Optimal metric for condition rating of existing buildings: is five the right number?

Félix Ruiz[a1], Antonio Aguado[a], Carles Serrat[b] and Joan R. Casas[a]
[a]Barcelona School of Civil Engineering (Department of Civil and Environmental Engineering),
[b]Barcelona School of Building Construction (Institute of Statistics and Mathematics Applied to the Building Construction),
Technical University of Catalonia-BarcelonaTECH

Abstract

In the context of the built environment in the recent years the concept of maintenance has changed from corrective to preventive maintenance. There is evidence that preventive maintenance is much more efficient than corrective maintenance, since severe deteriorations that may represent danger to people are avoided, and also money is saved. To make periodic inspections of the buildings is useful to quantify the extent to which deteriorations are severe or not, in order to facilitate decision making and prioritize interventions. To this purpose many scales have been used and are used to assess the severity of damage and degradation of the building components. But it appears evident that there is not consensus among users and these scales are different between them, with different number of degrees and metrics for the measurement of the condition state. The main goal of this paper is to calculate which is the optimal metric (which is the optimal number of degrees) of a severity scale of damages in buildings, so the corresponding scale could be of widespread and of common use among professionals, avoiding the problems of comparison between different evaluators. The proposed methodology to calculate the optimal metric of a scale can be also extended to other scopes.

Keywords: Scale, degree of severity, damages, direct assignment method, condition state, descriptive statistics.

1. Introduction

The interest in the evolution of the building and infrastructure stocks has been evolving during time, in some cases closely linked to the sustainable development debate (Kohler and Yang 2007). The topic of maintenance optimisation has been a focus of research interest for some time (Mazzuchi and René 2012). A central issue is the mortality of buildings. Lifetables of classical population dynamics (Klein and Moeschberger 2003) can be used for estimating the mortality of a sample of building and infrastructure stocks (Herz 1998; Schiller 2007).

In the same way, there are several authors who investigate construction defects, usually focused on failures in buildings due to lack of maintenance, design and construction errors, execution failure, material defect, inappropriate use, etc. (Frangopol 2011). Some of these defects are: humidity, defects in roofs, in natural stone coverings (Neno and de Brito 2012), in ceramic façade claddings (Silvestre and de Brito 2011), damage on the envelope of buildings (Flores et al. 2010; Rodrigues et al. 2011), damage in load-bearing rammed earth walls (Ruiz 2013), problems in the subsoil (Díaz et al. 2015), etc.

[1] Corresponding autor: Email: felix.ruiz@upc.edu
Other autors: Email: antonio.aguado@upc.edu // carles.serrat@upc.edu // joan.ramon.casas@upc.edu

On the other hand, the rapid industrialization and population migration of the last 30 years, has led to fast growing urbanization, doubling the building and partially the infrastructure stocks in very short periods (20-30 years) and this may happen more than once a century (Yang, 2006). The rate of growing is very high and it is not always known, how well these stocks are constructed.

In this context, the crucial indicator is the state of degradation of the different components of the stock and an objective indicator of this condition state (Kohler and Yang, 2007). In the same way, various asset management tools have been introduced to help asset managers in the difficult decisions regarding how and when to repair/replace their existing building stock cost-effectively (Flores and de Brito 2010; Elhakeem and Hegazy 2012). Similarly, "What is not defined can not be measured. What is not measured can not be improved. What is not improved, it is always degraded". This sentence is from Sir William Thomson, Baron Kelvin of Largs.

Although being from the nineteenth century, the sentence is still valid today. It is clearly demonstrated the importance of performing preventive maintenance in buildings, in order to prevent their degradation and the appearance of severe malfunctions and also in terms of economic efficiency. In the framework of maintenance, to make periodic inspections of buildings is useful to quantify the extent to which deficiencies are severe or not, in order to facilitate decision making and prioritize interventions.

All the referred shows the need of having a scale to assess the grade of severity of deterioration of constructive elements in buildings and to prioritize the interventions. In addition, the use of this scale also has implications in terms of economic efficiency.

On the other hand, it is important to highlight that, currently, there are many scales used to assess the grade of severity (intensity and extension) of damage (or condition state index) of the constructive elements in buildings. There is no consensus and these scales are different, with different number of degrees and metrics, according to the study to which they belong (Ruiz, 2014).

The study of the existing literature puts in evidence the following facts:

- In the definition of a scale and its number of degrees, the context of the scale (condition state of the buildings, earthquakes, degree of pain, intensity of wind, etc.) plays a decisive role, as well as other aspects, such as whether the degree is estimated through the direct assignment method or through quantitative indicators and mathematical algorithms, etc.

- If the number of degrees suitable for a scale is decided simply on the basis of the opinion of experts, (opinion based on experience, knowledge, etc.), due to the existence of different opinions among the experts, the result is different scales within the same sector, with different numbers of degrees. This is clearly the case in the context of the condition of buildings.

- There is not any methodological study that establish which is the optimal number of degrees of a scale in the context of condition of buildings, nor in any other context.

- The optimal situation in the area of the condition state of buildings would be the definition of a single scale of widespread and common use. This option is clearly more efficient than the current situation, with multiple scales, with different number of degrees. The aspect to discuss and the objective of this paper is whether to accomplish this goal (use a common scale in the context of the condition of buildings instead of many different scales) is possible or not.

The main objective of this paper is to calculate which is the optimal metric (the optimal number of degrees) of a severity scale of damages in buildings. This should be based on a methodological and scientific basis, trying to keep subjectivity at a minimum.

For this end, several tests were performed and various mathematical techniques were used, such as descriptive statistics (arithmetic means, standard deviations, densities, frequencies, histograms, etc.), binary logistic regression and clustering. In these tests 374 experts (mainly building engineers and architects) participated, and 12,342 responses were statistically analyzed.

2. General Methodology

To accomplish the mentioned objective, the methodology is based in the following steps:

- A critical review of the different scales currently in use.
- Propose an initial scale and apply it by using the direct assignment method.
- Select a group of 374 experts in building inspection and ask them to value a set of 33 images of constructive elements with different levels of deterioration, according to the initial proposed scale.
- Analyze statistically the 12,342 collected values from the answers of the experts.
- Determine, through the use of statistical analysis, which is the optimal number of degrees of the scale.

3. Review of existing condition scales

Among the existing scales, the criteria for selecting those reviewed here was based on two main aspects: a) They should be sufficiently representative and accepted, widely used in their respective fields; b) They belong to various fields of science, in order to get an approach with a wide perspective.

Some of the studied scales, among others, were the following: Beaufort (measuring wind intensity), Fujita-Pearson (intensity of a tornado), Saffir-Simpson (intensity of a hurricane), Richter (intensity of earthquakes), Modified Mercally (intensity of earthquakes), Mohs (hardness of a substance), VAS (degree of pain) (Von Korff *et al*. 1992), Norton (risk for pressure ulcers), Glasgow (grades of coma) (Gabbe *et al*. 2003), Likert (psycometric) (Payne *et al*. 2002), etc. It is important to say that all the scales studied in this part are commonly used worldwide.

The range of values of these scales is diverse. From 0 to 10 (VAS), from 0 to 12 (Beaufort ; Fujita-Pearson), from 1 to 5 (Saffir-Simpson ; Likert), from 5 to 20 (Norton), etc. Another important aspect is that in some scales variables are measured (Beaufort, Fujita, Saffir-Simpson, Douglas, Richter, etc.) which are easily measurable with the proper equipment, thus the level of certainty to assign values is very high. In the other scales (Mohs, VAS, Norton, Glasgow, etc.) attributes are measured, which in some cases are easily measurable, as happens in the Mohs scale, but in some other cases the measurement of the attributes may have high variability or subjectivity, as happens in the VAS scale.

In the context of the built environment, the level of certainty to assign values will not be a priori very high, since to assess the degree of damage of a component (either a beam, a balcony, a cornice, a bearing wall, etc.) may be subject to subjectivity.

Therefore, the scales that a priori can be more useful for our proposal are the Modified Mercalli scale, the VAS scale, the Norton scale and the Glasgow scale, especially the last 3 (associated with the field of medicine), since they have also some degree of uncertainty when assigning values in the diagnosis of people, some of them in emergency conditions. The field of medicine has important conceptual similarities with the field of diagnosis of building elements. Some of the similarities between diagnosis of human beings and diagnosis of buildings are the next ones:

1. To propose appropriate cure it is necessary first to develop an accurate diagnosis, in order to find out the causes of the dysfunctions.
2. The conceptual techniques to develop diagnosis in both scopes are similar, based on differential diagnosis methodology.
3. The main goal is the same: restore the health (of the human being or of the building).
4. Similar words are used: diagnosis, rehabilitation, therapy, etc.
5.- Both fiels use similar diagnostic tools: endoscopy, X ray, ultrasonic techniques, magnetic flux,…

To conclude this section, it is considered opportune, as it is a very topical issue, to mention the scale of terrorist alert. This scale in Spain, UK and Denmark has 5 degrees. In Belgium and Holland it has 4 degrees. In France and Italy it has 3 degrees. And in Germany it has no defined degrees (Ministerio del Interior 2015). Thus, in this topical issue there are also different scales, with different metrics or number of degrees.

In the case of construction, including bridges and buildings, the existing scales do not have specific names, in contrast to the previously nine studied scales that do have (Boufort, EVA, Richter, etc.). It is remarkable to highlight that unlike what happens in other areas of science where there are widely used and commonly accepted scales, in the field of buildings there is not a common scale for assessing the degree of deterioration of the constructed elements.

Table 1 shows a sample of the metrics (number of degrees) of 21 scales within the construction area, mainly in the field of buildings and bridges. Some of the used metrics in these scales are the following: 3, 4, 5, 6, 7, 9, 10, 11, 30, 70, 100.

Table 1. Sample of 21 metrics (number of degrees) of scales used in the field of construction

The metric "3" in the reference of Decret (2010) means that this scale has 3 degrees (the definitions of the three degrees are: no deficiency, mild deficiency, severe deficiency; but there are not numbers associated with each degree). In Roche (2018) it is defined the range of the scale (from 0.6 to 1.00). It is also explained the system to calculate the resulting value (which is always between 0.6 to 1.00). But nowhere says the number of degrees of the scale. In Brime (1999) is defined the range (1 to 4) and the number of degrees (30). The division between degrees is lineal, so each degree represents an increment of 0.13 respect the other (4/30).

As it can be seen in *Table 1*, the metrics are very varied, and there is no a concrete metric that is clearly more commonly used than the others. It is important to highlight that in none of these scales the used metric has been calculated based on a specific methodology. The metric has been decided based on the knowledge, experience and criteria of the authors of the scale. This aspect can be extended to the scales of any other field, in which the metrics (number of degrees) have not been calculated based on a specific methodology. Therefore, the contribution of this paper is precisely to propose a methodology to calculate which is the optimal metric for a certain type of condition assessment, in this case the condition rating of existing buildings. But it should be added that this proposed methodology can be applied to scales in other fields.

Another observed difference in the scales is the method of application to determine the value in the scale: direct assignment (DA) or application of mathematical functions or algorithms (AMF).

The first method has the advantage of being faster and easier to use by the technicians, while it has the disadvantage of greater variability. The second method has the disadvantage of being more laborious and complex to implement by the technicians, while it has the advantage of reducing the degree of variability.

4. Proposal of initial scale

To measure the grade of gravity (*G*) is proposed a scale ranging from value 0 (zero means that the constructive element is in perfect condition) to value 10 (extreme severity; it is not conceivable a greater severity; pathology in terminal phase; collapse may occur at any time). The proposed scale, presented in *Table 2*, explains in a general way what it means or represents each grade, in order to reduce the variability among different technicians when assigning values. Because the scale should be of application to any type of constructive element (walls, beams, columns, bearing walls, façades, etc.), definitions are necessarily generic (Ruiz, 2014).

Table 2. Proposal of gravity scale of construction elements in buildings

In order to facilitate the visualization of the distribution of severities of a constructive element, the proposal is to map this element with level curves corresponding to different *G* values, regardless of whether the studied element is a beam, or a whole floor, etc. For greater visibility, a color is associated to each *G* value.

Units of application

The concept of zone is proposed as the smallest unit to value, which is defined as a specific part of a constructive element, as for example an area of a wood floor, an area of a façade, an area of a reinforced concrete beam floor, an area of a balcony, etc. *Figure 1* shows some illustrative examples.

The next proposed unit is the system. The building is divided into different systems that encompass all parts composing it. In order to provide flexibility, the proposal is that the total number of systems (*S*) and the definition of them can be chosen by the technician that makes the study of the building, thus the proposed method is of general application. On the other hand, in order to propose a specific model to follow, it is considered appropriate to propose a reasonable division of building into the systems that are listed in *Table 3*.

Table 3. Definition of systems to be evaluated in the proposed scale

The last proposed unit is the entire building.

5. Application of the proposed scale and statistical analysis

In the direct assignment method (DA) based on the generic definitions of the scale, a certain degree of variability among technicians to assign values *G* is expected, since what is assessed are fuzzy attributes (degree of damage of a constructive element). Therefore, it is interesting to analyze the degree of variability to assign the values *G*, from which parameters such variability depends on, and what is the optimal number of degrees for this scale used by the DA method.

Thus, an experiment was carried out, consisting in the selection of 33 images of constructive elements with different degrees of damage that were shown to a population of 374 technicians, who should assign the grade of gravity *G* to each image, according to the generic definitions of the proposed scale. The objective is to analyze, among many other things, whether there is high or low variability between technicians when assigning *G* values.

From the collected data (12,342 *G* assigned values) a statistical analysis was carried out using various mathematical techniques such as descriptive statistics (arithmetic means, standard deviations, densities, frequencies, histograms, etc.), binary logistic regression and clustering, using for this the Minitab statistical analysis program. The *G* assigned values by the technicians who have completed the questionnaire are called *GA* values, in order to distinguish them from the *G* reference value or pattern (called *GR*).

The methodology consists of several related processes, which can be grouped into three main phases as schematically presented in *Figure 2*. The discriminant capacity of the scale means the probability of correct classification. Therefore, the probability that the technician assigns the same value *G* than the reference value *GR.*

Figure 2. Methodology for the validation of the proposed scale

5.1. Phase 1: Initial analysis

The criteria to decide on the number of images of constructive elements to show to technicians was based on the following aspects, mutually complementary:

- The number should not be very high, otherwise, it would be difficult to find technicians willing to complete the questionnaire.
- The number should be high enough to obtain sufficiently consistent and representative data to be statistically analyzed.
- The number should be enough to have representativeness of different types of constructive elements such as façades, wooden beams, reinforced concrete beams, balconies, metal sections, rammed earth walls, pillars, etc.

It was considered suitable the number of 33 images based on the above criteria. Furthermore, this number allows for each grade of the scale, eleven grades in total (from *G* = 0 to *G* = 10), to have 3 images, in order that the statistical analysis is balanced. It could be argued that, from a statistical point of view, it would be enough with one image for each grade, thus 11 images in total, due to the large number of technicians involved in the experiment. However, using 3 images instead of one for each grade, allows to collect not only the variability between technicians but also the variability of each technician for each grade. As an example, in *Figure 3* (left) it can be seen the images corresponding to *GR* = 0 (perfect condition), showing the cases of a building façade, and wooden and concrete beams.

Figure 3. Histogram of relative frequencies for the responses in the pictures Nos. 3, 17 and 25 (GR = 0)

The results of the arithmetic mean and the standard deviation of all the values are presented in *Table 4*. In this table it can be seen for each image, the corresponding *G* reference value or pattern (*GR*), the average of the *G* values assigned by the technicians who have completed the questionnaire ($\overline{GA}$), the standard deviation (σ_{GA}), and the difference $\overline{GA} - GR$.

Table 4. Arithmetic mean and standard deviation of collected values

A table was prepared for each *GR* value, showing the number of participants that have assigned a particular *GA* value each of the 3 images of one particular *GR* value. Similarly, each table has three related graphics, a graphic of the deviation for each of the three images belonging to the same *GR* value, a graphic showing the histogram of frequencies and a boxplot graphic (also called box-and-whisker diagram or plot) for the set of the three photographs. Thus there are 11 groups of tables and graphs, one for each *GR* value. As an example, in *Table 5* and *Figures 3*, *4* and *5* is presented the data corresponding to grade 0 (*GR* = 0). In *Table 5* "NV" means "no value". And it means that the technician didn't assign any value to that picture. The reason of this "NV" can be forgetfulness, not understand the image, etc.

Table 5. Frequency distribution of GA responses for GR = 0

Figure 4. Distribution of the deviation (Dev = GA - GR) for GR = 0

Figure 5. Response boxplot for the pictures Nos. 3, 17 and 25 (GR = 0)

From the figures, it can be observed how there is more dispersion between the answers given to the picture number 25 than to the picture number 3. Picture number 17 is an intermediate graphic between the other two. There is higher frequency of correct response in photograph number 3, with 325 technicians who have evaluated the picture exactly with *G* = 0. In contrast, in the photograph number 25, there is higher response frequency for *G* = 1 (170 technicians). This may be due to some existing small spots and shadows in the image that have induced many technicians to mark the value *G* = 1, instead of the reference value *GR* = 0. It is observed how to a less dispersion corresponds greater proportion of outliers. This is a consequence of the concentration of observations in a particular value (*GR* = 0).

Set of observed values

In *Figure 6* it can be seen the complete Boxplot with the values obtained from the survey, with 11 Boxplots, one for each *GR* value. The vertical discontinuous blue lines are the borders of each *GR* value. Discontinuous red horizontal segments mark the zone where the "boxes" of Boxplot should be located, in case that the *G* values assigned by technicians (*GA*) coincide with the *GR* reference values.

Figure 6. Boxplot of all obtained values according to observation number and GR value

Analysis by intervals of difference

The purpose of this part is to classify the participants in the survey according to the total difference in absolute value (*TD*) between *G* and *GR*. This allows to know what percentage of technicians has assigned gravity values similar to the *GR* reference values, what percentage has assigned different values, etc. Equation (1) is used to evaluate TD and *Figure 7* shows the graph of participants frequency by intervals of difference.

$$TD = \sum_{j=1}^{33} \left| Dev_j \right| = \sum_{j=1}^{33} \left| G_j - GR_j \right| \qquad (1)$$

Figure 7. Participant frequency plot by intervals of difference (TD)

The results show that 253 technicians (67.65% of total) have $TD \leq 40$, thus they have an average difference respect *GR* in each image less than one, or in other words, all their observations are on average in the interval $GR \pm 1$. Considering that the used scale consists of 11 grades (from 0 to 10), each grade represents the 9% of the scale. Since what is evaluated with this scale are fuzzy attributes (grade of damage of constructive elements), and it consists of 11 grades, one may conclude that an average difference respect to *GR* of 9% or lower is reasonable.

5.2. Phase 2: Data depuration

To obtain a consistent statistical analysis of the collected data, the possible outliers that can distort the analysis must be identified and eliminated. Once identified and analyzed the 365 outliers, from a total sample of 11,814 values, the Boxplot in *Figure 8* was obtained.

By making the first debugging of outliers, the sample is reduced to 11,449 values, of which 91 are now outliers, representing 0.79% of the total. Because it is less than 1%, it was decided to not continue debugging and consider the definitive data shown in *Figure 8*. It must also bear in mind that if these 91 outliers were removed, it would strongly influence the answers to *GR* = 0, eliminating the little dispersion they represent, such as observation 2 for *GR* = 0.

Figure 8. Boxplot after the first depuration of the outliers

Final data

In *Figure 9* the frequency histogram of the deviations for each of the *GR* values is presented. As shown in the graphs of *Figure 9*, the variations obey a Gaussian or Normal pattern except for *GR* = 0 and *GR* = 10, due to its extreme character. It also shows that for *GR* = 0, 1, 2, 3, 4, 5 and 6 there is an overestimation of the gravity value, while for *GR* = 8, 9 and 10 there is an underestimation thereof.

In the extreme *GR* = 0 is normal that deviations are positive, since by their extreme condition on the left side there can not be negative deviations. Likewise, it is normal that in the extreme *GR* = 10 deviations are negative, since by their extreme condition on the right side there can not be positive deviations. For *GR* values near the extremes occurs a similar effect, causing some asymmetry for these values of the Gaussian curves. Thus, for *GR* = 1 and 2 is normal to have some asymmetry to the right, with more frequency of positive deviations, while for *GR* = 8 and 9 is normal to have some asymmetry to the left, with more frequency of negative deviations. The fact that in *GR* intermediate values there is more frequency of positive deviations shows that technicians tended to evaluate the images of constructive elements in a conservative way, thus assigning *G* values somewhat higher than *GR*.

On the other hand, the dispersion is lower for the extreme *GR* values, indicating that there is a higher number of correct values assigned. This is due, on one hand, to the fact that they are extremes and, therefore, there is a smaller range of values from which to choose. Secondly, inspectors perceive more clearly and accurately the degree of gravity in the extreme values. Instead, dispersion is higher in intermediate values of *GR*, particularly in *GR* = 2, 3, 4, 5 and 6, because of more fuzzy definitions of the degree of gravity.

Figure 9. Histogram of frequencies of deviations (Dev) for each of the G values

Evaluation of the reliability of the initial scale

The next step is to know the percentage of correct classification of the scale, *P(ok)*. Thus, if the scale already has a probability of correct classification not less than a target value, it can be considered that the scale is reliable enough and has to not be improved. And conversely, if the probability is less than the required target, an improvement is necessary. .

A matrix *C* of dimension 11x11 is defined where the position c_{ij} for *i*, *j* = 0, ..., 10, is the cardinal, within the images with *GR = i*, with assignments *G = j* by the technicians. To simplify the notation it is denoted *GRi* and *Gj* respectively. In other terms,

$$c_{ij} = \text{card}\ \{GRi / Gj\}, \text{ for } i, j = 0, \ldots, 10 \quad (2)$$

The resulting initial matrix for the general scale is represented in *Table 6*, where it can be seen, for example, that from the 1,033 results of reference gravity *GR* = 3, 146 times have been classified as gravity *G* = 2 by technicians. Similarly, the resulting diagonal (highlighted in light blue) reflects the number of times a correct classification for each degree of severity was obtained.

Table 6. Initial Table for the General Scale

In order to calculate the probability of correct classification, or in other words, the reliability of the proposed scale, the Equation (3) is used, on *P(ok)* the probability of correct classification.

$$P(ok) = P\left(\bigcup_{k=0}^{10}(GRk \wedge Gk)\right) \quad (3)$$

Therefore, the probability of correct classification of the General Scale can be estimated by the Equation (4).

$$P(ok)=\frac{\sum_{k=0}^{10}c_{kk}}{\sum_{i=0}^{10}\sum_{j=0}^{10}c_{ij}}=\frac{3.673}{11.452}=0{,}3207=32{,}07\% \quad (4)$$

According to equation 2, c_{kk} represents the values of the diagonal of the matrix C.

Thus, the initial scale has a relatively low probability of correct classification and an improvement is needed. To achieve this goal, a clustering algorithm was used (Fisher, 1996), through which the number of degrees of the scale is progressively reduced, until a new scale is obtained with a number of degrees that allows a correct probability of assignment.

Evaluation of the specificity of the initial scale

It is also important to evaluate the specificity and power of the initial scale, since they serve to evaluate the error of classification of the scale, and corresponds to the calculation of two sources of classification error, according to the following definitions in the Equations (5) and (7):

a) Type I classification error: $\alpha_k = P(G \neq k \mid GR = k)$ (5)

Specificity: $1-\alpha_k = P(G = k \mid GR = k)$ (6)

Therefore, the Type I classification error are those cases in which, conditioned to a *GR* reference value, the *G* value assigned by the technician does not coincide with the *GR* value. Thus, the specificity $(1-\alpha_k)$ indicates the tendency of technicians to correctly assign a certain *GR* value.

b) Type II classification error: $\beta_k = P(GR \neq k \mid G = k)$ (7)

Power: $1-\beta_k = P(GR = k \mid G = k)$ (8)

Thus, the Type II classification error are those cases in which, conditioned on a classification value *G* by the technicians, the underlying objective value *GR* does not coincide with the assigned gravity. Thus, the power $(1-\beta_k)$ indicates the probability that an image to which the technicians have assigned a gravity *G*, this corresponds to the *GR* reference severity.

The improvement procedure will consist of minimizing Type I and Type II errors or, equivalently, maximizing the specificity and power, defined in the Equations (6) and (8).

In order to determine what degree of severity classifies with greater error and taking into account the interest in minimizing jointly both Type I and Type II errors, we propose the vector standard (α_k , β_k), e_k as a global measure of the classification error in degree *k* of severity, presented in the Equation (9):

$$e_k = \sqrt{\alpha_k^2 + \beta_k^2} \quad (9)$$

Considering the definition of standard deviation results the Equation (10), being therefore $\sigma(e_k)$ the standard deviation of the global error.

$$\sigma(e_k) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(e_k - \bar{e}_k\right)^2} \qquad (10)$$

5.3. Phase 3: Improvement of the scale and proposal of the optimal one

As it can be seen in *Table 7* and applying a clustering algorithm on the resulting data after the first depuration of the outliers, as the number of grades of the scale decreases, the reliability of the resulting scale increases. It can also be noted that for 5 grades (in the table highlighted in red), the standard deviation of the global error, $\sigma(e_k)$, is minimal, which means that the level of homogenization between the various grades comprising the scale is maximum, due to a similar probability of failure between the different grades.

Table 7. Simplification process of the scale

Figure 10 shows graphically this fact. The blue points indicate how the probability of correct classification increases as the number of grades of the scale reduces (tendency that is displayed with the blue arrow). The red points indicate how varies the standard deviation of the global error, $\sigma(e_k)$, as the number of grades of the scale reduces (tendency that is displayed by the red arrow), giving the minimum of $\sigma(e_k)$ for 5 grades (displayed with the borders in green). In contrast, to move to 4 grades (displayed with an orange circle) a significant increase of $\sigma(e_k)$ occurs.

Figure 10. Simplification process of the scale

From *Figure 10* and *Table 7*, an scale with 5 grades is the most suitable to use with the method of direct assignment. This solution balances the number of grades and the associated reliability, and minimize the standard deviation of the obtained global errors as well. In other words, the homogeneity of the global error in each of the grades $\sigma(e_k)$ is maximized.

Thus, the proposal is a scale of 5 grades (from "zero severity" to "extreme severity") with a probability of correct classification of 62.88%, instead of the initial scale of 11 grades (from $G = 0$ to $G = 10$) with a probability of 32.07%. For consistency with the used methodology, the proposal is the range of values from 0 to 4, and similarly, that the notation of this simplified grade or severity index is *GS*. The 5 different grades of the proposed scale, from $GS = 0$ to $GS = 4$, are defined in *Table 8*. For greater visibility, it is proposed that each *GS* value is associated to a color, in the same way it was previously proposed for the general scale.

Table 8. Proposal of simplified gravity scale of construction elements in buildings

According to the results obtained, it is appropriate to use the Simplified Scale when the degree of severity of constructive elements is assessed based on the direct assignment (DA) method.

But when other different methods than direct assignment are used, as methods based on indicators and mathematical expressions, which is expected to reduce very noticeably the degree of variability among technicians for obtaining *G* values, it is considered appropriate to use the General Scale. In order to confirm this expected reduction of variability, future studies such as this presented in this paper should be developed.

Similarly, it should be highlighted that *G* values (according to the General Scale of 11 grades) can be automatically translated to *GS* values (according to the Simplified Scale of 5 grades), as shown in *Table 9*.

Table 9. Correspondence between G and GS values

To conclude this section, it should be noted that in the assay the technicians have made direct assignment of *G* values only based on the generic definitions of the scale. But if the technicians, apart from the generic definitions of the scale, had also a catalog of images of constructive elements with their respective *G* values of reference (*GR*), in order to be used as orientation and additional information, it can be expected that the probability of accuracy of the technicians to assign *G* values would increase and would be higher to the resulting 32.07% of the current study. In the same way for the scale metric of 5 degrees, the probability of accuracy of the technicians to assign *GS* values, if they used a catalog of images, would increase and would be higher to the resulting 62,88% of the current study. Therefore, in further studies, authors propose to use this catalog when new survey processes will be carried out.

6. Conclusions

A severity scale of damage in buildings, of generalized and common use is needed in order to obtain a standardized method for deterioration grading in buildings where comparison between different countries and owners could be feasible. Based on a proposed initial severity scale of damages in buildings (named General Scale), with 11 grades of gravity (from 0 to 10), a survey involving 374 inspectors was launched, where the degree of severity of 33 images of constructive elements was ranked based on the direct assignment method and according to the General Scale. A total of 12,342 collected data have been gathered and statistically analyzed.

From the statistical analysis of the collected data it is possible to deduce, among other results, that in the General Scale there is a probability of correct classification of 32.07%, while in the named Simplified Scale with 5 degrees (*GS* = 0 to *GS* = 4) there is a probability of correct classification by the inspectors of 62.88%. For 5 degrees, the obtained standard deviation of the global error in the classification $\sigma(e_k)$ is also minimized.

In summary, it is proposed as the most accurate and reliable that in cases where the degree of severity of construction elements in buildings is valued by the direct assignment method, the appropriate scale metrics to be of 5 degrees.

The methodology followed in the presented study, although being applied to buildings, can be also generalized to other construction types and built assets as bridges, dams, power transmission lines, among others, in order to obtain the optimum number of degrees in the scale of deterioration.

References

Abaza, K.A., 2017. Empirical Markovian-based models for rehabilitated pavement performance used in a life cycle analysis approach. *Structure and Infrastructure Engineering*. 13 (5), 625-636.

Anderson, A., Rizzo, D.M., Huston, D.R. and Dewoolkar, M., 2017 Analysis of bridge and stream conditions of over 300 Vermont bridges damaged in Tropical Storm Irene. *Structure and Infrastructure Engineering*. 13 (11), 1437-1450.

BRIME, 1999. Review of current practice for assessment of structural condition and classification of defects. *Deliverable D2. BRIME PL97-2220*.

Decret 187/2010, 2010. Inspecció tècnica dels edificis d'habitatges. *Generalitat de Catalunya*.

Díaz, C., Cornadó, C., Santamaría, P., Rosell, J. R., Navarro, A., 2015. Actuación preventiva de diagnóstico y control de movimientos en los edificios afectados por subsidencia en el barrio de la Estación de Sallent (Barcelona). *Informes de la Construcción, 67(538): e089, doi: http://dx.doi.org/10.3989/ic.13.168*.

Elhakeem, A. and Hegazy, T., 2012. Building asset management with deficiency tracking and integrated life cycle optimisation. *Structure and Infrastructure Engineering* . 8 (8), 729-738.

Fisher, D., 1996. Iterative optimization and simplification of hierarchical clusterings. *Journal of Artificial Intelligence Research*. 4, 147-149.

Flores-Colen, I., de Brito, J. and Freitas, V., 2010. Discussion of Criteria for Prioritization of Predictive Maintenance of Building Façades: Survey of 30 Experts. Journal of Performance of Constructed Facilities**.** 24 (4) - August 2010.

Flores-Colen, I. and de Brito, J., 2010. A systematic approach for maintenance budgeting of buildings façades based on predictive and preventive strategies. Construction and Building Materials. 24 (9), 1718-1729.

Flourentzou, F., Brandt, E. and Wetzel, C., 2000. MEDIC - A method to predicting residual service life and refurbishment investment budgets. *Energy and Buildings*. 31 , 167-170.

Frangopol, D.M., 2011. Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges. *Structure and Infrastructure Engineering* . 7 (6), 389-413.

Gabbe, B.J., Cameron P.A. and Finche, C.F., 2003. The status of the Glasgow Coma Scale. *Emergency Medicine*. 15 (4), 353-360.

Geocisa, 2010**.** Metodología de cálculo de los índices de conservación y del índice de importancia de una estructura. Sistema de gestión de puentes de Geocisa. *Geocisa*.

Goyal, R., Whelan, M.J. and Cavalline, T.L., 2017. Characterising the effect of external factors on deterioration rates of bridge components using multivariate proportional hazards regression. *Structure and Infrastructure Engineering*. 13 (7), 894-905.

Herz, R., 1998. Exploring rehabilitation needs and strategies for drinking water distribution networks. *Proceedings of the 1st IWSA/AISE Conference, Prague, Czech Republic*.

Klein J.P. and Moeschberger M.L., 2003. Survival analysis: Techniques for censored and truncated data. *Springer-Verlag*.

Kohler, N. and Yang, D., 2007. Long-term management of building stocks. *Building Research and Information*, 35(4), 351-362.

Leaman, A. Fion, S. and Bordass, B., 2010. Building evaluation: practice and principles. *Building Research and Information*, 38(5), 564-577.

Lee, S.Y., Park, W., Ok, S.Y and Koh, H.M., 2011. Preference-based maintenance planning for deteriorating bridges under multi-objective optimisation framework. *Structure and Infrastructure Engineering*. 7 (7), 633-644.

León, J., 2006. Reflexiones en torno a la inspección de puentes. http://www.fhecor.es/img/proyectos/file/ 000000004000/inspeccionpuentes_es_4405.pdf

Liu, H., Wang, X., Jiao, Y., He, X. and Wang, B., 2017. Bridge residual service-life prediction through Bayesian visual inspection and data updating. *Structure and Infrastructure Engineering*. 13 (7), 955-965.

Martínez, J., 2016. Sistemas de gestión de puentes. Optimización de estrategias de mantenimiento. Implementación en redes locales de carreteras. Doctoral Thesis. *School of Civil Engineering of Madrid. Technical University of Madrid*.

Mazzuchi, T.A. and René, J., 2012. A Bayesian expert judgement model to determine lifetime distributions for maintenance optimisation. *Structure and Infrastructure Engineering* . 8 (4), 307-315.
Ministerio del Interior, 2015. Gobierno de España. Nivel de alerta antiterrorista. http://www.interior.gob.es/prensa/nivel-alerta-antiterrorista.
Neto, N. and de Brito, J., 2012. Validation of an inspection and diagnosis system for anomalies in natural stone cladding (NSC). *Construction and Building Materials, 30: 224-236, doi:http://dx.doi.org/10.1016/j.conbuildmat.2011.12.032.*
Payne, F., Harvey, K., Jessopp, L., Plummer, S., Tylee, A., Gournay, K., 2002. Knowledge, confidence and attitudes towards mental health of nurse working in NHS direct and the effects of training. Journal of Advanced Nursing, 40(5): 549-559.
Pellegrino, C., Pipinato, A. and Modena, C., 2011. A simplified management procedure for bridge network maintenance. *Structure and Infrastructure Engineering*. 7 (5), 341-351.
Roche, J.M., 2008. Monitoring Inequality among Social Groups: A Methodology Combining Fuzzy Set Theory and Principal Component Analysis. *Journal of Human Development and Capabilities (HDCA)*. New York, 9 (3), 427-452.
Rodrigues, F., Teixeira, J., Cardoso, J., 2011. Building envelope anomalies: A visual survey methodology. *Construction and Building Materials, 25(6): 2741-2750, doi: http://dx.doi.org/10.1016/j.conbuildmat.2010.12.029.*
Rodrigues, F., Matos, R., Di Prizo, M., Costa, A., 2018. Conservation level of residential buildings; methodology evolution. *Construction and Building Materials*. 172, 781-786, *doi: 10.1016/j.conbuildmat.2018.03.129.*
Ruiz, F., 2014. Escala de gravedad de daños en edificios: de la asignación directa a la contrastación estadística. Doctoral Thesis. *School of Civil Engineering of Barcelona. Technical University of Catalonia. Barcelona-Tech.*
Ruiz, F., 2013. Diagnosi de parets de tàpia. *L'Informatiu. Association of Building Engineering of Barcelona*. 335, 58-65. http://www.a16-01.com/pdfs/COL/2013/i335_58_65.pdf
Sadeghi, J.M. and Askarinejad, H., 2011. Development of track condition assessment model based on visual inspection. *Structure and Infrastructure Engineering*. 7 (11), 895-905.
SAMARIS, 2005. State of the art report on assessment of structures in selected EEA and CE countries. *Report SAM-GE-DE19. European Project SAMARIS (Sustainable and Advanced Materials for Road Infraestructure). V Framework Program.*
Schiller, G., 2007. Urban infrastructure – challenges for resource efficiency in the building stock. *Building Research and Information, 35(4), 399-411.*
Serrat, C., Gibert, V., Jordana, F., 2009. Survival analysis techniques applied to building maintenance. A: International Conference on Construction and Building Research. *1st International Conference on Construction and Building Research. School of Building Engineering of Madrid. Polytechnical University of Madrid, 1-12.*
Silvestre, J.D., de Brito, J., 2011. Ceramic tiling in building façades: Inspection and pathological characterization using an expert system. *Construction and Building Materials, 25(4): 1560-1571, doi: http://dx.doi.org/10.1016/j.conbuildmat.2010.09.039.*
Von Korff, M. Ormel, J., Keefe, F., Dworking, S., 1992. Grading the severity of chronic pain. *Elsevier B.V.*
Yang, D., 2006. International migration, remittances, and household investment: evidence from Philippine migrants exchange rate stocks. *National Bureau of Economic Researh. Cambridge.*
Zanini, M.A., Faleschini, F. and Pellegrino, C., 2017. Bridge residual service-life prediction through Bayesian visual inspection and data updating. *Structure and Infrastructure Engineering.* 13 (7), 906-917.

Scope of application	Measured concept	Metric (number of degrees and range of values)	Reference
Buildings	Deterioration grade	4 (1 to 4)	Flourentzou, et al. 2000
	Deterioration index on façades	7 (0 to 6)	Serrat, *et al*. 2009
	Deterioration index	100 (0 to100)	Elhakeem and Hegazy, 2012
	Deterioration qualification	3 (no numerical values)	Decret 187/2010
	Degradation level	8 (3 to 10)	Rodrigues *et al*. 2011
	Degradation level	5 (1 to 5)	Rodrigues *et al*. 2018
	Maintenance level	3 (I to III)	Rodrigues *et al*. 2018
	Housing adequacy	Not defined (0.600 to 1.00)	Roche, 2018
Bridges	Condition state	6 (1 to 6)	León, 2006
	State index	30 (1.0 to 4.0)	BRIME, 1999 Germany
	State index	70 (0 to 70)	BRIME, 1999 Austria
	Category of damage	7 (0 to 6)	SAMARIS, 2005
	Condition Value	6 (0 to 5)	Zanini *et at*. 2017; Pellegrino et al. 2011
	General Condition	10 (0 to 9)	Goyal *et al*. 2017
	Condition grade	5 (I to V)	Liu et al., 2017
	Condition state	4 (1 to 4)	Anderson *et al*. 2017
	Condition grade	5 (1 to 5)	Lee *et al*. 2011
	State index	10 (0 to 9)	Martínez, 2016
	Deterioration index	100 (0 to 100)	Geocisa, 2010
Pavements	Condition state	10	Abaza, 2017

		(1 to 10)	
Railway tracks	Severity level	3 (1 to 3)	Sadeghi and Askarinejad, 2011

Table 1. Sample of 21 metrics (number of degrees) of scales in the field of construction

G	Gravity	Definition
0	Null	The construction element is in perfect condition.
1	Very mild	First signs of very minor deterioration.
2	Very mild-mild	Very minor deterioration.
3	Mild	Slight deteriorations. Maintenance reviews of construction element are recommended to prevent increasing of gravity.
4	Mild-moderate	Deteriorations between mild and moderate. Recommended review of maintenance and superficial therapeutic treatments.
5	Moderate	Deteriorations of moderate entity. A maintenance review is necessary and to analyze the possibility, in the medium term, of applying therapeutic treatments to the constructive element to improve its durability.
6	Moderate-high	Deteriorations of between moderate and severe. Necessary therapeutic treatments between moderate and high, in medium-short term.
7	High	Existence of severe deteriorations, in advanced stage. Necessary therapeutic treatments of high entity, in short term, with the possibility of replacement of the damaged element. If the evaluated element is part of the general structure of the building, or a slab, it is advisable to evacuate the building or housing, and / or to take provisional measures of structural anesthesia, such as shoring or similar.
8	High-very high	Existence of severe deteriorations, in a very advanced stage. Necessary therapeutic treatments of high entity to very high, in very short term. Due to the severity of the damage, it is advisable to replace the damaged element. If the evaluated element is part of the general structure of the building, or a slab, it is necessary to evacuate the building or housing, and / or to take provisional measures of structural anesthesia, such as shoring or similar.
9	Very high	Existence of very severe deterioration, in stage between very advanced and terminal. The constructive element analyzed is so severely affected that the most suitable therapeutic treatment is its replacement. If the evaluated element is part of the general structure of the building, or a slab, it is necessary to evacuate immediately the building or housing. The damage is so severe that it is difficult to take provisional measures of structural anesthesia.
10	Extreme	Higher gravity is not conceivable. Existence of end-stage deterioration, and collapse can occur at any time. The analyzed constructive element is so devastated that practically the only possible therapeutic treatment is its replacement. If the evaluated element is the part of the general structure of the building, or a slab, it is necessary the immediate evacuation of the building or house.

Table 2. Proposal of gravity scale of construction elements in buildings

System	Description	Main constituent parts
1	Façades	Claddings, base material, cantilevers, cornices, windows and other practicable openings, railings, balustrades, ornamental elements, etc.
2	Vertical structure	Pillars, load walls, foundation, etc.
3	Horizontal structure	Beams, beam filling, vaults, arches, etc.
4	Roofs and inner courtyards	Roof tiles, pavements in flat roofs, waterproofing, thermal insulations, skylights, walls and practicable openings for inner courtyards, etc.
5	Interior building elements	Partitions, interior walls, practicable openings, pavements, interior claddings, etc.
6	Staircases	Walls, stair structure, steps, railings, etc.
7	Sewer facilities	Downpipes, drains, gutters, etc.
8	Other facilities	Electricity, water, gas, elevators, etc.

Table 3. Definition of systems to be evaluated in the proposed scale

Image	GR	$\overline{GA}$	σ_{GA}	$\overline{GA}-GR$	Image	GR	$\overline{GA}$	σ_{GA}	$\overline{GA}-GR$
1	**3**	2,59	1,19	-0,41	**18**	**2**	3,24	1,40	1,24
2	**5**	6,40	1,05	1,40	**19**	**9**	7,82	1,15	-1,18
3	**0**	0,20	0,71	0,20	**20**	**5**	5,72	1,52	0,72
4	**7**	7,39	1,17	0,39	**21**	**1**	0,49	1,01	-0,51
5	**9**	7,90	1,03	-1,10	**22**	**4**	5,74	1,52	1,74
6	**1**	1,92	1,53	0,92	**23**	**10**	9,81	0,61	-0,19
7	**3**	4,35	1,27	1,35	**24**	**7**	7,69	1,04	0,69
8	**2**	2,41	2,23	0,41	**25**	**0**	1,19	1,16	1,19
9	**8**	7,75	1,34	-0,25	**26**	**5**	4,99	1,37	-0,01
10	**7**	6,79	1,06	-0,21	**27**	**9**	8,51	1,18	-0,49
11	**10**	9,07	1,06	-0,93	**28**	**1**	1,72	1,13	0,72
12	**6**	7,16	1,28	1,16	**29**	**10**	8,38	1,19	-1,62
13	**4**	4,65	1,35	0,65	**30**	**8**	8,43	1,12	0,43
14	**3**	3,91	1,70	0,91	**31**	**2**	2,54	1,47	0,54
15	**6**	7,01	1,26	1,01	**32**	**4**	4,57	1,33	0,57
16	**8**	6,56	1,39	-1,44	**33**	**6**	6,16	1,25	0,16

17	**0**	0,73	1,36	0,73

Table 4. Arithmetic mean and standard deviation of collected values

	Pic.	0	1	2	3	4	5	6	7	8	9	10	DOUBTS	NV
GR=0	**No. 3**	325	36	4	4	0	2	0	0	1	0	0	1	1
	No. 17	255	45	25	25	8	10	1	2	0	0	0	2	1
	No. 25	104	170	52	30	10	4	2	1	0	0	0	1	0

Table 5. Frequency distribution of GA responses for GR = 0

G / GR	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	TOTAL
GR0	**656**	206	72	30	10	4	0	0	0	0	0	978
GR1	323	**369**	164	113	53	17	13	0	0	0	0	1052
GR2	117	180	**199**	241	166	88	62	12	0	2	0	1067
GR3	4	90	146	**225**	301	148	96	17	6	0	0	1033
GR4	0	6	41	100	**259**	258	241	119	35	10	1	1070

GR5	0	2	13	51	145	**217**	298	233	81	12	1	1053
GR6	0	0	0	11	24	134	**269**	315	216	87	16	1072
GR7	0	0	0	0	0	31	185	**356**	322	125	0	1019
GR8	0	0	0	0	26	52	138	220	**268**	265	74	1043
GR9	0	0	0	0	3	8	56	230	332	**335**	97	1061
GR10	0	0	0	0	0	1	3	65	171	244	**520**	1004
TOTAL	1100	853	635	771	987	958	1361	1567	1431	1080	709	**11452**

Table 6. Initial Table for the General Scale

											Number of grades	**% correct classification**	$\sigma(e_k)$
GR0	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	11	32.07%	0,2046
GR0	G1	G2	G3	G4	G5'		G7	G8	G9	G10	10	35.85%	0,2035
GR0	G1	G2	G3	G4	G5'		G7'		G9	G10	9	40.58%	0,1984
GR0	G1	G2'		G4	G5'		G7'		G9	G10	8	43.96%	0,1784
GR0	G1	G2'		G4'			G7'		G9	G10	7	49.79%	0,1535
GR0	G1	G2'		G4'			G7''			G10	6	58.10%	0,1397
GR0	G1'			G4'			G7''			G10	5	62.88%	0,0520
GR0	G1'			G4''						G10	4	76.91%	0,1379
GR0	G1'			G4'''							3	82.79%	0,1753
GR0'				G4'''							2	89.36%	0,0532
GR0''											1	100.00%	-

Table 7. Simplification process of the scale

GS	Gravity	Definition
0	Null	The construction element is in perfect condition.
1	Mild	Slight deteriorations. Maintenance reviews of construction element are recommended to prevent increasing of gravity.
2	Moderate	Deteriorations of moderate entity. A maintenance review is necessary and analyze the possibility, in the medium term, of applying therapeutic treatments to the constructive element to improve its durability.
3	High	Existence of severe deteriorations, in advanced stage. Necessary therapeutic treatments of high entity, in short term, with the possibility of replacement of the damaged element. If the evaluated element is the general structure of the building, or a slab, it begins to be advisable to evacuate the building or housing, and / or to take provisional measures of structural anesthesia, such as shoring or similar.
4	Extreme	Higher gravity is not conceivable. Existence of end-stage deteriorations, and collapse can occur at any time. The analyzed constructive element is so devastated that practically the only possible therapeutic treatment is its replacement. If the evaluated element is the general structure of the building, or a slab, it is necessary the immediate evacuation of the building or house.

Table 8. Proposal of simplified gravity scale of construction elements in buildings

<table>
<tr><td>G</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>GS</td><td>0</td><td colspan="3">1</td><td colspan="3">2</td><td colspan="3">3</td><td>4</td></tr>
</table>

Table 9. Correspondence between G and GS values

Figure 1. Images of zones of construction elements

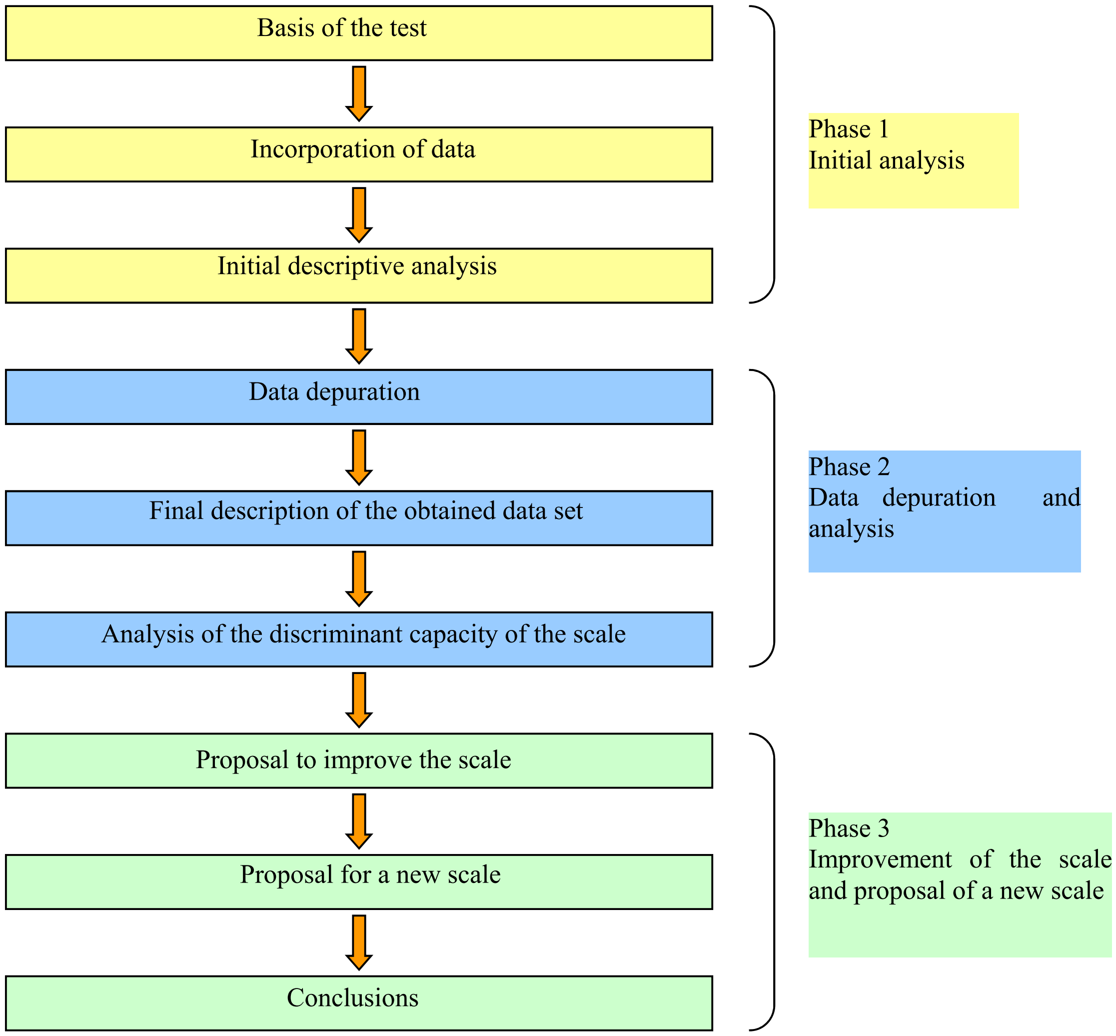

Figure 2. Methodology for the validation of the proposed scale

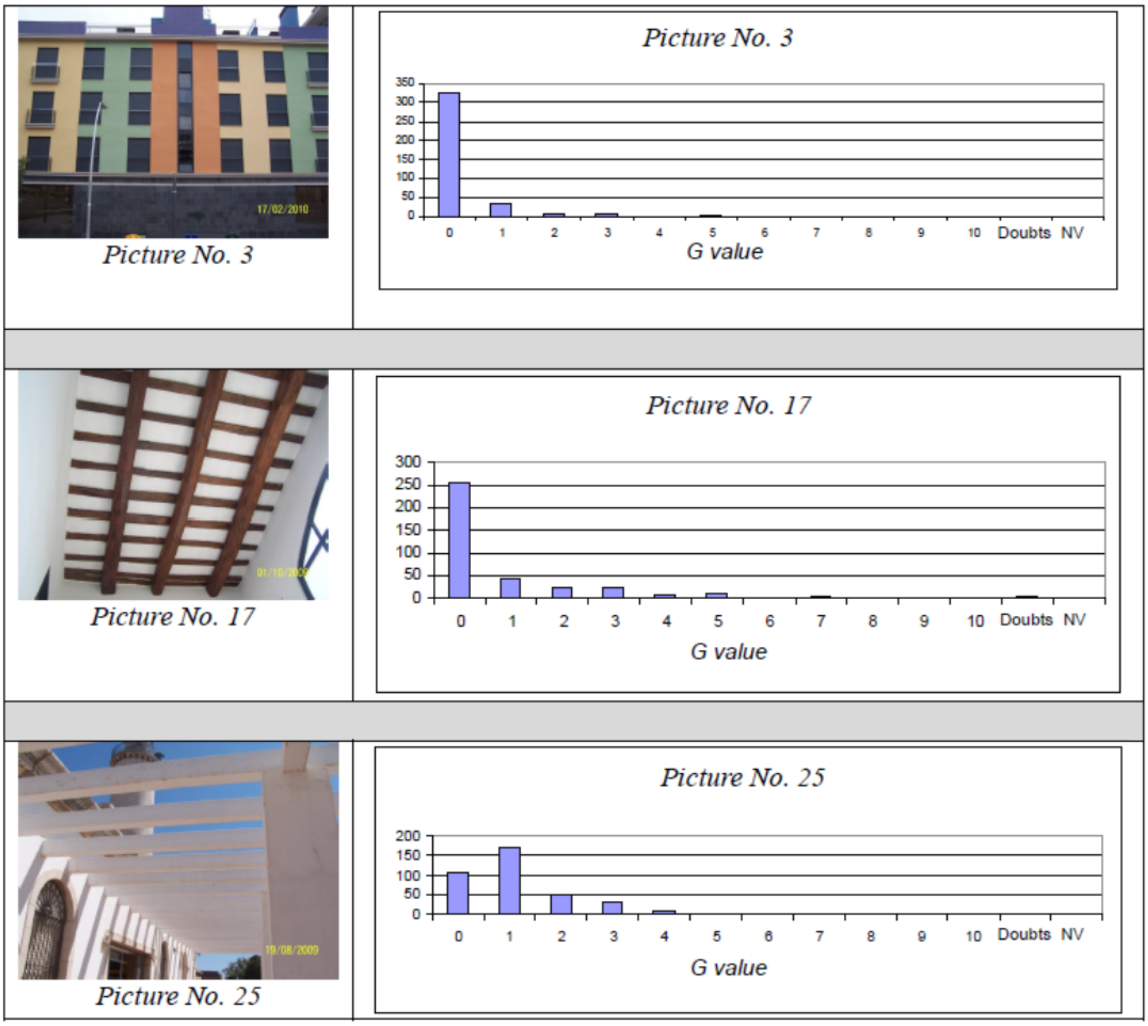

Figure 3. Histogram of relative frequencies for the responses in the pictures Nos. 3, 17 and 25 (GR = 0)

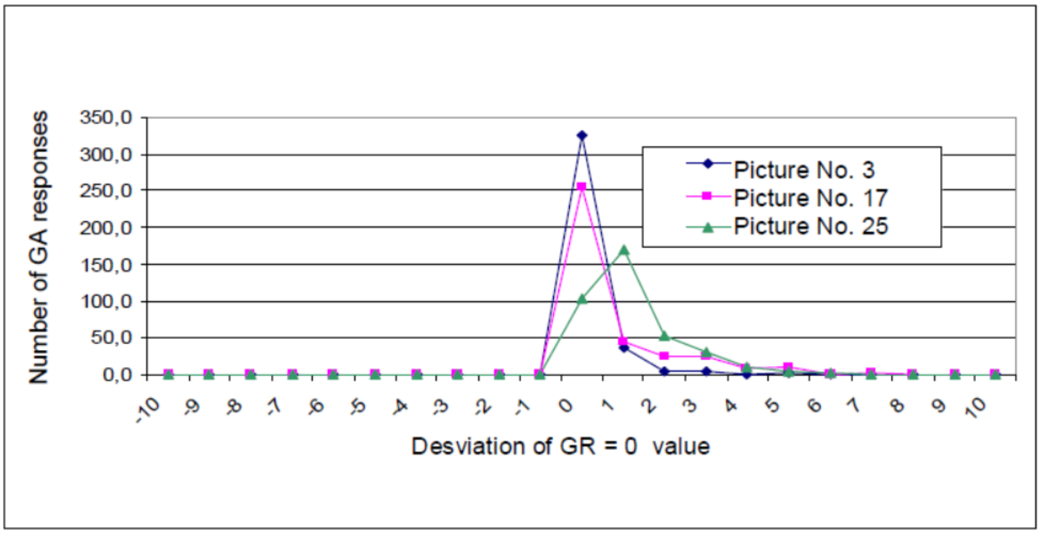

Figure 4. Distribution of the deviation (Dev = GA - GR) for GR = 0

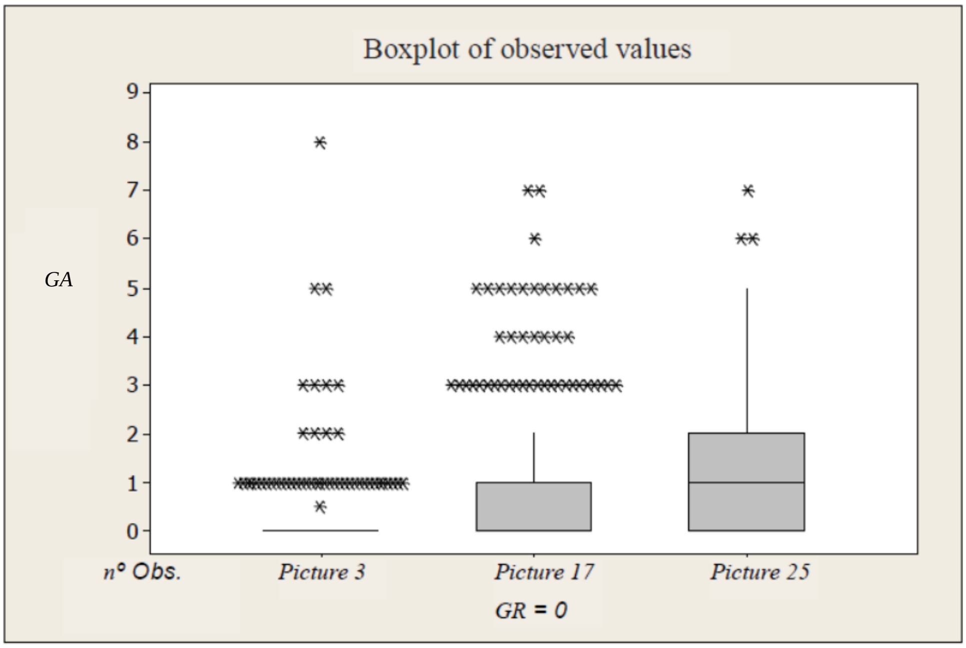

Figure 5. Response boxplot for the pictures Nos. 3, 17 and 25 (GR = 0)

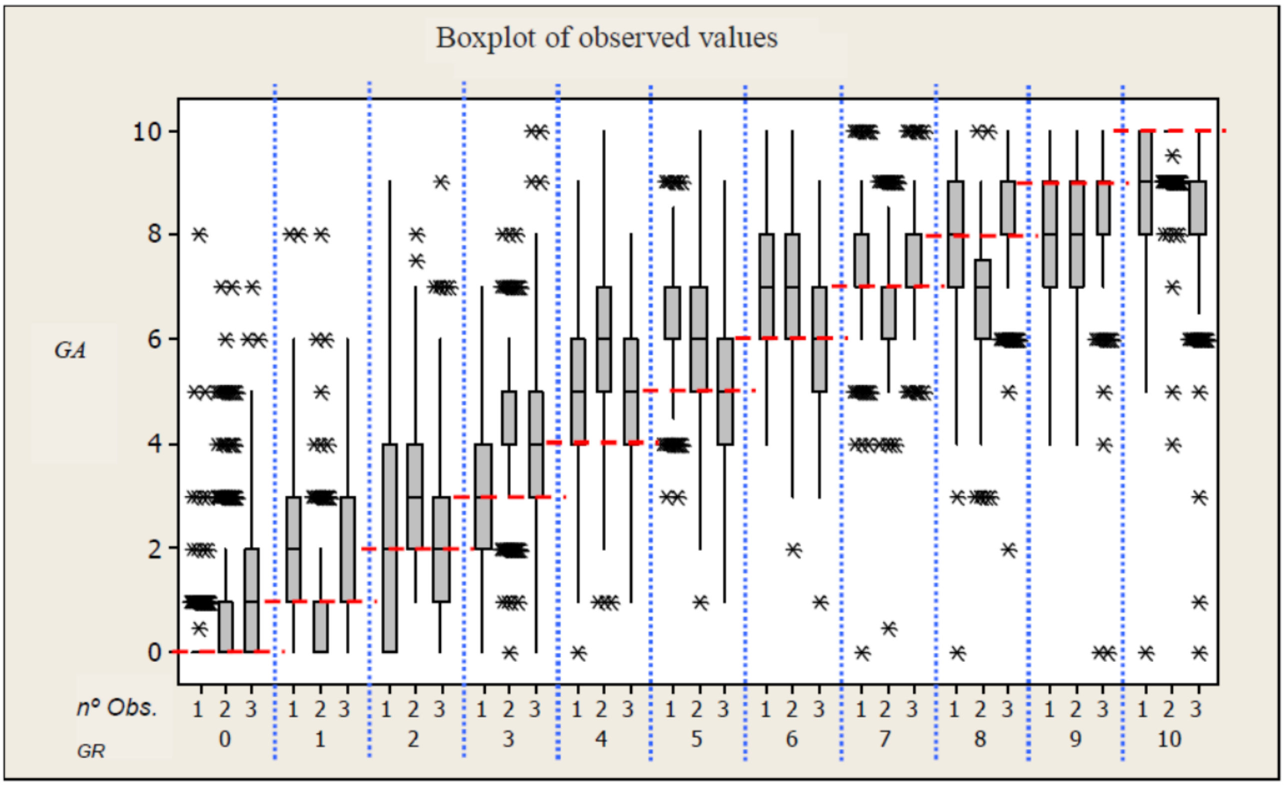

Figure 6. Boxplot of all obtained values according to observation number and GR value

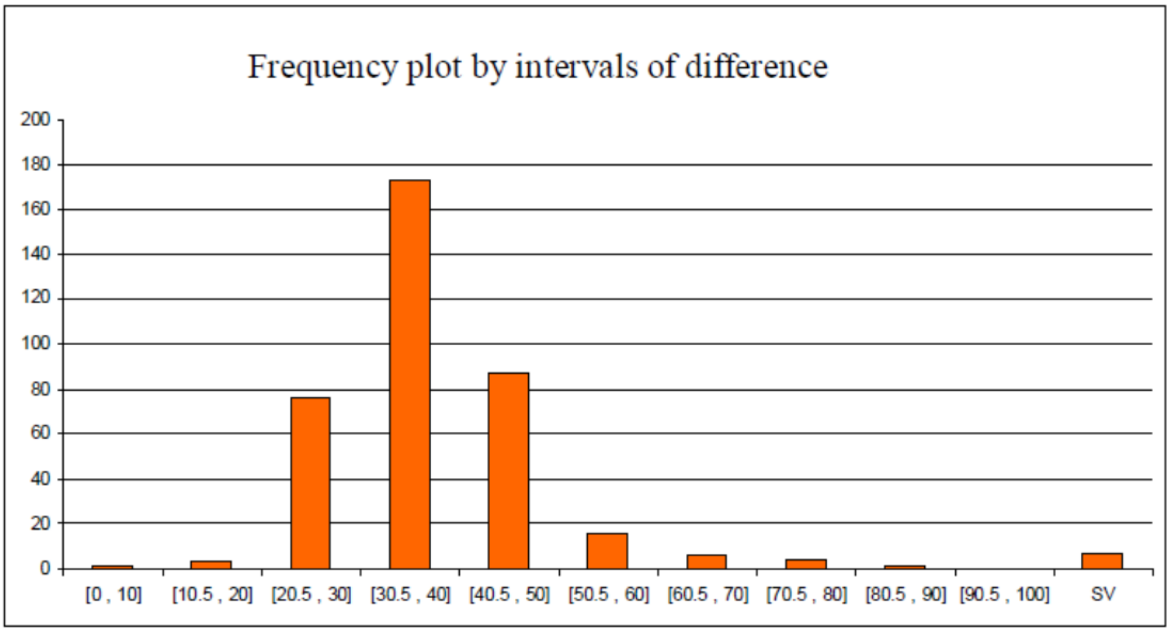

Figure 7. Participant frequency plot by intervals of difference (TD)

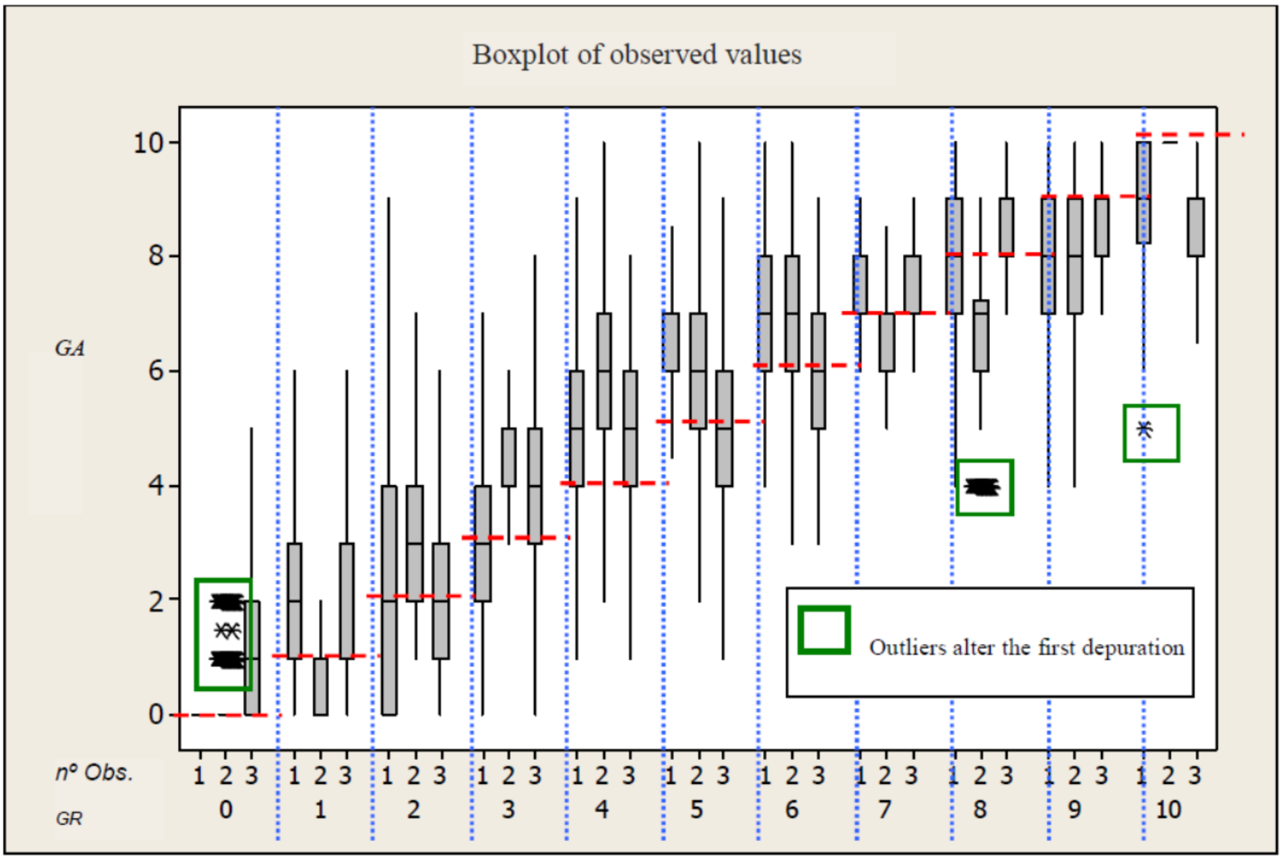

Figure 8. Boxplot after the first depuration of the outliers

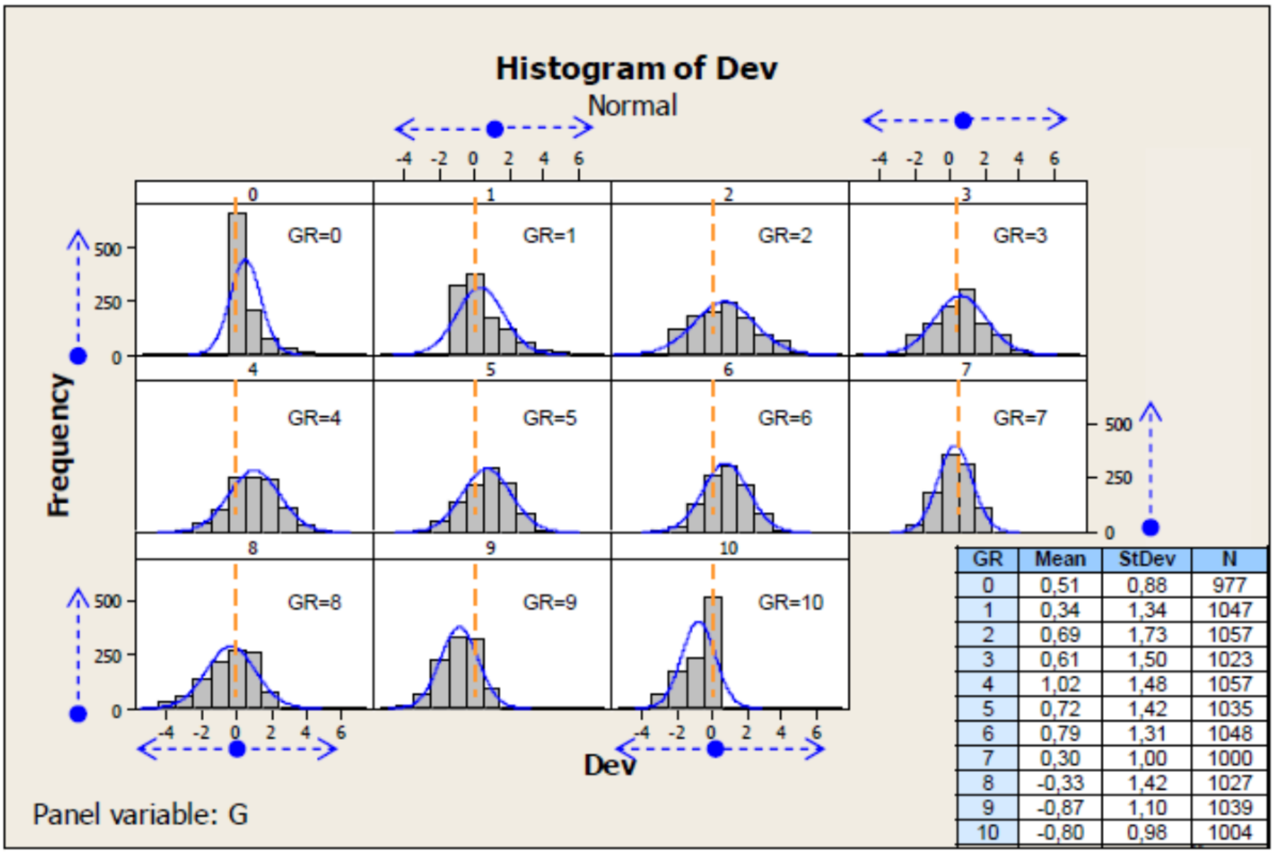

Figure 9. Histogram of frequencies of deviations (Dev) for each of the G values

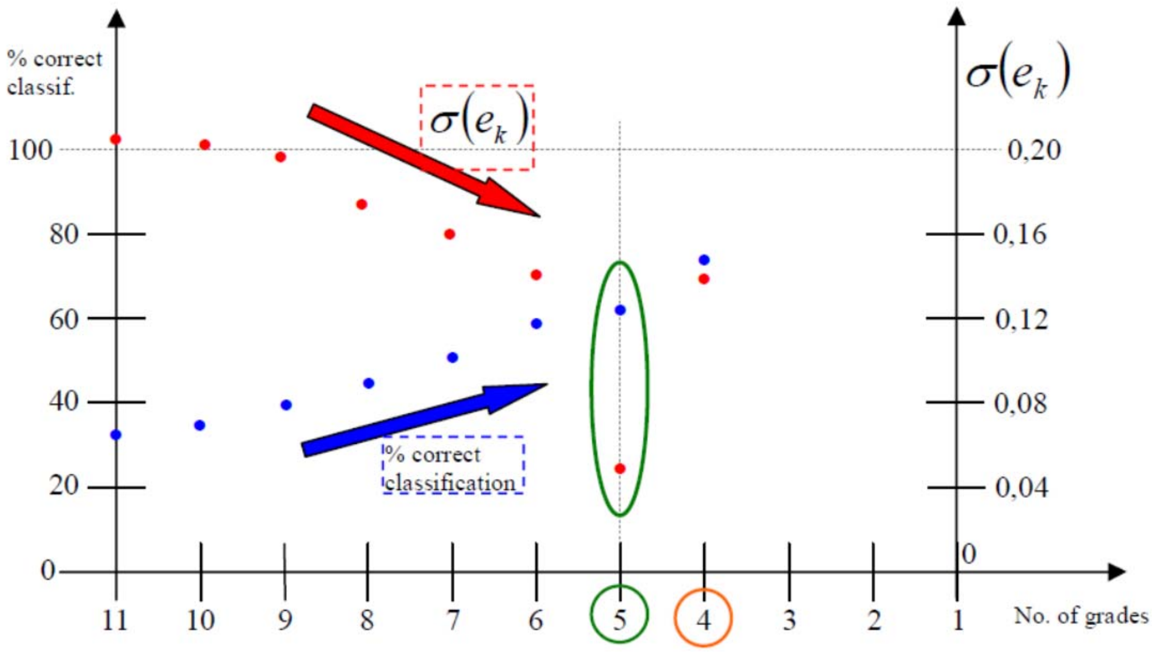

Figure 10. Simplification process of the scale