>
<td>0.109</td>
<td>0.94</td>
<td>.978</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.068</td>
<td>1.04</td>
<td>.979</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.333</td>
<td>0.70</td>
<td>.980</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.508</td>
<td>0.65</td>
<td>.991</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td></td>
<td>0.794</td>
<td>0.55</td>
<td>.995</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.065</td>
<td>0.98</td>
<td>.908</td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td></td>
<td>0.004</td>
<td>1.60</td>
<td>.966</td>
<td></td>
</tr>
</tbody>
</table>

The data was also plotted with the results from the CR10 scale on the x-axis. As can be seen from Figure 10 the obtained levels and shapes of the curves agree very well with what can be predicted from Figure 9. The results from this experiment show that the position of the verbal labels influences the shape of the psychophysical functions (both exponents and extra constants).

**An experiment on the importance of the visual design**

The scale values and their verbal expressions may yet be presented in many different ways. If they are spread out linearly with equal distances between the numbers the lowest expressions will visually be very close and the highest expressions visually very far from each other. If the expressions are placed logarithmically they will be parted by exactly the same visual distance.
Figure 10 — Blackness ratings for 10 subjects with four scales and ratings with the Borg CR10 scale on the X-axis.

In a study by Borg and Borg (1994) a CR scale was constructed after the principles described above in section 2, and based on Equation 3 with P=60 and K=8. The chosen verbal expressions and their corresponding scale values are given in Figure 11. Two visually different versions were constructed: one with the numbers visually linearly spaced (N); and one with the verbal labels with the same visual distance (L). If the visual design of the scale lacks importance one would expect no significant difference between responses on a linear or logarithmic version of the scale.

Figure 11 — The scale values for the verbal expressions of the CR60 scale used in the experiment.

<table>
<thead>
<tr>
<th>Expression</th>
<th>Scale value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nothing at all</td>
<td>0</td>
</tr>
<tr>
<td>Extremely weak</td>
<td>1.7</td>
</tr>
<tr>
<td>Very weak</td>
<td>2.8</td>
</tr>
<tr>
<td>Weak</td>
<td>4.6</td>
</tr>
<tr>
<td>Somewhat weak</td>
<td>7.7</td>
</tr>
<tr>
<td>Somewhat strong</td>
<td>13</td>
</tr>
<tr>
<td>Strong</td>
<td>22</td>
</tr>
<tr>
<td>Very strong</td>
<td>36</td>
</tr>
<tr>
<td>Extremely strong</td>
<td>60</td>
</tr>
<tr>
<td>Maximal</td>
<td>*</td>
</tr>
</tbody>
</table>
A new generation of scaling methods: level-anchored ratio scaling

A group of twelve persons, eight men and four women, served as subjects. The majority of the group had had some previous experience of psychophysical experiments. The modalities used as test material was blackness (as described above) and loudness. For loudness, a 1000 Hz tone was used with 7 stimuli from 40-100 dB SPL. Before the experiment started, two reference stimuli of 60 and 80 dB SPL were presented to the subject. Each tone was presented to the subject for 1 s. Stimuli were presented twice to each subject, in an individually randomized order, offered one at a time by the experimenter. The subjects were not allowed to ask for a repetition of a previous stimulus. The CR-scale and the written instructions were visible in front of the subject during the whole part of each test. Data have now been reanalyzed with focus not only on psychophysical functions, but also on differences between the two scales concerning level responses.

The results from the experiment are shown in Figure 12. For blackness there was a significant difference between the two scales for the first seven stimulus levels (10 - 70 %blackness NCS), $p = 0.05$, paired comparisons. For loudness, a significant difference was only found for the two lowest stimuli (40 and 50 dB).

Figure 12 — The results in log units from ratings of a) blakness and b) loudness with two visually different scales (N=12)

The psychophysical functions obtained with the two scales are shown in Table 3 (based on geometric means, and calculated by the method of least squares). For blackness, the exponent was about 1.1 for both scales and for loudness the exponent was close to 0.4 for both scales. For both scales and modalities there was, however, a need for a rather high $a$-value (of the same magnitude as for the fourth scale in the experiment above). This was regarded to be an artifact of the scale construction and the factor chosen in Equation 3.

It may be reasonable to believe that not only the words and numbers influence the subject to choose a certain answer, but also the visual “area” around an expression. From this experiment it is not very obvious how strong this influence is, but at least on the lower levels it had an effect on both modalities. This is, at least, a reason to be cautious and not change the visual design without previously testing it experimentally.
Table 3 — Psychophysical functions for two extreme scales of different design: 1) The numbers (N) linearly spaced; and 2) the verbal labels (L) linearly spaced.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Scale</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blackness</td>
<td>N</td>
<td>0.220</td>
<td>1.10</td>
<td>.972</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>0.023</td>
<td>1.61</td>
<td>.989</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>0.149</td>
<td>1.14</td>
<td>.959</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>0.0093</td>
<td>1.78</td>
<td>.989</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudness</td>
<td>N</td>
<td>25.6</td>
<td>0.37</td>
<td>.994</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>25.6</td>
<td>0.46</td>
<td>.998</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>25.1</td>
<td>0.39</td>
<td>.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>25.1</td>
<td>0.53</td>
<td>.999</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. **A SUMMARY OF STUDIES SUPPORTING CR-SCALING**

One of the main disadvantages with Stevens' "ratio scaling", and "no restrictions" of responses, is that it is not possible to obtain relevant intensity levels. This disadvantage is overcome in CR-scaling, since verbal anchors are placed in congruence with numbers on a ratio scale that covers a biologically and interindividually natural range of intensities. There are many studies supporting the value of CR-scaling. In this article we can only very briefly refer to some of these.

In the first experiment, above, on perceived exertion we showed that percentage estimations can function well to predict maximal performances. A drawback with such ratings is, however, that some people have difficulties in using percentages. A scale that is easy to use is the RPE scale, developed to facilitate ratings (R) of perceived (P) exertion (E) for most people. It functions very well as an interval scale, as shown in many studies (see Borg, 1998). CR-scaling can function for determinations of most kinds of experiences, including perceived exertion. It has the advantage over the RPE scale that it gives a more "true" picture of the real S-R-function, and thus also of a "dangerous strain" in risk assessments. The superiority of CR-scaling to VAS was shown by Neely et al. (1992), and to AME by Borg and Borg (1998). The "overall perceived exertion" is often complemented by determinations of fatigue, pain, breathlessness and dyspnea. The CR10 scale is now rather commonly used in assessments of pain and other somatic symptoms. A new field of application is in resistance training, both in athletes and patients, in order to monitor just the right exercise intensity (see Borg, 1998).

Since CR-scaling is based on fundamental principles to meet the demands of scaling most kinds of experiences, it has been applied in studies of common sensory perceptions, such as loudness and noise, colors, and taste. In a study on perceived bitterness CR-scaling was used by Marks, Borg and Westerlund (1992). Since some people are "non-tasters", and some also "super-tasters", perception of bitterness is a difficult area for diagnostic assessments (as recently stressed by
Bartoshuk'). By anchoring "maximum" on the CR10 scale in a conception of a maximal perceived exertion (for the sake of intersubjectivity) it was possible to differentiate "non-tasters" from ordinary "tasters". — CR-scaling is also used for the ratings of moods and emotions, and in "human factors" concerning perceived difficulty, e.g., the easiness to perceive symbols and follow instructions, or to estimate speech quality in telecommunication (Möller, 2000).

In most of these studies, the CR10 scale has been used, since in many situations it covers a sufficient range, and is judged to be rather easy to use. Sometimes, however, a more fine-graded scale is needed and then the CR100 scale — the "centiMax scale" (with responses expressed as cM-values) — may be preferable to use.

8. GENERAL DISCUSSION AND CONCLUSIONS

The verbally level-anchored ratio scaling methods that are presented in this article belong to a new generation of scaling methods that differ in several fundamental aspects from previous methods. The combination of "ratio scaling" and "category level anchoring" makes these methods very special, both with regard to theoretical principles and applications. There are several fundamental principles behind the new methods, of which it is especially important to stress the importance of "ratio scaling," and the verbal level anchoring, according to quantitative semantics and iterative trials, using "competent observers". The Range Model and the determinations of the size of the subjective dynamic range are other fundamental principles that are used, not only for corrections of obtained responses, but for constructions of the mere scales. For direct scaling of perceptual magnitudes, we consider it important that the fundamental principles are already integrated into the scale so that a direct response and interpretation of it is possible. There is, thus, a strong demand for congruence between the meaning of numbers and the verbal anchors. To this comes the need for one special reference point as a "fixed star" for evaluations and placements of all other landmarks and reference points. The perception of exertion and heaviness is used as such an important general point of reference that it can be used for intersubjective and intermodal purposes. Still it should be possible for each subject to give responses outside the given range on the verbally level-anchored scale in situations where an unexpectedly strong intensity is perceived. This helps to avoid truncation effects.

There is rather much empirical evidence from laboratory and field experiments showing the validity of the new scales. In the history of psychophysical scaling, this new methodology is the last in a four generation methodological development. The first psychophysical methods that were already used several hundred years ago, are the simple rating scales, or category rating scales. These are generally treated as scales giving responses belonging to a scale of ordinal character. Fechner never proposed any new scaling methods, and it may be possible that his idea about the psychophysical log-function was based on Weber's law together with a simple

---

1 see article by S. Carpenter in the Monitor, October 2000
category scale of five to seven categories. If the perceptual responses are placed according to such a scale with a kind of equal distance between them, the result of a logarithmic function becomes very probable. According to the principles we have presented in this article, this leads to the possible conclusion that a simple five- (or seven-) graded rating scale with verbal anchors placed symmetrically around a “mean” or “moderate” anchor, comes close to a logarithmic scale. Since a logarithmic scale is the inverse of an exponential scale, this should mean that by stretching out the category scale with a certain factor, a linear relation would be obtained between stimulus and responses. Some support of this has been found by Borg and Borg (1987).

To the second generation of scaling methods belong equisection scaling, giving some kind of “semi-interval” scales. Bisection is the simplest form of such methods, since the subject only has to try to set a stimulus intensity so it is perceived to be just halfway between a certain low and high intensity. The visual analogue scale (VAS) is classified by some researchers as an interval scale (and sometimes as a ratio scale). These kinds of methods do not have as good psychometric properties as ratio scaling. To the third generation of scaling methods belong Stevens’ ratio scaling. These methods are very fundamental for determinations of relative S-R-functions, but as we have argued and shown in this article, they do not give possibilities for any absolute determinations of levels of intensity.

The last, fourth, and now final generation of scaling methods are the verbally level-anchored ratio scales, also called CR scales, in which several fundamental principles are used and incorporated in the construction, so they can work as “semi-ratio” scales, permitting most kinds of descriptions and increasing the explanatory power. They can make most kinds of direct applications in medicine, sports, and ergonomy possible. It is our hope that these methods will overshadow all other kinds of scaling methods, and we hope that we do not hope in vain.

RÉSUMÉ:

On présente ici des principes, des méthodes et des expériences concernant l’échelonnement de rapport avec des niveaux d’ancrage verbaux. En tant que grands principes, on nommera: les méthodes d’échelonnement de rapport de Stevens; le Modèle de Variation d’après Borg; les déterminations d’amplitude de l’étendue dynamique subjective; la sémantique quantitative pour la sélection et l’emplacement des ancrages verbales; l’évitement des effets de marge; les validations psychophysologiques; une ancre toute spéciale, comme une “étoile fixe”, fondamentale pour l’intersubjectivité et pour la communication bilatérale. L’échelle qui en résulte est une échelle de rapport avec des niveaux d’ancrage, recouvrant une étendue assez naturelle d’intensités et dans laquelle les nombres et les ancrages verbales sont rendus congruents. Au sein du développement d’échelles psychométriques, elle représente la toute dernière génération envisageant l’échelonnement des sensations, perceptions, expériences et émotions. Elle a trouvé du support empirique dans plusieurs études réalisées aussi bien en laboratoire qu’au dehors. En plus, elle possède un champ d’applications assez large, portant sur le diagnostique clinique, l’entrainement e la rehabilitation, les sports et l’ergonomie.

A new generation of scaling methods: level-anchored ratio scaling

ABSTRACT

Principles, methods, and experiments in verbally level-anchored ratio scaling, also called CR-scaling is presented. To the main principles belong: Stevens' "ratio scaling" methods; the Range Model according to Borg; determinations of the size of the subjective dynamic range; quantitative semantics for the selection and placement of verbal anchors; the avoidance of end effects; psychophysiological validations; a special anchor, as a "fixed star", for intersubjectivity; and possibilities for a two-way communication. The final scale is a level-anchored ratio scale, that covers a natural range of intensities, and where there is a congruence in meaning between the numbers and the verbal anchors. In the psychometric scale development, this is the last generation of methods for scaling sensory perceptions, experiences, and emotions. Empirical support has been obtained in several laboratory and field studies. There is a vast application in clinical diagnostics, in training and rehabilitation, in sports, and ergonomics.

KEYWORDS: Psychophysical Scaling, Category-Ratio, CR Scale, Perceived Exertion, Intersubjectivity, Symptom Magnitude

REFERENCES


A new generation of scaling methods: level-anchored ratio scaling


ERRATA

Por lapso, não foi mencionado o nome da Prof. Doutora Luisa de Almeida Morgado como autora da recensão crítica da obra "Working with Piaget. Essays in honour of Bärbel Inhelder".

In the article by G. Borg, «A new generation of scaling methods: level-anchored ratio scaling» p.35 – table 1, the Ob. $W_{max}$ of group C should be 275 instead of 227. The correct table should read as follows

**Table 1**

<table>
<thead>
<tr>
<th>Group</th>
<th>R(100 W)</th>
<th>Pred. $W_{max}$ n = 1.0</th>
<th>Pred. $W_{max}$ n = 1.5</th>
<th>Ob. $W_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>46.5</td>
<td>215</td>
<td>167</td>
<td>175</td>
</tr>
<tr>
<td>B</td>
<td>32.5</td>
<td>398</td>
<td>212</td>
<td>225</td>
</tr>
<tr>
<td>C</td>
<td>24.0</td>
<td>417</td>
<td>259</td>
<td>275</td>
</tr>
</tbody>
</table>