Summer 2009

Constructed Response Answers in a Multiple-Choice Universe

Carl W. David
University of Connecticut, Carl.David@uconn.edu

Constructed Response Answers in a Multiple-Choice Universe

C. W. David*
Department of Chemistry
University of Connecticut
Storrs, Connecticut 06269-3060
(Dated: August 24, 2009)

I. ABSTRACT

A scheme is introduced which allows computer readable multiple choice forms used in traditional examinations to be employed for constructed response items.

II. INTRODUCTION

It behooves us to rigorize the testing procedure. We are faced with bigger classes, less help, and a growing demand for accountability concerning the product we are producing. Since almost no one denies that multiple choice (MC) testing is *a priori* flawed, it makes sense to examine properly using "constructed response" (CR) items for questions whose answers are numeric.

The main problem with such "constructed response" items is that they required hand grading, something inappropriate in this time of lessened resources and increased demands. A scheme for using machine gradable multiple choice forms for machine grading of constructed response items would seem to solve two separate problems, the hand grading problem of CR items, and the complaints about MC testing in a world in which nothing, post college, is multiple choice!

*Electronic address: Carl.David@att.net

III. THE SCHEME

A typical multiple choice (MC) question in freshman chemistry might be: "How many hydrogen atoms are contained in 35.0 grams of hydrogen gas?

1. $\sim 1 \times 10^{25}$
2. $\sim 2 \times 10^{25}$
3. $\sim 2 \times 10^{-25}$
4. $\sim 7.90 \times 10^{25}$
5. none of the above

"

Consider the same question in the herein newly proposed format:

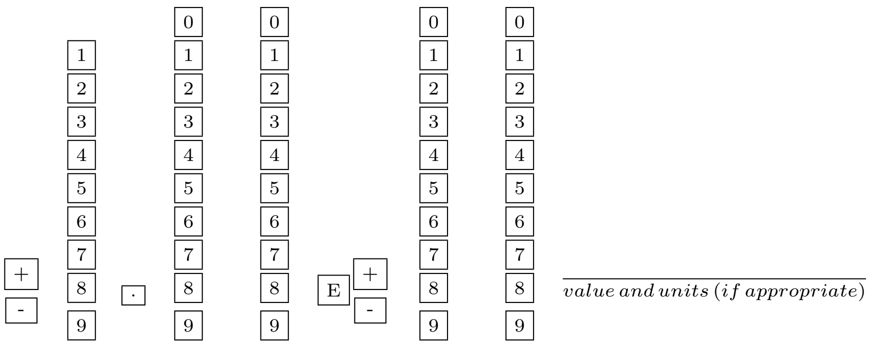
How many hydrogen atoms are contained in 35.2 grams of hydrogen (atomic mass of H = 1.008 $\frac{grams}{mole}$)?

1. In this and the following six questions report your answer in the form $ox.yz \times 10^{mn\ell}$ (remember that the initial sign is important!), o=:
   (a) o = +
   (b) o = -
2. Report your answer in the form $ox.yz \times 10^{mn\ell}$, x=:
   (a) x = 5
   (b) x = 1 or 6
   (c) x = 2 or 7
   (d) x = 3 or 8
   (e) x = 4 or 9
3. Report your answer in the form $ox.yz \times 10^{mn\ell}$, y=:
   (a) y = 0 or 5
   (b) y = 1 or 6
   (c) y = 2 or 7
   (d) y = 3 or 8
   (e) y = 4 or 9

Typeset by REVTEX

4. Report your answer in the form $ox.yz \times 10^{mn\ell}$, z=:
   (a) z = 0 or 5
   (b) z = 1 or 6
   (c) z = 2 or 7
   (d) z = 3 or 8
   (e) z = 4 or 9
5. Report your answer in the form $ox.yz \times 10^{mn\ell}$, m=:
   (a) m = +
   (b) m = -
   (c) not applicable
6. Report your answer in the form $ox.yz \times 10^{mn\ell}$, n=:
   (a) n = 0 or 5
   (b) n = 1 or 6
   (c) n = 2 or 7
   (d) n = 3 or 8
   (e) n = 4 or 9
7. Report your answer in the form $ox.yz \times 10^{mn\ell}$, ℓ =:
   (a) $\ell = 0$ or 5
   (b) $\ell = 1$ or 6
   (c) $\ell = 2$ or 7
   (d) $\ell = 3$ or 8
   (e) $\ell = 4$ or 9

   As a safety measure, so that if the scantron form is filled out incorrectly there's a record of your answer, write your answer here twice, once in the boxes, and once written in standard format (on the right).



(end of example)

I have chosen a scheme in which 7 responses are required items for each question which makes the constructed total answer worth seven more than an individual MC item. Alternatively, one could group all the constructed response items into a part whose answers were recorded on a separate machine readable form.

Of course, one could have, if publishers and digital scanners coöperated, 10 item multiple choice forms which would make the CR questions even better. The current scheme gives a form of partial credit for responses which are wrong in some significant detail, but, for example, have the correct sign for the power of ten.

We have earlier had a scheme for entering constructed response items into examinations, as pioneered by the mathematicians (I believe). The bottom part of our example shows this scheme. I've used it when erasure problems (cheating) threatened to inundate me. Unfortunately, there aren't scanners widely available for handling this kind of input.

Figure 1 shows an implementation programmed for Computer Assisted Testing which has been used locally by hundreds of students successfully. The pull down menus allow construction of an answer in the range $-9.99 \times 10^{-99} < 0 < 9.99 \times 10^{99}$ which covers all realistic cases in chemistry except zero itself. (An alternate form with a zero choice in the first digit was used then, but that was a dead giveaway. The choice 1 to 9 in the first digit insured proper scientific notation, hence its use in most questions.) In both cases, there are no hints to encourage "testmanship". Rather, the older spirit of being able to do problems completely, from beginning to end, is exemplified.

IV. DISCUSSION (ACTUAL USAGE)

In the above cited (example) question, the results were: 46 students correct out of 93 students on the constructed response question and 51 correct on the multiple choice version. In lecture and recitation, the (above noted) hydrogen based question had been used as the example (with numbers unchanged), so many students really had already seen it (those that attended, ∼60).

On an alternative problem, concerning aluminum reacting with oxygen (limiting reagent problem), the students scored 12 correct on the constructed response question, and 46 correct on the multiple choice version (again out of 93 papers submitted) with the starting amounts of aluminum and oxygen reversed.

Contrary to normal "educational experiments", all things really were equal here, with every student dealing with both questions at the same exam, under identical conditions.

Since this paper was submitted for refereeing, the scheme has been used for an entire summer semester. The students hated it; the grades were substantially lower than one would have expected if standard hand grading schemes had been used. However, the lack of controls in the summer semester "experiment" makes the results less convincing than the aforementioned tests, which were experimentally valid.

V. DISCUSSION (MULTIPLE CHOICE CRITICISMS)

Multiple choice questions (items, in testing parlance) suffer from so many defects that it is hard to know where to start in criticizing them.

The major criticism of them comes down to the idea that once out of school, students will never, ever, encounter a multiple choice problem in the real world of science, engineering, or anything remotely connected to the reality. Unless they take other examinations, the expertise they've developed in taking multiple choice examinations in various disciplines will serve them for naught.

Worse, from a teaching point of view, multiple choice items on an examination are intended to deceive students, something antithetical to the röle that most of us choose in our teaching/learning environment. Consider the following quotation:

"An important consideration in writing multiple choice questions is that the distracters are plausible. Poorly written distracters could easily cue a student to the correct answer [1]." Even the term "distracters" is upsetting. Intellectually, we are not in the business of distracting students into making mistakes; it is not our function, duty, or even desire. Our hypocrisy in attempting to do so is tantamount to fraud.

"Multiple-choice questions have been seriously maligned in recent years. While some of the criticism of the overemphasis on the results of standardized tests is justified, most criticisms of multiple-choice items are unreasonable. The most often heard criticism of this item type holds that multiple-choice items can be used only to measure low-level, rote-memory-type objectives. This is clearly incorrect. A good item writer can create question of any type to measure higher-order thinking skills. Essay or constructed response items have no inherent advantage over multiple-choice questions in this regard. Now, clearly, multiple choice questions are inadequate for some areas, such as evaluating how well a student can produce an original piece of writing or perform a solo or paint a picture of fix a carburetor. For virtually all other type of objective, multiple-choice items can server very well to evaluate student progress. This is not to imply that these questions should provide the only information related to student performance in any area. [2] ". This is utter nonsense. We note that, concerning this opinion, rote learning is not an issue in constructed response questions devoted to numerical answers.

Here is an example of the problems with multiple-choice items: Consider the following multiple choice question [3]:
Show that

$$\frac{1+\sqrt{5}}{2} = \sqrt{\left(\frac{3+\sqrt{5}}{2}\right)}$$

This identity came up in a molecular orbital problem, and we expect any high school graduate to be able to show that the statement is true.

Now, before continuing, I want to make sure you understand what I mean by "be able to show that the statement is true". I do not mean plucking something out of a multiple choice list. I mean actually writing lines of (what?) algebra/mathematics/arithmetic/whatever and arriving at a proof that the statement is true. A multiple choice question of the type:

$$\frac{1+\sqrt{5}}{2} =$$

1.
$$\sqrt{\left(\frac{3+\sqrt{5}}{2}\right)}$$

2.
$$\sqrt{\left(\frac{3-\sqrt{5}}{2}\right)}$$

3.
$$\sqrt{\left(\frac{5-\sqrt{3}}{2}\right)}$$

4.
$$\sqrt{\left(-\frac{3+\sqrt{5}}{2}\right)}$$

5. etc.

tells us almost nothing when the student gets it wrong, especially if we've been extraordinarily devious in constructing "distractors". Even getting the "right" answer assures us of nothing, since guesses are permissible in this environment! Worse yet, the calculator equipped student can evade the question completely, evaluating $\frac{1+\sqrt{5}}{2} = 1.68103\ldots$ and then evaluate each of the choices until s/he finds the "correct" one [4]. Multiple choice examination doesn't test what we think its testing, and such tests' results are useless from the point of view of knowing what a student knows!

In a piece for the New York Times, Ian Ayres quotes Dixit and Nalebuff's new book *The Art of Strategy* to give a multiple choice question's answers without the question itself, which they solve using a variant of "game theory". Thus, they give the choices

1. 4π sq.inches
2. 8π sq. inches
3. 16 sq. inches
4. 16π sq. inches
5. 32π sq. inches.

. With just a dash of *post hoc ergo propter hoc* reasoning they then work out what the question should have been and what the "right" answer was. Suffice it to say that their point, that thinking about choices from alternative points of view as if in a game, enhances our argument that multiple choice testing doesn't test what we desire to test!

VI. DISCUSSION(2)

A form of this manuscript was submitted to the Journal of Chemical Education during the last semester I was actually teaching. It appears to have been lost, since I've not heard from the journal in over 2 months.

I am therefore publishing it here, in the hopes that this last appeal for sanity in a world whose standards appears to be crumbling will result in someone, somewhere, listening.

The continuing decay of academic standards, coupled with a societal aversion to requiring hard work and effective study for meaningful grades continues apace. There is little question in my mind that multiple choice testing is one of the contributors to this decay. Computer Assisted Testing, which I espoused extensively during my academic tenure, is not likely to appear on the scale I thought required to rectify the situation. Perhaps this contribution will in some small way contribute to reversing the decline which imperils our future.

VII. DISCUSSION (DISCLAIMER)

There is no possibility that the method outlined here is unique. However, multiple searches on-line have proven fruitless to check for antecedents of this method. I apologize in advance if I have slighted someones prior claims.

It should be noted that the College Board (in its SAT tests, at least) have offered a form of constructed response item similar to the tableau shown at the bottom of the example (*op cit*), but they publish their own response fill-in forms, and therefore have much more latitude than normal teachers dealing with standard (digitec) forms.

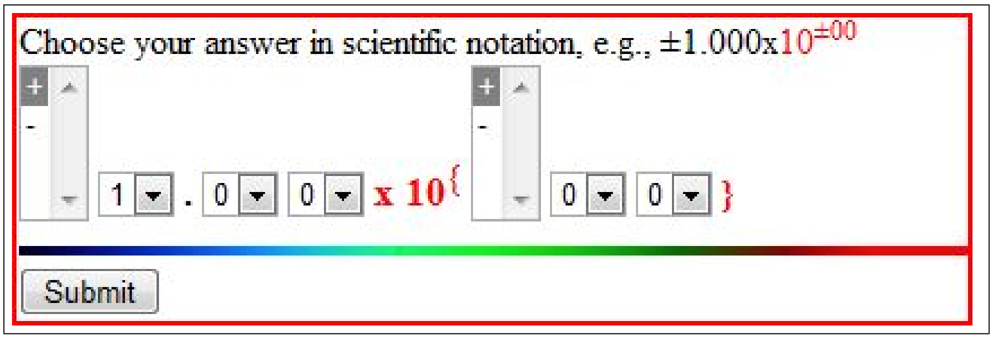


FIG. 1: An answer input scheme used in Computer Assisted Testing Mode

VIII. FIGURES

[1] K. Woodford and P. Bancroft, ACM Intl Conf. Proc. Series; Vol 106, Proc. 7th Australasian Conf. on Computing Educ., (2005) 42, 109-116
[2] Gary D. Phye, The Handbook of Classroom Assessment, Academic Press, page 276 (1997)
[3] This question actually came about from a Hückel computation done in two different ways by two different authors. We expect our students to be able to show that the two versions are actually equal to each other.
[4] Testmanship in the extreme