Physical Science Department
Unit-Level Assessment Liaison Report
Spring 2018

Liaison Project Start Date: Fall 2017
Liaison Report prepared by Allan Wilson

I. Department Buy-In and Outcome Definition

Prior to this year, the physical science department had been using standardized multiple choice exams from the ACS (American Chemical Society) which attempt to survey most of the big topics covered in a semester. This year, we transitioned to a free response assessment on one topic, developed in-house. This was done largely because the scores on the ACS exams were not changing significantly from semester to semester, and we wished to see a more in-depth picture of exactly what mistakes students were making than could be gleaned from their answers on a multiple choice exam.

We took some inspiration from the assessment results produced by Oakton College (see Appendix 1), and decided to focus on stoichiometry. We chose this topic because it is one of the most important topics in CHEM 121 (Basic Chemistry) and CHEM 201 (General Chemistry), addressed in multiple chapters in both courses. Additionally, the results from the ACS exams had shown that our students' understanding of this topic is "emerging" - they perform well on easy questions, but they struggle when confronted with more difficult ones.

II. Assessment Research and Design

An email was sent out to fulltime and adjunct faculty teaching chemistry, describing the thrust of the new assessment project and asking for sample questions. A test bank of possible questions was then made consisting of submissions from faculty, questions that I found in my own searches, and a few questions that I wrote myself (see appendix 2). The faculty were then asked to vote for one easy question (a multiple choice question about conceptual reaction stoichiometry), one question of moderate difficulty (a typical algorithmic problem of a sort frequently encountered in chemistry courses) and one difficult problem (a calculation problem which was not likely to resemble one they had encountered in class). There was one clear favorite for each class of problem, and these three questions formed our stoichiometry assessment (see appendix 3).

III. Pilot Assessment Tools and Processes

It was decided that there would be no pilot assessment. We were confident enough in our assessment tool that we felt we could administer the assessment without one.

IV. Administer Specific Assessment

The assessment was emailed to every professor for CHEM 121 and 201. Participation in the assessment was voluntary, and it was up to the individual professor how the assessment was incorporated into the course (as a stand-alone quiz, as part of the final exam, as extra credit, as an ungraded study activity, etc.). The only requests were that it be given in the final two weeks of the semester, and that it be given under "test-like" conditions (i.e. no group work or homework).

Professors who wanted to give students a grade for this assessment were instructed to first photocopy the exam responses and grade the photocopy. The original responses were given to me, and mixed with other sections of the same course to preserve the anonymity of both the students and the professors. I graded each exam according to a rubric that I developed after looking at a subset of student responses (see appendix 4).

V. Data Analysis

I performed the initial data analysis myself in the spreadsheet that was used for data collection. The data was then given to one of our data analysts, Fernando Miranda-Mendoza, for more sophisticated treatment (see appendix 5).

It was found that CHEM 201 students performed better on the test than CHEM 121 students, as expected, although the difference was not statistically significant. It was also found that the anticipated "easy" and "hard" questions were in fact easier and harder for students, based on the number of incorrect responses.

More interestingly, there were some common mistakes that could be identified, particularly on the most difficult problem. Students in both CHEM 121 and 201 struggled with this problem. It was a "real world" problem, giving students a chemical reaction used to remove the sulfur of a tarnished bracelet and asking students to perform calculations based on the change in mass of the bracelet. In the vast majority of problems students encounter, they are given the mass of a specific chemical; by giving them the mass of a real world object and asking them to discern what chemical is being referenced, the problem became much harder, and many students selected an incorrect substance.

VI. Supporting Evidence-Based Change (Use of Findings)

Students clearly need more practice with these more difficult real-world problems. Follow-up conversations with several professors indicated that a significant barrier to using more practice problems is that such questions are in very short supply. They are very difficult to write, and they are not commonly found in textbooks. I therefore set about creating a question bank of such questions, based on submissions from my fellow faculty, questions found from textbooks and online searches, and a few that I wrote myself (see appendix 6). It is anticipated that as professors become more confident writing and using these more challenging problems in their classes, they will add questions to the question bank.

Success Factors

The most important success factor has always been the participation of the faculty. Assessment projects in the physical science department enjoy broad support, and participation is high despite the voluntary nature of these projects. In addition to the large number of professors giving the assessment in their classrooms, I have benefitted greatly from helpful conversations with my colleagues, as well as participation in supplying possible assessment questions, choosing assessment questions, and making the question bank of real-world problems that was recommended by the results. I wish to extend my thanks to everyone in the department who helped make this work possible.

Another success factor was the previous assessment work on which this project was built. The results of the ACS exams highlighted stoichiometry as a topic in which student learning could be improved. As our first assessment project, it also helped establish the positive assessment culture that our department now enjoys.

Recommendations

As discussed above, the most important recommendation is for faculty to give more real-world problems in all units, particularly the ones that involve stoichiometry.

A more minor recommendation involves the assessment tool. The most difficult problem on the assessment was a two-part question. Students took the answer from part A and used it to calculate the answer for part B. The challenging part of the problem, however, was contained in part A. So almost all students got part B wrong, but in many cases this was because they had done the work for part B correctly, using the incorrect answer they had obtained for part A. Grading part B became very challenging, and the myriad possible mistakes made it difficult to discern any patterns, so ultimately the decision was made to disregard the results of this answer. In the future, I plan to avoid assessment problems with an A-B format.

Appendix 1: An assessment in quantum mechanics from Oakton College

Quiz: Quantum Mechanical Model of the Atom

1. Consider a sulfur atom (S) in its ground state. (a) Write the complete electron configuration. (b) Write the complete orbital diagram (in which arrows are used to represent electrons). In both cases, arrange the sublevels in order from lowest to highest energy (left to right).

2. (a) How many columns are in the d-block of the periodic table? ________
(b) If somehow in another universe, all the solutions to the Schrödinger equation were the same as in ours, but there were four possible spin options for an electron instead of two (so that four electrons could occupy any orbital rather than two), how many columns would there be in the d-block in that universe's periodic table? _______ Briefly give your reasoning (required).

3. Circle the correct answer, (a) – (e). Which of the following represent(s) a possible excited state?

I: [Ne] 3s23p2 II: [Ar] 4s13d54p3 III: [Kr] 5s24d145p1

(a) I only (b) II only (c) III only (d) II and III (e) I and III

4. In another universe, the solutions to the Schrödinger Equation were such that the following applied (see figure below). Note that a number of things are different in this universe, including the names of the sublevels that were chosen.

In the box below, write the complete electron configuration for zirconium (Zr), i.e. the element with atomic number 40 (Z = 40), in this alternate universe. Assume all other rules and restrictions related to quantum mechanics are the same in this universe as in ours. NOTE: Only work in the box below will be graded!

Appendix 2: Possible assessment questions sent to the department for voting

Possible Stoichiometry Assessment Questions

POSSIBLE EASY QUESTION #1

Which of the following pictorially represents the balanced chemical equation $N_2 + 3H_2 \quad 2NH_3$?

(a)

(b)

(c)

(d)

(e)

POSSIBLE EASY QUESTION #2

Iron combines with oxygen and water from the air to form rust. If an iron nail were allowed to rust completely, one should find that the rust weighs:

(a) less than the nail it came from.
(b) the same as the nail it came from.
(c) more than the nail it came from.
(d) It is impossible to predict.

What is the reason for your answer?

(a) It decomposes to other materials.
(b) It forms an oxide but the weight remains constant.
(c) The weight of oxygen can vary from place to place.
(d) Rust contains iron and oxygen.

POSSIBLE HARD QUESTION #1

A bracelet, originally made of pure silver, became tarnished over time with black silver sulfide (Ag_2S) forming on the surface. The bracelet was cleaned by converting the silver sulfide back to metallic silver using aluminum in the following reaction. The mass of the bracelet decreased by 0.0096 g in the cleaning process.

$3Ag_2S + 2Al \quad 6Ag + Al_2S_3$

(i) How many moles of sulfur (S) were removed from the bracelet when the silver sulfide (Ag_2S) was converted to aluminum sulfide (Al_2S_3)?

(ii) What mass of aluminum was used in the reaction?

POSSIBLE HARD QUESTION #2

Consider the reaction below, with unknown stoichiometry. If 6.2 g of iron is allowed to react completely in excess CO, and the mass of product obtained is 21.8 g, what is x?

$$Fe + x\,CO \quad Fe(CO)_x$$

The middle question (medium difficulty) will be a "standard" stoichiometry question requiring calculation. Since the difficult question will require calculations involving mass, the second question could either involve solutions (molarity to mole calculations) or gas laws (volume to mole calculations). Which would you prefer?

Appendix 3: The departmental stoichiometry assessment, Fall 2017

Stoichiometry Assessment

Physical Science Department, Harold Washington College, Fall 2017

#1. Circle the response that answers the question below.

Which of the following pictorially represents the balanced chemical equation N_2 + $3H_2$ $2NH_3$?

(a)

(b)

(c)

(d)

(e)

#2. Calculate the answer to the question below. Show your work.

If 50.0 mL of 2.0 M HCl solution reacts with excess magnesium according to the balanced equation below, calculate the moles of hydrogen gas that is produced.

2 HCl (aq) + Mg (s) $MgCl_2$ (aq) + H_2 (g)

#3. Calculate the answers to the two questions below. Show your work.

A bracelet, originally made of pure silver, became tarnished over time with black silver sulfide (Ag2S) forming on the surface. The bracelet was cleaned by converting the silver sulfide back to metallic silver using aluminum in the following reaction. The mass of the bracelet decreased by 0.0096 g in the cleaning process.

3 Ag2S + 2 Al 6 Ag + Al2S3

i) How many moles of sulfur (S) were removed from the bracelet when the silver sulfide (Ag2S) was converted to aluminum sulfide (Al2S3)?

ii) What mass of aluminum was used in the reaction?

Appendix 4: Assessment rubric

	Student	
	Score	
	Total possible	5
#1	A	-1
	B	-1
	C	0
	D	-1
	E	-1
#2	Correct 0	
	Math error	-1
	Wrong molarity	-1
	Molar Mass	-1
	Wrong stoich	-1
	Stoich not used	-1
	Assume eq. moles	-2
	Unintelligible	-2
	Blank	-2
#3	Correct	0
	Math error	-1
	Wrong substance	-1
	Wrong mole calc	-1
	Assume eq. moles	-2
	Unintelligible	-2
	Blank	-2

Appendix 5: Assessment results (questions from me, answers from Prof. Miranda-Mendoza)

Stoichiometry Assessment Results

- Here are some preliminary observations. I'm curious to see if you think they are in fact supported by the data. Have I missed something? Are the results statistically significant?

I ran several tests on the data you sent me to see if there were any significant differences (using standard statistical assumptions) between the two groups, 121 and 201. I used a robust statistical test called Welch's t-test which works well when comparing two groups with unequal sample sizes (like the two groups from this assessment, where the 121 group consists of 57 individuals while the 201 group consists of 147 individuals). Overall, no statistically significant differences were detected and there is no indication that the one group performed better than the other.

- 201 did better overall than 121: 2.62 vs. 2.39 (out of 5 possible points)

Not an indication of a significant difference (p-value = 0.3598)

- 201 did better on #3 than 121: 201 missed on average -1.43 points, whereas 121 missed -1.51 points (on a 2 point problem).

Not an indication of a significant difference (p-value = 0.5544)

- 201 did better on #2 than 121: -0.789 vs. -0.965 (2 point problem)

Not an indication of a significant difference (p-value = 0.2215)

- 201 did WORSE than 121 on #1: -0.158 vs. -0.14 (1 point problem). I'm particularly interested in this result. Is it statistically significant? If so, it might indicate that students forget some conceptual understanding of what a reaction describes if it is not reinforced in 201.

Even though the 201-group performed worse, here too there is no indication of a significant difference (p-value = 0.7617)

- Now to specific answers, for which I have a big nebulous question. I'm curious to know if the categories I identified for the rubric are good ones. For instance, it seemed to me that a significant fraction of students used a molar mass to solve problem number 2, which is a mistake. So it was a separate column in my rubric. When the math was done, students in 121 missed on average -0.28 points (out of 2) due to this error, and students in 201 missed -0.18 points. Is there a way to determine if these values are large enough to justify their own column in a rubric? Can I confidently say to the department, "it might be helpful to reinforce to students that not all mole calculations require the molar mass"? Does my question make sense? And I basically have this question about every column!

Yes, your question makes sense. One possible way to determine if a category is justified or not in your rubric is to look at the proportion of students that made the error specified in that category. We can then compare this proportion to a (hypothetical) probability that students made the error by just a random accident. A reasonable assumption is to take this probability to be 0.5 (50% chance of randomly making a specified error). A statistical test for proportions can be run to determine if the proportion of students that made an error is significantly different from a random accident. I ran tests for all the columns and the results indicate that the proportion of errors (in both groups) is significantly different than the probability of random guessing (0.5).

For example, for question 2, the proportion of students that made the molar mass error is 0.28 (or 28%) for the 121 group and 0.18 (or 18%) for the 201 group (Note: these values coincide with the average absolute values of negative points, since the molar mass error was penalized by taking one point off). A test of proportions strongly indicates that both, the proportion of 121 and 201 errors, are significantly different than 0.5 (for the 121 group the p-value= 0.001478 while for the 201 the p-value= 2.936x10^(-11)). Similar results were obtained on all the other columns. So, you have a justification for each of the types of errors that you considered.

Notice, however, that the approach above relies on having a good assumption as to what is the probability of a student making a given error by random accident. If you have a better estimate of this probability, for instance, if you know from previous experiences the frequency of seeing a given error, then I can run new tests to see if a given category is justified or not. Be aware that we cannot assume that the probability of making an error by random accident is 0. So, it is not appropriate to compare the proportions to 0.

Since the molar mass error is one of the highest, you can definitely inform your department that more attention should be paid on this category.

- On a smaller note, I told the committee that I THOUGHT that 121 students had higher rates of completely lost or "strange" answers. It turns out that 121 students on #3 missed on average -0.42 answers due to unintelligible responses, and -0.32 points due to blank responses. The 201 students missed -0.28 points to unintelligible answers, and -0.14 points due to blank answers. Are those differences significant? It looks like the other big mistake that lost them all the points for that problem - simply assuming that you have the number of moles mentioned in the balanced equation, or the mass provided by the periodic table - was equally prevalent in both 121 and 201 (-0.32 points vs. -0.33 points). I assume that small difference is not significant?

- As with other results, there is no indication of a significant difference for the average missed points in question #3, for the categories of “unintelligible” (p-value = 0.272), “blank” (p-value = 0.1102), or “assuming equal moles” (p-value = 0.8841).

- Lastly, an even more nebulous question. Based on the data collected, can you see any obvious questions that could be explored beyond the ones I've mentioned?

- I looked at the proportion of students that had a fully correct answer for each question (I noticed the “Correct” columns on your sheets). I thought some significant differences may pop out there, but interestingly the results of the tests once again did not yield a significant difference between the two groups. Hence, as it stands out now, both groups (121 and 201) appeared to have performed the same across all categories.

Let me know if you have any questions or would like me to look more into the data. Also, if you have a particular value for the hypothetical probability mentioned above, then let me know to rerun the test with a more appropriate value.

Fernando Miranda-Mendoza

Appendix 6: Stoichiometry question bank

"Real World" Stoichiometry Problems

A Note about the Questions Below:
Hi everyone! This set of questions came out of our department's assessment activities, primarily the stoichiometry assessment given in the Fall of 2017. In that assessment, it was found that students struggled to identify the pertinent reacting substances when problems were presented in a "real-world" fashion. For instance, compare the following two questions:

a) What mass of iron (III) oxide is produced from the reaction of 0.042 g of oxygen, according to the equation below?

b) Iron is the most widely used of all metals, with its high strength and low cost making it ideal for structural components in buildings and machinery. However, it is not indefinitely stable – over time it reacts with oxygen in the air, forming rust – iron (III) oxide – according to the equation below. Suppose you have an iron nail, originally with a mass of 2.213 g, that begins to rust. If its new mass is 2.255 g, what mass of iron (III) oxide is now present in the nail?

4 Fe (s) + 3 O2 (g) 2 Fe2O3 (s)

Even though both questions require the same stoichiometric calculations, the second is significantly harder for our students than the first. In particular, the fact that the masses are not explicitly linked to specific chemical substances causes a great deal of confusion. Many of you have indicated that one barrier to providing more practice in these sorts of questions is simply the difficulty in finding such problems. This problem set attempts to help with this issue.

Feel free to use any of these questions, in any format that you like. Make changes as you see fit to increase or decrease the difficulty level.

I hope to add more questions to this document over time. If you have other example questions that you think would be of interest to your fellow faculty, please send them to me, and I will add them to this question bank. The greater the variety in this resource, the more useful it will be!

Smiles!
Allan Wilson

Aspartame (NutraSweet) is an artificial sweetener. Once consumed, it breaks down into several compounds as described below. One of the products is phenylalanine, which poses health concerns for people who have a reduced ability to metabolize this compound (a condition known as phenylketonuria). One mole of aspartame (C14H18N2O5) reacts with two moles of water to produce one mole of aspartic acid (C4H7NO4), one mole of methanol (CH3OH) and one mole of phenylalanine.

a. What is the molecular formula of phenylalanine?

b. The primary source of aspartame in the United States is in diet soft drinks, where one can of diet soda

contains on average 180 mg of aspartame. What mass of phenylalanine is produced in the body after the consumption of 2 cans of soda?
c. The FDA has set the acceptable daily intake (ADI) for aspartame to be 50 mg of aspartame per kilogram of body weight (assuming you do NOT have phenylketonuria). How many cans of diet soda can a 175-pound person drink without exceeding the ADI for aspartame? (454 g = 1 pound)

Many famous sculptures are made of marble, which is composed of calcium carbonate ($CaCO_3$). At STP this substance is stable, but at higher temperatures it decomposes according to the equation below. Suppose a marble statue, originally weighing 92.36 pounds, is caught in a house fire. After it is rescued from the rubble, it is found to weigh 91.92 pounds. If we assume that all of the mass lost is due to the chemical reaction below (as opposed to physical processes such as breakage), what mass of calcium oxide is now present in the statue? (1.00 kg = 2.20 pounds)

$CaCO_3$ (s)   CaO (s) + CO_2 (g)

Approximately 46% of the electricity in the United States is produced from the burning of coal, which is mainly carbon. Most coal contains some sulfur (S_8) which is ultimately responsible for some acid rain in the eastern United States. Here's how:

1. In the power plant, the sulfur burns in the presence of oxygen to create sulfur dioxide gas.
2. Sulfur dioxide gas in the atmosphere combines with oxygen in a synthesis reaction to form sulfur trioxide gas.
3. Sulfur trioxide reacts with water vapor in the atmosphere to create sulfuric acid which dissolves in rain water and falls to the Earth.

If 0.0037% by mass of all the coal ore burned in the United States is actually sulfur,
(a) how many kilograms of coal ore have to burn to generate 1 kilogram of sulfuric acid?
(b) how many liters of sulfur trioxide gas does this amount to (assuming STP)?

On average, an acre of corn will remove 6 kilograms of phosphorus from the ground, and this needs to be replaced each year using fertilizer.
Imagine you inherit a farm. The farm is 340 acres and had corn planted the previous year. You must add fertilizer to the soil before you plant this year’s crop. At the store, you find a fertilizer which, according to the bag, has the molecular formula $Ca_3P_2H_{14}S_2O_{21}$. One bag contains 50 pounds of this fertilizer and costs $54.73. How much will it cost you to add the necessary fertilizer to your fields? (454 g = 1 pound)

Airbags in an automobile inflate during a collision to help protect the passengers from injury. The airbags are inflated by the chemical decomposition of sodium azide according to the reaction shown below. What mass of sodium azide is necessary to inflate a 60.0 L airbag at 25°C and 1.0 atm?

2 NaN_3 (s)   2 Na (s) + 3 N_2 (g)

The stone “icicles” found hanging from the ceilings of caves are called stalactites. They are formed gradually as dissolved calcium bicarbonate in rainwater precipitates out of solution according to the equation below.

$Ca(HCO_3)_2$ (aq) $CaCO_3$ (s) + H_2O (l) + CO_2 (g)

Not all of the calcium bicarbonate precipitates – the “percent yield” of this reaction is typically only 1%. If 50 mL of rainwater flows over a stalactite per hour, with a calcium bicarbonate concentration of 0.004 M, how much mass is added to the stalactite in one century?