

Ordinary Least Squares with Laboratory Calibrations: A Practical Way to Show Students that This Fitting Model may Easily Yield Biased Results When Used Indiscriminately

Juan M. Sanchez*

Department of Chemistry, University of Girona, Girona, Spain

*Corresponding author: juanma.sanchez@udg.edu

Abstract Analytical calibration using ordinary least squares (OLS) is the most widely applied response function for calibration in all type of laboratories. However, this calibration function is not always the most adequate and its indiscriminant use can lead to obtain biased estimates of unknowns. Students need to be taught about the practical requirements needed to obtain good results with OLS and when this fitting method is not accurate. Different experimental calibration curves were obtained in laboratory sessions using two common instrumental techniques: chromatography and atomic absorption spectrometry. After discussion seminars evaluating the data obtained by students, they were able to understand that linear fitting was not the most accurate model using atomic absorption spectrometry and a quadratic fitting provided most accurate estimates. Linearity was confirmed in chromatographic calibrations, but data presented heteroscedasticity, which is very common in calibrations done in chemical and biological analyses. A simple experiment was applied to show students how the use of the regression coefficients obtained by OLS with heteroscedastic data lead to highly biased estimates near the quantification limits of the calibration curve. The results obtained allowed to show students that, despite being widely used, OLS is not the most adequate fitting model to obtain accurate and precise results with many calibration methods routinely used in chemical and biological laboratories.

Keywords: calibration curves, linearity, homoscedasticity, biased results, outliers

Cite This Article: Juan M. Sanchez, "Ordinary Least Squares with Laboratory Calibrations: A Practical Way to Show Students that This Fitting Model may Easily Yield Biased Results When Used Indiscriminately." *World Journal of Analytical Chemistry*, vol. 5, no. 1 (2017): 1-8. doi: 10.12691/wjac-5-1-1.

1. Introduction

Many chemistry and biological experiments are devoted to the estimation of an unknown sample concentration. For this purpose, a well-designed and interpreted response function ($y=f(x)$, where y is the dependent variable, usually an instrumental signal, and x is the independent variable, commonly a concentration) is essential. Despite the critical importance of a correct definition and proper use of the calibration function, the practical aspects to consider when applying calibration curves in laboratory analyses and how to evaluate the utility of the parameters obtained are usually briefly showed to university students.

It is common to associate a calibration function with a linear function using ordinary least squares (OLS), a mathematical model that is implemented in the common computer software and scientific calculators. The function is defined by:

$$y = b_0 + b_1x$$

where b_0 and b_1 are the regression coefficients, called the intercept with the y -axis or origin (b_0) and the slope or

sensitivity (b_1). OLS estimates these coefficients with the goal of minimizing the sum of the square residuals generated between each experimental response (y_i) and its predicted y -value from the generated function ($\hat{y}_i = b_0 + b_1x_i$):

$$\sum res^2 = \sum (\hat{y}_i - y_i)^2.$$

During conversations along years with university students, they always claim that analytical chemistry lecturers are too repetitive about the calibration subject and the practical requirements needed to apply and use correctly a response function. They indicate that all scientific calculators can perform basic OLS calculations, which allow them to obtain in few minutes the three only parameters they are interested about: intercept, slope, and determination coefficient (R^2). This simple answer indicates that they can make many mistakes when doing calibrations. Unfortunately, the same idea is followed by many researchers.

There are many excellent textbooks containing calibration chapters where OLS is widely explained from a theoretical point of view. However, this topic should be taught differently when students have to face with calibrations in

a practical way, during laboratory sessions, where the main objective is to use the response function for obtaining accurate and precise estimations of unknowns in samples. Students must understand that the correct application of a calibration model must not only rely in the statistics used to build the curve, understanding the limitations of the assay and the instrument are also critical to the curve's success [1]. When manipulating laboratory data, students have to take into account the experimental parameters that can affect their results, to discern about the accuracy and precision of their results, and to be able to determine whether a linear response function is appropriate for a specific experiment.

When calibration functions are determined from experimental laboratory data, there is a common tendency to avoid the fact that any statistical function is based in the fulfillment of some requirements, which have been applied in the development of the model. In the specific case of OLS the requirements to be fulfilled are: (i) the relationship between the dependent (signal) and independent (concentration) variables is linear; (ii) there is no error for the independent variable or is less than one-tenth of the error in the dependent variable (i.e., errors in preparing standards are negligible compared to signal errors); (iii) errors in the dependent variable are normally distributed for each value of the independent variable; and (iv) variance of the dependent variable is constant at all values of the independent variable (homoscedasticity). Statistically accurate estimates can only be provided when all assumptions are fulfilled. Unfortunately, it is very common to face situations in laboratory calibrations where some of these requirements are not fulfilled, which can lead to significant bias and imprecision in concentration estimates [2,3].

Linear OLS fitting is the most widely used response function for calibrations in chemical and biological analysis for one main practical reason: many instruments usually show linear detector response along some orders of magnitude. However, students have to understand that, in practice, the linear function is not the only and universal option, and cannot be indiscriminately applied to all analyses. There are some detection techniques, such as atomic absorption spectrometry (AAS) [4] and immunoassay methods [5], which are rarely linear in their response or present very short linear ranges to be useful. In such situations, other response functions have to be chosen.

Another important aspect that students should know is that before an analytical method is implemented for routine use in a laboratory, it should be validated to demonstrate its fitness-for-purpose, which involves the verification of the calibration curve as a part of the suitability check carried out. The validation permits to identify and then control the factors responsible for signal variance, which could lead to biased estimation of unknowns. Once it has been established that the best response function follows a linear model and the confidence intervals for the regression parameters are well defined, an analytical method can be applied in routine analyses. This fact introduces another important topic to take into account when facing with analytical calibrations in the laboratory: method validation. The requirement of using validated methods in laboratories is very important and of common use in the industry, and it is also a

requirement for certified and accredited laboratories. However, the use of validated assay methods is not a routine in many teaching and research laboratories.

This study presents a discussion about the practical requirements to take into account for analytical calibrations done in chemistry and biological analyses, with emphasis in the most common problems and misconceptions observed during different university laboratory sessions where calibrations were used to determine estimates of unknowns.

2. Methods

210 students performed the laboratory sessions during the sixth term (out of eight) of the biotechnology degree at the University of Girona. The results were obtained during four academic years, from 2013/14 to 2016/17. Laboratory working groups consisted of 14-18 students, with 3-4 working groups every academic year.

Calibration curves obtained for five analytical methods using two different detection techniques were evaluated (liquid chromatography with ultraviolet detection, HPLC-UV, and AAS). HPLC methods are expected to produce linear response, whereas AAS is usually non-linear in its instrumental response. The study has been limited to external calibration curves; internal standard and standard addition calibration curves are out of the scope of this discussion. A total of 77 HPLC and 56 AAS calibration curves were obtained during the period evaluated. A minimum of six standards were used in each calibration, with concentration levels evenly distributed along the working range. At least two calibration curves were obtained for each method inside a working group. To allow assessing the precision at each calibration level, students were informed to prepare standards at the same concentration levels during an academic year. A discussion meeting at the end of all laboratory sessions was performed every academic year with all students, where calibration results were compared and discussed.

3. Results and Discussion

Before starting laboratory sessions, students were asked about their previous use of calibration curves. They had used calibration curves many times in previous laboratory subjects and the response function was always assumed to be linear, applying OLS with no preliminary and theoretical considerations. They calculated the slope, intercept, and determination coefficient as the only regression parameters of interest. This confirmed that students were used to work with OLS but not to consider whether linear calibration is the best option and to assess the validity of the calibration function. When students were asked about the assumptions required for an OLS fitting to be adequate, all of them answered that they know nothing about these requirements. Unfortunately, this answer was not surprising as they had indiscriminately applied OLS in all their previous laboratory experiments without a preliminary evaluation of the model and the requirements related, which happens in many laboratories [2].

The discussion sessions at the end of each academic year allowed evaluating the experimental results and detecting different mistakes and misconceptions in the calibration curves proposed by students. The majority of the mistakes and misconceptions found were associated to one of the following parameters: (i) homoscedasticity of the data; (ii) estimation of the goodness-of-fit; (iii) rejection of outliers; and (iv) treatment of the origin.

3.1. Homoscedasticity/Heteroscedasticity Evaluation

Many analytical and biological methods yield non-constant variances along the calibration range (heteroscedasticity) [2,6,7]. Therefore, the requirement of equal variances (homoscedasticity) is frequently not met in laboratory calibrations. From a practical point of view, the main experimental limitation for checking the assumption of variance homogeneity is the need to measure standard replicates, which is time consuming. This procedure is not usually applied for daily calibrations and it is only a requirement during method validation. To allow students to assess variance homogeneity, the measurements obtained at the same concentration level for different calibrations performed during the same academic year were used as replicates and the fulfilment of this requirement was evaluated during the discussion seminars.

Two procedures can be applied for assessing homoscedasticity. The most simple is by plotting the results of the different replicates in the calibration graph or by obtaining the plot of the residuals, which is usually more instructive. An increase in the dispersion of the replicates or residuals over the concentration range suggests heteroscedasticity. A most reliable option is to perform a statistical test. The most adequate is the Levene's test, designed to assess the equality of variances for more than two groups.

Plots of the residuals obtained for all HPLC-UV methods showed an increase in the dispersion of the residuals (Figure 1a), indicating unequal variances. The calculation of the relative standard deviations (RSD) of the replicates at each concentration level showed that this is the constant parameter across the curve ($RSDs=5.0\pm 0.4\%$ for the data in Figure 1a). Some studies have demonstrated that in heteroscedastic conditions, the instrument absolute errors (standard deviation, s , or variance, s^2) are usually proportional to the concentration [8-14], and RSD is the constant parameter across the curve instead of the variance [8,10]. The Levene's test was applied using the replicate measurements evaluated during discussion sessions and it confirmed that all chromatographic calibrations were heteroscedastic ($p<0.015$).

This result indicated that the concentrations calculated by students in their analyses by HPLC need to be evaluated carefully because when OLS is applied with heteroscedastic data, the true variance of the estimates can be biased, which may lead to obtain incorrect results [1,15]. It has been demonstrated that despite the proportional error for heteroscedastic data is maintained most of the range of practical interest, it becomes constant in the low-signal limit [11,12]. For this reason, with heteroscedastic linear calibrations, OLS models fails to obtain accurate estimates in the lower range of the calibration curves, near

the limit of quantification (LOQ), where precision loss can be as high as one order of magnitude [1,8,10,11,12]. In this situation, the best option to obtain accurate and precise estimates is to use weighted least squares (WLS) instead of OLS [1,6], which requires more complex calculations because weighting factors have to be determined at each concentration level. WLS calculations are not directly done with conventional scientific calculators and only statistical softwares introduce this option. Despite this, it is possible to prepare a macro in spreadsheets such as Excel to perform these calculations.

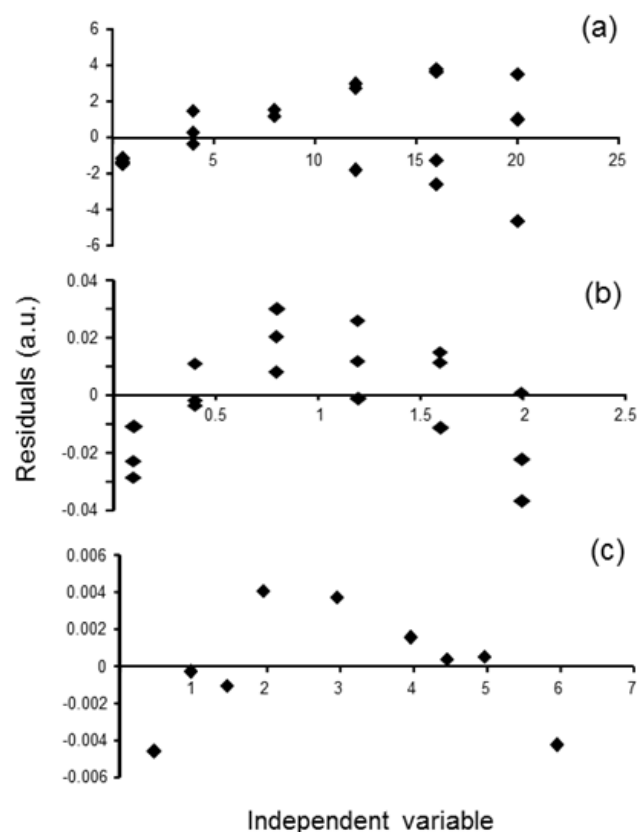


Figure 1. Plots of the residuals obtained for some of the calibration curves evaluated. (a) HPLC-UV method with replicate measurements; (b) atomic absorption spectrometry method with replicate measurements; (c) atomic absorption spectrometry method without replicate measurements

During the seminar sessions, both the OLS equation and the WLS equation were calculated for HPLC curves. The concentration estimates for all samples were determined with the two functions and the percentages of difference between them were calculated. For all samples with estimates at least one order of magnitude above the LOQ, the differences obtained ranged from -5 to +4%, which indicates that OLS gives accurate estimates for samples well above the LOQ. For samples with estimates near the LOQ, however, differences ranged from -15 to -120%, which indicates that highly negative biased results are obtained near the LOQ with OLS for heteroscedastic linear calibrations.

A simple experiment was done using one of the samples to be analyzed by HPLC to check this fact. For one of their samples, students were said to prepare a dilution of the aliquot obtained after sample treatment and just before instrumental analysis. The dilution factor was chosen to obtain a diluted aliquot at a concentration near

the LOQ. They had to analyze both the original aliquot and the diluted one and determine the concentration of the original sample from both results. The final concentration determined for each sample in these experiments should be the same independently of when the diluted or the undiluted aliquot was measured. The t-test for paired data confirmed that equivalent results were obtained between the results calculated using WLS and OLS for the undiluted aliquot ($p=0.644$, $n=21$, in one of the seminars). However, significantly smaller results were obtained applying OLS when the results of the diluted aliquot were measured ($p<0.001$). It was also found that the results obtained from diluted and undiluted samples were equivalent applying WLS ($p=0.826$), whereas significant differences were obtained with OLS ($p<0.001$).

In the case of AAS calibrations (Figure 1b), the plots of the residuals suggested homoscedasticity, which was confirmed by the Levene's test ($p=0.955$ for data in Figure 1b). A rule of thumb is that if a calibration is restricted to concentrations of up to about 50-100 times the detection limit, equal variances are usually expected; outside this range, unequal variances are obtained [16]. The short dynamic range used in the AAS calibrations (less than one order of magnitude) allowed obtaining equal variances.

3.2. Estimation of the Goodness-of-fit

This is probably the parameter showing most misconceptions between students, and many times between researchers. Linearity is a requirement with OLS but it is not correct to assume that all calibration curves are linear. Different procedures for assessing linearity of a calibration, such as graphical plots, statistical tests and numerical parameters, have been proposed [7,17-19]. Unfortunately, linearity assessment has been subject to different definitions and interpretations, which results in some of these procedures not being equivalent and, sometimes, contradictory results can be obtained.

The most frequently used numerical parameters in practice for assessing the goodness-of-fit are the Pearson's or linear correlation coefficient (R) and the determination coefficient (R^2). When students were asked about assessment of linearity in their previous calibrations, the answer obtained was always the same: they had only used R or R^2 and a minimum value of 0.99 was required to confirm linearity.

Correlation and regression are two concepts intimately related when applying OLS because the calculation and handling of data are similar, which has led to an indifferent use of these two parameters. Unfortunately, none of these parameters should be considered to determine linearity as they are a poor measure of the curve-fit quality [7,17,20-25]. Low R^2 values can be assumed for perfectly linear relationships depending on the fitness for purpose of the methods, and high R^2 values can easily be obtained with data non-linearly distributed along a calibration curve [8,20].

A practical option to check linearity in routine calibrations is the use of a graphical plot. The simplest procedure is by plotting the paired data and to carry out a visual inspection of the distribution of the data over the calibration line (Figure 2). However, the plot of the residuals can provide most clear information concerning

the goodness-of-fit (Figure 1). Sometimes, the graph of the standardized residuals or the studentized residuals is used, but the only advantage of using standardized values is in detecting possible outliers. Despite the practical utility of the residuals plots, it is not a potent tool to identify deviations from the linear regression model because no statistical test is involved. Moreover, some experience is required for a correct interpretation of the plot of the residuals. A large number of standards, with replicate measurements in case of heteroscedastic data, are recommended for a correct interpretation of the plot. It has to be taken into account that the inclusion of standards near the LOQ can easily disrupt the residuals plot, especially with heteroscedastic data.

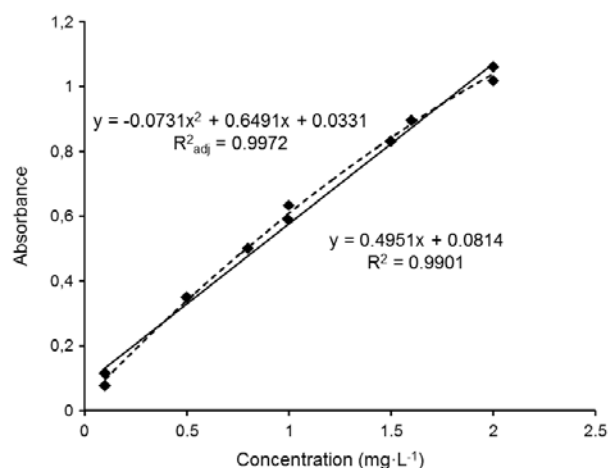


Figure 2. Calibration results obtained for the determination of Zn by atomic absorption spectrometry. Continuous line corresponds to the OLS fitting, whereas dashed line is for quadratic fitting

The evaluation of the plots of the residuals suggested that chromatographic methods evaluated are linear and heteroscedastic (Figure 1a). In the case of the AAS calibrations, residuals distribution gave a non-linear distribution (Figure 1b and 1c). It is well known that AAS instruments are non-linear in their response [4]. For practical applications, small dynamic ranges are used in calibrations with AAS detection to obtain near-linear curves. However, it has been demonstrated that a quadratic function provides most satisfactory fitting than OLS with AAS calibration curves with limited curvature [26]. In practically all AAS calibration curves evaluated in the present study a slight curvature was observed and, despite OLS fittings gave R^2 values ranging from 0.990 to 0.999, better adjustments were obtained with quadratic fittings (Figure 2).

The main drawback of applying OLS for non-linear curves is that results obtained from interpolation in linear models present systematic error [20]. To check this fact, a quality standard control was prepared by laboratory technicians at mid-scale level and was analyzed by all students. The percentage of agreement between the concentration of the standard control and the amount estimated with both linear and quadratic fitting was estimated (Figure 3). The quadratic fitting yielded non-biased results, whereas biased results, ranging from +5 to +7%, were obtained with OLS fitting. Similar bias errors and percentages were obtained in other studies comparing OLS with quadratic fitting for AAS calibrations [21].

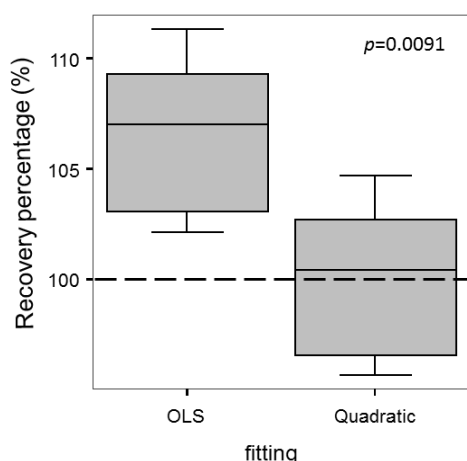


Figure 3. Box plots showing the recovery results obtained for a mid-scale quality standard control with linear calibration (OLS) and second-order calibration curve (quadratic)

Another useful graphical option is to draw the linearity or response factors plot. This plot presents the advantage that usually gives a better idea of the spread of the data points around the fitted straight line and also gives some indication on the percentage of error to be expected for estimated concentrations [22]. The response factor (RF) or sensitivity for each standard is obtained by calculating the signal-to-concentration ratio, and it is plotted against the concentration of that standard. To prevent leverage effects due to possible outliers, the median RF is calculated as the center value instead of the mean, and tolerance limits are determined by multiplying the median value by constant factors. The IUPAC has suggested a tolerance limit of $\pm 5\%$ for chromatography calibration curves [27]. Other guidelines accept tolerance limits $\pm 20\%$ [28].

In the case of chromatographic calibrations, linearity plots confirmed the goodness-of-fit of the linear function (Figure 4a). RF values were randomly distributed around the mean value, falling inside the set tolerance limits for chromatographic methods of $\pm 5\%$, except for a positive deviation that was observed for the first standard when it was prepared at a level near the LOQ. It has to be remembered that chromatographic calibrations presented heteroscedasticity and this implies that estimates near the LOQ give biased results when OLS fitting is used.

The linearity plots obtained for AAS methods (Figure 4b) confirmed a non-linear distribution. RF values were not randomly distributed around the mean value, falling many times out of the confidence limit. The percent relative errors of back-calculated concentrations of each standard (%RE) were determined by applying both the OLS and the quadratic functions (Table 1). It has been proposed that this value should not surpass a cut-off limit of $\pm 15\%$, and of $\pm 20\%$ near the LOQ, for a correct curve fitting [7,18,28]. The linear function gave non-random distribution of %RE with large bias at low concentration, usually surpassing the $\pm 20\%$ cut-off. However, quadratic functions gave random error distribution along the calibration curve, with acceptable percentage of error at all levels, including near the LOQ.

The most powerful way of assessing linearity is by the use of statistical tests. However, these tests require many experiments with the use of large number of replicates at each concentration level. For this reason, their use is not

common for routine calibrations and it is practically restricted to validation processes. There are two statistical test widely accepted for assessing linearity: analysis of variance lack-of-fit (LOF) and Mandell's test. The LOF test is more general [22] and requires 4-6 replicates at each experimental level to decrease the uncertainty significantly along the experimental range [22,29]. The Mandel test is limited because it only allows explaining the lack of linearity in the case of quadratic models. In this test, two residual errors are calculated, one for the OLS model and another for a quadratic regression function. An F-test is then applied to decide whether the quadratic regression is a better mathematical model than the linear regression using the null hypothesis that the two models are equivalent. The LOF test was performed during discussion sessions for all calibration curves and p -values obtained ranged from 0.081 to 0.195 for HPLC curves, which confirmed the linearity of the chromatographic calibrations. In the case of AAS calibrations, p -values were always < 0.01 .

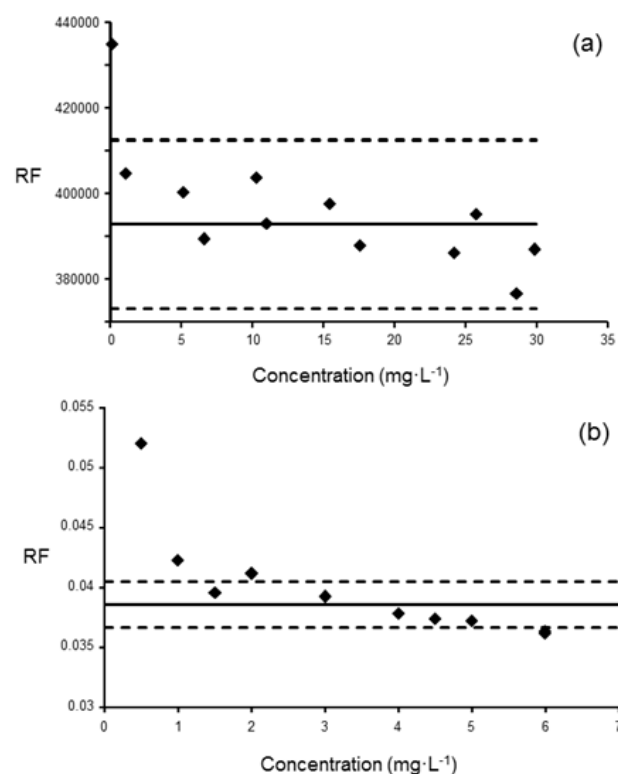


Figure 4. Response functions (RF) plots obtained for a chromatographic curve (a) and an atomic absorption curve (b). Continuous line indicates the mean RF value. Dashed lines are drawn for the $\pm 5\%$ cut-off values

3.3. Outliers

Calibration curves have to be inspected for possible outliers and points of influence [30]. A measurement is an outlier when this point is well separated from the other calibration points and it is due to a gross error when preparing the standard or performing the measurement. Sometimes, a suspected result is not an outlier because it is due to the variability in the measurements or it is due to the fact that the curve does not follow a linear trend in the range evaluated, which may happen with points at both extremes of the calibration range where curvature of instrument response is expected.

The common procedure followed by students for the elimination of discordant points in previous calibrations was by inspecting the scatter plot and removing those points that did not seem to follow a linear tendency, until a “good” R^2 value was obtained. Unfortunately, this wrong procedure is commonly applied in many laboratories.

Table 1. Back-calculated concentrations for the standards used in an atomic absorption spectrometry calibration method using the ordinary least squares and quadratic fitting models. %RE correspond to the percentage of error of the back-calculated value with respect to the concentration of the standard ($\%RE = 100 \frac{(BC-C)}{C}$)

Standard conc. (C) (mg·L ⁻¹)	Back-calculated (BC)		% RE	
	Linear	Quadratic	Linear	Quadratic
0.1	-0.01	0.11	-110 %	+10 %
0.3	0.06	0.25	-80 %	-17 %
0.5	0.54	0.52	+8 %	+4 %
0.8	0.84	0.79	+5 %	+1 %
1	1.02	1.02	+2 %	+2 %
1.3	1.29	1.32	-1 %	+2 %
1.5	1.51	1.47	+1%	-2 %
1.8	1.75	1.83	-3 %	+2 %

It is important to teach students that the decision of removing an experimental result should not be only based in the visual inspection of the data and checking simple parameters such as R^2 . Some statistical calculations and experimental confirmation with validated parameters or independent standard controls is required. When working in a laboratory with experimental results, it is needed to check if the suspect result is due to an instrumental error during the measurement or to some mistake during the preparation of the specific standard. The option of an instrumental error can be assessed by measuring another time the same standard. If the new measurement gives the same result, it is required to prepare a new standard at the same concentration to check if the mistake was in the preparation of the previous standard. This procedure allows solving the problem with outliers inside the linear range, but does not solve the problem with influence points in the extremes of the calibration curve. The main limitation is that it is time consuming and in some procedures, such as in many immunoassay methods, cannot be applied because all calibrators and samples are measured at the same time in multiple well plates.

Some statistical options have been proposed to assess possible outliers without the need to repeat experiments. The best option is having replicates for each standard. In this situation, a statistical test is applied to determine if one of the replicates of a standard is an outlier and can be omitted before applying the regression model. However, this is not the most common situation in routine calibrations. The most general situation is having a single value for each calibration standard. In this situation, a common solution is to perform an outlier test using the residuals to check whether there is a calibration point presenting an excessive residual value. A best option is to obtain the standardized residuals and check if one value is above a set value of ± 2 (because for normal distributions, 95% of the results are inside $\pm 2\sigma$).

In the present study, outliers were only evaluated for the chromatographic curves because AAS curves were not

linear. It was asked students to perform a visual inspection of their calibration curves and to indicate possible outliers. They suggested the presence of outliers in 11 curves because the removal of the suggested value yielded better R^2 values. During the seminar session, the standardized residuals were calculated for the calibrations with proposed outliers but the presence of outliers was only confirmed in one calibration.

3.4. Treatment of the Origin

When students were asked about the use of the zero-point calibration (0,0) for the regression fitting, around 50% answered that they had always introduced this point in their regression calculations because they had set the zero response of the instrument with a blank. Unfortunately, this can be a source of bias for two main reasons. First, analytical instruments have a background signal or noise, which is expected to be non-zero. Second, zeroing the instrument signal before doing an analysis is not the correct way to obtain the real value of the blank; it is required to run a replicates of a true blank with the method and this usually results in a signal value different from zero due to the random noise signal. Another problem is that absolute zero concentration can never be measured in chemical and biological analyses. If we prepare a calibration curve with a large number of points between blank and LOQ levels, we will see that there is always a curvature at low levels because there is a minimum concentration (limit of detection) from which the signal obtained does not differ from the noise of the instrument, and this value is always >0 .

I always suggest my students that expecting that the true relationship between two variables has to pass through the origin is not enough to force the estimated relationship through the origin [7]. A calibration function cannot be considered with intercept equal to zero unless it is demonstrated that b_0 is not significantly different from zero [7,28,31]. The significance of intercept values can be determined by different ways [7,25]. First, by applying the Students' t-test considering the null hypothesis $b_0=0$. Second, by calculating the confidence interval of the intercept and checking if this span zero. These two tests can be easily done with Excel for linear calibration models.

Table 2 shows the results obtained in one of the HPLC calibrations evaluated. In this case, the OLS fitting showed that b_0 can be considered zero as the t-test gave $p=0.273$, which means that the null hypothesis ($b_0=0$) is true. This result together with the fact that the HPLC methods evaluated present heteroscedasticity permitted to demonstrate students how the decision whether to use b_0 in their calculations has a significant effect on those determinations at low concentrations, near the LOQ. The calculation of the percent relative error of back-calculated concentrations (%RE) allows obtaining a measure of the error obtained when using the proposed regression equation, and, as indicated in a previous section, a $\pm 20\%$ cut-off value near the LOQ is usually set [7,18,28]. In the example shown in Table 2, the first standard was prepared at a concentration near the LOQ. The results obtained indicate that when b_0 is applied to back-calculate the concentration, bias obtained for the first standard was excessive (106%). On the other hand, the removal of b_0 in

back-calculations allowed obtaining good percentages of error at this level (6%). For those standards prepared well above the LOQ, the use of the intercept does not interfere significantly in the percentage of relative back-calculated error.

Table 2. Calibration results for one of the evaluated HPLC calibrations. Regression coefficients obtained by OLS: $b_0=79636$ ($sd_{b_0}=62666$); $b_1=399478$ ($sd_{b_1}=3602$); $R^2=0.9997$

x_i ($\text{mg}\cdot\text{L}^{-1}$)	y_i (peak area)	BC ^(a) ($y=b_0+b_1x$)	%RE ^(b)	BC ^(a) ($y=b_1x$)	%RE ^(b)
0.2	74810	-0.01	106	0.19	6
5.0	2661577	4.95	1	5.16	-3
10.0	4159177	10.21	-2	10.41	-4
15.0	6143993	15.18	-1	15.38	-3
25.0	10114050	25.12	-1	25.32	-1
29.0	11560281	28.74	1	28.94	1

^(a) BC: back-calculated concentration

^(b) %RE is the percent relative error of the back-calculated concentration.

The removal of the intercept, when it does not differ statistically from zero, in the calculations for linear calibrations with heteroscedastic data helps to minimize the loss of precision of heteroscedastic data near the LOQ and improves the accuracy of the results obtained at this level. However, significant bias can still be obtained. The same undiluted and diluted aliquots evaluated in the heteroscedastic section were also measured applying OLS with $y=b_1x$. The t-test for paired data yielded $p=0.027$, which confirms that there is still a significant bias for estimates near the LOQ, and WLS is required to obtain highly accurate results.

3.5. Confirmation of the Accuracy of Estimates

Once the best fitting model is found, a calibration curve is built to associate a response to a concentration. However, for quantitative calculations the accuracy of the result obtained must be confirmed. This requires the use of quality standard controls to confirm the response obtained through the calibration curve with a known concentration [1]. The analysis of some independent standard controls is needed (i) to check that the instrument response has not changed and regression coefficients are maintained inside the determined confidence intervals, and (ii) to confirm that no gross errors have been made in the preparation of stock solutions.

Linearity evaluation of the chromatographic curves indicated that 74 curves were linear. However, different results were obtained by different groups of students for replicates of the same sample, as observed during the presentation of their results in the discussion seminars. As an example, in one of the seminars seven linear calibration curves were confirmed after evaluating the linearity. Replicates of the same soft drink sample, which was used as quality control sample, were evaluated with each calibration. No differences were observed for the instrumental responses obtained for the replicates (RSD <5%). However, concentrations estimated with each calibration were not equal. One group reported a concentration 15% higher than the values reported by the other groups. The

revision of the different calibration functions showed that the sensitivity of the calibration function for the suspecting group ($b_1=341,070$, $sd_{b_1}=4,093$, $R^2=0.9995$) was smaller than those obtained by the other six groups (mean $b_1=380,731$, confidence interval= $\pm 5,150$). The application of the Grubbs' test to check if the suspected sensitivity value was an outlier gave a G-value of 2.164 (higher than the tabulated G-value of 2.097 for $n=7$ and 99% significance). A revision of the laboratory notebooks permitted to find that the suspected group made a mistake in the preparation of the stock solution, during the dissolution of the solid reagent. In the following steps, they did not make errors during the dilutions of the stock solution to prepare the calibration standards, which yielded a systematic proportional error in the reported values of the concentration for the standards. In these conditions, they obtained a linear calibration with a biased sensitivity.

Similar problems were found in 21 chromatographic calibrations (28.4%) along the period evaluated. This situation helped students to understand that (i) they cannot base their results only on a good linear calibration, and (ii) the importance of the analysis of standard controls.

4. Conclusions

As any other statistical function, calibration functions are mathematical models that only requires of some pairs of data to provide us some results. However, results obtained can only be accurate and precise when a correct selection of the experimental conditions and a good evaluation of the model are applied. During their formation, students must learn about all the practical considerations that may have a significant effect on their laboratory results. Moreover, they must be able to discern if their results are accurate and precise.

This study has been focused to sow students the practical aspects that should be considered when using calibrations in the laboratory, and how to assess the validity of the fitting model used. Students applied OLS to two analytical methods and found that, despite its extensive and indiscriminate use, OLS is not the best fitting method for many analytical calibrations.

The results obtained for the different linearity test showed that all AAS calibrations gave valid curves, with no outliers or gross errors found despite students considered many points as incorrect. This was due to the fact that AAS calibrations were not linear. Therefore, their "proposed" outliers were not biased results. For these calibrations, a quadratic function provided better fitting and most accurate estimates.

The evaluation of the HPLC methods allowed students to understand that in calibrations with linear response it is common to obtain heteroscedastic data, which results in OLS giving biased results for estimates at low levels, near the quantification limits.

Finally, the importance of checking the results with quality controls was also demonstrated. At the end of the sessions, students had a different consideration about calibration and understood that results obtained with conventional OLS can be many times biased and imprecise.

References

- [1] Zabell, A.P.R., Lytle, F.E. and Julian, R.K., "A proposal to improve calibration and outlier detection in high-throughput mass spectrometry", *Clinical Mass Spectrometry*, 2, 25-33, 2016.
- [2] de Souza, S.V.C. and Junqueira, R.G., "A procedure to assess linearity by ordinary least squares method", *Analytica Chimica Acta*, 552, 25-35, 2005.
- [3] Rozet, E., Ceccato, A., Hubert, C., Ziemons, E., Oprean, R., Rudaz, S., Boulanger, B. and Hubert, P., "Analysis of recent pharmaceutical regulatory documents on analytical method validation", *Journal of Chromatography A*, 1158, 111-125, 2007.
- [4] Welz, B., Sperling, M., *Atomic Absorption Spectrometry*, 3rd ed., Wiley-VCH, Weinheim, 2008.
- [5] Wild, D., *The immunoassay handbook: Theory and applications of ligand binding, ELISA and related techniques*, 4th ed, Elsevier, Oxford, 2013.
- [6] Van der Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K. and van der Werf, M.J., "Centering, scaling, and transformations: improving the biological information content of metabolomics data", *BMC Genomics*, 7, 142, 2006.
- [7] Raposo, F., "Evaluation of analytical calibration based on least-square linear regression for instrumental techniques: A tutorial review", *TRAC Trends in Analytical Chemistry*, 77, 67-185, 2016.
- [8] Kiser, M.M. and Dolan, J.W., "Selecting the best curve fit", *LC-GC North America*, 22, 112-117, 2004.
- [9] Johnson, E.L., Reynolds, D.L., Wright, D.S. and Pachla, L.A., "Biological sample preparation and data reduction concepts in pharmaceutical analysis", *Journal of Chromatographic Sciences*, 26, 72-379, 1988.
- [10] Almeida, A.M., Castel-Branco, M.M. and Falcao, A.C., "Linear regression for calibration lines revisited: weighting schemes for bioanalytical methods", *Journal of Chromatography B*, 774, 215-222, 2002.
- [11] Tellinghuisen, J., "Weighted least-square in calibration: What difference does it make?", *Analyst*, 132, 536-543, 2007.
- [12] Zeng, Q.C., Zhang, E., Dong, H. and Tellinghuisen, J., "Weighted least squares in calibration: Estimating data variance functions in high-performance liquid chromatography", *Journal of Chromatography A*, 1206, 147-152, 2008.
- [13] Tellinghuisen, J., "Least squares in calibration: weights, nonlinearity, and other nuisances", in: M.L. Johnson, L. Brand, eds., *Methods in enzymology*, Vol 454. Academic Press, San Diego, 259-285, 2009.
- [14] Gu, H., Liu, G., Wang, J., Aubry, A.F. and Arnold, M.E., "Selecting the correct weighting factors for linear and quadratic calibration curves with least-square regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay performance", *Analytical Chemistry*, 86, 8959-8966, 2014.
- [15] Marques-Marinho, F.D., Reis, I.A. and Vianna-Soares, C.D., "Construction of analytical curve fit models for Simvastatin using ordinary and weighted least squares methods", *Journal of the Brazilian Chemical Society*, 24, 1469-1477, 2013.
- [16] Mulholland, M. and Hibbert, D.B., "Linearity and the limitations of least squares calibration", *Journal of Chromatography A*, 762, pp 73-82, 1997.
- [17] Thompson, M. and Lowthian, P.J., *Notes on statistics and data quality for analytical chemists*, Imperial College Press, London, 2011.
- [18] Analytical Methods Committee, "Is my calibration linear?", *Analyst*, 119, 2363-2366, 1994.
- [19] Jurado J.M., Alcázar, A., Muñoz-Valencia, R.; Ceballos-Magaña, S.G. and Raposo, F., "Some practical aspects for linearity assessment of calibration curves as function of concentration levels according to the fitness-for-purpose approach", *Talanta*, 172, 221-229, 2017.
- [20] De Beer, J.O., De Beer, T.R. and Goeyens, L., "Assessment of quality performance for straight line calibration curves related to the spread of the abscissa values around their mean", *Analytica Chimica Acta*, 584, 57-65, 2007.
- [21] Van Looc, J., Elskens, M., Croux, C. and Beernaert, H., "Linearity of calibration curves: use and misuses of the correlation coefficient", *Accreditation and Quality Assurance*, 7, 281-285, 2002.
- [22] Reichenbacher, M. and Einax, J.W., *Challenges in Analytical Chemistry Assurance*, Springer, Heidelberg, 2011.
- [23] Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., De Jong, S., Lewi, P.J. and Smeyers-Verbeke, J., *Handbook of chemometrics and qualimetrics: Part A*, Elsevier, Amsterdam, 1997.
- [24] Van Arendonk, M.D. and Skogerboe, R.K., "Correlation coefficients for evaluation of analytical calibration curves", *Analytical Chemistry*, 53, 2349-2350, 1981.
- [25] Analytical Methods Committee, "Uses (proper and improper) of correlation coefficients", *Analyst*, 113, 1469-1471, 1988.
- [26] Ellison, S.L.R., Barwick, V.J. and Duguid Farrant, T.J., *Practical statistics for the Analytical Scientist: A bench guide*. Royal Society of Chemistry, Cambridge, 2009.
- [27] Bysouth, S.R. and Tyson, J.F. "A comparison of curve fitting algorithms for flame atomic absorption spectrometry", *Journal of Analytical Atomic Spectrometry*, 1, 85-87, 1986.
- [28] Ettre, L.S., "Nomenclature for chromatography", *Pure and Applied Chemistry*, 65, 819-872, 1993.
- [29] US-EPA, *SW-846 Test Method 8000C: Determinative chromatographic separations*, Revision3, Section 11.5.1, 2003.
- [30] Araujo, P., "Key aspects of analytical method validation and linearity evaluation", *Journal of Chromatography B*, 877, 2224-2234, 2009.
- [31] LCG/VAM, "Preparation of calibration curves: A guide to best practice", 2003. Available: <http://www.lcggroup.com/our-science/national-measurement-laboratory/publications-and-resources/good-practice-guides/preparation-of-calibration-curves-a-guide-to-best/> [Accessed Nov 16, 2017].
- [32] Dolan, J.W., "Calibration curves, Part I: To b or not to b?", *LC-GC North America*, 27, 224-230, 2009.