Automated CANVAS quiz for teaching Chemistry

Gergely Katona[1]
1. Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden.

Background

Practicing is an important part of teaching and learning. It is especially useful for developing skills that are considered routine and should not present a cognitive load for the students on the way towards more advanced studies. These skills prepare the ground for other areas that need complete attention and intense concentration. In the field of chemistry, this can be interpreting very large and very small numbers, converting and intuitively interpreting units and prefixes, naming and interpreting the names of compounds, perform routine calculations with concentrations and dilutions. The strategy used here was to give a large number of quizzes to the students in CANVAS with practicing and testing purposes. These can be used in large classes and the correction of the tests are automated without consuming human resources. It was expected that the students enjoy working with digital resources and they are more competent at routine tasks at the end of course. By blurring the boundary between practicing and testing the expected learning outcomes was very clear from the beginning of the course..

Theoretical support

There is an overwhelming evidence that long term retention is enhanced by repeated retrieval and practice.(1) As the authors describe:

"Testing at the university level provides an indirect bene-fit that complements the direct benefit that is discussed here. Many university courses require only one or two semester tests and a final exam, a practice that leads to the near universal phenomenon of students concentrating their study attempts just before the exams and not keeping up with the course [86,87]. Frequent quizzing (say, on a weekly or even a daily basis) forces students to stay current with the course by studying more regularly."

This is most recognizable during language training when building an initial vocabulary consist of retrieval of unknown words on the same day, but retrieval is also repeated periodically with several days in between. Teaching of basic chemistry is not fundamentally different, the students need an initial "vocabulary" of chemical compounds, have to understand the "grammar", the relationship between chemical formulas and nomenclature and "meaning" in the form of molar ratio. It is also important to give an emphasis of the context in which "words" of chemistry can be used.

In addition to the chemistry specific knowledge and skills, students are required extensive training in numerical literacy and practice unit conversions to start solving problems. It is also recognized that immediate feedback is less effective than delaying it. If nothing else repeating failed tests give more opportunity for retrieval and stronger retention.

As a teacher, one has to maintain the enthusiasm of the students and giving incentive for performing further retrievals. Being "almost there", "almost succeeding" is an excellent way of fueling the desire for further trials and exploited in different types of gamification.

The aim of the project was to provide nearly unlimited quiz items for repeated retrievals.

Digitization, implementation

The scripts for generating random multiple choice questions were developed in the python 3 language using the pandas and numpy libraries. The item creation process was digitalized with these supporting tools:

1. Converting computational float numbers to human readable format using Unicode notations to increase compatibility with a wide range of Unicode compatible educational tools. For example: The process converting 3.14E-04 to $3{,}14\times10^{-4}$.

Similarly, electric charges in chemistry have special notation: referring to iron and +3 charge yields the symbol Fe^{3+}.

2. Routines for unit conversion were used for unit conversion quiz generation, but were also incorporated in more complex quizzes. Internally, physical quantities were expressed by an arbitrarily chosen unit and displayed in a randomized fashion. For example a volume 2 L can be displayed randomly either as 2000 mL or 2×10^{-3} m^3.

3. Tools aimed at randomized associations between chemical formulas and chemical names. This includes random generation of ionic compounds with randomly selected cations and anions, which maintain charge neutrality. These can be used for nomenclature quizzing or provide further variations by randomly choosing between formulas and names in more complex quizzes.

Automatic multiple choice quiz generation involved the digitization of seemingly trivial teacher behavior:

1. Establishing ground truth. This can involve random combination of quantities that satisfy a given chemical equation or represent a valid association between two objects. Examples of ground truth:
    a. Volume= 3L, Concentration= 1M, amount of substance=3 mol.
    b. Järn(III)klorid = $FeCl_3$
2. Making sure that the distractor is different from the key. This is based on the underlying representation of the quantity rather than the displayed form.
3. Enforcing that two or more distractors are not the same by chance, but not very different either. Ideally, there is a tunable difficulty setting.
4. In chemistry, numbers are often chosen by convenience (calculation or practical). A 111 mL stock solution of 1.34 M concentration is never prepared in a laboratory. Flasks are available in sizes of 100 mL or similar convenient volume. We use the decimal system, therefore it is easier to calculate dilutions when 1 M concentration is prepared. Similarly, convenient dilution factors are also restricted in a human laboratory. For quiz generation purposes, it is better to choose discrete values from a list randomly, rather than generating uniform random numbers.
5. Similarly, distractors become more powerful if they differ by not a completely arbitrary number, but by a factor of 2, 1/3 or similar ratios.
6. For textual distractors the similarity was assessed by the Levenshtein distance(2). More difficult distractors were selected among large number of randomly selected distractors with lower Levenshtein distance.
7. The same ground truth can be probed by multiple randomly selected questions that ask about one missing quantity.

Communication between the python scripts and CANVAS was enabled by the GIFT markup language and the webservice GIFT2QTI. The output QTI format is widely used in different educational software. Items were randomized and mixed by the CANVAS question bank feature. The key and distractors were reordered randomly by the CANVAS quiz function.

The students had an intense Monday to Friday lecture series about stoichiometry. The students were offered the opportunity to receive extra points for the final examination if they completed at least some of the quizzes.

The deadlines of the quizzes were Tuesday and Thursday on the same week and the Monday and Wednesday of the following week. The students could retake the quiz unlimited times before the deadline and only their highest score was retained. Only the number of correct answer was revealed to the students. When the students

Table 1

	Unit conversion	Nomenclature	Solution mixing	Stock solution	Advanced stock solution	Total	Required correct answer
Quiz 1	25	40	3	2	0	70	42
Quiz 2	25	40	2	3	0	70	42
Quiz 3	10	20	15	15	0	60	42
Quiz 4	0	10	20	10	10	50	42
Total	60	110	40	30	10	250	168

requested help, they did not receive step by step instructions. Instead, they were asked to demonstrate how they were thinking and based on that, they were offered guiding questions. They were assumed to be responsible for solving the problems.

The composition of quizzes is detailed in Table 1. Unit conversion items tested number literacy, basic arithmetic and unit handling skills. Nomenclature items probed basic chemical knowledge and the recognition of the equivalence between chemical formulas and their name. Solution mixing and stock solution items were testing sharply more advanced chemical knowledge and problem solving skills. They could not be successfully completed without the routine use of aforementioned skills. The items within each groups were randomized from an automatically generated question bank consisting of 200 items each. The key and distractors were also randomly ordered.

Results

The analysis described here concerns the 82 students enrolled in the course Basic Analytical Pharmaceutical Chemistry. Figure 1 shows the number of students that successfully completed 0, 2, 3 and 4 quizzes respectively. There was not any student that completed only one quiz. Approximately half of the students reached the stringent 84% test limit enforced in the fourth quiz. This was achieved already less than two weeks into the course, which predicted a good outcome in the final examination for these students. Figure 2 shows the histogram of student highest scores for the respective quizzes. Quiz 1 and 2 had identical difficulty level and their deadline was separated by two days only. The near uniform shift in density towards improved results indicates that students at every level benefited from repeated retrievals. Quiz 3 represents a jump in difficulty and a substantial fraction of the students struggled initially to reach the 70% target of the correct answers. Nevertheless, half of the students quickly adapted to the even more difficult challenge represented by quiz 4. It is important to note that a failed quiz 4 with 60%-80% correct answers is still a promising result.

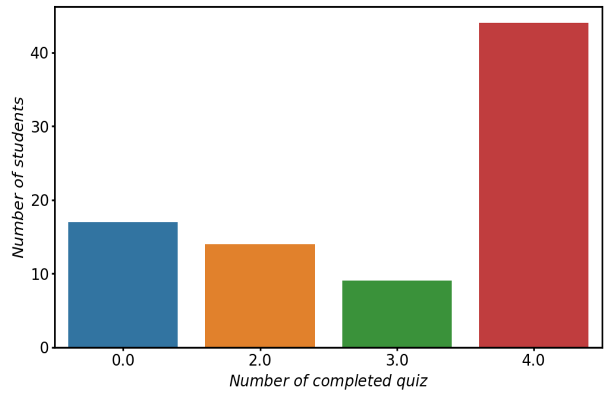

Figure 1 The number of students completing up to four quizzes.

Figure 3 shows the effect of repeated trials within quiz 4 for those students that repeated the quiz. These students included both those that needed better score for passing and those that returned to the quiz for practicing or other purposes. Each column represents the distribution of improvement or deterioration in the results compared to the previous trial. It appears that the fourth trial gave consistently the highest improvement and beyond that the benefit of further trials started to decline. Similar tendency was observed for quiz 3 with maximum benefit happening on the fourth trial (not shown).

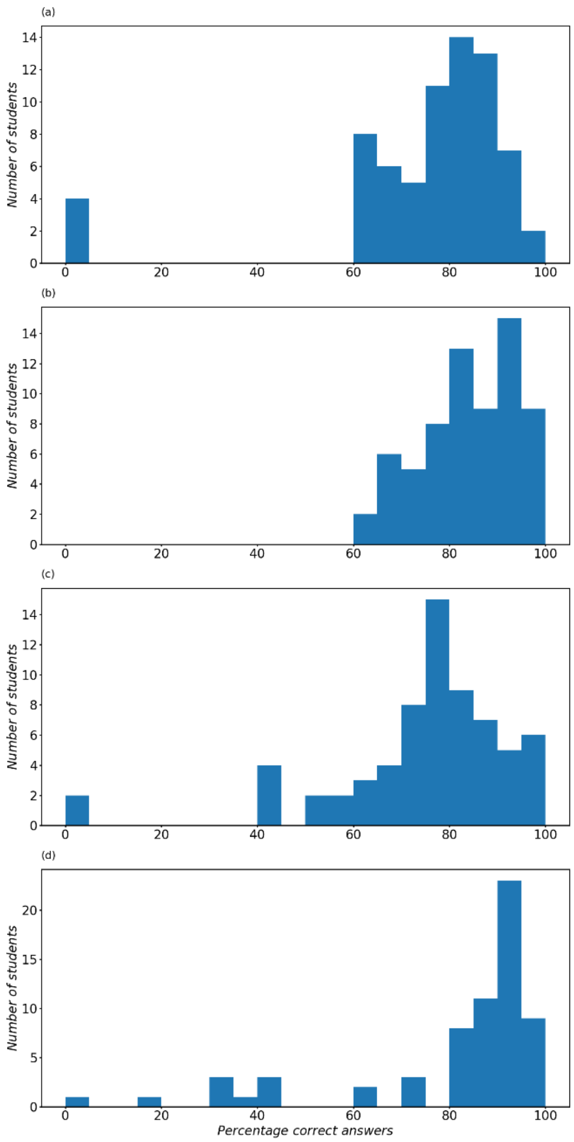

Figure 2 Histograms of student results for quizzes 1-4 (a-d).

Digitization in education tends to be promoted because it saves human resources and can be more accurate than a human teacher. If the 82 students complete four quizzes on average twice with 250 items total and the correction takes 10 sec/item, this means more than 3 weeks assistant time. This is comparable to drafting 1000 multiple choice items manually, which takes another two weeks to complete assuming that it takes 5 min to write a potentially flawed item with appropriate keys and distractors.

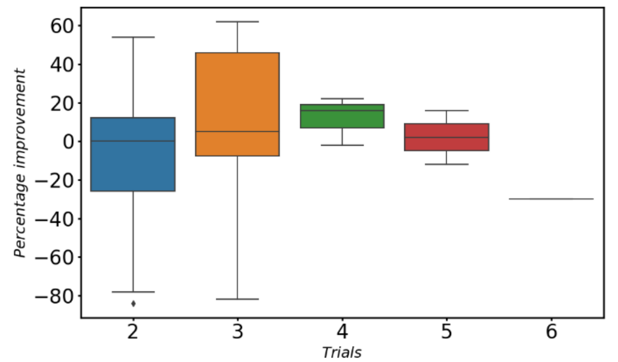

Figure 3 Boxplots of percentage improvement compared to the previous trial for students retaking quiz 4.

Lessons learned, conclusions

Authoring and correcting new multiple choice questions in chemistry does not require a human being. This process can be deconstructed into elementary steps and these steps can be efficiently mimicked by rule-based and probabilistic computational methods. It is possible to activate students that are only involved in vocational training very early in a course. For increasing the number of retrievals, it may be better to provide fine grained daily quizzes, but administrating quizzes in CANVAS is still not automatic and requires high level understanding of QTI format. In this study, I provided further evidence that repeated retrieval enhances learning.

Notes

The developed multiple access questions are freely available in the CANVAS system.

References

1. H. L. Roediger, 3rd, A. C. Butler, The critical role of retrieval practice in long-term retention. *Trends Cogn Sci* **15**, 20-27 (2011).

2. V. I. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Soviet Physics Doklady 10, 707 (1966).