World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Quantifying human error in visual endpoint detection during acid-base titration: An instructional study using independent student datasets

Hassan Darwish [1,*], Giovanna Vicentini [1], Johzeff Anderson [1], Diego Moran [1], Una Gibbons [2], Hiba Al Lawati [3] and Abdulhamid Alharthy [4]

[1] *Biology and Physical Sciences Department, School of Arts, Sciences, and Education, Ivy Tech Community College, South Bend, Indiana, USA.*
[2] *Royal Hospital, Oman (Former), Kilkenny, Ireland.*
[3] *School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Australia.*
[4] *Biomedical Sciences, School of Sport and Health Sciences, Cardiff Metropolitan University, UK*

World Journal of Advanced Research and Reviews, 2026, 30(01), 2155-2161

Publication history: Received on 13 March 2026; revised on 21 April 2026; accepted on 23 April 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.1.1089

Abstract

Visual indicators remain common in introductory analytical chemistry laboratories, but they also introduce subjectivity during endpoint recognition. In this study, human-observed endpoint variability was quantified by merging two independent student-generated triplicate datasets for the standardization of sodium hydroxide against 0.1000 M hydrochloric acid. Across six titrations, calculated NaOH molarity values ranged from 0.0996 to 0.1081 M, with a combined mean of 0.1041 M, standard deviation of 0.0035 M, relative standard deviation of 3.36%, and a 95% confidence interval of 0.1004-0.1078 M. Dataset-specific means were 0.1047 M and 0.1035 M, and no statistically significant difference was detected between the two small instructional datasets (Welch t-test, p = 0.732). The results demonstrate that visually judged endpoints can produce measurable trial-to-trial variability even when the same stoichiometric framework is followed. This manuscript therefore offers a classroom-ready model for teaching precision, bias, uncertainty, observer effects, and the analytical advantages of more objective endpoint detection.

Keywords: Acid-Base Titration; Endpoint Detection; Analytical Precision; Uncertainty; Phenolphthalein; Chemical Education; Visual Titration

1. Introduction

Acid-base titration is one of the most widely taught quantitative methods in chemistry because it is inexpensive, conceptually direct, and strongly connected to stoichiometry, equilibrium, and solution chemistry. [1-4] In teaching laboratories, it is often among the first experiments in which students convert experimental measurements into a calculated concentration. That role makes titration especially valuable not only for reinforcing chemical concepts but also for introducing broader ideas of analytical quality, precision, and uncertainty.[3,5]

Despite its pedagogical strength, indicator-based titration is inherently subjective. Visual endpoint detection depends on the analyst deciding when a transient color change has become sufficiently persistent to count as the endpoint. Small differences in color perception, room lighting, swirling rate, dropwise addition, waiting time, or tolerance for a faint pink color can shift the apparent endpoint away from the true equivalence point.[3,5-7] For experienced analysts, those differences may be modest; for novice students, however, they can become a major source of both random and systematic error.

* Corresponding author: Hassan Darwish

Modern analytical chemistry reduces this subjectivity through instrumental endpoint detection. Potentiometric titration can identify inflection behavior without relying on the observer's eye, while digital colorimetry and image-based approaches can reduce person-to-person variability in visually challenging samples.[8-10] Even so, visual acid-base titration remains deeply embedded in introductory laboratory curricula because it requires minimal instrumentation and provides a direct, intuitive experience of chemical change.

That continuing use creates an educational opportunity. Rather than treating visual endpoint judgment as a weakness to be ignored, instructors can use it as a practical demonstration of why repeated trials, descriptive statistics, confidence intervals, and careful technique matter.[11-14] Real student-generated variability is often more instructive than idealized textbook examples because it reveals that apparently simple experiments still contain interpretable measurement uncertainty.

The aim of the present study was therefore to quantify human-observed endpoint variability during a standard acid-base titration experiment by combining two independent student-generated triplicate datasets. The specific objectives were to (i) compare the molarity values derived from each dataset, (ii) evaluate combined precision statistics across all six trials, (iii) interpret the observed spread considering human endpoint judgment, and (iv) frame the exercise as an instructional model appropriate for Journal of Chemical Education readership.

2. Experimental section

Two independent student-generated triplicate datasets were incorporated into the present analysis. In both datasets, sodium hydroxide solution was standardized against 0.1000 M hydrochloric acid with phenolphthalein as the endpoint indicator. The neutralization reaction followed the expected 1:1 stoichiometric relationship:

$HCl + NaOH \rightarrow NaCl + H_2O$

In Dataset 1, 10.00 mL aliquots of HCl were titrated using NaOH volumes of 10.04, 9.25, and 9.39 mL. The accompanying student notes indicated that 30 mL deionized water was added to the Erlenmeyer flask prior to titration. In Dataset 2, the buret was first rinsed with deionized water and conditioned with NaOH solution. A 10.00 mL volumetric pipet was used to transfer HCl to an Erlenmeyer flask, followed by addition of 50 mL deionized water and three drops of phenolphthalein. Titration proceeded with continuous swirling until a faint pale pink color persisted for at least 30 s. The recorded NaOH volumes in Dataset 2 were 10.00, 9.50, and 9.50 mL.

All titrations were performed with standard teaching-laboratory volumetric glassware, including a 50 mL buret read to the nearest 0.01 mL and a 10.00 mL volumetric pipet. Although the two datasets differed slightly in procedural detail, both followed the same analytical principle and endpoint chemistry. The difference in dilution volume (30 mL versus 50 mL deionized water) does not alter stoichiometric equivalence, but it can influence the visual appearance and intensity of the endpoint, which is relevant to the current study because the focus is human recognition rather than reaction stoichiometry alone.

NaOH molarity was calculated using the relation M_HCl V_HCl = M_NaOH V_NaOH. Descriptive statistics included the mean, standard deviation (SD), relative standard deviation (%RSD), standard error, and 95% confidence interval (CI) of the mean. A Welch two-sample t-test was used as an exploratory comparison between datasets because the variances were not assumed to be identical and the sample sizes were small (n = 3 per dataset). Statistical analysis is not intended as a formal validation study; rather, it is meant to quantify the spread visible in authentic classroom data.

3. Results and discussion

The individual and derived titration results are summarized in Table 1, and the data set-level summary statistics are reported in Table 2. Across all six titrations, the calculated NaOH molarity values ranged from 0.0996 to 0.1081 M. The combined mean was 0.1041 M with an SD of 0.0035 M and a %RSD of 3.36%. The 95% CI of the combined mean extended from 0.1004 to 0.1078 M. These results show that a seemingly routine visual titration can produce measurable spread even when the underlying stoichiometric calculation is straightforward.

Dataset 1 produced a mean of 0.1047 M and a %RSD of 4.30%, whereas Dataset 2 produced a mean of 0.1035 M and a %RSD of 2.96%. The higher spread in Dataset 1 is visually evident in Figure 2 and Figure 4. One likely explanation is procedural control. Dataset 2 explicitly used an endpoint criterion requiring a pale pink color to persist for at least 30

s, whereas Dataset 1 was reported more succinctly from raw observations. This does not prove causation, but it is consistent with the general analytical principle that well-defined endpoint criteria can improve reproducibility. [3,5]

The combined mean also shows a modest positive bias relative to the nominal 0.1000 M expectation. The mean value was higher by approximately 4.12%, which is consistent with slight endpoint overshoot. In indicator-based strong acid-strong base titration, phenolphthalein is practical and widely used in instruction, but visually judged endpoints can drift beyond the exact equivalence point if the observer delays acceptance of the color change or adds one drop too many.[1,3,6] This distinction between accuracy and precision is pedagogically important: students may obtain an average value that appears close to the expected result while still displaying substantial trial-to-trial variability.

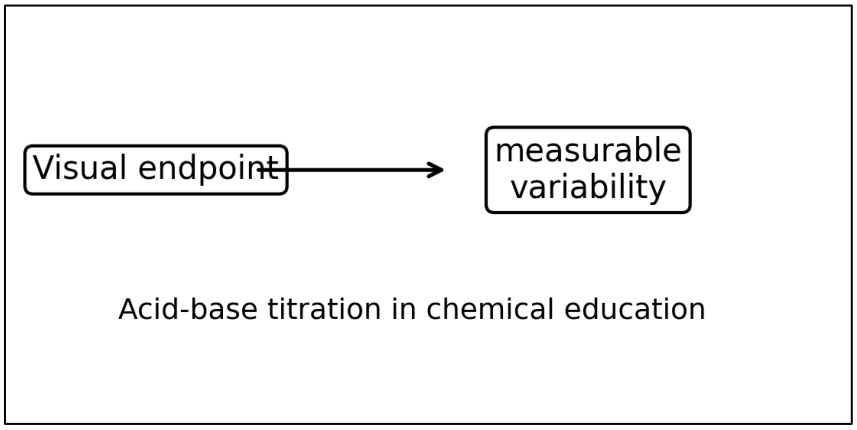
The graphical representations strengthen this interpretation. Figure 2 shows all six calculated NaOH molarity values and the dashed combined mean; the spread around the mean is immediately visible. Figure 3 presents the distribution as a histogram and shows that the values cluster near the mean but are broad enough to indicate real observer-dependent variation rather than exact repeatability. Figure 4 compares the two datasets with a boxplot and confirms substantial overlap between them, even though Dataset 1 is somewhat more dispersed. The exploratory Welch t-test did not detect a statistically significant difference between datasets ($t = 0.37$, $p = 0.732$), which is expected to be given the small sample size and should not be overinterpreted. Instead, the more useful conclusion is that both student groups produced values in the same general range while still showing measurable uncertainty.

From an educational perspective, this experiment is especially useful because it turns ordinary student records into a platform for discussing data quality. The calculation is simple enough for introductory students, yet the results naturally motivate conversation about repeatability, observer effects, bias, confidence intervals, and why analysts perform replicate measurements.[11-16] The exercise also creates a natural bridge to instrumental analysis. Instructors can ask students how the results might differ if the same titration were performed potentiometrically or with image-assisted color determination. That comparison helps students understand not only how instruments function, but why they are valuable.

This study has limitations. The sample size is small, the datasets were obtained in an instructional rather than method-validation environment, and the procedures were not fully identical. Consequently, the present work should not be interpreted as a definitive analytical comparison of endpoint methods. Its value lies instead in demonstrating that authentic classroom data can be analyzed rigorously enough to reveal meaningful measurement behavior. Additional replication, direct comparison with potentiometric detection, or explicit propagation of volumetric uncertainty would strengthen the work further and could form the basis of a follow-up teaching study. [8-10,17-20]

This type of dataset-driven analysis aligns strongly with current calls in chemical education to integrate authentic data interpretation into laboratory instruction, thereby helping students connect experimental practice with statistical reasoning and real-world analytical variability. These findings further support the use of authentic student-generated datasets as effective tools for reinforcing quantitative reasoning and analytical thinking in introductory chemistry laboratories.

3.1. TOC graphic



Visual endpoint judgment in an introductory acid-base titration produces measurable variability in calculated NaOH molarity values.

Table 1 Raw and Derived Titration Results from Two Independent Student Datasets

Dataset	Trial	Initial (mL)	Final (mL)	Used (mL)	Molarity (M)	Notes
1	1	47.95	37.91	10.04	0.0996	Raw data; 30 mL DI
1	2	36.93	27.68	9.25	0.1081	Raw data; 30 mL DI
1	3	18.30	8.91	9.39	0.1064	Raw data; 30 mL DI
2	1	50.00	40.00	10.00	0.1000	Calculated; 50 mL DI
2	2	50.00	40.50	9.50	0.1053	30 s pale-pink endpoint
2	3	50.00	40.50	9.50	0.1053	30 s pale-pink endpoint

Table 2 Summary Statistics for Each Dataset and for the Combined Six-Trial Analysis

Dataset	n	Mean (M)	SD (M)	%RSD	95% CI (M)	Bias
1	3	0.1047	0.0045	4.30	0.0935-0.1159	+4.70%
2	3	0.1035	0.0031	2.96	0.0959-0.1111	+3.53%
Combined	6	0.1041	0.0035	3.36	0.1004-0.1078	+4.12%

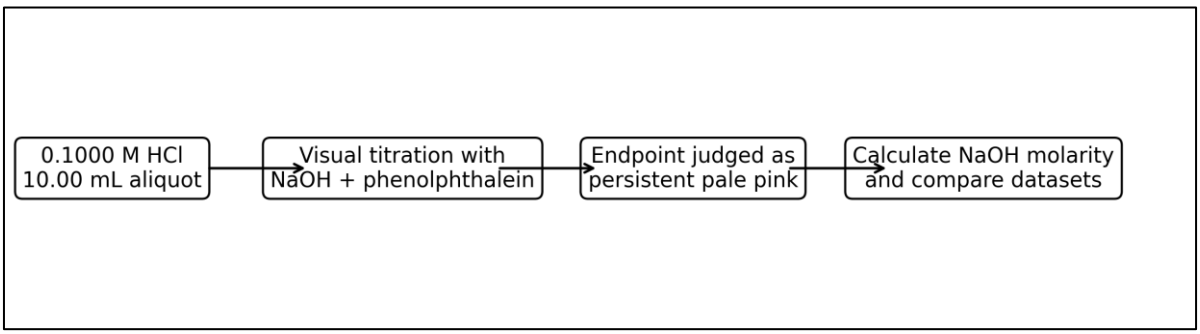


Figure 1 Instructional workflow used across the two student titration datasets. A known HCl aliquot was titrated visually with NaOH to a persistent pale pink phenolphthalein endpoint, and the resulting volumes were converted to NaOH molarity values for statistical comparison.

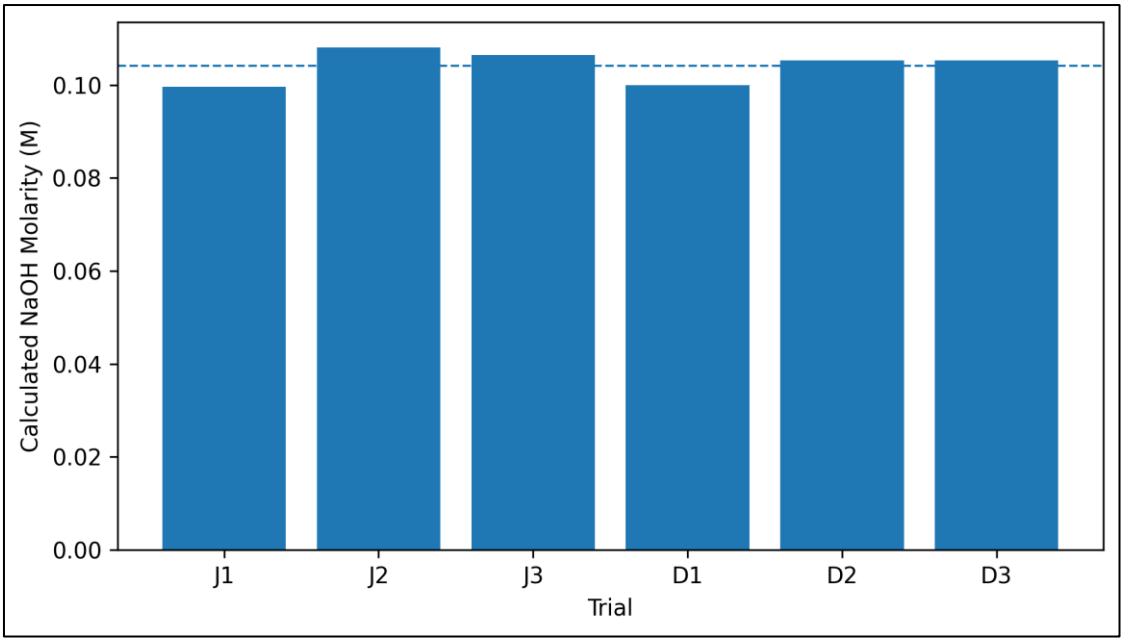


Figure 2 Calculated NaOH molarity values for all six trials. J1-J3 correspond to Dataset 1 and D1-D3 correspond to Dataset 2. The dashed horizontal line indicates the combined mean.

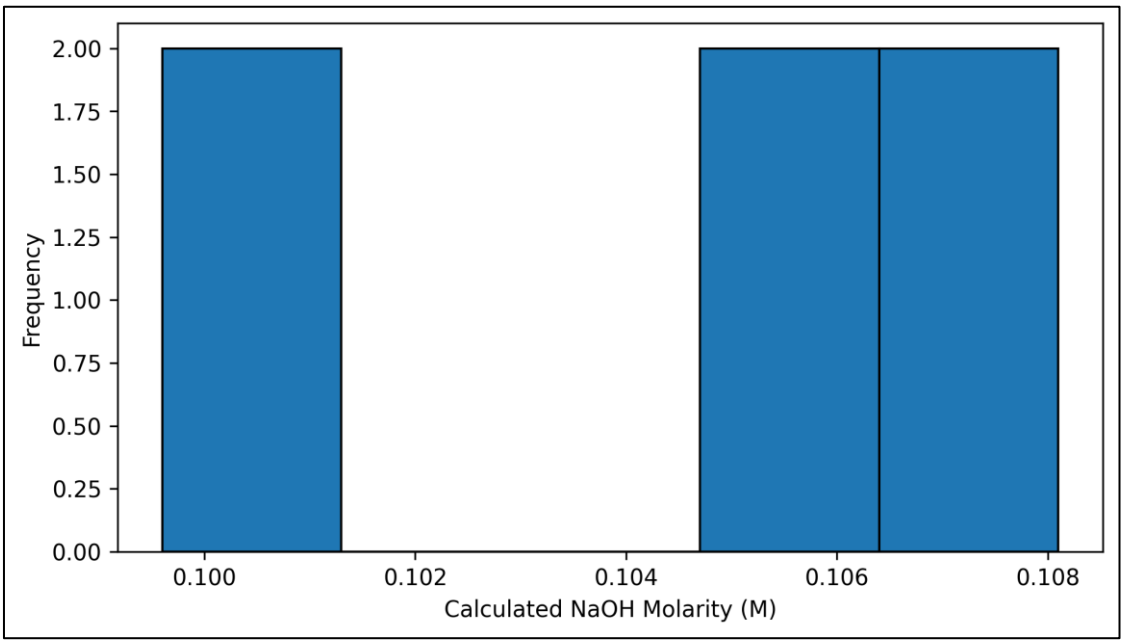


Figure 3 Histogram of calculated NaOH molarity values across both student datasets. The distribution clusters near the combined mean but remains broad enough to reveal observer-dependent spread within this small instructional dataset.

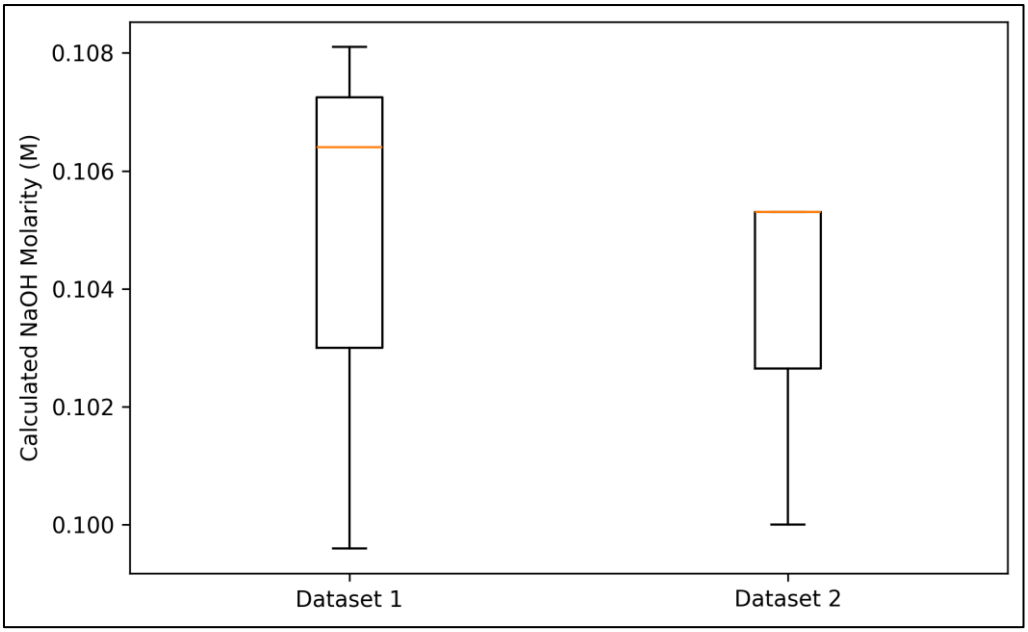


Figure 4 Boxplot comparison of the two student datasets. The distributions overlap substantially, although Dataset 1 shows somewhat greater spread than Dataset 2.

4. Conclusions

By combining two independent student-generated triplicate datasets, this study showed that visually determined endpoints in a standard acid-base titration produced a combined NaOH molarity of 0.1041 ± 0.0035 M (SD) with a relative standard deviation of 3.36%. The merged analysis demonstrates that human endpoint recognition is a measurable contributor to analytical variability. More importantly for instruction, the experiment provides a clear and accessible model for teaching precision, bias, confidence intervals, and motivation for instrumental endpoint detection. In that sense, the value of the exercise extends beyond the concentration result itself: it helps students see that chemistry measurements are not merely calculations, but observations shaped by technique, judgment, and uncertainty.

Compliance with ethical standards

Acknowledgments

The authors acknowledge the students whose laboratory records formed the basis of this instructional analysis and the teaching-laboratory environment that supported the experiment.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

Associated content

Supporting information may include the original student data sheets, the calculation workflow used to obtain NaOH molarity values from buret readings, and the spreadsheet or software output used for statistical analysis.

References

[1] Harris, D. C. Quantitative Chemical Analysis, 9th ed.; W. H. Freeman: New York, 2016.

[2] Skoog, D. A.; West, D. M.; Holler, F. J.; Crouch, S. R. Fundamentals of Analytical Chemistry, 9th ed.; Cengage Learning: Boston, 2014.

[3] Miller, J. N.; Miller, J. C. Statistics and Chemometrics for Analytical Chemistry, 8th ed.; Pearson: Harlow, 2021.

[4] Harvey, D. Modern Analytical Chemistry; McGraw-Hill: Boston, 2000.

[5] Taylor, J. R. An Introduction to Error Analysis, 2nd ed.; University Science Books: Sausalito, CA, 1997.

[6] Christian, G. D.; Dasgupta, P. K.; Schug, K. A. Analytical Chemistry, 7th ed.; Wiley: Hoboken, NJ, 2014.

[7] Vogel, A. I.; Mendham, J.; Denney, R. C.; Barnes, J. D.; Thomas, M. J. K. Vogel's Textbook of Quantitative Chemical Analysis, 6th ed.; Prentice Hall: Harlow, 2000.

[8] Vilaso-Cadre, J. E.; Benitez-Fernandez, D.; Lopez-Alvarez, I. A.; Tovar-Vazquez, F. Y.; Arada-Perez, M. A.; Reyes-Dominguez, I. A. Acid-Base Potentiometric Titration Using a Stainless Steel Electrode without Oxidative Treatment. Turk. J. Chem. 2023, 47 (4), 801-813.

[9] Berasarte, I.; Bordagaray, A.; Garcia-Arrona, R.; Ostra, M.; Reis de Araujo, W.; Vidal, M. Microscale Titration of Acetic Acid Using Digital Colorimetry and Paper-Based Analytical Devices. Talanta 2024, 276, 126254.

[10] Rubinson, K. A.; Rubinson, J. F. Contemporary Instrumental Analysis, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 2000.

[11] Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; Wenderoth, M. P. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. Proc. Natl. Acad. Sci. U.S.A. 2014, 111 (23), 8410-8415.

[12] Prince, M. Does Active Learning Work? A Review of the Research. J. Eng. Educ. 2004, 93 (3), 223-231.

[13] Bretz, S. L. Evidence for the Importance of Laboratory Courses. J. Chem. Educ. 2019, 96 (2), 193-195.

[14] Galloway, K. R.; Bretz, S. L. Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. J. Chem. Educ. 2015, 92 (7), 1149-1158.

[15] Pienta, N. J. Evaluating Laboratory Learning. J. Chem. Educ. 2003, 80 (7), 754-755.

[16] Ramette, R. W. Chemical Equilibrium and Analysis: A Practical Laboratory Approach. J. Chem. Educ. 1987, 64 (10), 885-887.

[17] Harris, H. H.; Fitzgerald, N.; Jones, S.; McLachlan, G. A Simple Automated Titration Experiment for the Introductory Analytical Laboratory. J. Chem. Educ. 2013, 90 (11), 1539-1542.

[18] Burke, K. A.; Greenbowe, T. J.; Windschitl, M. A. Developing and Using Conceptual Multiple-Choice Questions To Evaluate Student Learning in Analytical Chemistry. J. Chem. Educ. 1998, 75 (12), 1658-1661.

[19] Batzli, J. M.; Patrick, D. A. Measuring Student Learning in the Laboratory. J. Coll. Sci. Teach. 2010, 39 (6), 72-79.

[20] Domin, D. S. A Review of Laboratory Instruction Styles. J. Chem. Educ. 1999, 76 (4), 543-547.