

Interlaboratory comparisons for hardness tests: interpolation of assigned values according to loading charges

Louis-Jean Hollebecq
Scientific and technical Manager

Table of content

1	Introduction	3
2	Technical backgrounds.....	4
2.1	Symbols.....	4
2.2	Relationship between Brinell and Vickers test results and loads used for their measurements.....	5
2.2.1	Bibliography	5
2.2.2	Contributions of scientific knowledge for the determination of assigned values (X_{pt})	8
2.2.3	Contributions of scientific knowledge for the determination of the uncertainty on assigned values ($U_{X_{pt}}$)	10
2.2.4	Contributions of scientific knowledge for the determination of the standard deviation for the proficiency testing (σ_{pt}), the standard deviation of homogeneity (σ_H) and the standard deviation of repeatability (σ_{rpt})	10
2.3	The Monte-Carlo method	11
3	Design of experiments	11
3.1	Introduction	11
3.2	Selection of models	12
3.3	Determination of associated uncertainties	13
3.3.1	Uncertainties on input scales.....	13
3.3.2	Uncertainties on output scales	14
3.4	Procedure to determine output scale parameters from input scale parameters.....	15
4	Results and discussions.....	16
4.1	Uncertainties on input σ_H	16
4.2	Effects of lack of homoscedasticity of residues in regression	19
4.3	Assigned value X_{pt}	23
4.3.1	Introduction	23

4.3.2	Preliminary survey focused on 2020, 2021 and 2022 Vickers results.....	23
4.3.3	Results from other ILC.....	27
4.3.4	Conclusions for the determination of X_{pt}	29
4.4	Standard deviation for the assessment of the bias σ_{pt}	32
4.4.1	Introduction	32
4.4.2	Results of survey on Vickers hardness test results of years 2020, 2021 and 2022.....	33
4.4.3	Results from other ILC.....	35
4.4.4	Conclusions for the determination of σ_{pt}	37
4.5	Uncertainty $u_{X_{pt}}$ on the assigned value X_{pt}	38
4.6	Standard deviation of repeatability σ_{rpt}	39
4.6.1	Introduction	39
4.6.2	Results of survey on Vickers hardness test results of years 2020, 2021 and 2022.....	40
4.6.3	Results from other ILC.....	41
4.6.4	Conclusions for the determination of σ_{rpt}	43
4.7	Standard deviation of homogeneity σ_H	43
4.7.1	Introduction	43
4.7.2	Results of survey on Vickers hardness test results of years 2020, 2021 and 2022.....	43
4.7.3	Results from other ILC.....	44
4.7.4	Conclusion for determinations of σ_H	44
5	Proposals for calculation of output parameters from input scales	45
5.1	Recommendations for procedures to determine output scale parameters	45
5.2	Example of determination of output scale parameters from input scale parameters	46
5.3	Evaluation of these recommendations using the results of the years 2017 to 2023.....	47
6	Conclusions	49
7	References	49

Annex:

Tables of detailed results that were obtained in this study.

Abstract:

The possibility in a lab proficiency testing to assess hardness test results of a given Brinell or Vickers scale when an enough amount of test results is available for adjacent scales is investigated. 5 different methods are found to determine the assigned value and 2 different methods are found to determine the proficiency standard deviation, the repeatability standard deviation and the uncertainty on the assigned value. The best option depends on the interlaboratory testing conditions. A procedure is described to deal with the different possible options and to propose parameters to check the adequacy of each of them to help the choice of the most adapted one. An assessment of the results obtained with this procedure on CompaLab ILC results obtained during the 2017-2023 years was performed, leading to very small differences in the scoring of participants for available scales. When the size of the input data is large, output scoring is even likely to be more efficient than usual one.

1 Introduction

Interlaboratory tests are useful to check the proficiency of laboratories. However, for Brinell and Vickers hardness tests as defined in [1] and [2] for which several scales (corresponding to different loads) can be used, it is easy to get enough participants for some of these scales (typically HBW2,5/187,5 HV1 HV5 HV10 and HV30), but quite hard for most of the others. On the other hand, it is well known that test results obtained with different loads are quite similar but not exactly equal, as a well-known “indentation size effect” (ISE) applies for low loads. Moreover, performing proficiency tests request not only to determine an adequate assigned value but also to determine adequate reference values to determine the uncertainty of the assigned value and the acceptable scatter around the central value. As a matter of fact, ISO 13528 [3], which is the reference for the organisation of proficiency testing, proposes several options including the use of a model to determine reference values, but does not describe accurately how to manage this option. In details, CompaLab ILC provide several types of assessments:

- ✚ Bias, through z' scores, that request availability of a central value (X_{pt}), the uncertainty of this central value ($u_{X_{pt}}$), the homogeneity standard deviation (σ_H) and an interlaboratory standard deviation (σ_{pt});
- ✚ Repeatability, through z_r scores, that request availability of a repeatability standard deviation (σ_{rpt});
- ✚ Uncertainty claimed by the participant, through both a ζ score and a comparison to the standard deviation of reproducibility (σ_R), which request availability of respectively a central value (X_{pt}) and the uncertainty of this central value ($u_{X_{pt}}$), and the overall standard deviation of all results representing (σ_R).

To deal with these issues, this document:

- ✚ Explores the state of scientific knowledge about the relationship between the variations of measured Brinell and Vickers hardnesses according to the applied loads used to perform these measurements;
- ✚ Process the data of results of interlaboratory comparisons performed by CompaLab during the years 2017-2023;
- ✚ Propose methods to determine reference values needed to perform an interlaboratory comparison (i.e. homogeneity standard deviation (σ_H), central value (X_{pt}), uncertainty of this central value ($u_{X_{pt}}$), interlaboratory standard deviation (σ_{pt}), repeatability standard deviation (σ_{rpt})) for scales with a low number of participants (hereafter called *output scales*) from scales with a large number of participants (hereafter called *input scales*).

For doing so, several solutions are considered:

- ✚ Use general models to describe the relationship between hardness measurements and loads used to determine them;
- ✚ Use regression plotting to get the parameters of the models to determine the reference values for output scales from a certain number of input scales.

However, several problems need to be overcome, specifically in regression plotting:

- ✚ When a plotting of HD versus F is performed, the number of test results used to determine each HD result is different for each of the F values, and consequently, they should not be similarly weighted in the calculations of regressions;
- ✚ The existence of outliers in the test results is a major problem and needs to be dealt with. If all of them are used in the plotting, how to detect outliers if the totality of test results cannot be regarded as belonging to the same population?
- ✚ Participants are normally requested to produce their test results in repeatability conditions. But when results from several scales are used, this cannot anymore be satisfied. Deleterious statistical effects can appear if laboratories do not provide same number of results. This obviously occurs because they only provide results for scales for which they are equipped for, what differs from one lab to another;
- ✚ Formulas for calculating the uncertainty on regression parameters exist in statistical theories. However, these formulas are valid only if all residues are distributed on a same Gaussian distribution. It is here not the case because smaller loads obviously lead to greater standard deviations;
- ✚ When a model is likely to be relevant, how to check it as all available test results were used to build it up, and consequently, no more test results are available to check its validity.

All these issues are dealt with here after.

2 Technical backgrounds

2.1 Symbols and abbreviated terms

The symbols used in this document are listed in Table 1.

Table 1. List of symbols used in this document.

Symbol	Designation and comments
a_1 and a_2	Constants of the PSR model
d	Indentation size as defined in reference standards
D	Diameter of the ball in Brinell hardness test
E	Modulus of elasticity of the material
F	Load applied on the indenter during the hardness test <i>For ease of reading, throughout the whole document, F is expressed as in scale denominations and not in Newton as specified in the reference standards (e.g. 10 for HV10, 187,5 for HBW 2.5/187,5).</i>
h	Indentation depth
HBW	Brinell hardness test result. When it is related to a specific scale, the HBW symbol is completed with the related information as specified in ISO 6506-1
HD	Generic symbol for a hardness test result, that can be HBW or HV and any applied load
HD₀	Hardness measured with a load large enough to avoid any indentation size effect

Symbol	Designation and comments
HV	Vickers hardness test result. When it is related to a specific scale, the HV symbol is completed with the related information as specified in ISO 6507-1
k	Enlargement coefficient for computing an interval of confidence from a standard uncertainty
K	Constant linking HD, F and d^2 in the ISO 6506-1 and ISO 6507-1
Li	Constant of the PSR model expressed as function of F and HD
Me₁ and Me₂	Constants of the Meyer power law
n	Number of values
Nix	Constant of the ISE model expressed as function of F and HD
N_p	Number of participants to the PT
N_s	Number of input scales
P₁ and P₂	Constants of the Meyer power law model expressed as function of F and HD
r	Correlation coefficient of a regression
s_H	Estimate of σ_H (i.e. computed value of σ_H from the data of the ILC)
X_{pt}	Assigned value for the PT, as defined in ISO 13528
u_a, u_b and u_y	Standard uncertainties on regression coefficients a and b and on y when the regression is expressed in the form of $y = a \cdot x + b$
u_{xpt}	Standard uncertainty on assigned value for the PT, as defined in ISO 13528
Z_a and Z_{ua}	Parameters used to check whether the slope of a linear regression is significant or not
σ_H	Standard deviation of homogeneity of sets of samples sent to participants for the PT
σ_{pt}	Standard deviation for the proficiency testing, as defined in ISO 13528
σ_r	Standard deviation of repeatability
$\sigma_{rpt} - s_r$	Standard deviation taken as reference for the repeatability during the PT. σ_{rpt} (symbol chosen to be in line with ISO 13528) is also noted s_r where it is necessary to stress out that this value is an estimate of σ_r

In addition to these symbols, some acronyms are used as follows:

-  AV: assigned value;
-  IC: interval of confidence;
-  ILC: interlaboratory comparisons;
-  ISE: indentation size effect;
-  PT: proficiency tests.

2.2 Relationship between Brinell and Vickers test results and loads used for their measurements

2.2.1 Bibliography

Introduction:

The relationship between hardness test results and the loading used to determine them is an issue for a very long time. This is even more important since micro and nano indentation tests are performed to determine mechanical

properties of materials at microscopic scales, see [4]. Low charge hardness tests are particularly useful when very low quantities of material are available or to characterise coatings use to harden surfaces of mechanical pieces.

Indentation size effect (ISE):

In 1991, an “Indentation Size effect” (ISE) was identified by Atkinson [5], which abstract is as follows: “A recently-proposed description of “the indentation size effect” in terms of a size-dependent error in effective load is shown to be accurate and to be consistent with a finite plastic hinge zone at the perimeter of the indentation. The implication for the theory of indentation is that strain-hardening materials do not deform in a size-independent geometrically-similar manner. For Vickers hardness testing of metals, the notional plastic hinge zone is a few micrometers wide and has little effect on the hardness values when the diagonal length of the indentation is $> \sim 300 \mu\text{m}$, but causes a considerable increase in apparent hardness when the indentation is smaller. The magnitude of the effect is governed by the strain-hardening propensity of the indented metal.”

Li model:

In 1993, Li & al. [6] proposed a “proportional specimen resistance” (PSR) model which deals with elastic deformation and frictional effects that occurs during a hardness test, as described in Equation (1):

$$F = a_1 \cdot d + a_2 \cdot d^2 \quad (1)$$

where “ F ” is the applied load
 “ d ” is the size of the indentation
 a_1 and a_2 are constants.

Equation (1) can also be expressed in the form $F/d^2 = a_1/d + a_2$, which might be more relevant in some circumstances because:

- ✚ The equation of definition of Vickers hardness provided in [2] is of the form $HD = k \cdot F/d^2$;
- ✚ On the other hand, the equation of definition of Brinell hardness provided in [1] is of the form $HD = k \cdot F / \left(D^2 \cdot \left(1 - \sqrt{1 - d^2/D^2} \right) \right)$ with $0,24 \leq d/D \leq 0,6$. In an ILC, all participants find test results in the same range (typically better than $\pm 5\%$), leading to very similar d/D ratios for all of them (typically better than $\pm 0,1\%$). Then, the equation of definition for Brinell tests can also be regarded as close to the form $HD = k \cdot F/d^2$. In this formula, a_1 characterizes the load dependence of hardness and a_2 is a load independent constant.

With the alternate form of Equation (1), it can then be seen that $HD = k \cdot a_1/d + k \cdot a_2$. In this equation, $k \cdot a_2$ can be seen equal to HD_0 , the measured hardness when F is big enough to avoid any ISE, and $k \cdot a_1/d$ can be seen as modelling the ISE.

Nix and Gao model:

In 1998, Nix and Gao proposed a model to quantify this ISE, see [7]. They showed that the ISE for crystalline materials can be accurately modelled using the concept of geometrically necessary dislocations. The model led to the following characteristic form for the depth dependence on the hardness:

$$HD/HD_0 = \sqrt{1 + h^*/h} \quad (2)$$

where “HD” is the hardness for a given indentation depth “h”
 “HD₀” is the hardness for a hypothetical indentation of infinite depth
 “h*” is a characteristic length depending on the shape of the indenter, the shear modulus and HD₀.

This paper extensively discusses the role of strain, of strain gradient and movements of dislocations in the plastically deformed area of the metal while indentation takes place.

Meyer power law:

With regard to geometries of indentations this law was further transformed into a so-called Meyer power law, according to Equation (3) as follows.

$$F = Me_1 \cdot d^{Me_2} \quad (3)$$

where “F” is the load applied to perform the hardness test,
 “d” is the size of the indentation,
 Me₁ and Me₂ are constants.

This Meyer power law can also be written in the form $\log(F) = Me_1 \cdot \log(d) + Me_2$, that is particularly useful to determine Me₁ and Me₂ constants from experimental data.

The exponent Me₂ (Meyers index, named as “n” in the publication) is a measure of the ISE effect. When Me₂ = 2, the test result is independent to the applied load (because equations of definition of Brinell and Vickers hardnesses provided in [1] and [2] are of the form $HD = k \cdot F/d^2$ in which k is a constant). In most cases of micro and nano hardness tests, n > 2 corresponding to an ISE that increase the measured HD when the load decreases. Sometimes, Me₂ < 2. These situations are described as “reverse ISE”, as they cannot be described by the classical models exposed here upper.

Other contributions:

In 2002, Swadener & al. [8] found out that, for a spherical indenter, hardness is not affected by depth, but increases with decreasing sphere radius. They based a correlation for pyramidal and spherical indenter shapes on geometrically necessary dislocations and work-hardening. They extended the Nix and Gao indentation size effect model for conical indenters to indenters of various shapes and compared to the experimental results.

In 2003, X.J. Ren & al. [9] put in evidence a correlation between ISE and HD/E (hardness to modulus of elasticity ratio).

In 2013, Budarsia [10] took advantage of the possibility of steels to produce a large panel of HD/E ratios (as HD can be easily controlled by heat treatments without significantly changing E) to determine Me₁ and Me₂ constants of the Meyer power law as function of HD/E. This study is also of special interest for us because it uses materials similar to what CompaLab uses for its proficiency tests. Budarsia found a good adequacy of his results with upper described models and a linear relationship between a₂ of PSR model and HD/E ratio.

In 2022, Xiaozhen Li & al. [11] discussed related effects of ISE and of some imperfections (i.e. surface defects and geometrical defects of the indenter) on the HD measurements. They showed that these imperfections are likely to cause reverse ISE occurrence even when an usual ISE is expected.

Unsurprisingly, no article considers the impact of ISE and PSR on the performance of proficiency tests concerning hardness tests. As a matter of fact, the usual and easy way to organise PT on hardness tests is to request participants to use the same scale, with the disadvantages exposed in § 1. However, the literature enlightens our issue in several ways, as follows.

2.2.2 Contributions of scientific knowledge for the determination of assigned values (X_{pt})

The models describing the relationship between F and HD can be used to determine X_{pt} for an output scale from a series of input scales. However, these models can be more or less conflicting, depending on which phenomenon is predominant for the conditions in which the PT was organised. As conclusions on this issue, the following guidelines can be drawn:

- ✚ When the applied load is large enough to produce indentations size larger than 0,3 mm, we can expect that no ISE occurs and $HD \rightarrow HD_0$. This situation is the easy one, where HD can be regarded as independent from F ;
- ✚ When an ISE occurs, the 2 models described respectively by Equation (1) and Equation (3) are competing. Determining X_{pt} from data of other scales imply to check which model is the best adapted to the actual situation;
- ✚ Following [12], for Brinell test, the $0,102 \cdot F/D^2$ ratio (where D is the diameter of the ball) can vary between 1 and 30, leading to geometrical effects in ISE. For this reason, input scales must all be of the same $0,102 \cdot F/D^2$ ratio, and the output scale must also be of this same ratio;
- ✚ Unfortunately, the d values are not available for the ILC organiser. Only F and HD values are available for the test data processing. These d values can obviously be computed from HD and F . ILC parameters could then be computed for them. However, it is then needed to transform them back into HD values, what is easy for X_{pt} but not easy at all for standard deviations. Consequently, Equations (1), (2) and (3) need to be transformed into relationship between HD and F to be useful for ILC.

Model of Li & al.:

Combining Equation (1) and equation $HD = K \cdot F/d^2$ (K constant) enables us to find Equation (4) as follows:

$$HD = HD_0 + a_1/d, \text{ then, by replacing } d \text{ with its expression from } HD = K \cdot F/d^2:$$

$$HD = HD_0 + \sqrt{a_1/K \cdot F} \cdot \sqrt{HD}, \text{ which can be written in the form:}$$

$$(\sqrt{HD})^2 - \sqrt{a_1/K \cdot F} \cdot \sqrt{HD} - HD_0 = 0,$$

which is a polynomial of degree 2 where \sqrt{HD} stands as the variable.

Classical formulas of algebra can be used to find out the (relevant) root of this equation, as follows:

$$HD = \frac{a_1}{4 \cdot K \cdot F} \cdot \left(1 + \sqrt{1 + \frac{4 \cdot HD_0 \cdot K \cdot F}{a_1}} \right)^2$$

that can be reformulated as follows:

$$HD = \frac{Li}{F} \cdot \left(1 + \sqrt{1 + \frac{HD_0 \cdot F}{Li}} \right)^2 \quad (4)$$

where "HD" is the hardness test result

" HD_0 " is the hardness measured with a load large enough to avoid any indentation size effect,

" F " is the load applied to perform the hardness test,

Li is a constant.

It shall be noted that the Li constant is homogeneous to HD.F, that is to say depends on the units that are used to express both HD and F.

It can be seen from that:

- ✚ When F is very large, $HD_0.F/Li \gg 1$, $HD = HD_0$;
- ✚ When F is very low, $HD_0.F/Li \ll 1$, $HD = 4.Li/F$;
- ✚ When $F/Li = 1$ (or $F = Li$), $HD = (1 + \sqrt{1 + HD_0})^2$.

Consequently, 2 ways can be explored to find out the Li constant:

1. Determine HD_0 by which any increase of F does not decrease the HD measured value. Then compute $HD = (1 + \sqrt{1 + HD_0})^2$. Then determine from the data the corresponding F value and determine Li from F value;
2. Plot HD as a function of F for very low charges. When F is low enough, the plot becomes a straight line and determine Li from the equation $HD = 4.Li/F$.

We need to choose F for which $F.HD_0/Li = 440$ to get $HD/HD_0 = 1,1$, which is the max ratio encountered during the performances of ILC (see § 4). In the interval $F.HD_0/Li \in [440; +\infty[$, Equation (4) can be approached with the Equation $Log(HD) = Log(HD_0) + 0,868/\sqrt{F.HD_0/Li}$, that is to say a straight line when the abscissa is plotted as $1/\sqrt{F}$ and the ordinate is plotted $Log(HD)$. This approximation can be reasonably used from $F.HD_0/Li = 15$ (error on HD less than 0,5%), corresponding to a ratio $HD/HD_0 = 1,667$.

Model of Nix & al.:

Considering Equation (2):

- ✚ Considering that, thanks to geometrical properties of indentations, h and h^* of Equation (2) are proportional to d^2 and to a hypothetical d^{*2} corresponding to h^* ;
- ✚ Considering a hypothetical hardness HD^* corresponding to the indentation depth h^* , from Equation (2) we can find that $HD^* = HD_0.\sqrt{2}$;
- ✚ Considering F^* , hypothetical loading force corresponding to HD^* , so that $d^{*2} = K.F^*/HD^*$;
- ✚ Considering that $\frac{h^*}{h} = \frac{d^*}{d}$, which is exactly true for Vickers testing and an approximation for Brinell testing;
- ✚ And combining Equation (2) and the upper enables us to find Equation (5) as follows:

$$\frac{h^*}{h} = \sqrt{\frac{F^*.HD}{F.HD_0.\sqrt{2}}}$$

then

$$\left(\frac{HD}{HD_0}\right)^2 = 1 + \sqrt{\frac{F^*.HD}{F.HD_0.\sqrt{2}}}$$

then

$$HD^4 - 2.HD_0^2.HD^2 - \frac{F^*.HD_0^3}{F.\sqrt{2}}.HD - HD_0^4 = 0$$

that can be reformulated as follows (using $Nix = F^*/\sqrt{2}$):

$$HD^4 - 2.HD_0^2.HD^2 - Nix.HD_0^3.HD/F - HD_0^4 = 0 \tag{5}$$

where "HD" is the measured hardness for a given load "F"

"HD₀" is the hardness measured with a load large enough to avoid any indentation size effect,

Nix is a constant depending on the shape of the indenter, the modulus of elasticity and HD₀.

This polynomial of degree 4 is not easy to handle. However, all coefficients are independent from F except the slope, which varies as function of $1/F$.

Model of power law:

Combining Equation(3) and equation $HD = k.F/d^2$ (k constant) enables us to find Equation (6) as follows:

$$\log(HD) = (P_1 - 1/Me_2). \log(F) + P_2/Me_2 \quad (6)$$

where "HD" is the hardness test result

" HD_0 " is the hardness measured with a load large enough to avoid any indentation size effect,

" F " is the load applied to perform the hardness test,

P_1 , P_2 and Me_2 are constants.

Conclusion concerning models:

In conclusion of the upper, only Equation (6) is easy to handle because it can be plotted as a straight line as soon as logarithmic scales are used for the abscissas and the ordinates.

To deal with Equation (4) and Equation (5), HD as function of $1/F$ and $\log(HD)$ as function of $1/\sqrt{F}$ should be plotted.

2.2.3 Contributions of scientific knowledge for the determination of the uncertainty on assigned values ($u_{x_{pt}}$)

$u_{x_{pt}}$ is composed of two contributions:

- ✚ A contribution coming from the adequacy of the model to the situation encountered during the PT performance;
- ✚ A contribution coming from the uncertainties related to the data used in the model and how they combine each other in this model (typically, how uncertainties on data used to plot a regression straight line combine to produce a given uncertainty on a given X_{pt} somewhere on the regression straight line).

Concerning the first one, most of publications listed here upper show very good correlation coefficients on their figures. It shall be noticed that tests that were used to plot them are likely to have been obtained in repeatability conditions, that is to say, in far better conditions than the reproducibility ones that we normally face during a PT performance. On the other hand, the range of the parameters considered in these publications is in most cases equal or wider than encountered in PT performance (for example, the F range studied varies in a ratio of 1 to 1000 in both cases). As a conclusion, no formal way to determine the contribution of adequacy of models on uncertainties can be drawn for bibliography but, at this stage, this contribution seems to be minor.

Concerning the second one, the issue is completely depending on statistical techniques and will be dealt with further on in § 3.3.2.

2.2.4 Contributions of scientific knowledge for the determination of the standard deviation for the proficiency testing (σ_{pt}), the standard deviation of homogeneity (σ_H) and the standard deviation of repeatability (σ_{rpt})

As all studies described in bibliography were conducted within a single laboratory (or, if not, no information is available about this respect), no relevant direct information can be taken from scientific literature concerning this issue. However, it shows the main causes for ISE to occur (movements of dislocations, surface of the test specimens, good or bad condition of the indenter, ...) that can easily be classified in material dependent and/or laboratory dependent and/or machine dependent and/or operator dependent issues, which can more or less affect each of

the standard deviations relevant for proficiency testing. Conditions can significantly vary from a PT to another, in accordance with PT provider instructions to participants. For example, surface preparation of test specimens may be performed by the PT provider or let to participants. According to the option, the related scatter will be included in the homogeneity standard deviation or in the interlaboratory standard deviation.

In the same way, no relevant direct information can be taken from scientific literature concerning the issue of homogeneity of samples. In particular, no conclusion can be made about whether the standard deviation of homogeneity has some chance to be independent to loading conditions or not. However, the relationship between movements of dislocations in the metal and ISE should be considered with respect to this issue.

As a conclusion, no valuable information was found in the literature concerning any possible indentation size effect on standard deviations used in a PT. However, it is well known that uncertainty on hardness tests increase with decreasing load. We can then guess that the standard deviations relevant for a PT follow the same way.

2.3 The Monte-Carlo method

The Monte-Carlo methods are a large category of algorithms that use random numerical realisations of a given model. They are often used to solve mathematical or physical problems, difficult or impossible to solve by other methods. For a survey of the history and applications of the Monte-Carlo methods, see for example [13].

Hard calculations are needed to solve several of the issues of this document (interval of confidence on σ_H , uncertainties on regressions for which residues are not uniformly distributed), that could not be solved with another method Monte-Carlo.

However, using Monte-Carlo methods requests to use a model that represents reasonably well the situations that we want to deal with. To achieve this, an appropriate modelling is needed. This is obviously not a problem in the present case.

Using the Monte-Carlo methods also requests to use random input values. When several random values are necessary to produce one Monte-Carlo result and when correlations between them apply in the phenomenon to be modelize, these correlations must be incorporated in the input values of the computations. That can be a bit difficult to do properly. In our case, only one random variable is needed and no correlation is to be feared.

To assure the validity of the conclusions, the random series need to be numerous enough, depending on many factors. In our study, we computed series of 10^6 to 10^8 numbers for each situation. Each of these series was divided in sub-groups. This enables us to compute the standard deviation of the parameters that we are determining. This standard deviation is then used to determine an interval of confidence (IC) for each of the determinations, with an enlargement coefficient equal to 2. We decided to stop the Monte-Carlo processes when we considered that the IC is small enough for each particular issue to solve. These IC are provided in the results when relevant and drove the rounding of the results that are provided by this study.

3 Design of experiments

3.1 Introduction

For each of the parameters needed to conduct an ILC (see § 1), the following steps were followed:

- ✚ Compare several models that can adequately represent the phenomenon and select the most adequate one;

- ✚ In accordance with the selection, find a model to determine the associated uncertainties, in order to be able to validate the selected model;
- ✚ Define a procedure on how to determine output scale parameters from input scale parameters with respect to the number of available input scales and possibly with respect to how they combine.

3.2 Selection of models

Forms of appropriate models for X_{pt} were extensively discussed in § 2.2.2. The results of this survey were used to select which models to check.

For standard deviations, several models were checked:

- ✚ Power law expressed in its logarithmic form, in order to plot linear regression;
- ✚ Linear laws;
- ✚ Constant laws (i.e. the parameter is not significantly affected by the scale).

To plot the curves, several options are possible:

1. Use the participants results altogether;
2. Use the regressions of each individual participant;
3. Use the parameters obtained from the data processing of each of the input scales;
4. Use the parameters obtained from the data processing of each of the input scales weighted with the number of test results used for each of them.

In **option 1**, outliers cannot be detected (because it cannot be decided whether a significant gap is due to a gross error or to an ISE effect). As outlying results have a very strong effect on ILC results, this option was abandoned.

Option 2 consists in:

1. Compute the regression parameters participant per participant;
2. Determine the global regression parameters by averaging them from all regression results.

This method has the advantage of erasing the “lab effect” in the regression. However, it significantly lowers the size of the data available to compute parameters (only results from participants that have provided results for several scales can be taken into account). This option was then also abandoned.

Option 3 consists in:

1. Process the test results scale per scale as usually performed;
2. Plot the parameters determined for each scale as function of the applied load.

This option has the advantage of avoiding the deleterious effect of outliers. However, the parameters of each scale are similarly weighted in the regression even if they were not obtained from the same number of participants, and consequently, are not affected by the same uncertainty.

Option 4 consists in same operations than option 3, but at step 2, the value obtained on each scale is input in the regression as many times as the number of test results used to obtain them. This option keeps the advantages of option 3 but suppresses its disadvantage. That is why this option 4 was used to plot regression lines as exposed in § 4.

3.3 Determination of associated uncertainties

3.3.1 Uncertainties on input scales

Uncertainties on parameters for input scales are needed to check the validity of the model used to determine output scale parameters. If the regression line does not cut satisfactorily the segments representing the uncertainties on input scale parameters, the selected model cannot be regarded as satisfactory.

In CompaLab PT concerning hardness tests, each participant receives 3 samples and provides 2 test results for each sample, i.e. 6 test results per participant. Consequently, for input scales σ_{pt} values are computed as follows:

1. Compute the standard deviation obtained by each participant on each sample (3 SD for each participant);
2. Compute the repeatability of the participant from these 3 standard deviations, using the classical formula

$$s_r = \sqrt{\sum s_i^2 / 3};$$

3. Compute σ_{rpt} with Algo S (see [3]) applied to the series of s_r of each participant.

The resulting number of degrees of freedom is approximately $(2 - 1) \cdot 3 \cdot N_p = 3 \cdot N_p$, where N_p is the number of participants (the approximation comes from the use of the robust Algo S, that replaces possible outliers by virtual values computed from the initial data).

In the same way, for input scales σ_H values are computed as follows:

1. Compute the mean value obtained by each participant on each sample (3 \bar{x} per participant);
2. Compute the homogeneity standard deviation for each participant from these 3 mean values (using the classical formula $s_{H,i} = \sqrt{\sum (\bar{x}_i - \bar{\bar{x}})^2 / 2}$);
3. Compute w_H with Algo S (see [3]) applied to the series of s_H of each participant;
4. Compute σ_H with the classical formula, as follows in Equation (7):

$$s_H = \sqrt{w_H^2 - \sigma_{rpt}^2 / 3} \quad (7)$$

The resulting number of degrees of freedom is approximately $(3 - 1) \cdot N_p = 2 \cdot N_p$, where N_p is the number of participants.

Uncertainty on X_{pt} values:

$u_{X_{pt}}$ values are provided by the data processing of input scales in accordance with recommendations of ISO 13528 (classical formula with $k = 2$ as enlargement coefficient and an extra coefficient of 1,25 because of the use of Algo A).

Uncertainty on σ_{pt} values:

σ_{pt} values are computed in accordance with recommendations of ISO 13528 (i.e. Algo A on mean values of participants). Consequently, the resulting number of degrees of freedom is approximately $N_p - 1$. The approximation is also due to the use of Algo A. The classical formula for the IC on a standard deviation estimate is:

$$s = \sigma \sqrt{\frac{\chi_{n-1}^2(P)}{n-1}} \quad (8)$$

where s is the estimate of a standard deviation computed from the series,
 σ is the standard deviation to be estimated,
 P is the is the cumulated probability used for the IC
 and n is the number of values used for computing the standard deviation.

The uncertainties on σ_{pt} values were computed with the Equation (8) in which:

- ✚ n value is $N_p - 1$, where N_p is the number of participants;
- ✚ P values chosen equal to 0,025 and 0,975, corresponding to a bilateral IC of 95%.

A numerical simulation with the Monte-Carlo method confirmed the validity of this method of determination.

Uncertainty on σ_{rpt} values:

σ_{rpt} values are computed as explained here upper, with a number of degrees of freedom equal to $3 \cdot N_p$. We can use Equation (8) in which:

- ✚ n value is $3 \cdot N_p$, where N_p is the number of participants;
- ✚ P values chosen equal to 0,025 and 0,975, corresponding to a bilateral IC of 95%.

A numerical simulation with the Monte-Carlo method confirmed the validity of this method of determination.

Uncertainty on σ_H values:

σ_H values are computed as explained here upper, with a number of degrees of freedom equal to $2 \cdot N_p$. We can use Equation (8) in which:

- ✚ n value is $2 \cdot N_p$, where N_p is the number of participants;
- ✚ P values 0,025 and 0,975, corresponding to a bilateral IC of 95%.

However, numerical simulations with the Monte-Carlo method did not confirm the validity of this method of determination because of the effect of the $\sigma_{rpt}^2/3$ term in Equation (7). A study about this issue was then performed, which results are provided in § 4.1.

3.3.2 Uncertainties on output scales

Introduction

The uncertainty on X_{pt} in output scales is the only one that is needed to process the data in a proficiency test in accordance with ISO 13528 (see § 2). Uncertainties on other parameters are useful for checking the validity of the selected model but are not needed to assess the results of participants.

Uncertainty on output values from regressions

As, for output scales, X_{pt} is determined by using a regression plot, we need to refer to statistical knowledge to determine $u_{x_{pt}}$. When the regression curve of (x_i, y_i) couples is a straight line and residues belong to a same Gaussian distribution which mean value is 0 and which standard deviation is σ_{res} , the standard uncertainties on regression parameters are:

- ✚ On the slope:

$$u_a = \frac{s_{res}}{\sigma_{x_i} \cdot \sqrt{n_v - 1}} \quad (9)$$

- ✚ On the ordinate at origin:

$$u_b = s_{res} / \sqrt{N_s} \quad (10)$$

where s_{y_i} is the estimated variance of y_i values,
 σ_{x_i} is the variance of x_i values,
 and N_s is the number of input scales.

It shall be noted that:

1. Equation (10) is simplified from the general equation because abscissas in our study are related to hardness scales, which are exactly defined and not coming from an estimation. Consequently, σ_x and s_x are always exactly same;
2. Even when many (x_i, y_i) couples introduced in the regression as suggested as in option 4 of § 3.2, these values are not independent, and then n_v must be limited to the number of input scales.

Standard uncertainty on a Y_i value computed with the regression parameters for an abscissa X_i is:

$$u_{Y_i} = \sqrt{u_a^2 \cdot (X_i - \bar{X})^2 + u_b^2} \quad (11)$$

Effect of the absence of homoscedasticity of residues

In practice, in our study, for technical reasons, the standard deviations of residues are larger for scales with low charge than scales with high charge. The hypothesis of homoscedasticity is not fulfilled and Equation (11) cannot be validly used to compute u_{Y_i} . To address this problem, we performed numerical simulations using the Monte-Carlo method. Results are provided in § 4.2.

3.4 Procedure to determine output scale parameters from input scale parameters

Several models are proposed and discussed in § 4. According to the type of test (Brinell or Vickers, the latter having low charges and normal charges scales), to the products (soft in the range of HD 150-250 or hard in the range of HD 500-750) and to the years, the number of available input scales can vary from 1 to 6 or more. Of course, the suitability of the model selected to determine parameters for output scales strongly depends on this number.

Typically, the number of constants of a possible model cannot be more than the number of available input scales. For example, if a polynomial of degree 2 is used as model, at least 3 input scales need to be used and checking the adequacy of it requests at least 4 input scales. Of course, models with more parameters are likely to provide better estimates of output parameters than those with less parameters. The general rule is then that a model that includes n parameters should not be used if less than $n-1$ input scales are available. This general rule may be not followed in some cases which are discussed in § 4. In these cases, recommendations are provided to check the validity of the selected model.

The performance of CompaLab ILC supposes to determine X_{pt} , $u_{X_{pt}}$, σ_{pt} , σ_{rpt} and σ_H . For each of these parameters, the following steps were followed:

-  Overview of the results of the related CompaLab ILC between 2017 and 2023;
-  Evaluation of the adequacy of several possible models for each of these ILC;
-  Proposal of recommendations to select the most appropriate model for each type of situation.

In order to assure total comparability of results, all ILC parameters were re-computed with the last version of the software, what can lead to little differences with the formerly published ones, in particular concerning σ_H .

For ILC concerning low hardness of 2023, samples from 2020 and 2021 were used. Thanks to that, a larger range of input scales and test results are available for these ILC, in particular for Vickers hardness tests for which low load scales are also available.

The list of PT sessions which results were used in this study is enclosed in Table 2 as follows:

Table 2. List of PT sessions used in this study.

Product	Type of hardness	Year	Input scales
Hard metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000
Hard metal	HV	2017	HV5 - HV10 - HV30
Soft metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000
Soft metal	HV	2017	HV0,1 - HV0,3 - HV1 - HV10
Hard metal	HBW	2018	HBW 2,5/187,5 - HBW 10/3000
Hard metal	HV	2018	HV5 - HV10 - HV30
Soft metal	HBW	2018	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000
Soft metal	HV	2018	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30
Hard metal	HBW	2019	HBW 2,5/187,5 - HBW 10/3000
Hard metal	HV	2019	HV5 - HV10 - HV30
Soft metal	HBW	2019	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000
Soft metal	HV	2019	HV0,1 - HV0,1 - HV1 - HV5 - HV10 - HV30
Hard metal	HBW	2020	HBW 2,5/187,5
Hard metal	HV	2020	HV5 - HV10 - HV30
Soft metal	HBW	2020	HBW 1/30 - HBW 2,5/187,5 - HBW 2,5/187,5 - HBW 10/3000
Soft metal	HV	2020	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30
Hard metal	HBW	2021	HBW 2,5/187,5
Hard metal	HV	2021	HV5 - HV10 - HV30
Soft metal	HBW	2021	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000
Soft metal	HV	2021	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30
Hard metal	HBW	2022	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000
Hard metal	HV	2022	HV5 - HV10 - HV30
Soft metal	HBW	2022	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000
Soft metal	HV	2022	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30
Hard metal	HBW	2023	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000
Hard metal	HV	2023	HV5 - HV10 - HV30

In conclusion, a procedure is proposed to help the selection of the most adapted model for each encountered situation.

4 Results and discussions

4.1 Uncertainties on input σ_H

It was identified in § 3.3.1 that uncertainties related to the determination of σ_H cannot be properly computed with Equation (8). This issue was studied using the Monte-Carlo method. The corresponding results are provided in Figure 1 and detailed results in Table A.1.

When σ_H/σ_{rpt} is less than a certain ratio depending on N_p , the numerical simulation shows that the lower limit of the IC95% is 0, whatever σ_H , see Figure 1.

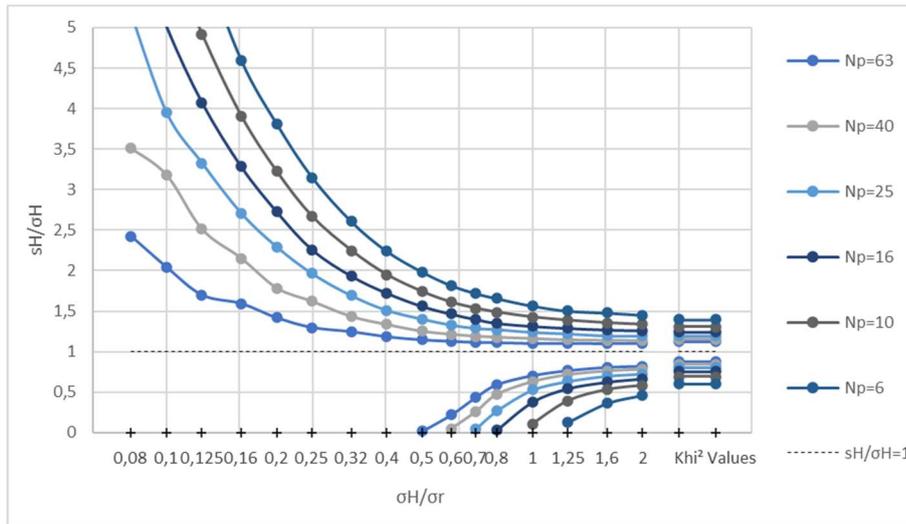


Figure 1: Lower and upper limits of the IC95% of s_H/σ_H for N_p participants, 3 samples per participant, 2 test results per sample.

Several statements can be made from these results:

1. Equation (8) cannot be used to determine s_H/σ_H ratios when $\sigma_H/\sigma_r < 2$. When $\sigma_H/\sigma_r \geq 2$, the $\sigma_{rpt}^2/3$ term in Equation (7) becomes negligible and Equation (8) can be used to determine s_H/σ_H ratios. However, this situation is normally never encountered in practice;
2. Here upper, the Monte-Carlo method was used to check whether the distribution of s_H/σ_H ratios can be found with Equation (8). However, in practice, we get s_H values and we need an IC on σ_H . Consequently, we need to know σ_H/s_H rather than s_H/σ_H . Of course, upper limits of s_H/σ_H are lower limits of σ_H/s_H and vice-versa. However, when $s_H = 0$, σ_H/s_H ratios become infinite, and no upper limit can be computed by this method. As these situations occur for “low values” of σ_H/σ_r , we can solve the problem in the following way: when $s_H = 0$, knowing N_p we can find from Figure 1 an absolute maximum for σ_H/σ_r and deduce an upper limit for σ_H from the estimate of σ_r ;
3. In the same way, for each particular situation, we need σ_H/σ_r ratio to locate the abscissa to use in Figure 1, but the figure that is actually available is s_H/s_r (computed from estimations of σ_H and σ_r).

A second round of numerical simulations was then performed to determine limits of IC for σ_H as a function of s_H/s_r and N_p . Because:

1. IC on σ_H appear to be very large;
2. They are needed only to check whether their impact on σ_{pt} is not too large;
3. It almost never happens for hardness tests;
4. Reference standards do not even raise the question of the reliability with which σ_H is determined;

we only considered empiric formulas that can only be used in the context of this study.

However, this issue is quite important for the reliability of PT programs and will be more comprehensively studied in a further work.

Anyway, performing the Monte-Carlo method to this issue requests the generation of a set of σ_H/σ_r random values for which s_H/σ_H and s_H/s_r ratios are computed and plotted to each other. Of course, the way in which this set is constituted influences the final results, especially the ends of the plotted curves. For this study, we decided to use

random σ_H/σ_r values distributed on a log normal distribution with $\sigma_H/\sigma_r = 0,8$ as mean value and 25% as standard deviation (it produces 99,7% of σ_H/σ_r random values in the interval $[0,15;4,5]$).

The results of it are displayed in Figure 2 to Figure 4:

- ✚ Figure 2 shows clouds of s_H/σ_H values obtained for N_p from 6 to 63. It can be seen that values obtained are scattering around 2 segments. The first one joins the dots $[0;0]$ and $[1;1]$. The second one follows the equation $x = 1$ for $x > 1$. The scatter decreases when N_p increases. When σ_H and s_H are low, their ratio behaves oppositely in Figure 1 and in Figure 2;
- ✚ Figure 3 shows clouds of σ_H/s_r values obtained for N_p from 6 to 63. Envelope curves of the clouds can be used as acceptable upper limits of confidence for σ_H , as function of s_H/s_r , which values are available for the PT provider. On these figures, centiles 2,5% and 97,5% were computed for each section of σ_H/s_r values, and a regression curve in the form $IC = B \cdot 10^{(A \cdot s_H/s_r)}$ is drawn;
- ✚ Figure 4 shows the evolution of coefficients A and B of the Equation $IC = B \cdot 10^{(A \cdot s_H/s_r)}$ as function of N_p . A linear regression applies, enabling the formulation of empirical equations to determine upper and lower IC limits for σ_H as function of s_H/s_r and N_p , for 3 samples and 2 test results per sample for each participant.

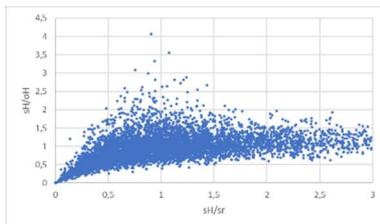


Figure 2.a: s_H/σ_H obtained on series with $N_p=6$ participants, 3 samples per participant, 2 test results per sample

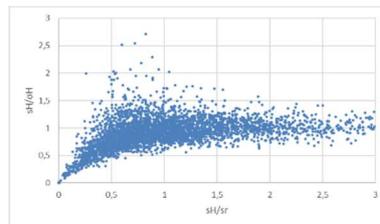


Figure 2.b: s_H/σ_H obtained on series with $N_p=16$ participants, 3 samples per participant, 2 test results per sample

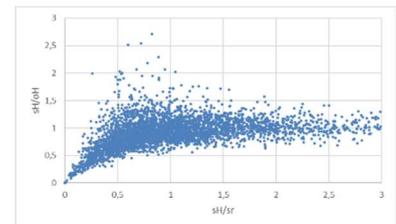


Figure 2.c: s_H/σ_H obtained on series with $N_p=63$ participants, 3 samples per participant, 2 test results per sample

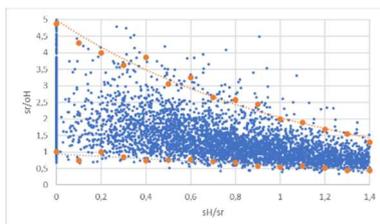


Figure 3.a: σ_H/s_r obtained on series with $N_p=6$ participants, 3 samples per participant, 2 test results per sample

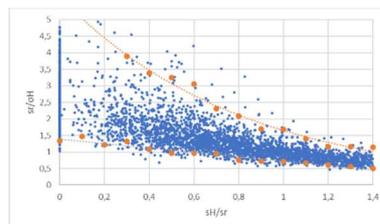


Figure 3.b: σ_H/s_r obtained on series with $N_p=16$ participants, 3 samples per participant, 2 test results per sample

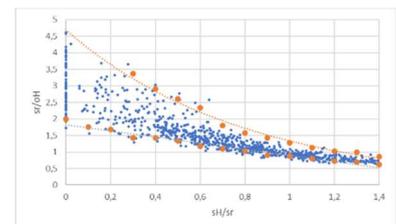


Figure 3.c: σ_H/s_r obtained on series with $N_p=63$ participants, 3 samples per participant, 2 test results per sample

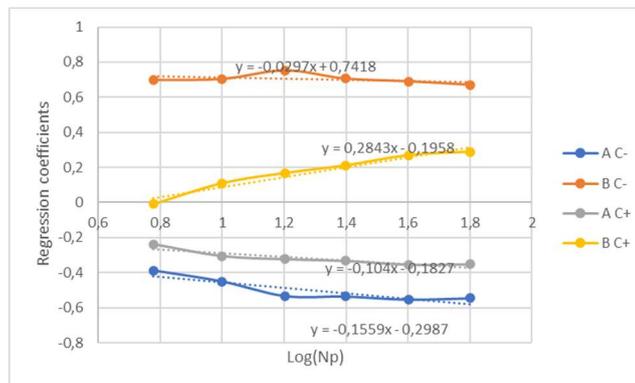


Figure 4: s_r/σ_H 2,5% and 97,5% centiles as function of $\log(N_p)$, 3 samples per participant, 2 test results per sample

As a conclusion, the determination of limits for σ_H can be performed as follows:

1. Determine s_H and s_H/s_r using Equation (7);
2. If $s_H = 0$, $IC(\sigma_H) = [0; IC^+]$ where IC^+ is computed with Equation (12);
3. If $0 < s_H/s_r < 1,5$, $IC(\sigma_H) = [IC^-; IC^+]$ where IC^- is computed with Equation (13) IC^+ is computed with Equation (12);
4. If $s_H/s_r > 1,5$, $IC(\sigma_H) = [IC^-; IC^+]$ where IC^- and IC^+ are computed with Equation (8).

$$IC^+(\sigma_H) = s_r \cdot 10^{((0,1 \cdot \log(N_p) + 0,18) \cdot s_H/s_r + (-0,28 \cdot \log(N_p) + 0,2))} \quad (12)$$

Where:

$IC^+(\sigma_H)$ is the upper limit 97,5% of the interval of confidence on the homogeneity standard deviation and N_p is the number of participants to the PT program.

$$IC^-(\sigma_H) = s_r \cdot 10^{((0,16 \cdot \log(N_p) + 0,3) \cdot s_H/s_r + (0,03 \cdot \log(N_p) - 0,74))} \quad (13)$$

Where:

$IC^-(\sigma_H)$ is the lower limit 2,5% of the interval of confidence on the homogeneity standard deviation and N_p is the number of participants to the PT program.

In practice, s_H/s_r often belongs to the interval $[0,5;1,5]$ and N_p in the interval $[10;25]$. Consequently, with respect to the upper, σ_H is most often underestimated and the IC on it is generally quite large.

4.2 Effects of lack of homoscedasticity of residues in regression

In our study, for technical reasons, the standard deviations of residues are larger for scales with low charge than scales with high charge. The hypothesis of homoscedasticity is not fulfilled and Equation (11) cannot be validly used to compute u_{Y_i} . To address this problem, we performed numerical simulations using the Monte-Carlo method to check the validity of the computed $u_{x_{pt}}$ values, for a selection of situations that represent the usual ones encountered during the CompaLab PT programs. Both Brinell and Vickers usual scales roughly follow a geometric progression, as shown in Table 3:

Table 3. Geometric progression of Brinell and Vickers usual input scales.

Brinell		Vickers	
Subsequent scales	Corresponding ratio	Subsequent scales	Corresponding ratio
30 and 187,5	6,25 ($\cong 10^{0,8}$)	0,1 and 0,3	3 ($\cong 10^{0,5}$)
187,5 and 750	4 ($\cong 10^{0,6}$)	0,3 and 1	0,33 ($\cong 10^{0,5}$)
750 and 3000	4 ($\cong 10^{0,6}$)	1 and 5	5 ($\cong 10^{0,7}$)
-	-	5 and 10	2 ($\cong 10^{0,3}$)
-	-	10 and 30	3 ($\cong 10^{0,5}$)

It follows that x values used in regressions that use $\log(F)$ as abscissas distribute quite regularly. When Nix or Li model are used, residues remain related to the logarithmic progression of scales, while abscissas are to be transformed from $\log(F)$ to $1/F$ and $1/\sqrt{F}$ respectively (see Table 4).

It can be seen from results (see § 4.4 and § 4.6) that residues can be modeled with a power law model (equation of the form $\sigma_{res,F} = a \cdot 10^{-\log(F)/b}$ where a and b are constants, F is the load applied during the hardness test).

We can expect that these larger residues for low charge have two effects:

- ✚ Decreasing the centre of gravity of the regression, because results the results of low charges introduce more scatter than results from high charges;
- ✚ Enlarging the intervals of confidence of regression parameters.

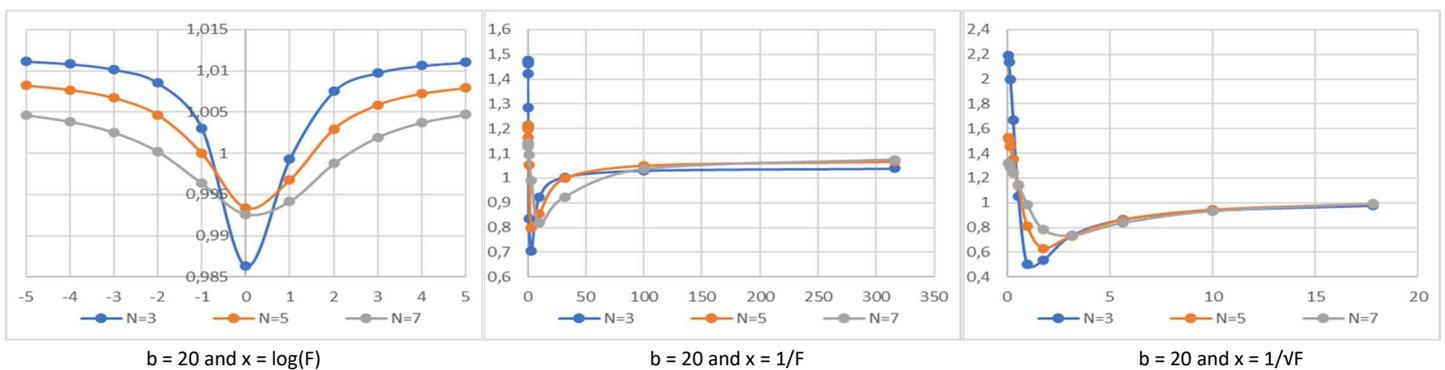
Numerical simulations using the Monte-Carlo method were performed to check this. These simulations involved regressions computed from 3, 5 and 7 “y” results centred on 0, with standard deviations on residues on “y” varying as $10^{x/b}$ with b values varying from 3 to 20. As an example, the case for which with b = 10 is shown in Table 4:

Table 4. Typical example of abscissas and related standard deviations of residues encountered in the ILC, for which b = 10.

x = log(F)	-5	-4	-3	-2	-1	0	1	2	3	4	5
x = 1/F	316,2	100	31,62	10	3,162	1	0,316	0,1	0,031	0,01	0,003
x = 1/√F	17,78	10	5,623	3,162	1,778	1	0,562	0,316	0,178	0,1	0,056
σ_{res}	3,162	2,512	1,995	1,585	1,259	1	0,794	0,631	0,501	0,395	0,316

Option 4 of § 3.2 leads to compute regressions with groups of same (x,y) values, each of them representing an input scale. The number of (x,y) couples is intended to weight the input values, and consequently, differ for each group. However, no significant differences were found between the simulations where the number of test results are same for all groups and the situations where the number of test results are distributed symmetrically around a central value (situations that are usually encountered in practice). The calculations were then simplified into single couples, each of them representing a full group of (x,y) couples.

Results are provided in Figure 5 here after. They are displayed as ratios $u_y/u_{y,th}$ where u_y is the result from the Monte-Carlo calculations and $u_{y,th}$ is the theoretical uncertainty that can be computed with Equation (11). $u_y/u_{y,th} = 1$ then signifies that no difference occurs between theoretical and actual u_y , i.e. the lack of homoscedasticity has no effect on uncertainty on the y-value in the corresponding case.



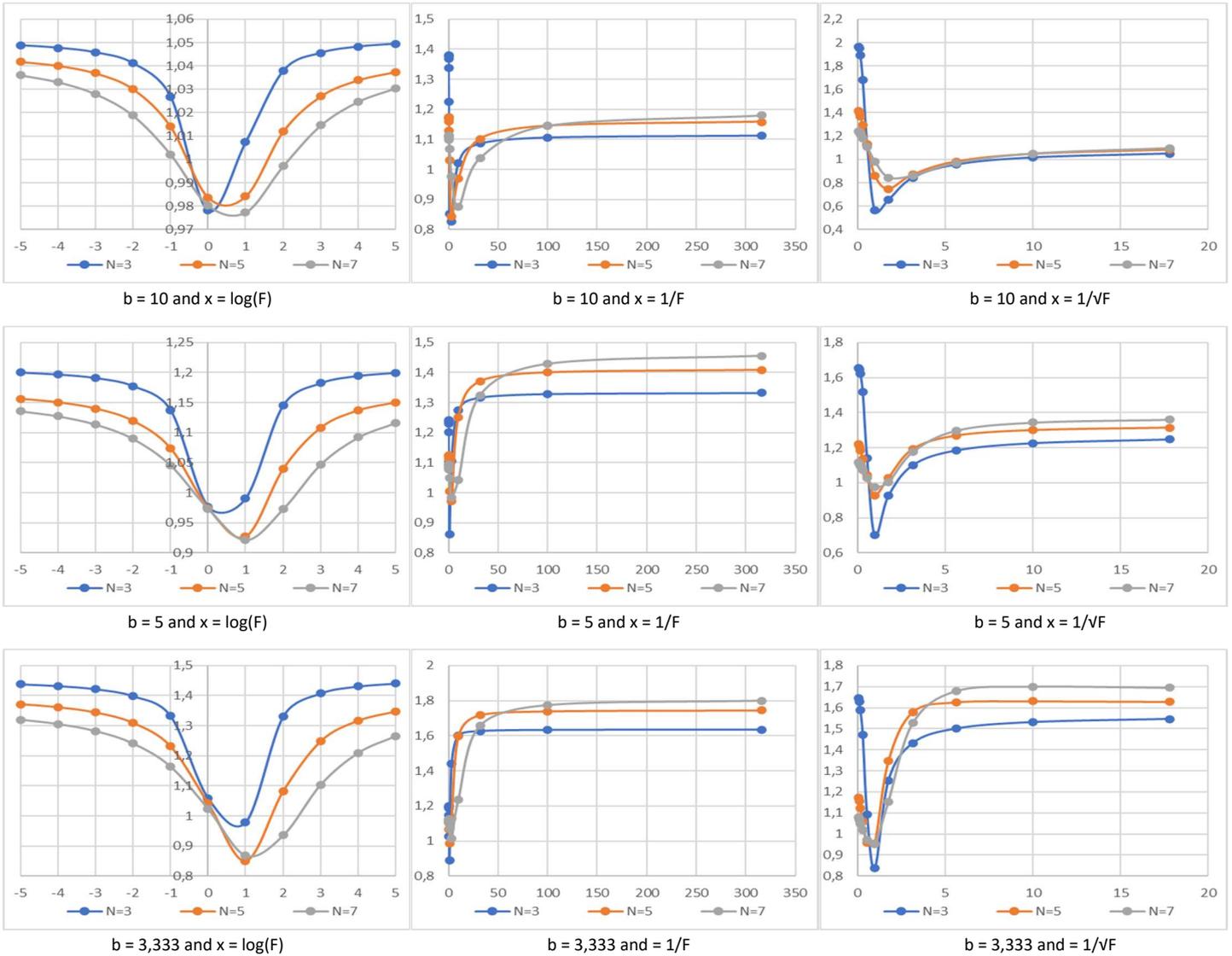


Figure 5: u_y from Monte-Carlo calculations over u_y from Equation (11) as function of N (number of x values used in the regression) b and the form of the abscissas.

Conclusions when abscissas are in the form of log(F) (use of the power law model):

The effect of varying residues as function of F on the global uncertainty as computed with Equation (11) can be evaluated with the empirical Equation (14), as follows:

$$\frac{u_y}{u_{y,th}} = \text{Max} \left(0,8; K \cdot \left(A - \frac{B \cdot \text{Sign}(x - \Delta x)}{x - \Delta x} \right) \right) \tag{14}$$

- where u_y is the true uncertainty on y_i values,
- $u_{y,th}$ is the theoretical uncertainty on y_i values, computed with Equation (11),
- K is a coefficient representing the global increase of uncertainties, computed with Equation (14).a,
- A is a coefficient computed with Equation (14).b
- B is a coefficient computed with Equation (14).c
- Δx is a coefficient representing the move of the centre of gravity of x_i , computed with Equation (14).d.
- and $\text{Sign}(x - \Delta x)$ is the sign of $x - \Delta x$, i.e. -1 if $x - \Delta x < 0$ and +1 if $x - \Delta x > 0$.

$$K = 1 + 10^{((0,1/b-0,12).N-2.\log(b)+0,8)} \quad (14).a$$

$$A = 1 + (0,73.\log(N) + 0,06). \left(\frac{1}{b} - 0,042\right) \quad (14).b$$

$$B = (2.\log(N) - 0,34). \left(\frac{1}{b} - 0,04\right) \quad (14).c$$

$$\Delta x = N. \left(\frac{1,45}{b^2} - \frac{1,05}{b} + 0,002\right) \quad (14).d$$

where N is the number of results included in the regression,
 b characterizes the variation of residues on y values (as $10^{x/b}$).

A minimum conventional value of 0,8 is introduced in Equation (14) because the main term $A - (B.\text{Sign}(x - \Delta x))/(x - \Delta x)$ is not valid when x is too close to the centre of the distribution, i.e. when $x \approx \Delta x$.

No significant effect (i.e. $u_y/u_{y,th} \approx 1$) occurs when $b > 20$. In practice, results showed that b is generally more than or equal to 10, for which corrections are between 0,98 and 1,05. When b is lower than 10, Equations (14) should be used to estimate the uncertainties on the computed values.

Conclusions when abscissas are in the form of 1/F or 1/VF (use of the Nix or Li model):

When the Nix or the Li models are used, we are facing 3 types of situations:

1. F is greater than the max F of the input scales. $u_y/u_{y,th}$ is then huge (up to 2) but u_y remains acceptable because the related abscissa remains very close to the central x_0 value. When such u_y are needed, coefficients $u_y/u_{y,th}$ equal to 1,5 for the Nix model and 2 for the Li model can be used whatever b and N ;
2. F belongs to the interval of F values used to compute the regression coefficients. Then, $u_y/u_{y,th}$ are usually between 0,9 and 1,1. A value pertaining to this interval can then be adopted for calculation of u_y , whatever b and N ;
3. F is lower than the min F of the input scales. $u_y/u_{y,th}$ becomes asymptotic to values that can be read from the right part of the corresponding figures and displayed in Table 5. However, x -values are then huge, leading to u_y too large to be relevant in most cases, especially for the Nix model. This is because, in those cases, all y -values used in the regression are cramped in the left part of the figure. This issue is also dealt with in § 5.2 with respect to results coming from actual ILC.

Table 5. Values of $u_y/u_{y,th}$ for large values of 1/F or 1/VF.

	x = 1/F (Nix model)				x = 1/VF (Li model)			
	b = 20	b = 10	b = 5	b = 3,33	b = 20	b = 10	b = 5	b = 3,33
N = 3	1,05	1,1	1,35	1,6	1	1,05	1,25	1,55
N = 5	1,05	1,15	1,4	1,7	1	1,05	1,3	1,63
N = 7	1,05	1,2	1,45	1,8	1	1,05	1,35	1,7

4.3 Assigned value X_{pt}

4.3.1 Introduction

Following the statements of § 2.2, several models were checked by using the results of the recent CompaLab ILC:

- ✚ The power law model, plotting $\log(HD)$ as a function of $\log(F)$;
- ✚ All results of a single ILC are coming from a same product. Then, differences between the HD results are not very important (usually less than 10%) and $\log(HD)$ is almost proportional to HD. Moreover, handling standard deviations related to log values is a bit more complicated than those from direct values. Then, plotting HD as a function of $\log(F)$ was also checked in order to check whether the power law model can be simplified thanks to that;
- ✚ The Nix & al. model, plotting HD as a function of $1/F$;
- ✚ The Li & al. model, plotting $\log(HD)$ as a function of $1/\sqrt{F}$.

4.3.2 Preliminary survey focused on 2020, 2021 and 2022 Vickers results

A first survey of how models fit was performed on Vickers results of years 2020, 2021 and 2022 for which a large quantity of results (respectively 166, 218 and 139), and of input scales (HV0,1 HV0,3 HV1 HV5 HV10 and HV30) are available. The results of this first survey are available in Figure 6.a to c.

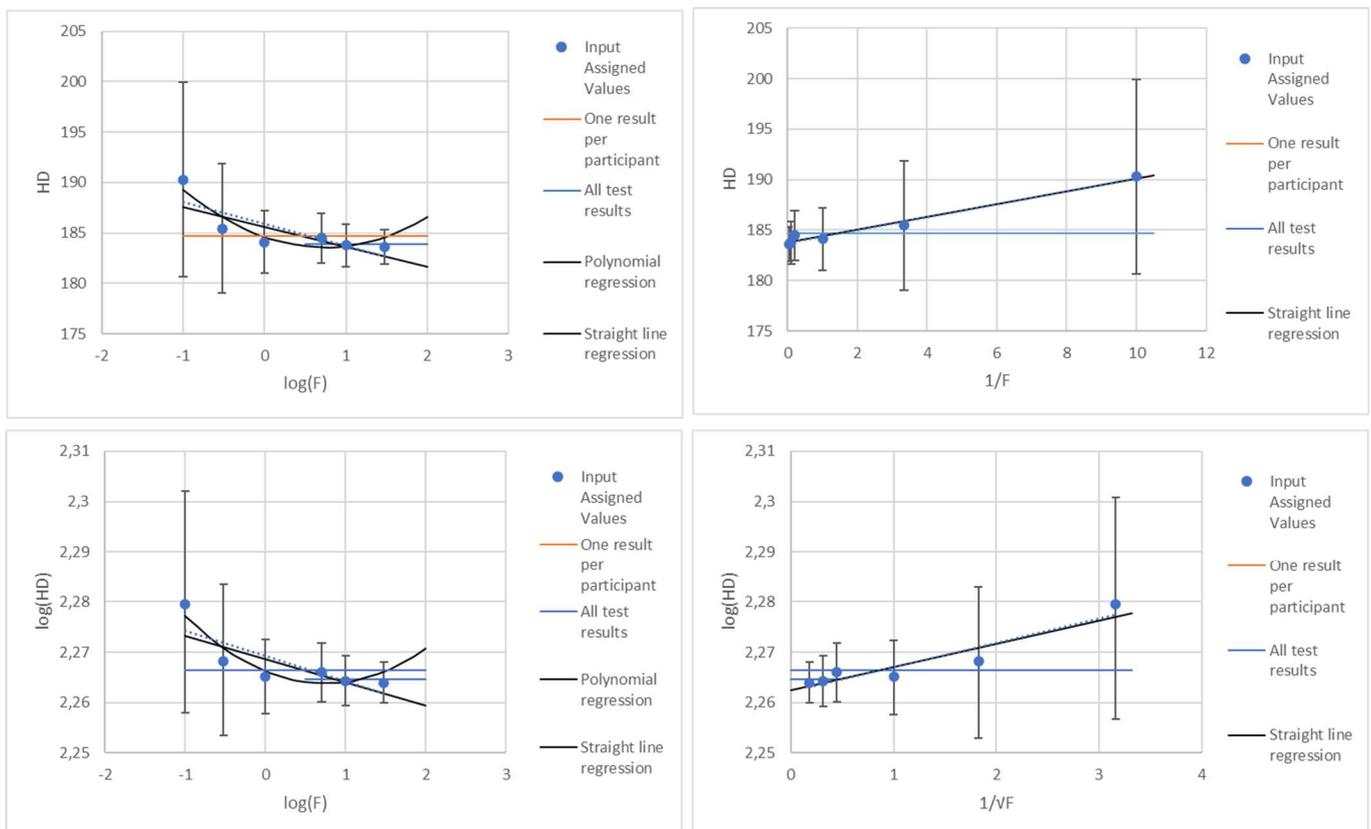


Figure 6.a: Assigned values as function of scales – 2020 Vickers hardness test results.

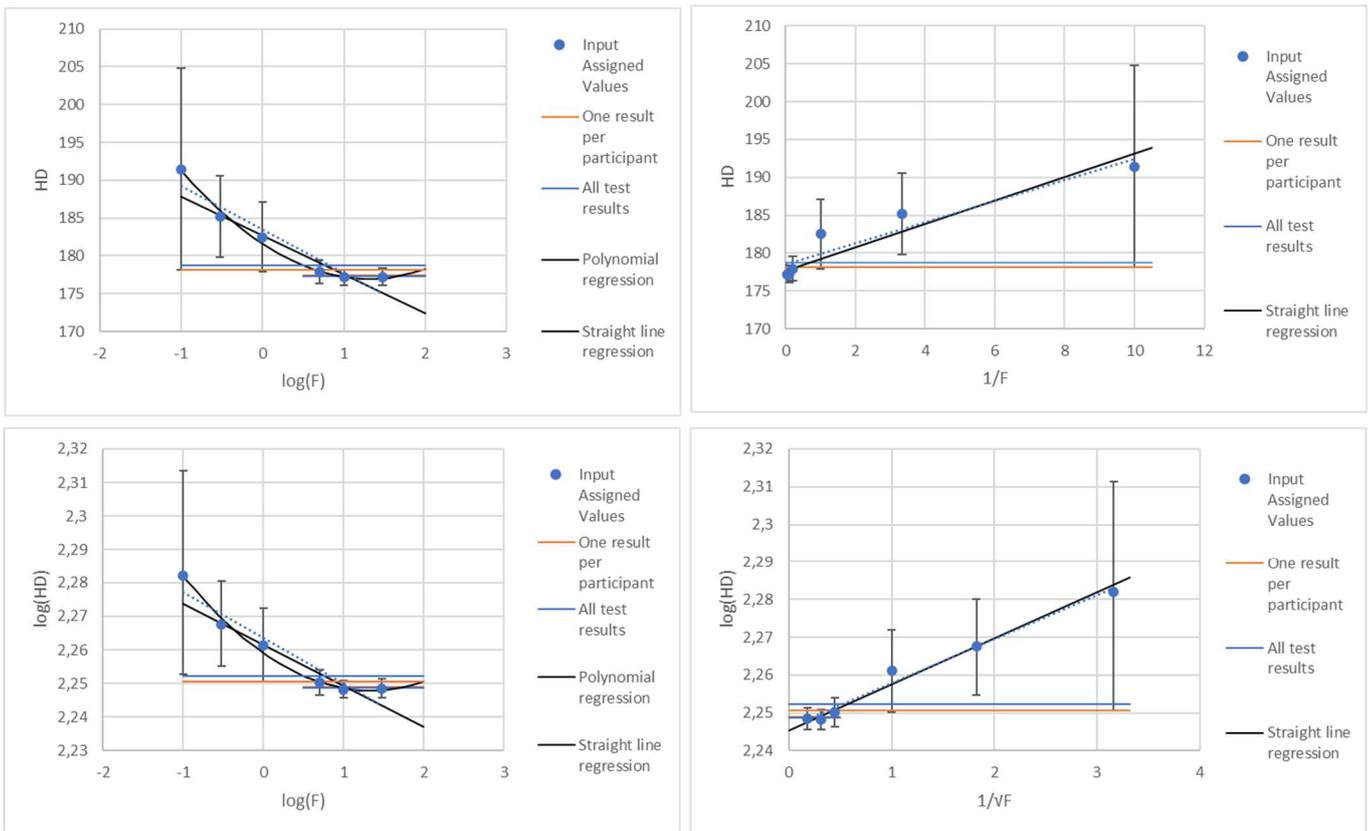


Figure 6.b: Assigned values as function of scales – 2021 Vickers hardness test results.

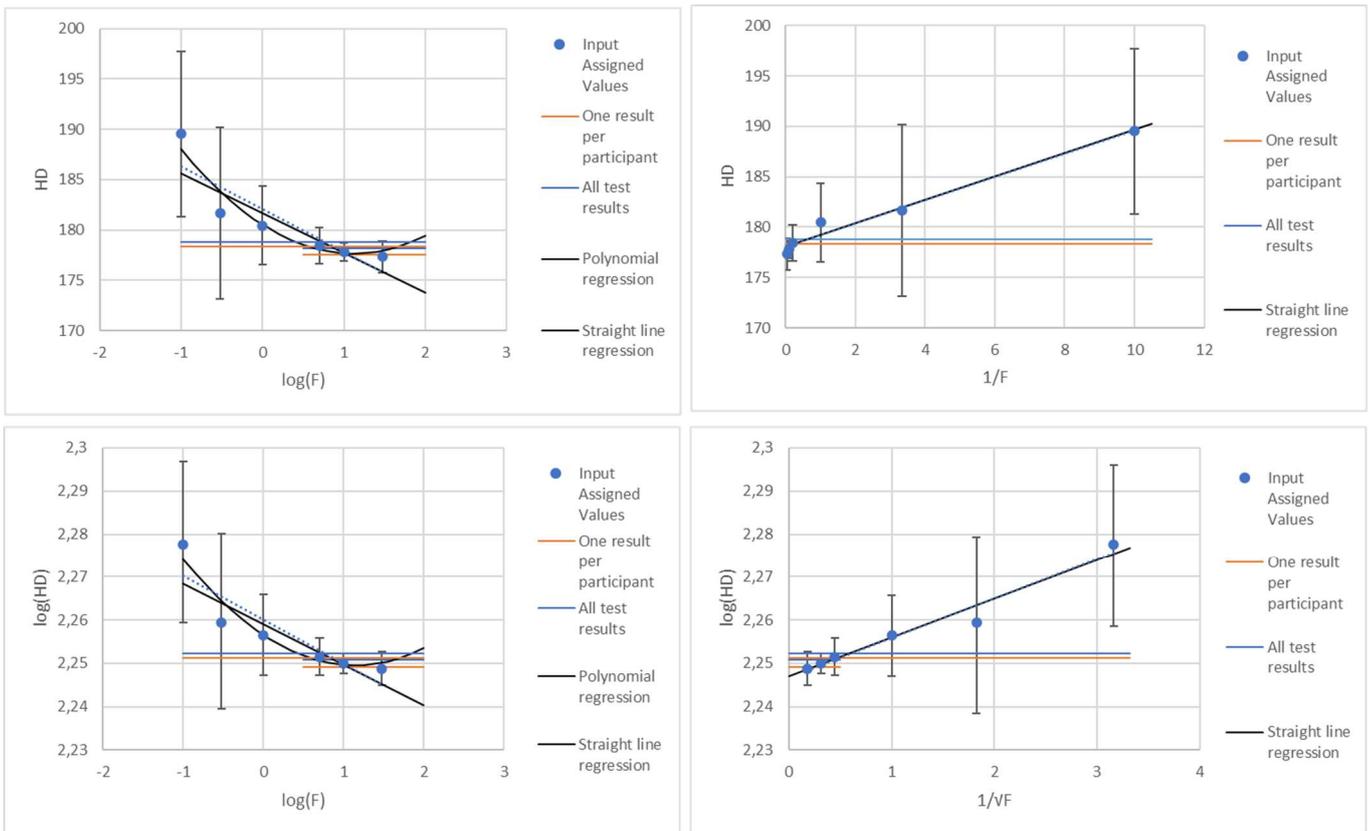


Figure 6.c: Assigned values as function of scales – 2022 Vickers hardness test results.

It can be seen from these figures that:

1. No significant differences occur between the “pure” power law model and the approximated one (not using log of ordinates);
2. Slight differences occur between the straight-line regressions using weighted input assigned values (option 4 of § 3.2), in black on the figures, and the unweighted ones (option 2 of § 3.2), in blue dotted lines on the figures. It makes then sense to use weighted regressions rather than unweighted ones;
3. No significant differences occur between mean values obtained from all tests together and means values that were obtained by using only one result per participant (blue and orange horizontal lines superpose themselves, so that no difference can be seen between the two). It makes then sense to use the “one result per participant” option, that avoids any possible covariance within same participant’s results. Moreover, the related computed uncertainties are then more reliable;
4. Unsurprisingly, slight differences occur between overall mean values computed from usual scales (short lines for which $0,5 \leq \log(F) \leq 2$, $0 \leq 1/F \leq 0,2$ and $0 \leq 1/\sqrt{F} \leq 0,5$) and from all scales (long lines) including low load scales (HV0,1 for which $\log(F) = -1$, $1/F = 10$ and $1/\sqrt{F} \approx 3,15$). In all cases, such overall mean values fit well to input assigned values of usual F values (i.e. HV5 and upper). To check this, a test of significance of the slope of the regression line compared to uncertainties on input assigned values was performed and results are provided here after;
5. Consistent with the conclusion 4 here upper, a significant ISE only applies to “low charge” Vickers hardness, i.e. HV1 and lower;
6. For the power law model, an approximation using a polynomial of degree 2 fits significantly better than a simple straight-line regression. However, upper scales (typically HV30) appear then to be in the increasing zone of the polynomial, what is obviously irrelevant (indentation size effect is supposed to lower HD when the load is increasing until a constant HD_0 which is reached for high values of loading). The minimum of the polynomial can be computed and regarded as HD_0 value. This approach is detailed here after;
7. Contrarily, despite the Li & al. model forecasts a polynomial of degree 2 (however with coefficients linked together, see § 2.2.2), straight-line regressions are obviously enough to describe properly the ISE in the context of ILCs;
8. In the Li & al. model, the usual scales for which $0 \leq 1/F \leq 0,2$, are cramped in the left part of the figures. Moreover, these scales usually provide most of the available data. We could then expect the uncertainties on slope coefficients to be not good. But this does not seem confirmed by the figures. To check this, a comparison between models was performed using regression coefficients. Corresponding results are provided here after.

Test of significance of the slope of the regression line compared to uncertainties on input assigned values:

The ratio Z_a of Equation (15) was constructed to check whether the slope of the regression is significantly different from 0 or not.

$$Z_a = \frac{a \cdot (Max_F - Min_F)}{\sqrt{\frac{\sum_i n_i \cdot u_{xpt,i}^2}{\sum_i n_i}}} \quad (15)$$

Where Z_a is a score of significance for the slope,

a is the slope of the regression straight line,

Max_F and Min_F are respectively the maximum and the minimum values of the abscissas used for the regression,

n_i is the number test results used to determine X_{pt} of the i^e input scale
 $u_{X_{pt},i}$ is the uncertainty of the i^e input scale.

In this ratio:

- ✚ The term $a \cdot (Max_F - Min_F)$ represents the range of variation of the ordinates of the regression within the interval of calculation of it;
- ✚ The $\sqrt{\sum_i n_i \cdot u_{X_{pt},i}^2 / \sum_i n_i}$ term represents a weighted quadratic mean value of the uncertainties on the assigned values of input scales.

With respect to usual practices, we can consider that the slope is significant when $|Z_a| > 2$. When a significant slope is detected, we can conclude that a significant ISE occurs within the scales of the ILC. However, as the ISE effect is assumed to decrease HD measured as the load increases, we will only consider cases where the sign of Z_a confirms this, i.e. $Z_a > 2$ for abscissas in $\log(F)$ and $Z_a < -2$ for abscissas in $1/F$ and $1/\sqrt{F}$.

Test of significance of the slope of the regression line compared to its own uncertainty:

The Z_a coefficient then compares the variations of HD within input scales to the uncertainties on input scale parameters. Another test which compares the slope to its uncertainty would make sense, as follows.

$$Z_{ua} = \frac{a \cdot (Max_F - Min_F)}{u_a} \quad (16)$$

Where Z_{ua} is a score of significance for the slope,
 a is the slope of the regression straight line,
 Max_F and Min_F are respectively the maximum and the minimum values of the abscissas used for the regression,
 u_a is the uncertainty of the slope as computed with Equation (9).

With respect to usual practices, we can consider that the slope is significant when $|Z_{ua}| > 2$. When a significant slope is detected, we can conclude that a significant ISE occurs within the scales of the ILC.

Z_{ua} score is more efficient than Z_a when the number of scales and test results is important (i.e. more than 3).

HD₀ determinations according to the models:

In both Li & al. and Nix & al. models, HD₀ are the ordinate at origin of the regression line. Consequently, both their values and, if necessary, the uncertainties on these values are easy to determine.

In the power law model:

- ✚ No HD₀ exists in the straight-line regression;
- ✚ HD₀ is the minimum ordinate of the polynomial, that can easily be determined as follows.

If the regression produces an equation of the form: $y = a_2 \cdot x^2 + a_1 \cdot x + a_0$, the minimum ordinate is obtained for:

$$\log(F) = -a_1 / (2 \cdot a_2) \quad (17)$$

and as HD₀ is the ordinate corresponding to this abscissa $-a_1 / (2 \cdot a_2)$, i.e.:

$$HD_0 = -\frac{a_1^2}{4 \cdot a_2} + a_0 \quad (18)$$

Where HD_0 is the hardness free of ISE,
 a_0 , a_1 and a_2 are the coefficients of the regression polynomial of degree 2 (cf. equation here upper).

When the polynomial of degree 2 equation is used, F_0 for which $\log(F_0) = -a_1/(2 \cdot a_2)$ shall be determined, and the polynomial law $HD = a_2 \cdot (\log(F))^2 + a_1 \cdot \log(F) + a_0$ shall be used only for $F < F_0$. When $F \geq F_0$, HD shall be chosen as equal to HD_0 as computed with Equation (18).

Of course, HD_0 can also be determined from the overall mean value of test results for which no significant ISE occurs, i.e. for HV5 and upper in the present case.

Comparison of results obtained from models:

Results of Z_a , HD_0 and r^2 for Vickers hardness in ILC of years 2020, 2021 and 2022 are provided in Table 6.

Table 6. Results of Z_a , HD_0 and r^2 for Vickers hardness in ILC of years 2020, 2021 and 2022.

Parameter	Year	Power law polynomial	Power law straight line	Overall test results	Nix & al. HD = a/F+b	Li & al. log(HD) = a/√F+b
Z_a	2020	-	-2,28	-	+2,95	+2,74
	2021	-	-7,3	-	+8,8	+8,8
	2022	-	-5,3	-	+6,4	+6,1
HD_0	2020	183,6	-	183,9±0,7	183,8±0,2	183,0±1,5
	2021	177,0	-	177,4±0,5	177,7±1,2	175,9±1,7
	2022	177,7	-	177,5±0,6	178,0±0,5	176,6±1,4
r^2	2020	0,849	0,600	-	0,979	0,975
	2021	0,990	0,850	-	0,955	0,979
	2022	0,932	0,790	-	0,966	0,980

It can be concluded from these data that:

1. Nix and Li models seem to be slightly better than the power law one even when ameliorated with a polynomial regression;
2. All ways of computing HD_0 provide very similar results. If necessary, the uncertainty for HD_0 computed with the polynomial law can be estimated by excess with its value from straight line regression;
3. Z_a coefficient as defined in Equation (15) is relevant to check whether a significant ISE occurs in the results of the ILC.

4.3.3 Results from other ILC

For all other results of ILC for which at least 3 input scales are available, we checked each of the conclusions of § 4.3.2. Table 7 encloses the results of this check:

- ✚ Only conclusions that differ to those of § 4.3.2 are mentioned in the table;
- ✚ When only 3 input scales are available, any polynomial regression is irrelevant. No conclusion was then provided about polynomial regression in those cases.

Table 7. Check of conclusions of 4.3.2 for other ILC for which at least 3 input scales are available.

Product	Type of hardness	Year	Input scales	Conclusion of check
Hard metal	HV	2017	HV5 - HV10 - HV30	The signs of Z_a coefficients are opposite to what is expected. All input AV fit with the overall mean value
Soft metal	HV	2017	HV0,1 - HV0,3 - HV1 - HV10	Power law straight line shows a significantly better r^2 than Li and Nix. Polynomial regression is irrelevant
Hard metal	HV	2018	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2018	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Overall mean value is obviously the most relevant option (r^2 coefficients are not good for all of the regressions)
Soft metal	HV	2018	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30	Nothing specific to note
Hard metal	HV	2019	HV5 - HV10 - HV30	A notable ISE occurs even if all scales are more than HV1
Soft metal	HBW	2019	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	No notable ISE occurs, including for the HBW 1/30 (9,8 Newton) scale
Soft metal	HV	2019	HV0,1 - HV0,1 - HV1 - HV5 - HV10 - HV30	Nothing specific to note
Hard metal	HV	2020	HV5 - HV10 - HV30	Power law straight line shows a significantly better r^2 than Li and Nix
Soft metal	HBW	2020	HBW 2,5/187,5 - HBW 2,5/187,5 - HBW 10/3000	Overall mean value is obviously the most relevant option (r^2 coefficients are not good for all of the regressions)
Hard metal	HV	2021	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2021	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	An ISE occurs within the scales
Hard metal	HBW	2022	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Overall mean value is obviously the most relevant option (r^2 coefficients are not good for all of the regressions)
Hard metal	HV	2022	HV5 - HV10 - HV30	A slight ISE occurs
Soft metal	HBW	2022	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	A slight ISE occurs
Hard metal	HBW	2023	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Overall mean value is obviously the most relevant option (r^2 coefficients are not good for all of the regressions)
Hard metal	HV	2023	HV5 - HV10 - HV30	A notable ISE occurs even if all scales are more than HV1

In the same way, we checked conclusions of § 4.3.2 that are applicable for ILC where only one or two input scales are available. Table 8 encloses the results of this check.

Table 8. Check of conclusions of 4.3.2 for which only 1 or 2 input scales are available.

Product	Type of hardness	Year	Input scales	Conclusion of check
Hard metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000	No significant ISE occurs
Soft metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000	No significant ISE occurs
Hard metal	HBW	2018	HBW 2,5/187,5 - HBW 10/3000	A slight ISE is likely to occur ($ Z_a =1,9$). However, the overall mean value fits the AV of both input scales
Hard metal	HBW	2019	HBW 2,5/187,5 - HBW 10/3000	A slight ISE might occur ($ Z_a =1,5$). However, the overall mean value fits the AV of both input scales
Hard metal	HBW	2020	HBW 2,5/187,5	The overall mean value fits the AV of the input scale
Hard metal	HBW	2021	HBW 2,5/187,5	The overall mean value fits the AV of the input scale

4.3.4 Conclusions for the determination of X_{pt}

Results of Table 7 and Table 8 enable to set up a typology of the encountered situations:

- Case 1. 4 to 6 input scales are available, including low load scales, typically Vickers hardness on soft metal, that include scales from HV0,1 to HV30. Figure 6.a to c provide examples of this case. In those cases, an ISE obviously occurs, that can be well described by the polynomial model, the Li model and the Nix model. In most cases, the straight-line model is not as good as the others;
- Case 2. 3 to 5 input scales are available for which an ISE occurs, but the polynomial regression is not obviously relevant. Figure 7 shows a typical example of this situation;
- Case 3. 3 to 5 input scales are available, and no ISE seems to occur, or the slope is of the unexpected sign. In those cases, the use of the overall mean value is likely to be a good solution to determine X_{pt} for output scales. Figure 8 shows a typical example of this situation;
- Case 4. 2 input scales are available, typically for Brinell test on hard metal. In general, it is not possible to draw a definite conclusion on whether an ISE occurs or not. Figure 9 shows a typical example of this situation;
- Case 5. Only 1 input scale is available, typically for Brinell test on hard metal. In general, the overall mean value fits well with the AV of the input scale. Figure 10 shows a typical example of this situation.

In general, the examination of the graphs and of r^2 and Z_a values enables to select the most appropriate model. This examination can also include more global information like $s(\Delta z_i)$ as used in and § 5.3. In cases 3, 4 and 5, the use of the overall mean value, based on a large data, is generally the most adapted solution to determine X_{pt} of output scales.

However, in cases 4 and 5, as no clear conclusion can be made on whether an ISE occurs, extrapolation of X_{pt} should be limited to adjacent scales.

In the Li model, all computations were conducted using $\log(HD)$ as ordinates. However, as shown for the power law option, no significant differences are likely to occur between the processing with HD and with $\log(HD)$, because HD/HD_0 is always lower than 1,1.

The limit of indentation size of $d = 0,3$ mm that was found in the literature (see § 2.2.1) upper which no ISE is supposed to occur can also be used to decide whether X_{pt} can be taken as a constant or not. d can be computed from the general equations of standard and are reminded in Table 9. This can be very useful for Brinell tests, for which the same HD_0 can be a priori used for all scales for soft metals and for all scales except HBW 1/30 scale for hard metals.

These conclusions do not take into account the approximations due to the models. We stated in § 2.2.2 that for both Li and Nix models, formulas could be approached by straight lines provided that F is large enough to do so. In any case, for these models, the straightness of the plots needs to be checked to make sure that approximations remain valid for the considered F values.

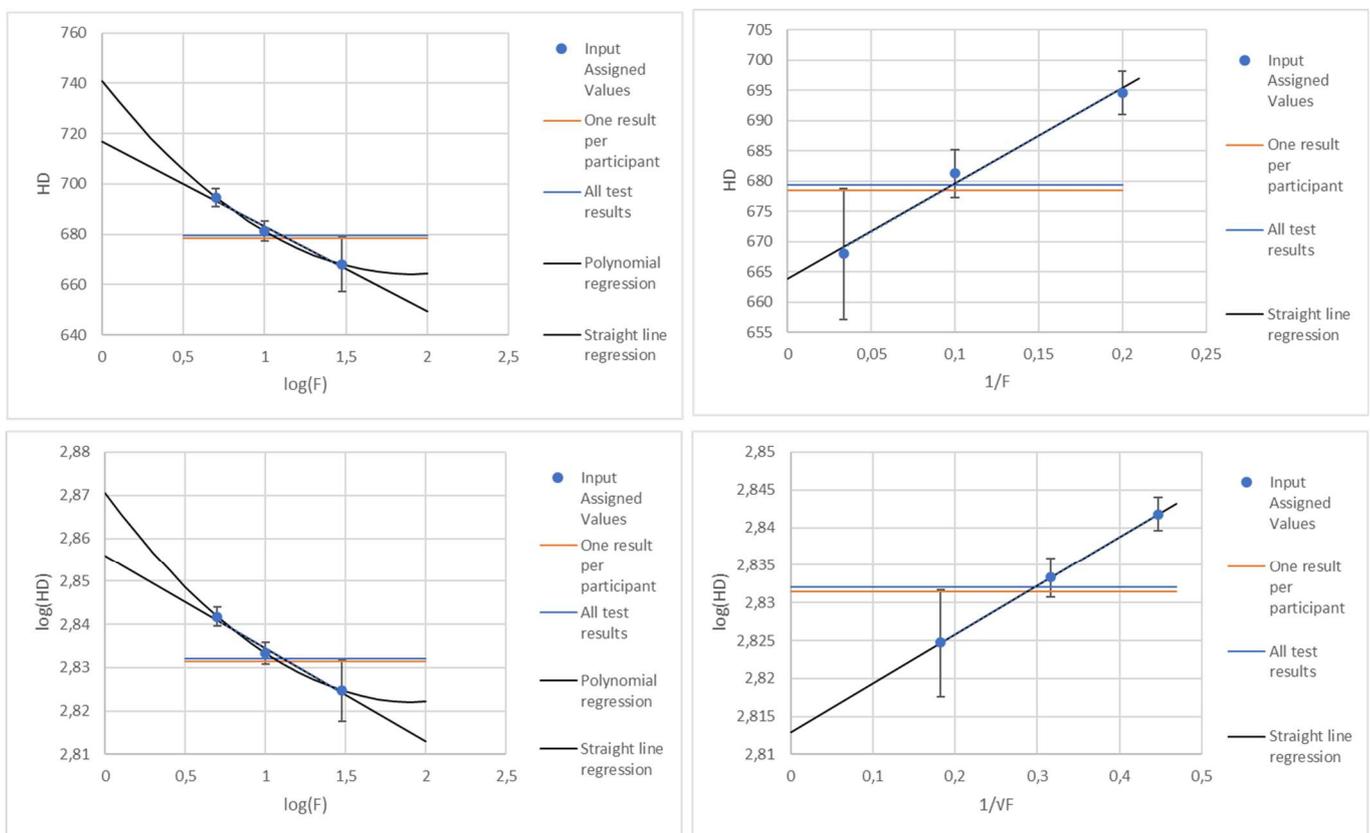


Figure 7: Assigned values as function of scales – 2019 Vickers hardness test results on hard metal.

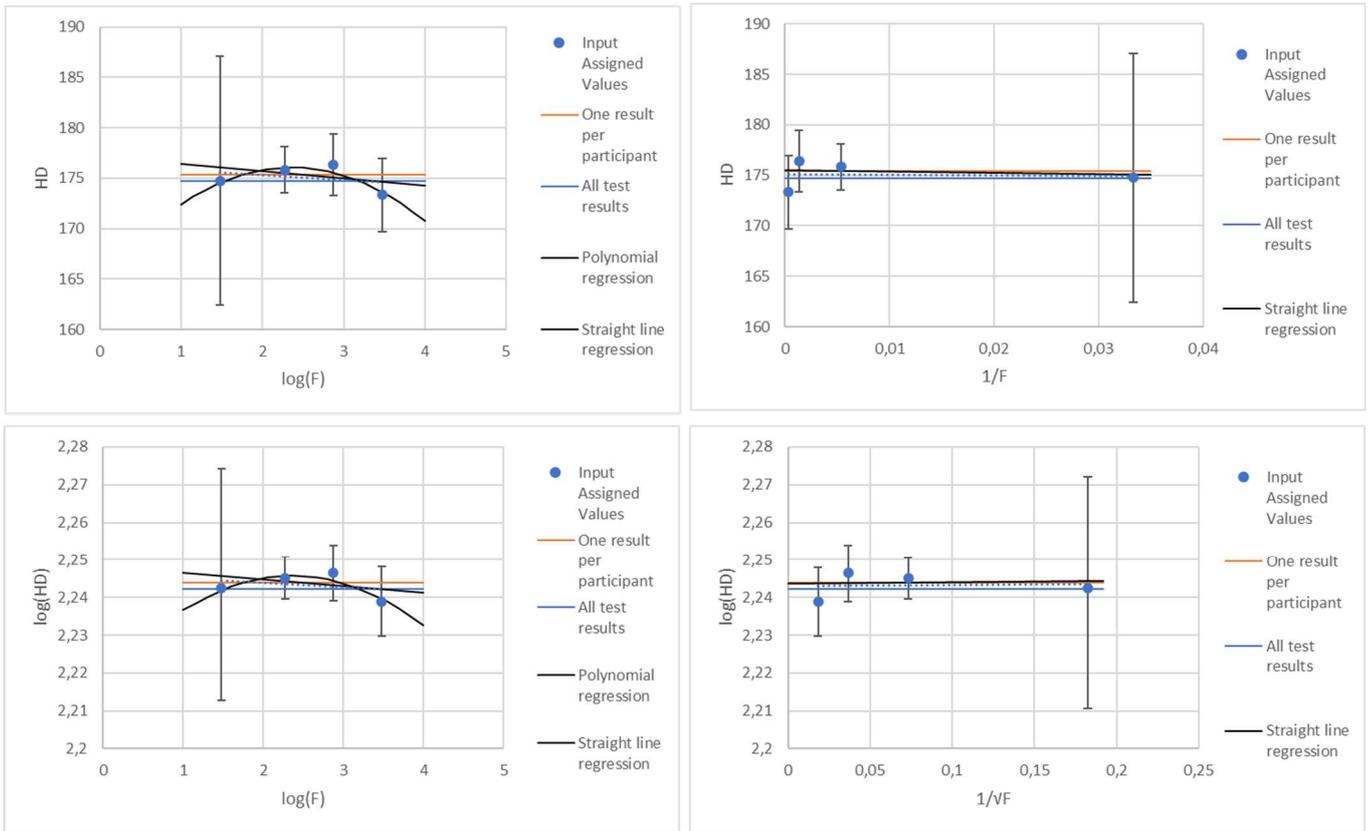


Figure 8: Assigned values as function of scales – 2019 Vickers hardness test results on soft metal.

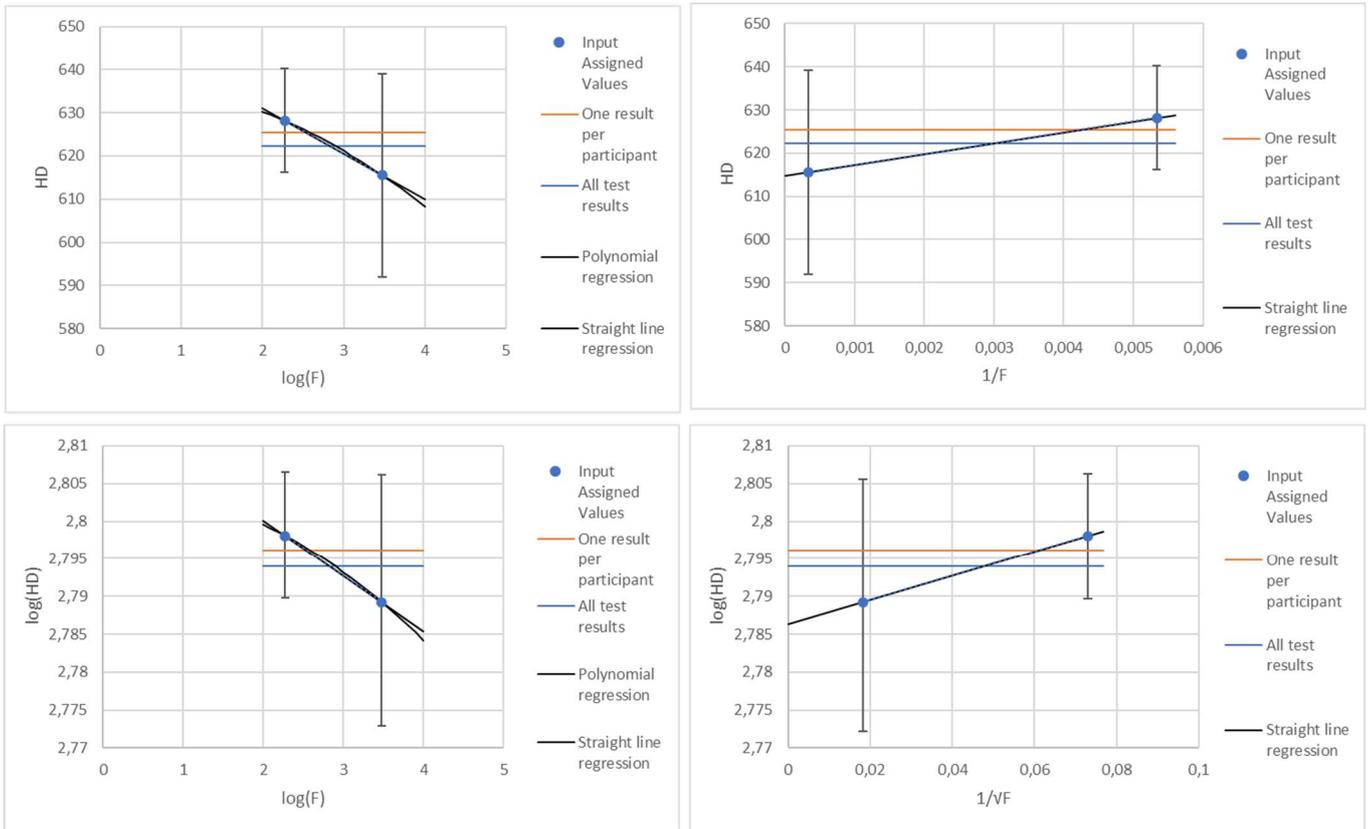


Figure 9: Assigned values as function of scales – 2019 Brinell hardness test results on hard metal.

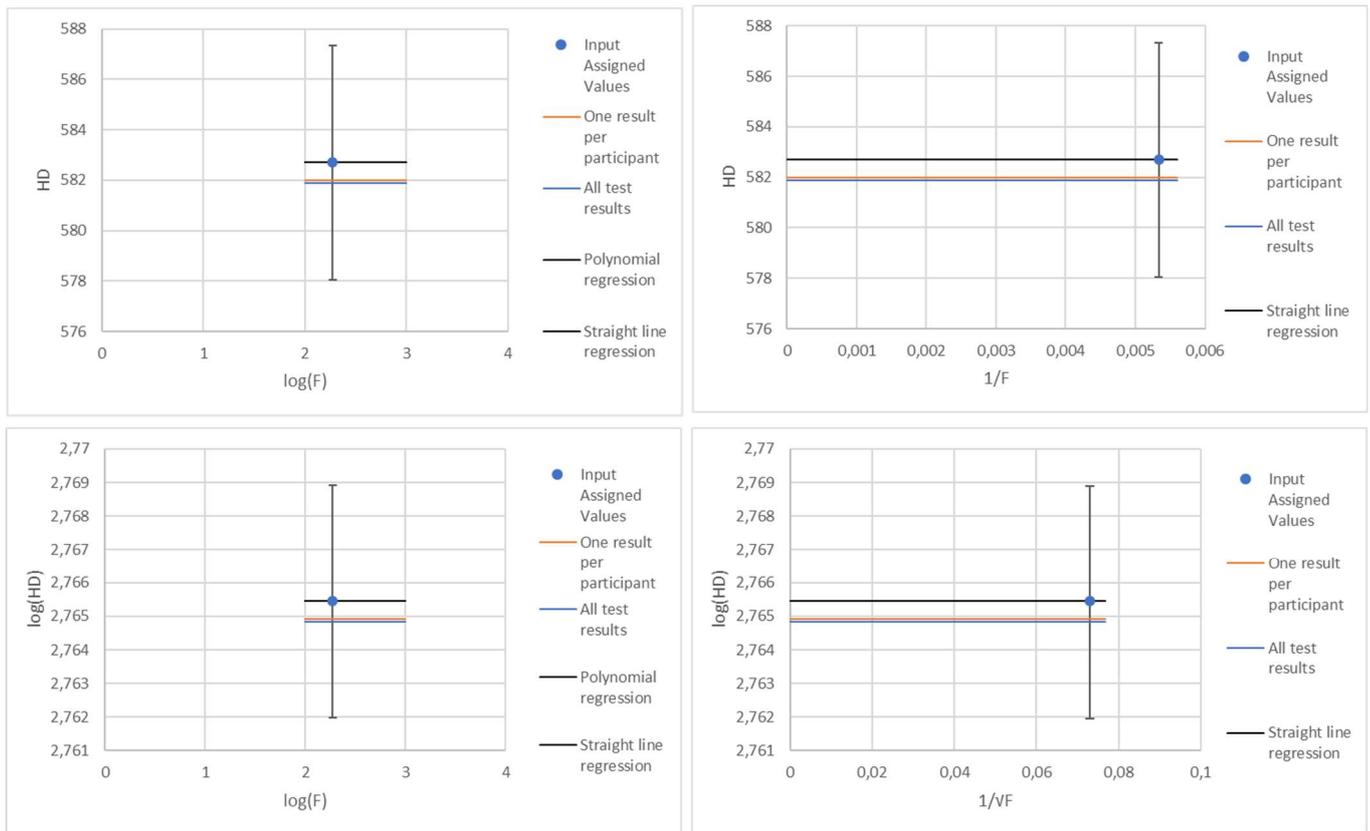


Figure 10: Assigned values as function of scales – 2020 Brinell hardness test results on hard metal.

Table 9. Limit HD values for $d \geq 0,3$ mm.

	Vickers			Brinell
	HV5	HV10	HV30	HBW 1/30
HD max	103	206	618	415

4.4 Standard deviation for the assessment of the bias σ_{pt}

4.4.1 Introduction

Even if the attention is usually mainly focused on the quality of determination of X_{pt} (including in ISO 13528, that cares about $u_{X_{pt}}$ but not about uncertainties related to standard deviations), the main source of inaccuracy in the determination of the scores used to trigger alerts is the lack of quality in the determination of σ_{pt} . For this reason, much attention was paid to this issue.

Normal probability plots of σ_{pt}/X_{pt} ratios were performed for checking whether:

- ✚ σ_{pt} values can be supposed to be rather constant whatever X_{pt} or rather proportional to X_{pt} ;
- ✚ σ_{pt}/X_{pt} ratios depend on the product (soft or hard) and on the type of hardness tests (Brinell or Vickers);
- ✚ σ_{pt}/X_{pt} ratios can be regarded as normally distributed.

For guidance concerning normal probability plots and their interpretation, see [14].

Results are provided in Figure 11.

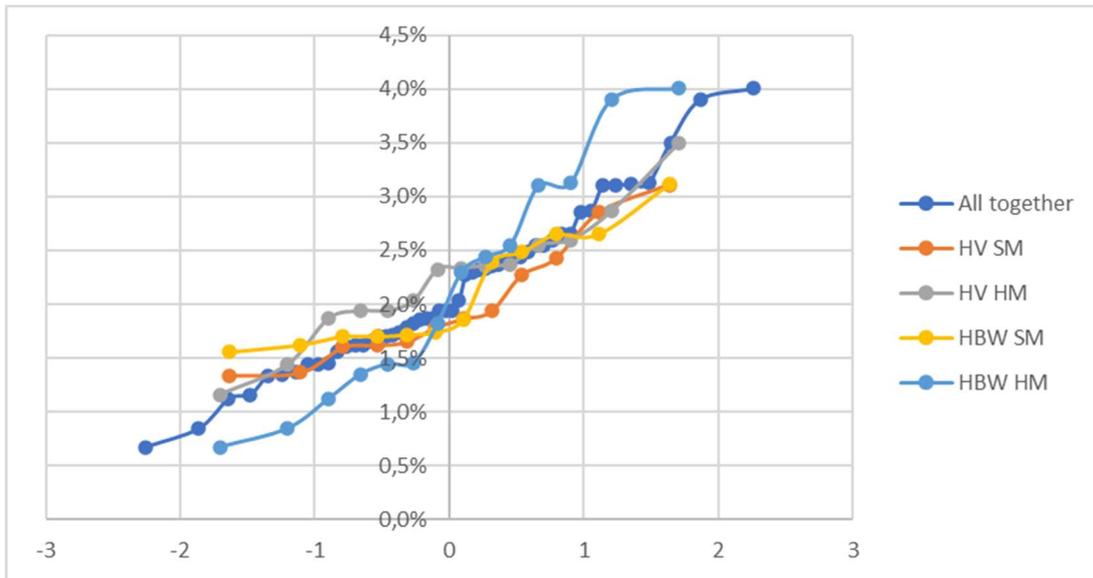


Figure 11: Normal probability plots of σ_{pt}/X_{pt} determined during CompaLab hardness tests ILC between 2017 and 2023.

It can be seen from the figure that:

1. σ_{pt} values can be regarded as proportional to X_{pt} and a central value for σ_{pt}/X_{pt} is about 2% all scales, all products and both Brinell and Vickers hardness tests all together. This is quite important because provides a general guidance on the range in which we can expect σ_{pt} to be, and even a related interval of confidence;
2. HV SM and HV HM curves are more horizontal than HBW SM and HBW HM. This means that HV ratios are more centred around the 2% value than HBW ratios;
3. In the same way, HV SM and HBW SM curves are more horizontal than HV HM and HBW HM. This means that SM (soft metals) ratios are more centred around the 2% value than HM (hard metals) ratios.

4.4.2 Results of survey on Vickers hardness test results of years 2020, 2021 and 2022

In the same way than for X_{pt} , a first survey was made on Vickers results of years 2020, 2021 and 2022 for the same reasons than for X_{pt} .

Figure 12.a to c provide the results of σ_{pt} determined during the ILC concerning Vickers hardness tests in years 2020, 2021 and 2022.

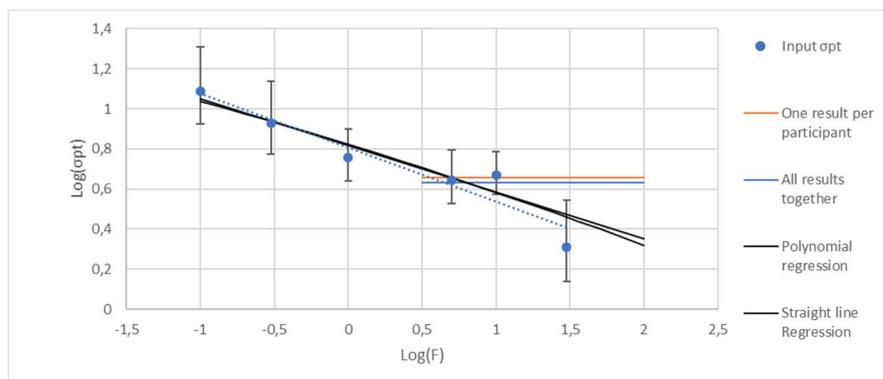


Figure 12.a: Standard deviation for assessment of the bias, as function of scales – 2020 Vickers hardness test results on soft metal.

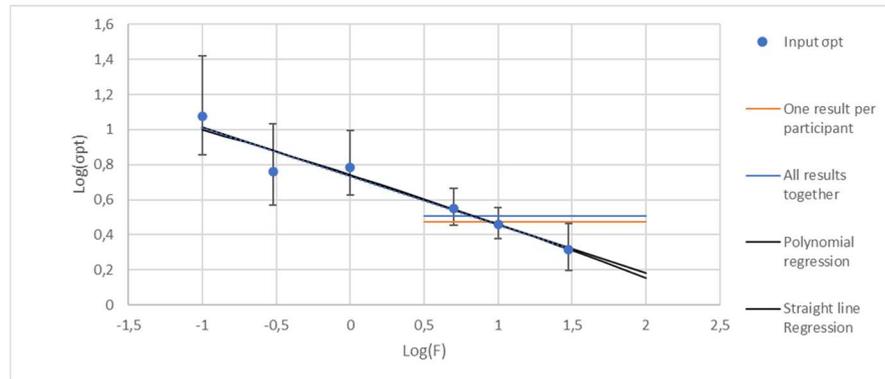


Figure 12.b: Standard deviation for assessment of the bias, as function of scales – 2021 Vickers hardness test results on soft metal.

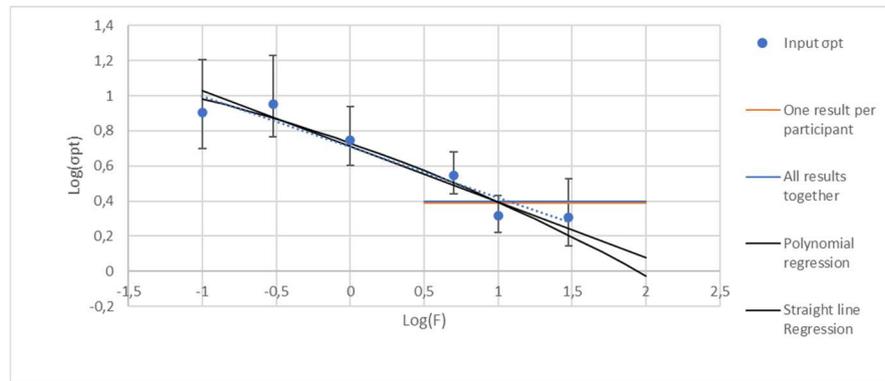


Figure 12.c: Standard deviation for assessment of the bias, as function of scales – 2022 Vickers hardness test results on soft metal.

It can be seen from these figures that:

1. An ISE obviously occurs for σ_{pt} as well as for X_{pt} . Determination of σ_{pt} as the overall mean value from all scale is irrelevant. No need to compute any kind of Z_a to confirm it. Variations of σ_{pt} as function of scales are quite more important than for X_{pt} (1 to 6 to be compared to 1 to 1,1);
2. Slight differences occur between the straight-line regressions using weighted input assigned values (option 4 of § 3.2), in black on the figures, and the unweighted ones (option 2 of § 3.2), in blue dotted lines on the figures. It makes then sense to use weighted regressions rather than unweighted ones;
3. No significant differences occur between mean values obtained from all tests together and means values that were obtained by using only one result per participant (blue and orange horizontal lines superpose themselves, so that no difference can be seen between the two). It makes then sense to use the “one result per participant” option, that avoids any possible covariance effect within same participant’s results. Moreover, the related computed uncertainties are then more reliable;
4. The linear regression and the polynomial regression provide results very close to each other. No need to use polynomials to predict correctly σ_{pt} from input scales.

It is then possible to compute any σ_{pt} for an output scale using the Equation (19), as follows:

$$\log(\sigma_{pt,F}) = a \cdot \log(F) + b \tag{19}$$

Where σ_{pt} is the standard deviation used for assessing the proficiency of participants of scale F
 a and b are the regression coefficients.

Equation (19) can also be formulated as follows:

$$\sigma_{pt,F} = 10^{(a \cdot \log(F) + b)}$$

4.4.3 Results from other ILC

For all other results of ILC for which at least 3 input scales are available, we checked each of the conclusions of § 4.4.2. Table 10 encloses the results of this check (only conclusions that differ to those of § 4.4.2 are mentioned in the table).

Table 10. Check of conclusions of § 4.4.2 for other ILC for which at least 3 input scales are available.

Product	Type of hardness	Year	Input scales	Conclusion of check
Hard metal	HV	2017	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HV	2017	HV0,1 - HV0,3 - HV1 - HV10	Slope of the regression line has the unexpected sign. Polynomial regression is irrelevant
Hard metal	HV	2018	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2018	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	No significant ISE occurs
Soft metal	HV	2018	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30	Nothing specific to note
Hard metal	HV	2019	HV5 - HV10 - HV30	Slope of the regression line has the unexpected sign. Polynomial regression is irrelevant
Soft metal	HBW	2019	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Correlation coefficient is not very good. Polynomial regression is irrelevant
Soft metal	HV	2019	HV0,1 - HV0,1 - HV1 - HV5 - HV10 - HV30	Nothing specific to note
Hard metal	HV	2020	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2020	HBW 2,5/187,5 - HBW 2,5/187,5 - HBW 10/3000	Slope of the regression line has the unexpected sign. Polynomial regression is irrelevant
Hard metal	HV	2021	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2021	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Correlation coefficient is not very good
Hard metal	HBW	2022	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Correlation coefficient is not very good
Hard metal	HV	2022	HV5 - HV10 - HV30	No significant ISE occurs. Polynomial regression is irrelevant
Soft metal	HBW	2022	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Polynomial regression is irrelevant
Hard metal	HBW	2023	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	No significant ISE occurs. Polynomial regression is irrelevant
Hard metal	HV	2023	HV5 - HV10 - HV30	Polynomial regression is irrelevant

Contrarily to § 4.4.2, the use of a calculation of Z_a can be useful to decide whether the slope of the straight line is significant or not. In this case, Equation (15) must be reconsidered because the ordinates are plotted as logarithms and the related uncertainties need to be transformed into logarithms as well. This raises two difficulties:

1. The uncertainties on estimates of standard deviations follow square roots of χ^2 distributions, that are not symmetrical;
2. The transformation of ordinates into their logarithms also introduce asymmetry in the limits of intervals of confidence.

Luckily, despite these two asymmetries, the range of IC is kept whatever the coefficient of enlargement, as shown in Table 11. It is reminded that in the case of the estimation of σ_{pt} , n is equal to the number of participants for the scale.

Table 11. Relative range of intervals of confidence of estimates of standard deviations as function of n (number of data used for the calculation of the estimate) and the level of confidence of the interval of confidence.

k_G	IC	n							
		8	10	13	16	20	25	32	40
3	99,73%	1,117	1,090	1,067	1,053	1,042	1,033	1,025	1,020
2,5	98,76%	1,094	1,072	1,054	1,043	1,034	1,026	1,020	1,016
2	95,45%	1,075	1,058	1,043	1,034	1,027	1,021	1,016	1,013
1,5	86,64%	1,060	1,046	1,034	1,027	1,022	1,017	1,013	1,010
1	68,27%	1,049	1,038	1,028	1,023	1,018	1,014	1,011	1,009
0,5	38,29%	1,043	1,033	1,025	1,020	1,016	1,012	1,009	1,008

For example, for $n = 8$, the actual IC95% is equal to $2 \times 1,075 = \pm 2,15.u$ instead of the $\pm 2.u$ that would occur if the distribution were Gaussian. However, even if the extents of IC are kept whatever k , the central value is not located at the centre of the IC, and the distance between the two changes (reduces) with “ n ”.

Thanks to the results displayed in Table 11, Equation (15), that only deals with the extents of uncertainties, can be used to determine the Z_a coefficients.

In the same way, we checked conclusions of § 4.4.2 that are applicable for ILC where only one or two input scales are available. Table 12 encloses the results of this check.

Table 12. Check of conclusions of § 4.4.2 for which only 1 or 2 input scales are available.

Product	Type of hardness	Year	Input scales	Conclusion of check
Hard metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000	Slope is not significantly different from 0
Soft metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000	Slope is likely to be different from 0, but the conclusion is not robust because of lack of input scales
Hard metal	HBW	2018	HBW 2,5/187,5 - HBW 10/3000	Slope is not significantly different from 0
Hard metal	HBW	2019	HBW 2,5/187,5 - HBW 10/3000	Slope is not significantly different from 0
Hard metal	HBW	2020	HBW 2,5/187,5	The overall mean value fits with σ_{pt} of the input scale
Hard metal	HBW	2021	HBW 2,5/187,5	The overall mean value fits with σ_{pt} of the input scale

4.4.4 Conclusions for the determination of σ_{pt}

Results of Table 7 and Table 8 show that a polynomial regression is never relevant. The following typology of the encountered situations can be set up:

- Case 1. 3 to 6 input scales are present, and a significant decreasing slope is present. In those cases, a straight-line regression can be computed to link $\log(\sigma_{pt})$ and $\log(F)$. Figure 12.a to c here upper show examples of this situation;
- Case 2. 3 to 6 input scales are present, and no significant decreasing slope is present. In those cases, σ_{pt} computed from the test result per participant is relevant. Figure 13.a and b show examples of this situation;
- Case 3. Only 1 or 2 input scales are present. In those cases, σ_{pt} computed from the test result per participant is likely to be relevant. However, the extrapolation should be limited to adjacent output scales because of lack of confidence due to lack of input scales. Figure 14 shows an example of this situation.

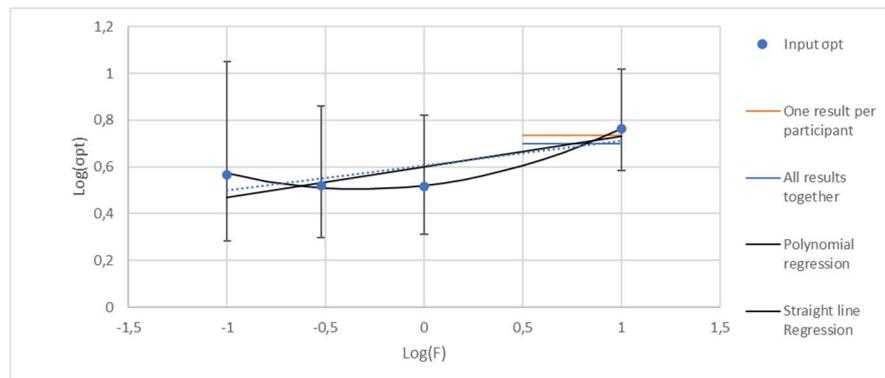


Figure 13.a: Standard deviation for assessment of the bias, as function of scales – 2017 Vickers hardness test results on soft metal.

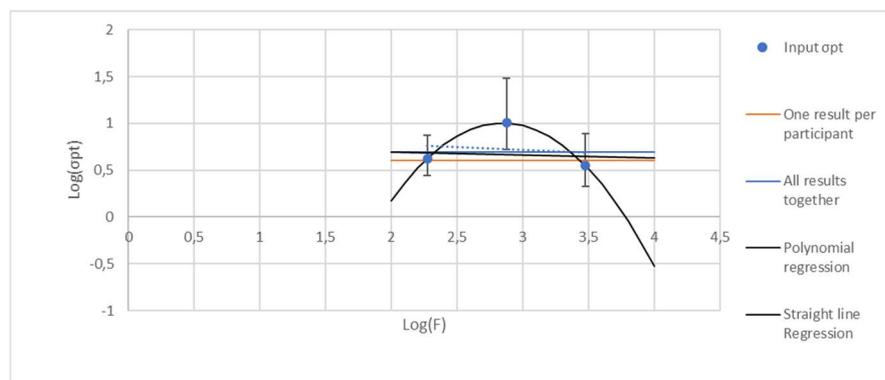


Figure 13.b: Standard deviation for assessment of the bias, as function of scales – 2022 Brinell hardness test results on hard metal.

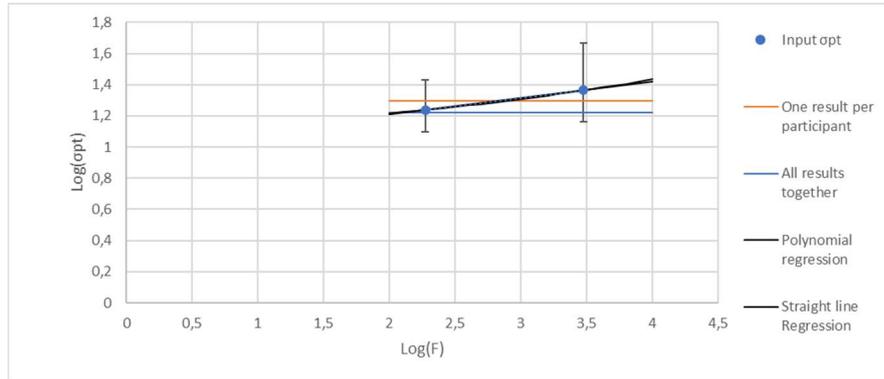


Figure 14: Standard deviation for assessment of the bias, as function of scales – 2019 Brinell hardness test results on hard metal.

In all cases, the examination of the graphs and the use of the regression coefficient enables to select the most appropriate model. This examination can also include more global information like $s(\Delta z_i)$ as used in and § 5.3.

4.5 Uncertainty $u_{X_{pt}}$ on the assigned value X_{pt}

ISO 13528 proposes to compute $u_{X_{pt}}$ from σ_{pt} using the equation (20), as follows:

$$u_{X_{pt}} = 1,25 \cdot \frac{\sigma_{pt}}{\sqrt{p}} \quad (20)$$

Where $u_{X_{pt}}$ is the uncertainty on X_{pt} ,
 σ_{pt} is the standard deviation used for assessing the bias of participants
 p is the number of participants.

When σ_{pt} can be regarded as independent from scales (Case 2 and Case 3 of § 4.4.4), $u_{X_{pt}}$ can then also be regarded as independent from scales. It can then be computed with the Equation (21), as follows:

$$u_{X_{pt}} = \sqrt{\frac{\sum_i n_i \cdot u_{X_{pt,i}}^2}{\sum_i n_i}} \quad (21)$$

Where $u_{X_{pt}}$ represents a weighted quadratic mean value of the uncertainties on the assigned values of input scales,
 n_i is the number test results used to determine X_{pt} of the i^e input scale
 $u_{X_{pt,i}}$ is the uncertainty of the i^e input scale.

When σ_{pt} cannot be regarded as independent from scales (Case 1 of § 4.4.4), we cannot use Equation (19) because of the term $\sqrt{N_p}$ of Equation (20) which is different for each scale. We then need to compute the regression of $u_{X_{pt}}$ against $\log(F)$, in the same way than in § 4.4.4. All statements of § § 4.4.4 also apply to this computation.

We get then the Equation (22) (quite similar to Equation (19)) to compute any $u_{X_{pt}}$ for an output scale, as follows:

$$\log(u_{X_{pt,F}}) = a \cdot \log(F) + b \quad (22)$$

Where $u_{X_{pt}}$ is the standard uncertainty used for assessing the proficiency of participants of scale F
 a and b are the regression coefficients.

Equation (22) can also be formulated as follows:

$$u_{X_{pt,F}} = 10^{(a \cdot \log(F) + b)}$$

When the polynomial option is selected, only the left half of the parabola is used. For F values equal to or more than the minimum of the parabola, $X_{pt} = HD_0$ and $u_{x_{pt}}$ can be chosen equal to u_{HD_0} .

4.6 Standard deviation of repeatability σ_{rpt}

4.6.1 Introduction

Normal probability plots of σ_{rpt}/σ_{pt} ratios were performed in order to check whether:

- ✚ σ_{rpt} values can be supposed to be rather constant whatever σ_{pt} or rather proportional to σ_{pt} ;
- ✚ σ_{rpt}/σ_{pt} ratios depend on the product (soft or hard) and on the type of hardness tests (Brinell or Vickers);
- ✚ σ_{rpt}/σ_{pt} ratios can be regarded as normally distributed.

For guidance concerning normal probability plots and their interpretation, see [14].

Results are provided in Figure 15.

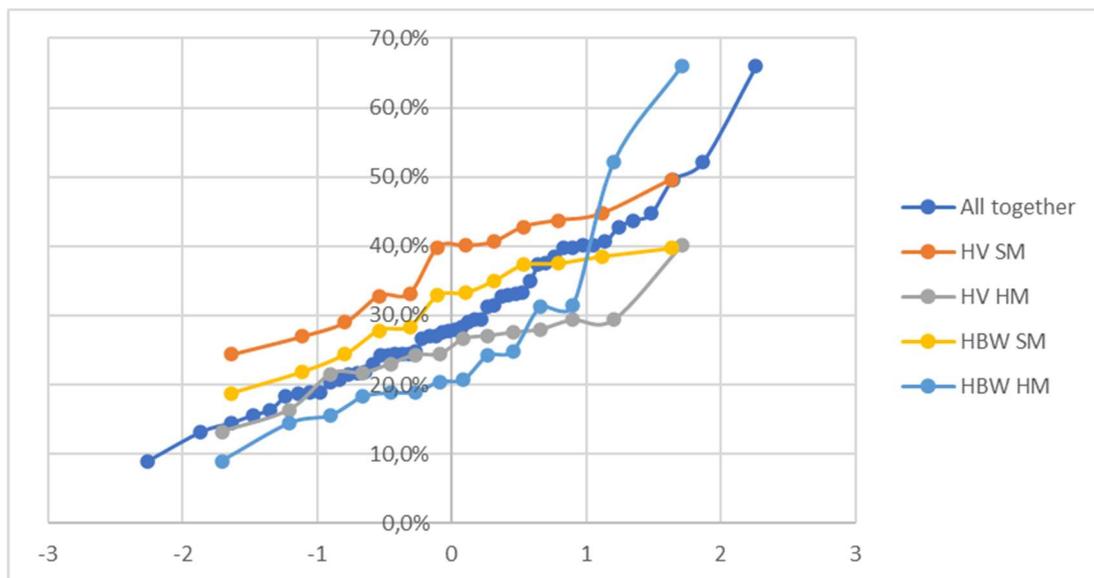


Figure 15: Normal probability plots of σ_{rpt}/σ_{pt} determined during CompaLab hardness tests ILC between 2017 and 2023.

It can be seen from the figure that:

1. σ_{rpt} values can be regarded as proportional to σ_{pt} and a central value for σ_{rpt}/σ_{pt} is about 28% all scales, all products and both Brinell and Vickers hardness tests all together. This is quite important because provides a general guidance on the range in which we can expect σ_{rpt} to lie, and even a related interval of confidence;
2. HV SM and HV HM curves are above HBW SM and HBW HM ones by about 5%;
3. HV SM and HV HM curves are more horizontal than HBW SM and HBW HM. This means that HV ratios are more centred around their central values than HBW ratios;
4. HV SM and HBW SM curves are above than HV HM and HBW HM ones by about 15%;
5. HV SM and HBW SM curves are more horizontal than HV HM and HBW HM. This means that SM (soft metals) ratios are more centred around their central value than HM (hard metals) ratios.

4.6.2 Results of survey on Vickers hardness test results of years 2020, 2021 and 2022

In the same way than for X_{pt} , a first survey was made on Vickers results of years 2020, 2021 and 2022 for the same reasons than for X_{pt} .

Figure 16.a to c provide the results of σ_{rpt} determined during the ILC concerning Vickers hardness tests in years 2020, 2021 and 2022.

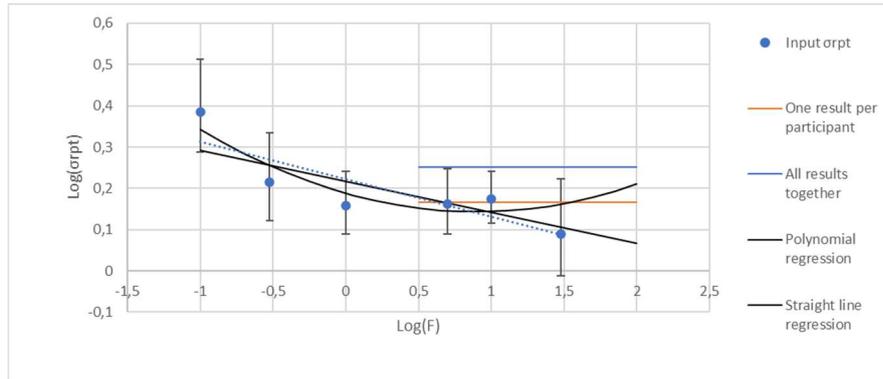


Figure 16.a: Repeatability standard deviation, as function of scales – 2020 Vickers hardness test results on soft metal.

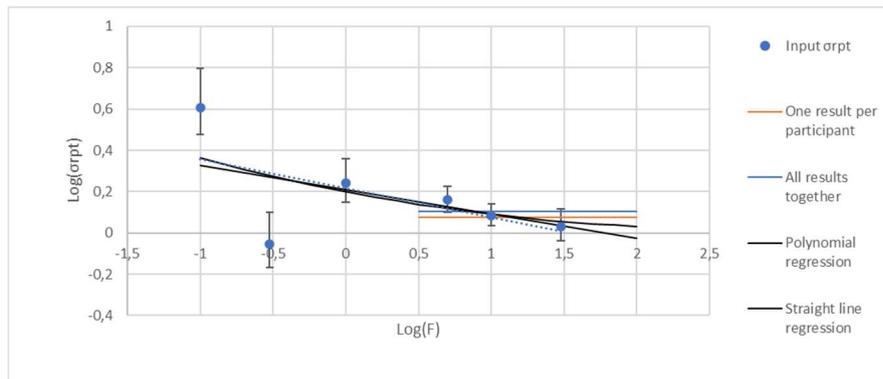


Figure 16.b: Repeatability standard deviation, as function of scales – 2021 Vickers hardness test results on soft metal.

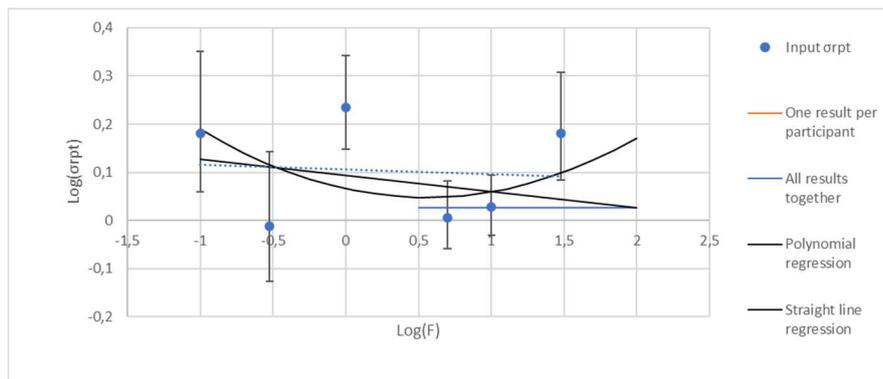


Figure 16.c: Repeatability standard deviation, as function of scales – 2022 Vickers hardness test results on soft metal.

It can be seen from these figures that:

1. In 2020 and 2021, an ISE obviously occurred for σ_{rpt} as well as for X_{pt} and σ_{pt} . Determination of σ_{pt} as the overall mean value from all scale is irrelevant. However, 2022 figures seem erratic: adoption of a constant $\log(\sigma_{rpt}) = 0,1$ seems the less irrelevant;

2. Slight differences occur between the straight-line regressions using weighted input assigned values (option 4 of § 3.2), in black on the figures, and the unweighted ones (option 2 of § 3.2), in blue dotted lines on the figures. It makes then sense to use weighted regressions rather than unweighted ones;
3. In 2021 and 2022, no significant differences occur between mean values obtained from all tests together and means values that were obtained by using only one result per participant (blue and orange horizontal lines superpose themselves, so that no difference can be seen between the two). In 2020, the means values obtained without redundancy of participants is obviously more relevant. It makes then sense to use the “one result per participant” option, that avoids any possible covariance effect within same participant’s results. Moreover, the related computed uncertainties are then more reliable;
4. The linear regression and the polynomial regression provide results are either very close to each other or irrelevant. No need to use polynomials to predict correctly σ_{pt} from input scales.

In most cases, it is then possible to compute any σ_{rpt} for an output scale using the Equation (23), as follows:

$$\log(\sigma_{rpt,F}) = a \cdot \log(F) + b \tag{23}$$

Where σ_{rpt} is the standard deviation used for assessing the proficiency of participants of scale F
 a and b are the regression coefficients.

Equation (23) can also be formulated as follows:

$$\sigma_{rpt,F} = 10^{(a \cdot \log(F) + b)}$$

With same justification than in § 4.4.3, a computation of Z_a with Equation (15) can help to decide whether Equation (23) should be used or not.

4.6.3 Results from other ILC

For all other results of ILC for which at least 3 input scales are available, we checked each of the conclusions of § 4.6.2. Table 13 encloses the results of this check (only conclusions that differ to those of § 4.6.2 are mentioned in the table).

Table 13. Check of conclusions of § 4.6.2 for other ILC for which at least 3 input scales are available.

Product	Type of hardness	Year	Input scales	Conclusion of check
Hard metal	HV	2017	HV5 - HV10 - HV30	Slope is not significantly different from 0
Soft metal	HV	2017	HV0,1 - HV0,3 - HV1 - HV10	Same situation than “Soft metal HV 2022”
Hard metal	HV	2018	HV5 - HV10 - HV30	Slope is not significantly different from 0
Soft metal	HBW	2018	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Nothing specific to note
Soft metal	HV	2018	HV0,1 - HV0,3 - HV1 - HV5 - HV10 - HV30	One scale (HV5) is a kind of outlier
Hard metal	HV	2019	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2019	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Same situation than “Soft metal HV 2022”

Product	Type of hardness	Year	Input scales	Conclusion of check
Soft metal	HV	2019	HV0,1 - HV0,1 - HV1 - HV5 - HV10 - HV30	Nothing specific to note
Hard metal	HV	2020	HV5 - HV10 - HV30	Slope is not significantly different from 0
Soft metal	HBW	2020	HBW 2,5/187,5 - HBW 2,5/187,5 - HBW 10/3000	Slope is not significantly different from 0
Hard metal	HV	2021	HV5 - HV10 - HV30	Slope is not significantly different from 0
Soft metal	HBW	2021	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Same situation than "Soft metal HV 2022"
Hard metal	HBW	2022	HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Slope is not significantly different from 0
Hard metal	HV	2022	HV5 - HV10 - HV30	Nothing specific to note
Soft metal	HBW	2022	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Nothing specific to note
Hard metal	HBW	2023	HBW 1/30 - HBW 2,5/187,5 - HBW 5/750 - HBW 10/3000	Slope is not significantly different from 0
Hard metal	HV	2023	HV5 - HV10 - HV30	Slope is not significantly different from 0

In the same way than in § 4.4.2, the use of a calculation of Z_a can be useful to decide whether the slope of the straight line is significant or not. For same reasons than in § 4.4.3, Equation (15), that only deals with the sizes of uncertainties, can be used to determine the Z_a coefficients.

In the same way, we checked conclusions of § 4.4.2 that are applicable for ILC where only one or two input scales are available. Table 14 encloses the results of this check.

Table 14. Check of conclusions of § 4.6.2 for which only 1 or 2 input scales are available.

Product	Type of hardness	Year	Input scales	Conclusion of check
Hard metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000	Slope is of the unexpected sign and is close to limit to be significant
Soft metal	HBW	2017	HBW 2,5/187,5 - HBW 10/3000	Use of a constant is compatible with the results of the 2 scales
Hard metal	HBW	2018	HBW 2,5/187,5 - HBW 10/3000	Slope is of the unexpected sign and is close to limit to be significant
Hard metal	HBW	2019	HBW 2,5/187,5 - HBW 10/3000	Slope is of the unexpected sign and is close to limit to be significant
Hard metal	HBW	2020	HBW 2,5/187,5	The overall mean value fits with σ_{rpt} of the input scale
Hard metal	HBW	2021	HBW 2,5/187,5	The overall mean value fits with σ_{rpt} of the input scale

4.6.4 Conclusions for the determination of σ_{rpt}

Results of Table 13 and Table 14 show that a polynomial regression is never relevant. They enable to set up a typology of the encountered situations:

- Case 1. 3 to 6 input scales are present, and a significant decreasing slope is present. In those cases, a straight-line regression can be computed to link $\log(\sigma_{rpt})$ and $\log(F)$;
- Case 2. 3 to 6 input scales are present, and no significant decreasing slope is present. In those cases, σ_{rpt} computed from the test result per participant is relevant;
- Case 3. Only 1 or 2 input scales are present. In those cases, σ_{rpt} computed from the test result per participant is likely to be relevant. However, the extrapolation should be limited to adjacent output scales because of lack of confidence due to lack of input scales.

In all cases, the examination of the graphs and the use of the regression coefficient enables to select the most appropriate model. This examination can also include more global information like $s(\Delta zi)$ as used in § 5.3.

4.7 Standard deviation of homogeneity σ_H

4.7.1 Introduction

In accordance with conclusions of § 3.3.1, we could expect that no Gaussian distribution can be found for σ_H . Normal probability plots confirmed this expectation.

4.7.2 Results of survey on Vickers hardness test results of years 2020, 2021 and 2022

In the same way than for X_{pt} , a first survey was made on Vickers results of years 2020, 2021 and 2022 for the same reasons than for X_{pt} .

Figure 17.a to c provide the results of σ_H determined during the ILC concerning Vickers hardness tests in years 2020, 2021 and 2022.

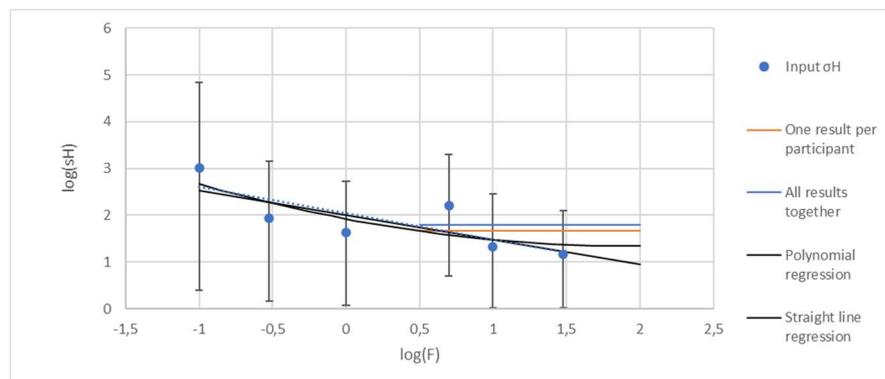


Figure 17.a: Homogeneity standard deviation, as function of scales – 2020 Vickers hardness test results on soft metal.

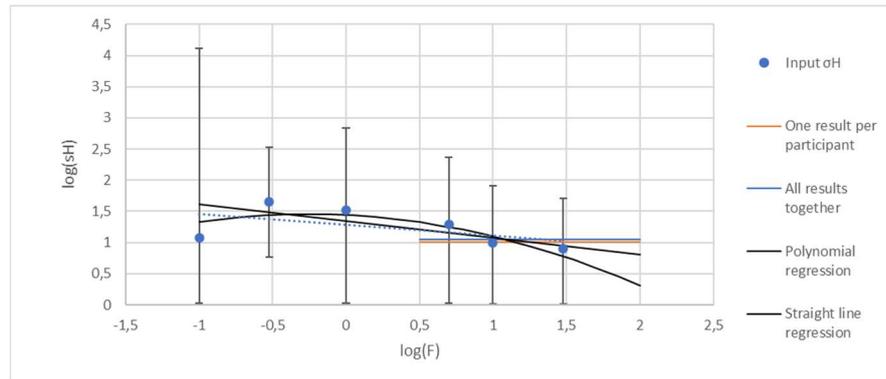


Figure 17.b: Homogeneity standard deviation, as function of scales – 2021 Vickers hardness test results on soft metal.

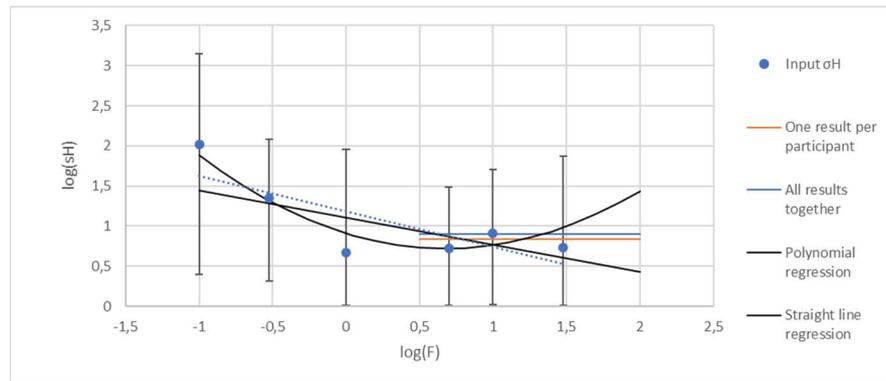


Figure 17.c: Homogeneity standard deviation, as function of scales – 2022 Vickers hardness test results on soft metal.

It can be seen from these figures that an ISE seems to occur for σ_H . However, because of the phenomena described in § 3.3.1, the uncertainties on σ_H are so huge that the slope cannot be regarded as significant. Moreover, the use of the Z_a parameter might be problematic because of probable asymmetries that occur for the distributions of uncertainties. In all cases, σ_H computed from all values fits with the results of all input scales.

On the technical point of view, σ_H is linked to the product and to the characteristic. Product is obviously exactly same. An argument can be developed to decide whether different scales of a same hardness test are different characteristics or not. If the answer were “yes”, then, by definition of homogeneity, σ_H is same whatever the scales. However, low charge hardness tests can detect local differences in hardness (typically in hardened layers) than high charge hardness tests. In the other hand, one hardness test is not necessarily made of only one hardness measurement. In fact, laboratories are supposed to describe in their procedures how many measurements shall be performed to make one hardness test result, and these procedures can request different numbers of measurements according testing conditions, among which the used scale can be considered. Consequently, increased scatter of low charge testing can be compensated by increasing the number of measured used to determine one test result.

4.7.3 Results from other ILC

All ILC from 2017 to 2023, for both soft metals and hard metals and both HV and HBW show the same situation. In particular, most of the lower limits of IC are 0 or very close to 0. In all cases, σ_H computed from all values fits with the results of all input scales with their uncertainties.

4.7.4 Conclusion for determinations of σ_H

Overall σ_H values from input scales should be used in all cases for output scales.

As a matter of fact, for input scales, σ_H values are only used in the determination of σ_{pt} . For output scales, σ_{pt} are computed from input scales, taking into account σ_H at the input stage. Consequently, σ_H of output scales is only needed for information and has no impact on assessing parameters.

5 Proposals for calculation of output parameters from input scales

5.1 Recommendations for procedures to determine output scale parameters

The following steps are recommended to determine assessing parameters of output scales:

Step 1: determination of input parameters

Perform the statistical analysis of input scales as usual. Care should be taken to use same options for treatment for all scales. For example, if a standard deviation of rounding is introduced, the same rounding should be applied for all scales. If necessary (i.e. if different options are needed according to the scales), 2 statistical analyses may be performed:

- ✚ One to be used for the assessment of test results of participants;
- ✚ One to be used for the determination of output scale parameters, using the same options for all input scales.

It may be valuable to perform some statistical analyses on scales with a number of participants too low to meet the requirements of internal procedures, but useful to increase the number of input scales even if the related uncertainties are large. In any cases, this adds information for the further treatment.

Also perform a statistical analysis all results together, using only one test result per participant (typically the median value of all the test results from each of the scales in which it provided results). This statistical analysis is intended to provide the overall values of output parameters when necessary.

Step 2: determination of output X_{pt}

Plot X_{pt} and related $u_{X_{pt}}$ from input scales, as function of $\log(F)$, X_{pt} as function of $1/F$ and $\log(X_{pt})$ as function of $1/\sqrt{F}$, and overall X_{pt} as computed at step 1, as shown in Figure 6.a to c.

Build up a weighted series of $n_i \cdot X_{pt,i}$ where i are each of the input scales. Compute the regression coefficients for:

- ✚ The polynomial (degree 2) regression for X_{pt} as function of $\log(F)$, F_0 and HD_0 with Equations (17) and (18) (only the decreasing part for which $F < F_0$ shall be used);
- ✚ The straight-line regression for X_{pt} as function of $\log(F)$;
- ✚ The straight-line regression for X_{pt} as function of $1/F$;
- ✚ The straight-line regression for X_{pt} as function of $1/\sqrt{F}$.

Compute r^2 , Z_a and Z_{ua} (see Equations (15) and (16)), u_a and u_b (see Equations (9) and (10)) for each of the plots.

Decide which model is the most adapted to determine output scale parameters as function of figures, r^2 , Z_a and Z_{ua} results and the number of parameters of the model compared to the number of input scales.

Step 3: determination of output σ_{pt} and σ_{rpt}

Plot $\log(\sigma_{pt})$ and $\log(\sigma_{rpt})$ as function of $\log(F)$, as shown in Figure 12 and Figure 16 respectively.

Build up the weighted series of $n_i \cdot \sigma_{pt,i}$ and $n_i \cdot \sigma_{rpt,i}$ where i are each of the input scales. Compute the regression coefficients for the straight-line regressions of $\log(\sigma_{pt})$ and $\log(\sigma_{rpt})$ as function of $\log(F)$.

Compute $r^2 Z_a$ and Z_{ua} (see Equations (15) and (16)), u_a and u_b (see Equations (9) and (10)) for each of the plots.

Decide whether a decreasing straight line or a constant straight-line (i.e. σ is independent to F) is the most adapted model to determine output scale parameters as function of figures, r^2 , Z_a and Z_{ua} results.

Step 4: determination of output u_{xpt} :

If it was decided at step 3 that σ_{pt} is independent to F , use Equation (21) to determine u_{xpt} of output scales. If it was decided to use a decreasing straight-line to determine output σ_{pt} , use Equation (22) to determine u_{xpt} .

Step 5: determination of σ_H :

Use the overall σ_H determined at step 1 as output value for σ_H .

Step 6: determine output scale parameters:

Use options decided at steps 2 to 5 to compute any output scale parameters and Equation (11) to compute the related uncertainties. Decide whether these output parameters can be used with respect to these uncertainties and how its fits on the figures plotted at steps 2 and 3.

5.2 Example of determination of output scale parameters from input scale parameters

As an example, this procedure was applied to results of 2020 Vickers tests on low hardness metal. 2 types of scales were checked:

- ✚ Output results on scales that were used as input scales, in order to check how output results are in line with input results;
- ✚ Output results on scales that are not input scales, in order to check whether output results make sense.

It provided the following results:

Output results on scales used as input scales:

A comparison between computed parameters and input ones was performed, which results are provided in Table 15 here after.

Table 15. Comparison of input and output parameters for 2020 Vickers tests on low hardness metal.

Scale	X_{pt}				u_{xpt}				σ_{pt}		σ_{rpt}	
	Polyn.	Nix	Li	Input	Polyn.	Nix	Li	Input	Output	Input	Output	Input
0,1	189,3	190,1	189,2	190,3	1,4	1,1	2,4	4,8	11,2	12,2	1,96	2,43
0,3	186,6	185,88	186,6	185,5	1,1	0,46	1,4	3,2	8,67	8,51	1,80	1,64
1	184,6	184,40	185,0	184,1	0,9	0,32	0,9	1,6	6,55	5,7	1,65	1,44
5	183,6	183,89	183,9	184,5	2,0	0,30	0,7	1,2	4,51	4,41	1,46	1,45
10	183,6	183,83	183,6	183,8	2,0	0,30	0,7	1,2	3,84	4,67	1,39	1,49
30	183,6	183,79	183,4	183,6	2,0	0,30	0,6	0,8	2,97	2,03	1,28	1,23

It is reminded that the recommendation to compute σ_H is to adopt the overall value, i.e. in this case, 1,67.

These results show good consistency between computed results and input results. In most cases uncertainties on computed results is better than input ones. As a matter of fact, computed ones are based on a significantly larger data.

Output results on scales that are not input scales:

A comparison between output parameters was performed, which results are provided in Table 16 here after.

Table 16. Output parameters for 2020 Vickers tests on low hardness metal.

Scale	X_{pt}			$U_{X_{pt}}$			σ_{pt}	σ_{rpt}
	Polyn.	Nix	Li	Polyn.	Nix	Li		
0,01	197,8	247,0	203,4	4,8	15,0	7,8	19,1	2,33
0,5	185,6	185,0	185,8	0,9	0,4	1,2	7,70	1,74
20	184,1	183,8	183,4	2,0	0,3	0,6	3,26	1,32
100	183,6	183,8	183,2	2,0	0,5	0,6	2,25	1,17

This comparison shows that polynomial model should be preferred for low load scales. This confirms the statement of § 3.3.2. For high load scales, Li and Ni models seem to provide better approximations, even if the polynomial model provides quite acceptable results as well.

5.3 Evaluation of these recommendations using the results of the years 2017 to 2023

These recommendations were applied to the totality of the available results of the years 2017 to 2023. Results of these evaluations are displayed in Table 17.

Table 17. Evaluation of the recommended procedures with the results of the years 2017 to 2023.

Legend of the columns:

-  Product: SM is metal in the range of 150-250 HV, HM is metal in the range of 550-700 HV;
-  Hardness: HV is Vickers hardness, HBW is Brinell hardness;
-  Nb of input scales: Number of input scales;
-  Nb of input values: Number of input test results;
-  Option for X_{pt} : Option that was used to determine X_{pt} , x^2 = polynomial, x = straight line, Nix = Nix model, Li = Li model, Cst = constant;
-  Option for σ_{pt} and option for σ_{rpt} : options that were used to determine σ_{pt} and for σ_{rpt} , x = straight line, Cst = constant;
-  $m(\Delta z_i)$ and $s(\Delta z_i)$: respectively mean value and standard deviation of differences between z scores (concerning bias) computed from initial calculations on input scales and computed from output parameters;
-  $m(\Delta z_{ri})$ and $s(\Delta z_{ri})$: respectively mean value and standard deviation of differences between zr scores (concerning repeatability) computed from initial calculations on input scales and computed from output parameters;
-  1: percentage of participants that got a more severe (rank difference = 1) with output parameters than with input parameters (example: z_{input} = "Warning" while z_{output} = "Action"). The number provides the extent of the difference (for example "2" when z_{input} = "No alert" while z_{output} = "Action"). In the same way, in columns "-1" and "-2", percentage of participants that got a less severe with output parameters than with input parameters.

Year	Product	Hardness	Nb of input scales	Nb of input V.	Option for X_{pt}	Option for σ_{pt}	m(Δz_i)	s(Δz_i)	Percentage of participants with differences of alerts between input and output parameters					Nb of input V.	Option for σ_{pt}	m(Δz_{ri})	s(Δz_{ri})	Percentage of participants with differences of alerts between input and output parameters				
									-2	-1	0	1	2					-2	-1	0	1	2
2023	HM	HV	3	96	x	a.10 ^b	0,01	0,15	0,0	1,0	95,8	3,1	0,0	84	a.10 ^b	0,02	0,25	0,0	6,0	85,7	8,3	0,0
2023	HM	HBW	4	39	x	Cst	0,00	0,25	0,0	0,0	97,4	2,6	0,0	36	a.10 ^b	0,03	0,26	0,0	5,6	91,7	2,8	0,0
2022	SM	HV	6	141	x ²	a.10 ^b	0,11	0,70	0,0	2,8	94,3	2,8	0,0	135	a.10 ^b	0,03	0,29	0,0	4,4	91,9	3,0	0,7
2022	SM	HBW	4	65	Nix	Cst	-0,02	0,40	0,0	1,5	96,9	1,5	0,0	65	a.10 ^b	-0,01	0,18	0,0	3,1	93,8	3,1	0,0
2022	HM	HV	3	80	Nix	Cst	0,00	0,13	0,0	0,0	100,0	0,0	0,0	71	a.10 ^b	-0,01	0,24	0,0	1,4	95,8	2,8	0,0
2022	HM	HBW	3	32	x	a.10 ^b	0,11	0,50	0,0	0,0	96,9	3,1	0,0	27	Cst	-0,18	0,53	0,0	7,4	92,6	0,0	0,0
2021	SM	HV	6	223	x ²	a.10 ^b	0,01	0,12	0,0	0,4	99,6	0,0	0,0	202	a.10 ^b	-0,01	0,45	0,0	2,0	93,1	4,5	0,5
2021	SM	HBW	4	90	x ²	a.10 ^b	0,09	0,30	0,0	4,4	93,3	2,2	0,0	82	Cst	0,01	0,29	0,0	1,2	92,7	6,1	0,0
2021	HM	HV	3	47	x	a.10 ^b	0,01	0,20	0,0	2,1	97,9	0,0	0,0	47	Cst	0,18	0,30	0,0	0,0	91,5	8,5	0,0
2021	HM	HBW	1	14	Cst	Cst	0,19	0,24	0,0	14,3	85,7	0,0	0,0	14	Cst	0,02	0,02	0,0	0,0	100,0	0,0	0,0
2020	SM	HV	6	143	x ²	a.10 ^b	0,02	0,20	0,0	0,7	98,6	0,7	0,0	139	a.10 ^b	0,01	0,16	0,0	0,0	97,8	2,2	0,0
2020	SM	HBW	4	33	Cst	Cst	-0,09	0,44	0,0	0,0	97,0	3,0	0,0	30	Cst	0,13	0,18	0,0	0,0	93,3	6,7	0,0
2020	HM	HV	3	45	x	a.10 ^b	0,00	0,23	0,0	2,2	95,6	2,2	0,0	43	Cst	-0,08	0,13	0,0	9,3	90,7	0,0	0,0
2020	HM	HBW	1	10	Cst	Cst	0,12	0,41	0,0	10,0	90,0	0,0	0,0	9	Cst	-0,23	0,18	0,0	11,1	88,9	0,0	0,0
2019	SM	HV	6	98	x ²	a.10 ^b	0,00	0,12	0,0	0,0	100,0	0,0	0,0	93	a.10 ^b	0,00	0,06	0,0	0,0	96,8	3,2	0,0
2019	SM	HBW	4	41	x	a.10 ^b	0,08	0,39	0,0	2,4	95,1	2,4	0,0	38	a.10 ^b	0,08	0,27	0,0	0,0	89,5	10,5	0,0
2019	HM	HV	3	65	x	Cst	0,91	2,32	0,0	6,2	93,8	0,0	0,0	57	a.10 ^b	0,00	0,00	0,0	0,0	100,0	0,0	0,0
2019	HM	HBW	2	25	Cst	Cst	-0,09	0,36	0,0	0,0	100,0	0,0	0,0	25	Cst	0,15	0,52	0,0	4,0	76,0	8,0	12,0
2018	SM	HV	6	96	x	a.10 ^b	0,01	0,17	0,0	1,0	99,0	0,0	0,0	95	a.10 ^b	0,01	0,29	0,0	1,1	92,6	6,3	0,0
2018	SM	HBW	3	38	Cst	Cst	0,06	0,38	0,0	0,0	97,4	2,6	0,0	35	a.10 ^b	-0,23	1,39	0,0	2,9	97,1	0,0	0,0
2018	HM	HV	3	79	x	a.10 ^b	0,00	0,02	0,0	0,0	98,7	1,3	0,0	74	a.10 ^b	0,00	0,01	0,0	0,0	100,0	0,0	0,0
2018	HM	HBW	2	24	Cst	Cst	-0,01	0,42	0,0	4,2	91,7	4,2	0,0	19	Cst	-0,05	0,43	0,0	26,3	68,4	5,3	0,0
2017	SM	HV	4	30	x	Cst	-0,05	0,60	0,0	3,3	96,7	0,0	0,0	23	Cst	-0,43	1,88	0,0	4,3	91,3	0,0	4,3
2017	SM	HBW	2	8	Cst	Cst	-0,10	0,52	0,0	0,0	100,0	0,0	0,0	7	Cst	-0,02	0,30	0,0	0,0	100,0	0,0	0,0
2017	HM	HV	3	36	Cst	a.10 ^b	-0,06	0,34	0,0	0,0	97,2	2,8	0,0	35	Cst	0,10	0,31	0,0	5,7	88,6	5,7	0,0
2017	HM	HBW	2	14	Cst	Cst	-0,29	0,78	0,0	0,0	92,9	7,1	0,0	14	Cst	-0,11	0,36	14,3	0,0	78,6	7,1	0,0

These results show that the differences in scoring of participants are quite low. In most cases, differences in alerts are due to scores close to the limits (example 1,95 and 2,05 close to each other but separated by the limit = 2). In a very small number of cases, a major difference occurs (i.e. “No alert” and “Action signal”). This is encountered in 2 types of situations:

1. When many input scales are present with many input results (example: 2021-SM-HV). In those cases, the differences always are in the “+” direction. That can let us suppose that output parameters are more efficient than the input ones, and that output results are more reliable than input ones;

2. When few input scales are present, what lead the choice to “constants” for much or all parameters, leading to a lower efficiency of output parameters than input ones. In those case, the statistical officer should decide not to use output results for the assessment of participants.

No formal evaluation of the relevance of the models of § 2.2 could be performed, because the “true” values cannot be known, but the very good correlation between output parameters determined by using the different proposed models when an enough size of input data is available tends to prove that they are relevant for performing PT.

6 Conclusions

This study showed that it is possible to assess hardness test results of a given scale when an enough amount of test results is available for adjacent scales. To achieve this:

- ✚ 5 different methods (i.e. $HD=a.\log(F)^2+b.\log(F)+c$, $HD=a.\log(F)+b$, $HD=a/F+b$, $HD=a/\sqrt{F}+b$, $HD=a$) can be used to determine the assigned value for a given F scale;
- ✚ 2 different methods (i.e. $HD=a.10^{b.F}$, $HD=a$) can be used to determine the proficiency standard deviation, the repeatability standard deviation and the uncertainty on the assigned value for a given F scale.

The person responsible for the statistical treatment shall choose the best options, that depend on the ILC conditions (i.e. the number of available input scales, the overall number of test results and how the input results fit with what is expected for relationship between measured hardness and load). This study provides a recommended procedure to deal with the options and proposes parameters to check the adequacy of each of the options to help the choice of the most adapted one.

An assessment of the results obtained with this procedure on CompaLab ILC results obtained during the 2017-2023 years was performed, leading to very small differences in the scoring of participants for available scales. When the size of the input data is large, output scorings are likely to be more efficient than usual ones.

For Brinell scales, most of available input test results are from the HBW 2,5/187,5 scale, but a lower amount of test results is usually available for HBW 1/30, HBW 5/750 and HBW 10/3000 scales. The proposed extrapolating procedure is likely to help the assessment of these other scales.

For Vickers scales, many test results are usually available for HV5, HV10 and HV30, and HV0,1 HV0,3 and HV1 on low hardness metals. The proposed extrapolating procedure is likely to help the assessment of other scales ranging from HV0,01 to HV100.

7 References

- [1] ISO 6506-1:2014 – Metallic materials – Brinell Hardness test – Test method
- [2] ISO 6507-1:2018 – Metallic materials – Vickers Hardness test – Test method
- [3] ISO 13528:2022 - Statistical methods for use in proficiency testing by interlaboratory comparison
- [4] B.W. Mott, Micro-indentation Hardness Testing, Butterworths Scientific Publications 272 pages, 1956
- [5] M. Atkinson, “Origin of the size effect in indentation of metals”, International journal of mechanical sciences Vol. 33 Issue 10 pp. 843 to 850, 1991
DOI: [https://doi.org/10.1016/0020-7403\(91\)90006-O](https://doi.org/10.1016/0020-7403(91)90006-O)

- [6] H. Li and R.C. Bradt, The microhardness indentation load/size effect in rutile and cassiterite single crystals, Journal of material science Vol. 28 pp. 917 to 926, 1993
DOI: <https://doi.org/10.1007/BF00400874>
- [7] W.D. Nix and H. Gao "Indentation size effect in crystalline materials: a law for strain gradient plasticity", Journal of the Mechanics and Physics of Solids Vol. 46 N°3 pp. 411 to 425, 1998
DOI: <https://doi.org/10.1016/j.jisolsolstr.2013.05.002>
- [8] J.G. Swagener, E.P. George, G.M. Pharr, "The correlation of the indentation size effect measured with indenters of various shapes", Journal of the Mechanics and Physics of Solids Vol 50 N°4 pp. 681 to 694, 04/2002
DOI: [https://doi.org/10.1016/S0022-5096\(01\)00103-X](https://doi.org/10.1016/S0022-5096(01)00103-X)
- [9] X.J. Ren, R.M. Hooper, C. Griffith, J.L. Henshall, "Indentation size effect in ceramics: correlation with H/E", Journal of science letters Vol. 22 pp. 1105 to 1106, 2003
DOI: <https://doi.org/10.1023/A:1024947210604>
- [10] N. Budarsia, Indentation size effect ISE of Vickers hardness in steels: correlation with H/E, Applied mechanics and materials Vol 91 pp. 23 to 28, 2013
DOI: <http://dx.doi.org/10.4028/www.scientific.net/AMM.391.23>
- [11] Xiaozhen Li, Weixu Zhang, Zhelin Dong, Zhiguo Wang, Dingjun Li and Jianpu Zhang, "Study of size effects in indentation tests", Coatings Vol. 12 Issue 12, 2022
DOI: <http://dx.doi.org/10.3390/coatings12121962>
- [12] ISO 6506-1:2014 Metallic materials – Brinell hardness tests – Test method
- [13] David Luengo, Luca Martino, Mónica Bugallo, Víctor Elvira and Simo Särkkä, "A survey of Monte Carlo methods for parameter estimation" EURASIP Journal on Advances in Signal Processing, Article 25, May 2020
DOI: <https://doi.org/10.1186/s13634-020-00675-6>
- [14] L.J. Hollebecq, Appropriate rankits to use for normal probability plots and standard deviation probability plots, CompaLab technical publications, 10/2023
DOI: <https://www.compalab.org/medias/files/publication-interne-valeurs-t-en-2023-10.pdf>

Annex:

Detailed test results that were obtained for this study

**Table A1: Lower and upper limits of the IC95% of s_H/σ_H
for N_p participants, 3 samples per participant, 2 test results per sample (enlargement coefficient for uncertainty: $k = 2$)**

$N_p=63$			$N_p=40$			$N_p=25$		
σ_H/σ_r	Lower limit	Upper limit	σ_H/σ_r	Lower limit	Upper limit	σ_H/σ_r	Lower limit	Upper limit
2	0,821±0,006	1,103±0,006	2	0,783±0,008	1,137±0,006	2	0,727±0,008	1,193±0,008
1,6	0,806±0,006	1,100±0,006	1,6	0,763±0,008	1,141±0,006	1,6	0,700±0,008	1,197±0,006
1,25	0,766±0,008	1,102±0,008	1,25	0,717±0,010	1,147±0,006	1,25	0,634±0,008	1,222±0,010
1	0,705±0,010	1,101±0,008	1	0,631±0,010	1,162±0,010	1	0,528±0,016	1,241±0,010
0,8	0,593±0,016	1,110±0,010	0,8	0,474±0,022	1,180±0,012	0,8	0,272±0,030	1,275±0,010
0,7	0,434±0,026	1,115±0,012	0,7	0,261±0,038	1,192±0,014	0,7	0,045±0,018	1,290±0,012
0,6	0,221±0,032	1,127±0,012	0,6	0,051±0,022	1,215±0,016	0,6	0,001±0,002	1,331±0,014
0,5	0,019±0,012	1,146±0,018	0,5	0,000±0,000	1,252±0,018	0,5	0,000±0,000	1,404±0,014
0,4	0,000±0,000	1,184±0,024	0,4	0,000±0,000	1,336±0,022	0,4	0,000±0,000	1,512±0,026
0,32	0,000±0,000	1,247±0,034	0,32	0,000±0,000	1,434±0,034	0,32	0,000±0,000	1,693±0,028
0,25	0,000±0,000	1,299±0,052	0,25	0,000±0,000	1,621±0,046	0,25	0,000±0,000	1,965±0,038
0,2	0,000±0,000	1,418±0,078	0,2	0,000±0,000	1,786±0,064	0,2	0,000±0,000	2,295±0,050
0,16	0,000±0,000	1,593±0,096	0,16	0,000±0,000	2,154±0,074	0,16	0,000±0,000	2,710±0,066
0,125	0,000±0,000	1,698±0,146	0,125	0,000±0,000	2,513±0,112	0,125	0,000±0,000	3,329±0,078
0,1	0,000±0,000	2,041±0,182	0,1	0,000±0,000	3,181±0,134	0,1	0,000±0,000	3,949±0,118
0,08	0,000±0,000	2,423±0,218	0,08	0,000±0,000	3,509±0,204	0,08	0,000±0,000	5,167±0,128
$N_p=50$			$N_p=32$			$N_p=20$		
σ_H/σ_r	Lower limit	Upper limit	σ_H/σ_r	Lower limit	Upper limit	σ_H/σ_r	Lower limit	Upper limit
2	0,806±0,006	1,120±0,008	2	0,761±0,006	1,176±0,008	2	0,702±0,008	1,217±0,008
1,6	0,788±0,008	1,116±0,006	1,6	0,734±0,008	1,166±0,008	1,6	0,667±0,008	1,230±0,008
1,25	0,735±0,008	1,130±0,010	1,25	0,677±0,010	1,184±0,008	1,25	0,598±0,010	1,246±0,008
1	0,659±0,012	1,129±0,008	1	0,593±0,012	1,190±0,010	1	0,447±0,018	1,278±0,010
0,8	0,516±0,022	1,156±0,010	0,8	0,387±0,026	1,223±0,010	0,8	0,121±0,030	1,299±0,010
0,7	0,340±0,030	1,157±0,012	0,7	0,180±0,032	1,248±0,010	0,7	0,013±0,010	1,347±0,012
0,6	0,101±0,026	1,191±0,014	0,6	0,009±0,008	1,274±0,016	0,6	0,000±0,000	1,392±0,016
0,5	0,001±0,002	1,193±0,016	0,5	0,000±0,000	1,316±0,014	0,5	0,000±0,000	1,481±0,018
0,4	0,000±0,000	1,267±0,022	0,4	0,000±0,000	1,417±0,024	0,4	0,000±0,000	1,611±0,024
0,32	0,000±0,000	1,302±0,030	0,32	0,000±0,000	1,554±0,032	0,32	0,000±0,000	1,796±0,030
0,25	0,000±0,000	1,447±0,042	0,25	0,000±0,000	1,766±0,044	0,25	0,000±0,000	2,111±0,038
0,2	0,000±0,000	1,643±0,074	0,2	0,000±0,000	2,062±0,054	0,2	0,000±0,000	2,510±0,044
0,16	0,000±0,000	1,813±0,090	0,16	0,000±0,000	2,418±0,074	0,16	0,000±0,000	2,953±0,064
0,125	0,000±0,000	2,184±0,124	0,125	0,000±0,000	2,952±0,098	0,125	0,000±0,000	3,751±0,080
0,1	0,000±0,000	2,506±0,156	0,1	0,000±0,000	3,571±0,128	0,1	0,000±0,000	4,541±0,106
0,08	0,000±0,000	3,255±0,202	0,08	0,000±0,000	4,379±0,150	0,08	0,000±0,000	5,630±0,128

N_p=16		
σ_H/σ_r	Lower limit	Upper limit
2	0,667±0,006	1,261±0,008
1,6	0,627±0,008	1,270±0,008
1,25	0,546±0,012	1,290±0,010
1	0,380±0,020	1,314±0,010
0,8	0,034±0,014	1,352±0,010
0,7	0,000±0,000	1,400±0,012
0,6	0,000±0,000	1,466±0,016
0,5	0,000±0,000	1,560±0,016
0,4	0,000±0,000	1,724±0,020
0,32	0,000±0,000	1,934±0,026
0,25	0,000±0,000	2,259±0,038
0,2	0,000±0,000	2,728±0,044
0,16	0,000±0,000	3,284±0,054
0,125	0,000±0,000	4,074±0,074
0,1	0,000±0,000	5,010±0,082
0,08	0,000±0,000	6,180±0,120

N_p=13		
σ_H/σ_r	Lower limit	Upper limit
2	0,637±0,006	1,291±0,008
1,6	0,582±0,010	1,303±0,008
1,25	0,500±0,012	1,337±0,010
1	0,264±0,026	1,359±0,010
0,8	0,006±0,008	1,418±0,012
0,7	0,000±0,000	1,454±0,012
0,6	0,000±0,000	1,519±0,014
0,5	0,000±0,000	1,640±0,014
0,4	0,000±0,000	1,790±0,020
0,32	0,000±0,000	2,075±0,028
0,25	0,000±0,000	2,455±0,034
0,2	0,000±0,000	2,914±0,046
0,16	0,000±0,000	3,558±0,060
0,125	0,000±0,000	4,366±0,076
0,1	0,000±0,000	5,463±0,094
0,08	0,000±0,000	6,789±0,116

N_p=10		
σ_H/σ_r	Lower limit	Upper limit
2	0,583±0,008	1,335±0,006
1,6	0,530±0,008	1,355±0,008
1,25	0,391±0,016	1,387±0,010
1	0,106±0,024	1,426±0,010
0,8	0,000±0,000	1,488±0,012
0,7	0,000±0,000	1,535±0,012
0,6	0,000±0,000	1,610±0,016
0,5	0,000±0,000	1,740±0,018
0,4	0,000±0,000	1,950±0,020
0,32	0,000±0,000	2,248±0,030
0,25	0,000±0,000	2,669±0,034
0,2	0,000±0,000	3,224±0,046
0,16	0,000±0,000	3,909±0,066
0,125	0,000±0,000	4,919±0,072
0,1	0,000±0,000	6,016±0,084
0,08	0,000±0,000	7,379±0,108

N_p=8		
σ_H/σ_r	Lower limit	Upper limit
2	0,534±0,008	1,387±0,008
1,6	0,462±0,010	1,414±0,010
1,25	0,294±0,018	1,431±0,010
1	0,029±0,014	1,479±0,010
0,8	0,000±0,000	1,559±0,012
0,7	0,000±0,000	1,606±0,014
0,6	0,000±0,000	1,708±0,016
0,5	0,000±0,000	1,826±0,018
0,4	0,000±0,000	2,069±0,024
0,32	0,000±0,000	2,391±0,024
0,25	0,000±0,000	2,884±0,036
0,2	0,000±0,000	3,537±0,044
0,16	0,000±0,000	4,202±0,054
0,125	0,000±0,000	5,274±0,066
0,1	0,000±0,000	6,516±0,090
0,08	0,000±0,000	8,131±0,096

N_p=6		
σ_H/σ_r	Lower limit	Upper limit
2	0,463±0,010	1,452±0,008
1,6	0,364±0,012	1,478±0,008
1,25	0,129±0,024	1,503±0,010
1	0,000±0,000	1,565±0,010
0,8	0,000±0,000	1,657±0,010
0,7	0,000±0,000	1,720±0,012
0,6	0,000±0,000	1,818±0,014
0,5	0,000±0,000	1,985±0,018
0,4	0,000±0,000	2,241±0,022
0,32	0,000±0,000	2,606±0,028
0,25	0,000±0,000	3,152±0,026
0,2	0,000±0,000	3,812±0,042
0,16	0,000±0,000	4,598±0,042
0,125	0,000±0,000	5,888±0,066
0,1	0,000±0,000	7,218±0,074
0,08	0,000±0,000	9,019±0,096