A Modern Twist on an Old Measurement: Using Laboratory Automation and Data Science to Determine the Solubility Product of Lead Iodide

Alexander Norquist[1*], Gabe Jones-Thompson[1], Keqing He[1], Thomas J. Egg[2], Joshua Schrier[2*]

[1] Department of Chemistry, Haverford College, Haverford PA, 19041, USA
[2] Department of Chemistry, Fordham University, The Bronx, NY 10458, USA

* Email: anorquis@haverford.edu
* Email: jschrier@fordham.edu

Abstract

Laboratory automation and data science are valuable new skills for all chemists, but most pedagogical activities involving automation to date have focused on upper-level coursework. Herein, we describe a combined computational and experimental laboratory suitable for a first-year undergraduate general chemistry course, in which these topics are introduced in the context of determination of the solubility equilibrium constant of lead iodide. Students analyze their data using a logistic regression analysis, which has a physical interpretation in terms of the solubility equilibrium expression and its stoichiometric coefficients. In addition to laboratory automation, data visualization, and data fitting skills, students also practice core laboratory skills such as the preparation of stock solutions using a volumetric flask and use of micropipettes. To keep the lab affordable, we demonstrate the use of a low-cost 3d-printed liquid dispensing robot to perform the automated experiment, in addition to a commercial liquid-handling robot. Example pre- and post-lab computational notebooks are provided in both Mathematica and Python programming languages.

Keywords: solubility product, equilibria, automation, randomized experimentation, logistic regression

Introduction

Data science and laboratory automation are rapidly becoming essential tools in modern experimental chemistry.[1,2] As such, it is important that we teach these skills throughout the chemistry curriculum, but efforts to date have primarily been in upper-level courses. Recent articles in this *Journal* have discussed training machine learning for spectroscopy-related problems in physical and analytical chemistry.[3–6] High-throughput experimentation has been discussed in the context of second-year organic chemistry and upper-level biochemistry courses.[7–10] The closely related idea of autonomous experimentation[1]—colloquially known as "self-driving labs"— has been discussed in the context instrumental analysis or advanced elective courses.[11–17]

How could data science and automation be incorporated in the first-year undergraduate laboratory experience? Aspects of experiment planning algorithms may be taught as instructions for humans,[18] but in general, some amount of computer programming is needed. High-level programming environments, such as Mathematica, enables first-year undergraduates to perform non-trivial tasks related to scientific computing and chemistry[19] and training neural networks to identify chemical glassware identification.[20] Our goal is to incorporate data analysis into the experimental laboratory experience. Additionally, we want to change student mindsets away from thinking about results comprised of a single experiment to thinking about dozens or even hundreds of parallel experiments. Automation is merely a means to that end, so this first student experience emphasizes *using* automation to solve a problem rather than the technical aspects of *building* devices.

In this paper, we describe a combined computational and experimental laboratory experience for first-year undergraduate students that introduces ideas of high-throughput experimentation and data analysis in the context of determining a solubility equilibrium constant. Equilibrium and equilibrium constants are foundational parts of introductory chemistry curriculum, so this synchronizes with existing

coursework. Solubility and solubility product equilibrium determination experiments are a common part of both introductory and advanced undergraduate pedagogical laboratory experiences. (The subtle distinction between these two ideas has been discussed previously;[21] our focus shall be on the latter.) Past articles in this *Journal* have discussed pedagogical experiments for the determination of the solubility product,[22,23] and its dependence upon ionic strength.[24] A variety of methods are used, including gravimetric measurements,[22,25,26] complexometric titration,[27] colorimetry,[28] potentiometry,[29] gamma-emission from ^{65}Zn,[30] and benchtop NMR.[31] In general, these labs have students perform a small number of experiments, and perform relatively simple numerical calculations to analyze their data. A notable exception is the upper-level quantitative analysis or physical chemistry lab described by McGarvey, in which students pool data and perform a nonlinear least squares regression to determine parameters in a Debye-Hückel model of the ionic strength effect.[27]

Our experiment takes a very different approach to determining the solubility equilibrium constant. Instead of slow, *quantitative* analyses of each reaction, a student need only quickly and *qualitatively* determine if a precipitation occurred or not. The experimental steps require manipulations commonly taught in first year laboratory programs (use of volumetric glassware, preparation of solutions with specific concentrations) plus the use of automatic pipettors. Automated experimentation allows each pair of students to generate a statistically large sample (as many as 96 experiments), enabling the use of a logistic regression model to infer the solubility product constant. In addition to emphasizing the underlying topic of equilibria, it engages students in thinking about experimental precision and sources of error. The experiment can be performed with a range of equipment from manual pipetting to a $800 home-made liquid dispenser to a commercial liquid handler, and thus can be adapted to the resources available at any institution. The experiment introduces elementary ideas of

computer programming and data analysis, without the need for prior student experience in these topics. This experiment is designed to follow a classroom discussion of equilibrium.

Theory

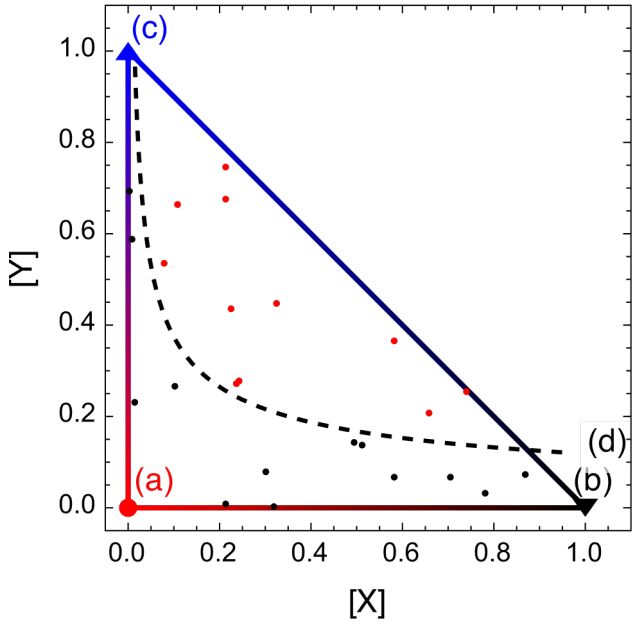


FIGURE 1: Illustration of concentration space, convex hull, random experiment, and equilibrium concepts. The stock solutions—denoted (a), (b), (c)—correspond to points in the concentration space plane. Mixtures between any pair of stock solutions define a solid line. The convex hull defined by the stock solutions—in this case, the triangle with vertices (a), (b), (c)—define the possible compositions that can be achieved by mixing the stock solutions. Random experiments are illustrated by the points. Experiments whose composition is above the dashed equilibrium line (d) are supersaturated and result in crystal formation (red); points below the equilibrium line are undersaturated and will not result in crystal formation (black). The equilibrium line (d) best separates the two types of observations.

This experiment is best described in terms of the concentration space formalism.[32,33] Figure 1 shows a schematic illustration of this approach. Each stock solution is represented as a vector, whose components denote the concentration of a species of interest. For the precipitation of PbI_2 from aqueous Pb^{2+} and I^- salts, we ignore the counter ions, so the relevant dimensions are $[Pb^{2+}]$ and $[I^-]$ concentrations. Pure solvent (in this case water) has a concentration of {0,0} corresponding to the origin. The mixture, $\boldsymbol{m}$, of any two stock solutions, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, is defined in terms of the *volume fraction*

of each of the solutions, v_1 and v_2, which must satisfy $0 \leq v_1 \leq 1$ and $0 \leq v_2 \leq 1$ and $v_1 + v_2 = 1$. For a binary mixture this simplifies to $v_2 = 1 - v_1$. The composition of the mixture is on a line between the two stock solutions, $\boldsymbol{m} = v_1\boldsymbol{s}_1 + v_2\boldsymbol{s_2}$. These conditions means that the mixture is a *convex combination* of the stock solutions.[34] More generally, the set of all possible mixtures that can be made from the set of stock solutions is comprised of all of the possible convex combinations of the available solutions, $\boldsymbol{m} = v_1\boldsymbol{s}_1 + v_2\boldsymbol{s_2} + \cdots v_n\boldsymbol{s}_n$, where the volume fractions each obey $0 \leq v_i \leq 1$ and collectively obey $\sum_i v_i = 1$. The possible mixtures are contained within the *convex hull* of the stock solution points; one way to visualize this is as region whose boundary is defined by stretching an elastic material over the stock solution points. This corresponds to the triangle (a-b-c) in Figure 1. Sampling possible compositions corresponds to generating random points within the convex hull.

Equilibrium expressions, such as K_{sp}=[Pb^{2+}][I^-]^2, can also be expressed concentration space;[32,33] this is especially easy to visualize in a two- or three-dimensional problem like the one considered here. Rearranging this into an expression for [I^-] lets us plot an *equilibrium curve* in concentration space (dashed line in Figure 1). Points above the equilibrium curve correspond to mixtures whose reaction quotient, *Q*, is greater than the equilibrium constant, and thus result in precipitation. Determining the equilibrium curve (and its underlying determinative parameter, the equilibrium constant) corresponds to finding a function that separates compositions observed to precipitate or not. This is the standard machine learning problem of supervised classification.

The equilibrium expression lends itself to treatment as classification problem using a logistic regression.[35] Starting with $K_{sp} = [A]^a[B]^b$, take the logarithm of both sides and subtract to yield $0 = a\ \ln[A] + b\ln[B] - \ln K_{sp}$. The terms on the right-hand side define the function to fit, *f([A], [B]*). The unknown parameters correspond to the stoichiometric constants, *a* and *b,* and the equilibrium constant, ln K_{sp} If the stoichiometric constants (*a* = 1, *b* = 2) are known, they need not be included as fitting

parameters. (Bishop described the use of a logarithmic transform to obtain graphical estimates of the stoichiometric coefficients in solubility product experiments.[36]) Reaction outcomes are described by an indicator variable, *y*, which takes a value of 1 if precipitation occurs and 0 if no precipitation occurs. One fits the logistic function $y = 1/(1 + e^{-f([A],[B])})$ to match the outcomes by varying *a, b,* and ln K_{sp}.to best separate the outcome classes. Standard libraries for performing logistic regression and obtaining these parameters exist for many programming languages.[37,38] As the parameters can all be multiplied by a constant value and obtain the same fit, it is necessary to normalize the values relative to one of the parameters—in this case, setting the stoichiometric coefficient of Pb^{2+} as the unit is a natural choice.

Learning Objectives

- Apply knowledge of solution mixing and equilibria in an experimental setting.
- Introduce laboratory automation (for synthesis and characterization) and learn where this is an appropriate strategy.
- Introduce ideas about error and precision in laboratory experimentation.
- Develop and practice basic computer and data handling skills.

Materials

Lead(II) nitrate ($Pb(NO_3)_2$, 99+ %) and potassium iodide (KI, 99 %) were purchased from Sigma-Aldrich and used without further purification, (approximately 0.330 g of $Pb(NO_3)_2$ and 0.166 g of KI are needed per group). Deionized water was used in these reactions. Equipment needs consist of two 100 mL volumetric flasks, two ~20 mL vials, three SBS microplates, and one or two 100 uL micropipettes per group. Two different robot systems were used for testing the experiment. The Sidekick

is an open-source, low-cost liquid dispenser;[39] the device used in the experiment was built by the two undergraduate coauthors (KH and GJ). A Hamilton NIMBUS 4 enclosed 4-channel liquid handler was also used. Alternative hardware options are discussed below. Student exercises were performed using Mathematica 13.2;[40] a free version of this software is available for non-commercial use on the Raspberry Pi computer.[41] New users of Mathematica at academic institutions are encouraged to consult their information and technology department as licensing is controlled at an institutional level. We have also provided versions of these student exercises as Python 3 Jupyter notebooks; this requires some additional installation of modules for interactivity described in the Instructor Guide.

Hazards and Disposal

Caution: *Lead compounds are poisonous; wear gloves and wash your hands thoroughly after use.* Lead is a well-known toxic substance whose primary toxicity mode is by ingestion.[42] Follow usual laboratory best-practices of safety eyewear, lab coats, gloves, and handwashing after lab to minimize lead exposure. Lead-containing waste should be properly disposed of in a dedicated waste container.

Caution: *Use caution with automated liquid handlers by keeping hands away from the moving parts during operation.* Larger liquid handlers, like the NIMBUS, are often inside a safety interlock cabinet to prevent the user from reaching inside during operation; this should be used, if applicable. This is less of a concern with the Sidekick, whose motors provide too little torque to cause pinch hazards.

Description of the Laboratory

The laboratory experience consists of a pre-lab computational exercise, an experimental laboratory period, and a post-lab data analysis period, depicted schematically in Figure 2.

The target student population are first-year undergraduate general chemistry students, without any specific pre-requisites in computer programming or mathematics. About half of the students had previous exposure to Mathematica through undergraduate calculus courses, with the others having no prior experience. Mathematica and Python notebooks, laboratory protocols, and scripts facilitating the automation are provided in the electronic Supporting Information. The pre-lab experience provides a theoretical discussion of concentration space, reaction design, and some basic programming activities. The in-lab exercises provide opportunities to introduce micro pipettors, volumetric solution preparation and the use of automated experiment platforms. The post-lab activities focus on observation, reaction evaluation, data entry and data analysis.

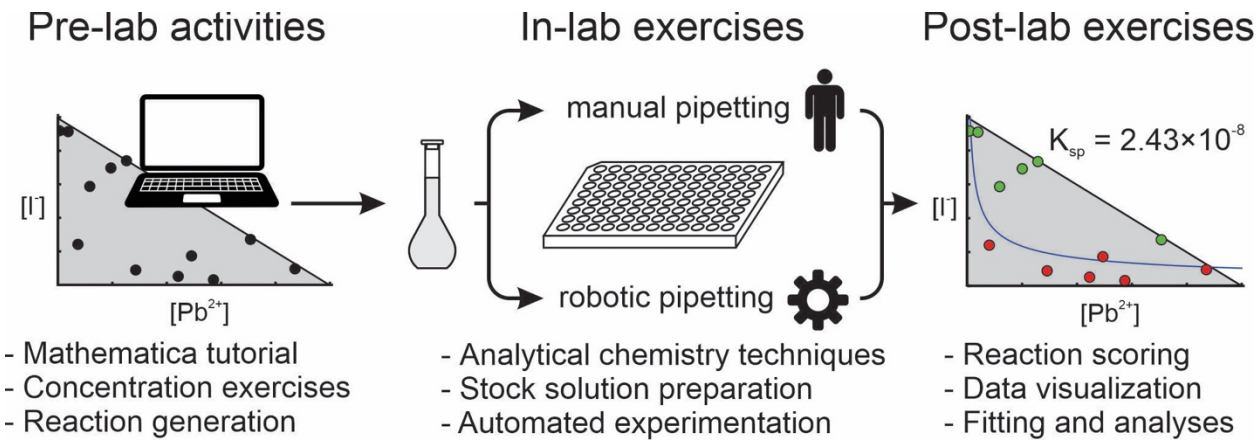
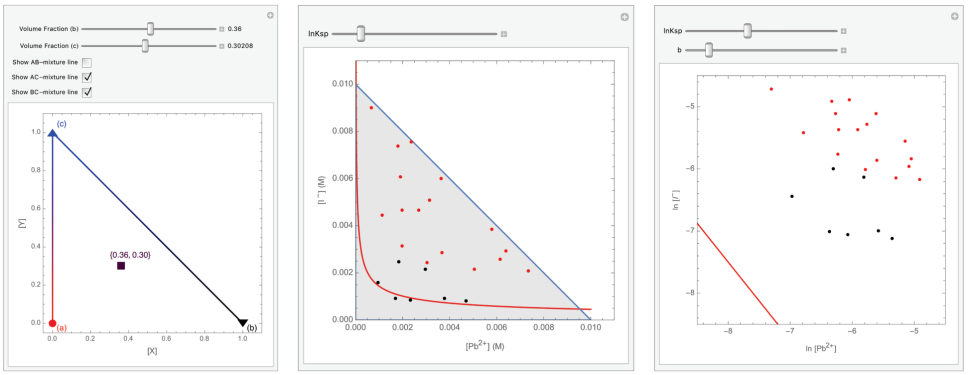


FIGURE 2: Schematic of experimental workflow.

Pre-lab computational activity. An interactive Mathematica 13.2 notebook is provided to the students for a pre-lab activity introducing the key concepts. This exercise covers the following topics: (i) Introduction to Mathematica (variables, functions, plotting); (ii) Concentration space representation of mixtures and equilibria; (iii) Generation of randomized reactions; (iv) Plotting and fitting of hypothetical data. (We also developed a Python 3 version of these exercises, which covers similar material, but the interactivity is less polished.)

A brief prelude introduces computer programming concepts of variables and functions, links to Mathematica tutorial videos, , and supplementary reading resources. The emphasis is on students learning how to *read* program code and to make *small modifications* to the code; they are not required to write large sections of code. In places, the text provides additional optional explanations or calculation methods that involve ideas from linear algebra to engage students with more advanced mathematics experience.



a b c

FIGURE 3: Interactive activities contained in the pre-lab assignment: (a) solution mixing; (b) equilibrium curve; (c) equilibrium curve after taking the logarithm of the concentrations.

The first part of the notebook contains interactive visualizations which do not require the students to enter any program code, but instead to explore how the system responds by manipulating slider bars. See Figure 3. The first activity introduces the idea of solution mixing and convex hulls (Fig 3a). Students are instructed to see how selecting different volume fractions results in different final mixtures. An interactive animation of the equilibrium curve in terms of the linear concentration (Fig 3b) and logarithm of concentration (Fig 3c) is used to build student intuition about data-fitting. The slider changes the fitting parameters, allowing students to adjust the equilibrium curve (red line) to best divide the two sets of outcomes in an example dataset. Finally, students perform a numerical logistic regression fit on an example data set and plot the results.

The second part of the notebook guides students through generating their own experimental plan. Students construct a convex hull for three stock solution concentrations (a lead source, and iodide source, and water), sample random concentrations within this space, and convert those concentrations to stock solution volumes. Students each generate 96 random experiments to perform, exporting them as a comma-separated-value (CSV) file, for use in both the both the manual- and automated-experiments. Each row specifies an experiment, described in terms of three columns indicating the volumes of the three liquids (water, Pb^{2+} (aq), I^- (aq)). TAs perform file conversions into robot specific formats using file conversion scripts provided in the Supporting Information.

In-lab exercises. This experiment is designed to be viable for a range of experimental configurations involving 96-well microtiter plates. A detailed, step-by-step lab procedure is provided in the Supporting Information. The laboratory begins with a demonstration of the proper use of a micropipette, as many students have little to no experience with this equipment and it is a useful laboratory skill for subsequent molecular biology and biochemistry courses. The instruction includes the proper way to pick up and drop a tip, setting the volume of the micropipette, and accurately dispensing the desired volume. The instruction is most effective when each student has a micropipette in hand. Commercial micropipettes are commonly available, but 3d-printed micropipettes are sufficiently accurate.[43] Each student performed 12 or 24 manual experiments, depending on time, from their randomly generated 96-reaction dataset generated in the pre-lab exercise.

A low-cost approach to automation used the Sidekick liquid dispenser robot, which can be built for under $800 USD in about 4 hours.[39] The Sidekick uses digital dispensing pumps which are pre-calibrated to dispense 10 uL per cycle, minimizing the need for extensive volumetric or gravimetric calibration. A higher-cost approach is to use a general liquid handler robot, which is essentially a

computer-controlled micropipette on a Cartesian axis gantry capable of both aspirating samples and dispensing them. However, for this laboratory requires only liquid dispensing. We used a Hamilton NIMBUS 4 with disposable 300-uL tips, which retails for approximately $70,000. Many other liquid handlers are capable of the experiments described here, including the OpenTrons OT-2 (which can be purchased for approximately $10,000 USD[44]) and pipette-based liquid handlers constructed from Lego™ [45,13] or laser-cut parts.[46]

Students are divided into pairs, with each student prepared one 100 mL of 0.010 M stock solution (either Pb^{2+} or I^-) using $Pb(NO_3)_2$ or KI and deionized water. Students calculate the mass of either $Pb(NO_3)_2$ or KI required for their respective stock solution. Stock solution volumes of 100 mL suffice for 144 experiments, comprising both manual- and automated experiments. In order to be efficient in lab, hand pipetting, automated NIMBUS and Sidekick experiments are conducted concurrently. Completing 96 automated-experiments required approximately 20 – 30 minutes. As such, the first two groups to prepare their stock solutions started with the automated experiments, under the supervision of the TA or instructor. Each group performed one complete set of 96 automated experiments according to the plan they generated, on one of the two robots described above. The remaining groups began with hand-pipetting experiments. As automated experiment runs completed, additional groups shifted from hand-pipetting to automated experiments. Each student performed at least 12 hand-pipetting reactions, based upon their 96 random reaction dataset, with additional experiments performed if time permitted. Seven of the fourteen students completed 12 hand-pipetting experiments, while the remaining seven completed 24 reactions. At the conclusion of the experiment, students cover and label their SBS plates. Lead-containing and non-lead-containing (KI solution) aqueous waste were collected separately in waste containers.

Precipitation can be slow, especially if the reaction composition is near the equilibrium curve (i.e., only slightly supersaturated). This means observations taken at the end of the laboratory period are likely to result in significant false negative results. To avoid this, digital photographs were taken of each reaction plate after 24 hours using a cell phone camera, and shared with students via Google drive, before disposing of the plates. Reaction plates were stored at room temperature in the laboratory. No additional temperature control was employed.

Our emphasis was on giving students hands-on experience with solution preparation, manual micropipettes, and laboratory automation. However, the experiments could be conducted using a teleoperated laboratory, such as the Carnegie Mellon University Cloud Lab.[47] Xie et al. discuss nuances of teaching remote-operated wet labs.[48]

Post-lab data analysis. Students score the outcomes by examining the photograph, noting the presence or absence of any quantity of yellow precipitate. An example is shown in Figure 4. Reaction scoring criteria were discussed prior to each student evaluating their own reaction results. The instructor described the main characteristics of crystal formation (color, habit, and yield), using examples from a projected image of a reaction plate. The distinction between crystals and bubbles was intentionally addressed, with a focus on round versus angular (for distinguishing bubbles from crystals) and colorless versus yellow (for distinguishing the presence of lead iodide). Example differences can be observed in wells A8 (bubble) vs A9 (small crystallites) in Figure 4. Reaction yield was discussed in the context of both distance between the point in concentration space and the equilibrium curve, and different in the reaction quotient, Q, and equilibrium constant, K. We considered automated scoring (e.g., computer vision, plate readers) but preferred having the students score outcomes by eye, both for operational simplicity and because it enabled student discussions about error sources.

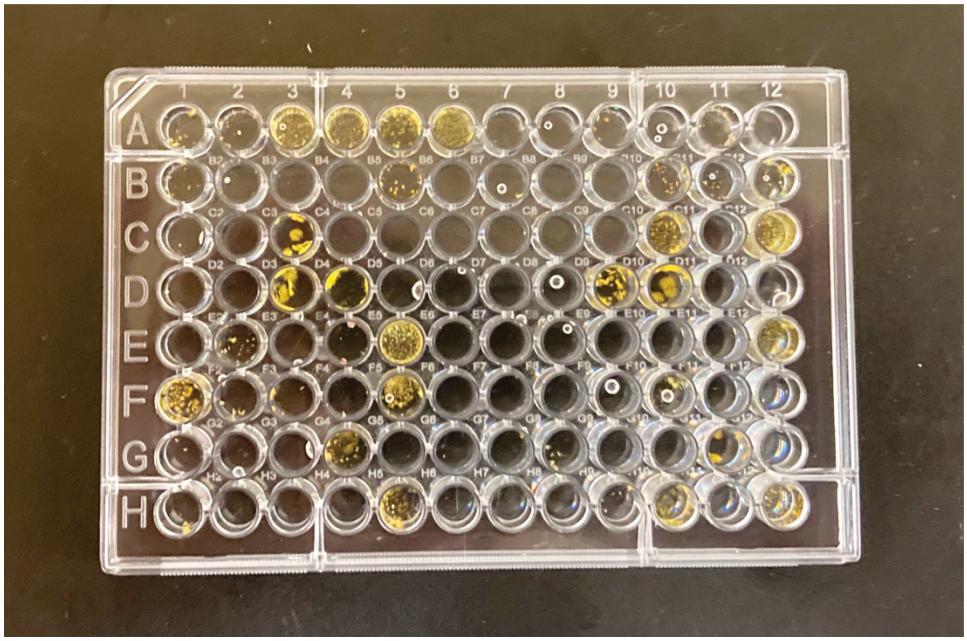


Figure 4. Photograph of 96 reactions.

To facilitate student scoring and comparisons, we developed a worksheet that students can use to categorize their results before entering them into a spreadsheet. Students used a figure of a blank 96-well plate, coloring in reactions with precipitates (provided in the SI). In pairs, scores (1 or precipitation, 0 for no precipitation) were added as a fourth column to their datasets. All reactions were scored.

Students then imported and analyzed their data. We developed a series of instructor slides that guide students through the process (using a sample dataset) as they follow along with each step. Each student then performed steps to separate the different reaction outcomes, plot results, fit the data, extract K_{sp} values, and analyze 'misfit' reactions. Students reported calculated K_{sp} values and the number of misclassified experiments for each batch of experiments (hand pipetting vs automated experiments). Finally, students submitted Mathematica notebooks and CSV data files after completing their analyses.

Results and Discussion

Chemical system selection. We considered three commonly used systems—lead iodide, barium sulfate, and silver sulfate. The desired properties are that the stock solutions should not be too dilute (to minimize possible sources of error in the stock solution), that observational information be accurate (facilitated by colored products), and that the reactions be relatively fast (to enable reactions to reach equilibrium before solvent evaporation affects outcomes). The stock solution concentrations are selected such that approximately half of random reactions sampled within the convex hull result in precipitation. We eliminated barium sulfate because of the challenge in making the dilute stock solution needed (which would increase error). A colored product facilitates accurate determination of the presence or absence of precipitates; Lead iodide is yellow, while silver sulfate is white, making lead iodide the superior choice. Although avoiding hazardous materials is desirable, there is ample precedent for using lead iodide in pedagogical experiments.[23] The use of microscale chemistry minimizes waste generation, but developing greener alternatives could be an area for future development.

Reliability and sources of error. Typical first-year laboratory experiments involve performing a single—or at most triplicate—experiment, and students expect perfect results. Laboratory experiments are generally designed to 'work' on a single trial, meaning the absence of a product indicates the student failed to correctly perform the appropriate sequence of steps. In contrast, automation lets students take a statistical approach to the problem, allowing small variations to be averaged out across many trials. Automation lets students quickly generate a large dataset to see statistically meaningful trends, without the tedium of repetitive manipulations. Additionally, students can directly observe how the initial distance from equilibrium is manifested in the time until an observable outcome, the quantity of product, and the likelihood of misclassification.

Pre-lab discussions focused on: (i) differences in 'control' between traditional and high-throughput experiments; (ii) sources, qualitative estimation and propagation effects of noise and errors; and, (iii) variations in fitted results given a finite data sample. Intuitions gained by the interactive fitting activities depicted in Figure 3 helped illustrate how many possible solutions exist. Post-lab discussions focused on errors associated with observations (the same 96 reactions being scored and entered by two different students) and differences in control and error of the various experimental methods (hand-pipetting vs automated experiments using different liquid handlers). Discussions were focused around calculated K_{sp} values, as each student performed the data analyses for their hand-pipetted and automated experiments during the post lab meeting. The instructor asked for the lowest and highest calculated K_{sp} values for each experiment type. Qualitative differences between experiment type were quickly identified. Hand-pipetting resulted in the largest range of values, whereas NIMBUS experiments resulted in the smallest range. See Figure 5. These qualitative results were discussed in the context of expected precision for each experiment type. Differences were also observed between partners evaluating the same 96 experiments. The use of the printed 96-well plate scoring template enabled students to quickly determine if their observations differed or if the discrepancy resulted from data entry errors.

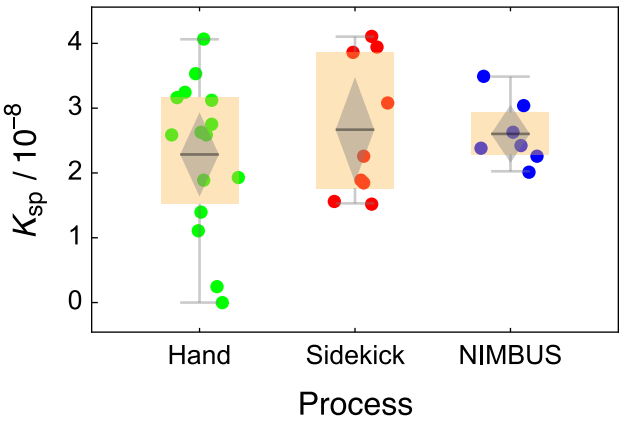


FIGURE 5: Student measurements of K_{sp} by different processes. Shaded boxes indicate the limits of top and bottom quartiles, whiskers indicate the total range of data, black horizontal lines indicate the mean, and the shaded diamond indicates the 95% confidence interval limits.

Accuracy of the Determined Equilibrium Constant. Fifteen students performed hand-pipetting experiments, 9 used the Sidekick, and 7 used the Nimbus (one student performed automated experiments on both machines). The determined solubility constants are shown in Figure 5. Hand-pipetting experiments show a wide range of results and outlier points. Mean and standard error of the measured K_{sp} values from three methods were $(2.29 \pm 0.31) \times 10^{-8}$, $(2.7 \pm 0.4) \times 10^{-8}$, and $(2.60 \pm 0.19) \times 10^{-8}$, respectively. All methods agree with the mean prediction, but unsurprisingly the variances are greater for the hand-pipetting reactions. This is comparable to the value of 3.1×10^{-8} determined by titration of a 0.01 M $Pb(NO_3)_2$ solution by Goodman and Petrucci[23] and values in the range of $2.68 \times 10^{-8} - 5.90 \times 10^{-8}$ (depending on variations in the ionic strength) determined spectroscopically by Green et al.[49]

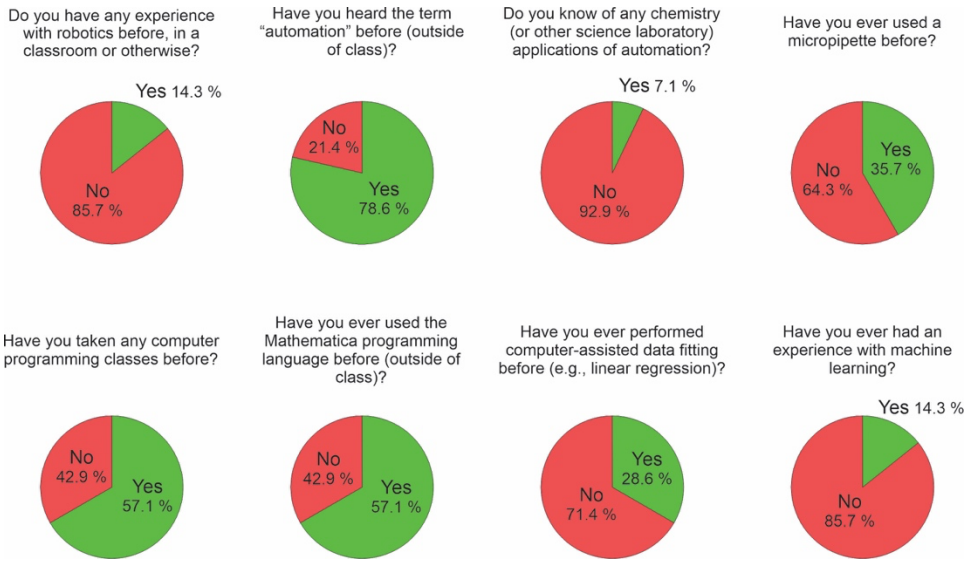


FIGURE 6: Pre-lab student survey responses from the student participants.

Student Feedback. We surveyed students before and after the lab to assess their preparation and perception of the experience. The pre-lab survey comprised 8 yes/no questions about prior preparation; the results are depicted in Figure 6. Most students reported no prior classroom or extracurricular experience in robotics, data fitting, or machine learning. While most students had heard about automation before, they did not know that it was applicable to chemistry. Approximately half of the students had prior experience in computer programming and in Mathematica specifically from previous mathematics coursework. We did not characterize the extent of these prior experiences, but the even distribution in experience levels suggests that dividing the students into pairs would allow for peer-instruction.

The post-lab survey comprised 10 questions scored on a 5-point Likert scale (with 5 is "Very strongly agree, 4 is "Agree", etc.); the results are summarized in Table 1. Students had a positive response about their perceived learning of the underlying course content (solution mixing and equilibria), perceived acquisition of technical skills (micropipetting, data analysis, programming), and desires to learn more about the topics. Students had a stronger perception of improved ability to read

program code than they did of writing code. Overall, the experience created a strong desire by the students to learn more about laboratory automation and data science, and they enjoyed the experience. Having students perform both "traditional" and automated experimentation, stimulated them to think and discuss about the relative advantages and disadvantages of different methods, so that they can think critically about where this might be applicable in their future work. A more advanced class might follow up on this discussion by reading relevant papers, such as Ref. [50], which discuss these issues in the context of organic chemistry.

TABLE 1. Post-lab student perception survey. [a] Mean and standard deviation, scored on a 5-point Likert Scale, with 5 as "Very strongly agree".

Post-lab student perception	**Outcome**[a]
Understand solution mixing better	4.3 ± 0.6
Understand equilibria better.	3.9 ± 0.7
Understand the tradeoffs between manual and automated experimentation better.	4.1 ± 0.8
Learned some skills that will be useful for my future studies.	4.6 ± 0.5
Learned about new ways to fit experimental data to solve problems.	4.2 ± 0.7
More able to read Mathematica code.	4.2 ± 0.7
More able to write Mathematica code.	3.7 ± 1.0
Interested in learning more about laboratory automation.	4.3 ± 0.7
Interested in learning more about data science.	4.3 ± 0.7
Enjoyed this lab	4.7 ± 0.6

Instructor observations. First, the familiarity with Mathematica varied widely across the student participants. This made live coding prohibitively slow. Debugging code became an impediment

to understanding and to progress. Instead, we shifted to an alternative approach in which students are asked to modify provided code. This enabled them to gain an understanding of the code and personalize their efforts while avoiding needlessly long steps. Alternatively, students could be provided with the code and an asynchronous video explanation, which they could follow at their own pace. The pre-lab meeting could then be used for debugging purposes. Second**,** most student participants had no experience with manual pipettors. We found that it was critically important to have a 'training step' in which each student held a pipettor while an instructor demonstrated how it works. A few minutes practicing pipetting water as a group helps avoid many errors in the actual experiments. Third, the student participants appreciated engaging with as many aspects of the automated experiments as possible. For example, each group recalibrated the Sidekick robot, even though it was not absolutely necessary. The inclusion of such steps, if practical or reasonable, provides a more active experience for the students.

Conclusion

In this paper, we describe a combined experimental and computational laboratory experience for first-year undergraduate students that introduces micropipetting skills, laboratory automation, and data analysis in the context of solubility equilibria determination. Students with little or no prior experience in these advanced techniques were able to successfully use and apply them in a lab course and reported a positive learning experience. This is a first step towards training a chemistry workforce that is aware of automation and data-science methods, and can think critically about sources of error and limitations of these approaches. Students reported a desire to learn more about these emerging aspects of chemical research practice. They also acquired discrete skills (use of volumetric flasks, solution preparation,

pipetting, use of robotic liquid handlers, data import/export, data visualization and fitting), which are useful for future coursework and research.

While we focused on first-year undergraduate students, this experience could be adapted for other student populations. For example, if more time was spent scaffolding the programming and data analysis aspects, this could also be used in a high-school chemistry course. Physical or analytical chemistry courses could incorporate the nonideality of the ionic system,[27,51] adapting the PbI_2 solubility lab of Green et al.[49] Analytical chemistry students could engage more extensively with the equipment by building and calibrating the Sidekick robot dispensing system[39] before applying it to this problem. An instrumental/quantitative analysis or computer-programming-for-chemists course might incorporate basic image analysis, such as counting the number of yellow precipitate pixels in each cell and relating this to the distance of each point from the equilibrium curve. (Zhang et al. recently described teaching these methods in the context of a senior undergraduate or beginning graduate bioanalytical chemistry course,[52] which could serve as a model.)

Acknowledgements

We thank the organizers and participants of the ACS Division of Physical Chemistry "Innovative Teaching in Physical Chemistry" symposium (Spring 2023 ACS National Meeting) for helpful comments about this work. This study is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001118C0036. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA. JS acknowledges support by the National Science Foundation under Grant No. PHY-22265110 and resources of the MERCURY consortium (http://mercuryconsortium.org/) supported by National Science Foundation Grant No. CNS-2018427.

Corresponding Authors

Joshua Schrier
Email: jschrier@fordham.edu
ORCID: 0000-0002-2071-1657

Alexander J. Norquist
Email: anorquis@haverford.edu
ORCID: 0000-0003-1040-5880

Co-authors

Gabe Jones-Thompson
ORCID: 0000-0001-8339-791X

Keqing He

Thomas J. Egg
ORCID: 0009-0003-1651-4751

Notes

The authors declare no competing financial interest.

Supporting Information

The Mathematica exercises and example data are archived online at (https://notebookarchive.org/2023-06-2tlv1o1) as well as present in the Electronic Supporting Information.

Mathematica notebooks (pre-lab and post –lab exercises) (.nb)
Python notebooks (pre-lab and post-lab exercises) (.ipynb)
Lab procedure (.docx)
Scripts for converting student experiments into inputs for Sidekick and Hamilton (.nb and .ipynb)
Drawing of a big 96-well plate to assist in scoring (pdf)
Student questionnaires (pre- and post-lab) (docx)
Example data file of dispense instructions and crystallization results (.csv and .xlsx)
Instructor Guide (.docx)

References

(1) Stach, E.; DeCost, B.; Kusne, A. G.; Hattrick-Simpers, J.; Brown, K. A.; Reyes, K. G.; Schrier, J.; Billinge, S.; Buonassisi, T.; Foster, I.; Gomes, C. P.; Gregoire, J. M.; Mehta, A.; Montoya, J.; Olivetti, E.; Park, C.; Rotenberg, E.; Saikin, S. K.; Smullin, S.; Stanev, V.; Maruyama, B. Autonomous Experimentation Systems for Materials Development: A Community Perspective. *Matter* **2021**, *4* (9), 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.

(2) Yano, J.; Gaffney, K. J.; Gregoire, J.; Hung, L.; Ourmazd, A.; Schrier, J.; Sethian, J. A.; Toma, F. M. The Case for Data Science in Experimental Chemistry: Examples and Recommendations. *Nat Rev Chem* **2022**, *6*, 357–370. https://doi.org/10.1038/s41570-022-00382-w.

(3) St James, A. G.; Hand, L.; Mills, T.; Song, L.; Brunt, A. S. J.; Bergstrom Mann, P. E.; Worrall, A. F.; Stewart, M. I.; Vallance, C. Exploring Machine Learning in Chemistry through the Classification of Spectra: An Undergraduate Project. *J. Chem. Educ.* **2023**, *100* (3), 1343–1350. https://doi.org/10.1021/acs.jchemed.2c00682.

(4) Thrall, E. S.; Lee, S. E.; Schrier, J.; Zhao, Y. Machine Learning for Functional Group Identification in Vibrational Spectroscopy: A Pedagogical Lab for Undergraduate Chemistry Students. *J. Chem. Educ.* **2021**, *98* (10), 3269–3276. https://doi.org/10.1021/acs.jchemed.1c00693.

(5) Revignas, D.; Amendola, V. Artificial Neural Networks Applied to Colorimetric Nanosensors: An Undergraduate Experience Tailorable from Gold Nanoparticles Synthesis to Optical Spectroscopy and Machine Learning. *J. Chem. Educ.* **2022**, *99* (5), 2112–2120. https://doi.org/10.1021/acs.jchemed.1c01288.

(6) Lafuente, D.; Cohen, B.; Fiorini, G.; García, A. A.; Bringas, M.; Morzan, E.; Onna, D. A Gentle Introduction to Machine Learning for Chemists: An Undergraduate Workshop Using Python Notebooks for Visualization, Data Processing, Analysis, and Modeling. *J. Chem. Educ.* **2021**, *98* (9), 2892–2898. https://doi.org/10.1021/acs.jchemed.1c00142.

(7) Hoogenboom, R.; Meier, M. A. R.; Schubert, U. S. The Introduction of High-Throughput Experimentation Methods for Suzuki–Miyaura Coupling Reactions in University Education. *J. Chem. Educ.* **2005**, *82* (11), 1693. https://doi.org/10.1021/ed082p1693.

(8) Lee, J.; Schmink, J. R.; Berritt, S. Introduction of Low-Barrier High-Throughput Experimentation in the Undergraduate Laboratory: Suzuki—Miyaura Reaction. *J. Chem. Educ.* **2020**, *97* (2), 538–542. https://doi.org/10.1021/acs.jchemed.9b00794.

(9) Wathen, B.; Lanehart, E.; Woodis, L. A.; Rojas, A. J. Buchwald–Hartwig Amination, High-Throughput Experimentation, and Process Chemistry: An Introduction via Undergraduate Laboratory Experimentation. *J. Chem. Educ.* **2021**, *98* (3), 996–1000. https://doi.org/10.1021/acs.jchemed.0c01355.

(10) Wentland, M. P.; Raza, S.; Gao, Y. 96-Well Plate Colorimetric Assay for Ki Determination of (+/-)-2-Benzylsuccinic Acid, an Inhibitor of Carboxypeptidase A: A Laboratory Experiment in Drug Discovery. *J. Chem. Educ.* **2004**, *81* (3), 398. https://doi.org/10.1021/ed081p398.

(11) Nel, M.; Potgieter, K.; Alimi, O. A.; Nel, A. L.; Meijboom, R. The Automated Synthesis of Aspirin: An Undergraduate Practical Activity. *J. Chem. Educ.* **2022**, *99* (11), 3773–3779. https://doi.org/10.1021/acs.jchemed.2c00503.

(12) Baird, S. G.; Sparks, T. D. What Is a Minimal Working Example for a Self-Driving Laboratory? *Matter* **2022**, *5* (12), 4170–4178. https://doi.org/10.1016/j.matt.2022.11.007.

(13) Saar, L.; Liang, H.; Wang, A.; McDannald, A.; Rodriguez, E.; Takeuchi, I.; Kusne, A. G. The LEGOLAS Kit: A Low-Cost Robot Science Kit for Education with Symbolic Regression for

Hypothesis Discovery and Validation. *MRS Bulletin* **2022**, *47* (9), 881–885. https://doi.org/10.1557/s43577-022-00430-2.
(14) Tan, S. W. B.; Naraharisetti, P. K.; Chin, S. K.; Lee, L. Y. Simple Visual-Aided Automated Titration Using the Python Programming Language. *J. Chem. Educ.* **2020**, *97* (3), 850–854. https://doi.org/10.1021/acs.jchemed.9b00802.
(15) Kosenkov, Y.; Kosenkov, D. Computer Vision in Chemistry: Automatic Titration. *J. Chem. Educ.* **2021**, *98* (12), 4067–4073. https://doi.org/10.1021/acs.jchemed.1c00810.
(16) Famularo, N.; Kholod, Y.; Kosenkov, D. Integrating Chemistry Laboratory Instrumentation into the Industrial Internet: Building, Programming, and Experimenting with an Automatic Titrator. *J. Chem. Educ.* **2016**, *93* (1), 175–181. https://doi.org/10.1021/acs.jchemed.5b00494.
(17) Vargas, S.; Zamirpour, S.; Menon, S.; Rothman, A.; Häse, F.; Tamayo-Mendoza, T.; Romero, J.; Sim, S.; Menke, T.; Aspuru-Guzik, A. Team-Based Learning for Scientific Computing and Automated Experimentation: Visualization of Colored Reactions. *J. Chem. Educ.* **2020**, *97* (3), 689–694. https://doi.org/10.1021/acs.jchemed.9b00603.
(18) Schrier, J.; Tynes, M. F.; Cain, L. Determining the Activity Series with the Fewest Experiments Using Sorting Algorithms. *J. Chem. Educ.* **2021**, *98* (5), 1653–1658. https://doi.org/10.1021/acs.jchemed.1c00043.
(19) Sharma, A. K.; Thuermer, C.; Ruan, V. Learning Programming through Chemistry in a First-Year Scientific Computing Course. In *Teaching Programming across the Chemistry Curriculum*; ACS Symposium Series; American Chemical Society, 2021; Vol. 1387, pp 43–56. https://doi.org/10.1021/bk-2021-1387.ch004.
(20) Sharma, A. K. Laboratory Glassware Identification: Supervised Machine Learning Example for Science Students. *JOCSE* **2021**, *12* (1), 8–15. https://doi.org/10.22369/issn.2153-4136/12/1/2.
(21) Clark, R. W.; Bonicamp, J. M. The Ksp-Solubility Conundrum. *J. Chem. Educ.* **1998**, *75* (9), 1182. https://doi.org/10.1021/ed075p1182.
(22) Wolthuis, E.; Pruiksma, A. B.; Heerema, R. P. Determination of Solubility: A Laboratory Experiment. *J. Chem. Educ.* **1960**, *37* (3), 137. https://doi.org/10.1021/ed037p137.
(23) Goodman, R. C.; Petrucci, R. H. Is the Solubility Product Constant? Introductory Experiment in Solubility Equilibrium. *J. Chem. Educ.* **1965**, *42* (2), 104. https://doi.org/10.1021/ed042p104.
(24) Carmody, W. R. Variation of the Solubility Product Constant with Ionic Strength. *J. Chem. Educ.* **1959**, *36* (3), 125. https://doi.org/10.1021/ed036p125.
(25) Sawyer, A. K. Solubility and Ksp of Calcium Sulfate: A General Chemistry Laboratory Experiment. *J. Chem. Educ.* **1983**, *60* (5), 416. https://doi.org/10.1021/ed060p416.
(26) Bonomo, R. P.; Tabbì, G.; Vagliasindi, L. I. Solubility and Solubility Product Determination of a Sparingly Soluble Salt: A First-Level Laboratory Experiment. *J. Chem. Educ.* **2012**, *89* (4), 545–547. https://doi.org/10.1021/ed2003143.
(27) McGarvey, D. J. A Data-Pooling Laboratory Activity to Investigate the Influence of Ionic Strength on the Solubility of CaSO4·2H2O(s). *J. Chem. Educ.* **2020**, *97* (2), 517–521. https://doi.org/10.1021/acs.jchemed.9b00817.
(28) Peterson, B. H. The Solubility Product Constant for Copper Iodate: An Experiment. *J. Chem. Educ.* **1957**, *34* (12), 612. https://doi.org/10.1021/ed034p612.
(29) Grabowski, L. E.; Goode, S. R. Determining a Solubility Product Constant by Potentiometric Titration To Increase Students' Conceptual Understanding of Potentiometry and Titrations. *J. Chem. Educ.* **2017**, *94* (5), 636–639. https://doi.org/10.1021/acs.jchemed.6b00460.
(30) Lyndrup, M. L.; Robinson, E. A.; Spencer, J. N. The Solubility of Zinc Oxalate. A Physical Chemistry Experiment. *J. Chem. Educ.* **1972**, *49* (9), 641. https://doi.org/10.1021/ed049p641.

(31) Pollock, D. W.; Truong, G. T.; Bonjour, J. L.; Frost, J. A. Solubility from the Femtoscale to the Macroscale. *J. Chem. Educ.* **2018**, *95* (6), 962–969. https://doi.org/10.1021/acs.jchemed.7b00965.
(32) Novak, I. Geometrical Description of Chemical Equilibrium and Le Châtelier's Principle: Two-Component Systems. *J. Chem. Educ.* **2018**, *95* (1), 84–87. https://doi.org/10.1021/acs.jchemed.7b00665.
(33) Schrier, J. Solution Mixing Calculations as a Geometry, Linear Algebra, and Convex Analysis Problem. *J. Chem. Educ.* **2021**, *98* (5), 1659–1666. https://doi.org/10.1021/acs.jchemed.0c01456.
(34) Boyd, S. P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK ; New York, 2004.
(35) Bernard, E. *Classic Supervised Learning Method*. Introduction to Machine Learning. https://www.wolfram.com/language/introduction-machine-learning/classic-supervised-learning-methods/ (accessed 2023-05-09).
(36) Bishop, J. A. The Solubility Product: A Proportionality Constant. *J. Chem. Educ.* **1954**, *31* (11), 574. https://doi.org/10.1021/ed031p574.
(37) *sklearn.linear_model.LogisticRegression*. scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (accessed 2023-06-06).
(38) *Multinomial logistic regression - MATLAB mnrfit*. https://www.mathworks.com/help/stats/mnrfit.html (accessed 2023-06-06).
(39) Keesey, R.; LeSuer, R.; Schrier, J. Sidekick: A Low-Cost Open-Source 3D-Printed Liquid Dispensing Robot. *HardwareX* **2022**, *12*. https://doi.org/10.1016/j.ohx.2022.e00319.
(40) Wolfram Research. Mathematica, Version 13.2. https://www.wolfram.com/mathematica.
(41) *Wolfram + Raspberry Pi Project: A Wolfram Engine on Every Raspberry Pi*. https://www.wolfram.com/raspberry-pi/ (accessed 2023-05-09).
(42) Wani, A. L.; Ara, A.; Usmani, J. A. Lead Toxicity: A Review. *Interdisciplinary Toxicology* **2015**, *8* (2), 55–64. https://doi.org/10.1515/intox-2015-0009.
(43) Brennan, M.; Bokhari, F.; Eddington, D. Open Design 3D-Printable Adjustable Micropipette That Meets the ISO Standard for Accuracy. *Micromachines* **2018**, *9* (4), 191. https://doi.org/10.3390/mi9040191.
(44) *OT-2 Liquid Handler | Opentrons Lab Automation from $10,000 | Opentrons*. https://opentrons.com/products/robots/ot-2/ (accessed 2023-05-09).
(45) Gerber, L. C.; Calasanz-Kaiser, A.; Hyman, L.; Voitiuk, K.; Patil, U.; Riedel-Kruse, I. H. Liquid-Handling Lego Robots and Experiments for STEM Education and Research. *PLOS Biology* **2017**, *15* (3), e2001413. https://doi.org/10.1371/journal.pbio.2001413.
(46) Li, E.; Lam, A. T.; Fuhrmann, T.; Erikson, L.; Wirth, M.; Miller, M. L.; Blikstein, P.; Riedel-Kruse, I. H. DIY Liquid Handling Robots for Integrated STEM Education and Life Science Research. *PLOS ONE* **2022**, *17* (11), e0275688. https://doi.org/10.1371/journal.pone.0275688.
(47) *CMU Cloud Lab | A Future of Science Initiative*. https://cloudlab.cmu.edu/ (accessed 2023-04-04).
(48) Xie, C.; Li, C.; Ding, X.; Jiang, R.; Sung, S. Chemistry on the Cloud: From Wet Labs to Web Labs. *J. Chem. Educ.* **2021**, *98* (9), 2840–2847. https://doi.org/10.1021/acs.jchemed.1c00585.
(49) Green, D. B.; Rechtsteiner, G.; Honodel, A. Determination of the Thermodynamic Solubility Product of PbI2 Assuming Nonideal Behavior: An Analytical Chemistry Experiment. *J. Chem. Educ.* **1996**, *73* (8), 789. https://doi.org/10.1021/ed073p789.
(50) Christensen, M.; Yunker, L. P. E.; Shiri, P.; Zepel, T.; Prieto, P. L.; Grunert, S.; Bork, F.; Hein, J. E. Automation Isn't Automatic. *Chem. Sci.* **2021**, *12* (47), 15473–15490. https://doi.org/10.1039/D1SC04588A.

(51) Denbigh, K. G. The Theory of Solubility Product. *J. Chem. Educ.* **1941**, *18* (3), 126. https://doi.org/10.1021/ed018p126.
(52) Zhang, B.; Frkonja-Kuczin, A.; Duan, Z.-H.; Boika, A. Workshop on Computer Vision for Bioanalytical Chemists: Classification and Detection of Amoebae Using Optical Microscopy Image Analysis with Machine Learning. *J. Chem. Educ.* **2023**, *100* (2), 539–545. https://doi.org/10.1021/acs.jchemed.2c00631.

Graphical Table of Content