Metacognition Experiments as Classroom Assessment by Teachers

In my class for teachers, my first attempt to train beginning teachers in metacognition experiments as a classroom assessment activity began as a dismal failure. The main responses were confusion, bafflement, and frustration. Things changed when I provided examples and when my class decided to plan and conduct their own metacognition experiment. I posted copies of some metacognition experiments from college freshmen testing their own learning. My prospective teachers then decided to design and carry out a metacognition experiment on themselves as a class, as if it was an experiment activity in a class they were teaching. They spent about 20 minutes of planning the experiment below. Later, we wrote up the results:

The Class Metacognition Experiment (BIOL 455 "*How Students Learn*", Fall 2012):

Learning with folded lists in comparison to a standard study method.

What we did (Introduction). First, the class chose to test a learning method that their students would not already know. Next they developed an experiment to test whether they would learn something better using the new method or with standard methods used by many high school students. They carried out the experiment, and we evaluated the results.

Methods.

Learning task: letters with sounds in the Cyrillic (Russian) alphabet.

Time learning: 30 minutes on each of two learning tasks (1st half of alphabet vs. 2nd half).

Learning methods: a comparison of two methods, (1) drawing in a folded list (see the instructions posted) versus (2) any common method they wished to use. At least 10 letters and sounds had to be learned using each method. The folded list was for letters in the first half of the Cyrillic alphabet (from the Wikipedia Cyrillic Script table), and their other method for letters in the second half of the alphabet. All students used the folded list method first, and then their alternative.

"Standard learning method": "Standard" learning methods chosen included (1) flashcards (the two who had 90% or above with the "Standard Method", plus a third whose flashcard learning did poorly), (2) writing "over and over and over" (one who had a 65% with the "Standard Method"), (3) drawing letters in the air multiple times, and (4) simply staring at the letters and trying to remember them.

Test for effectiveness: to reproduce their letters with their sounds at the beginning of the next class period (two weeks away, as there was a holiday period after the day of experimental design). The measure would be the percentage of (i) correct Cyrillic letters with (ii) their correct sound represented through printed letters in English.

Results.

Scores were higher with the folded list (Table 1). The median score was about 90% using the folded list, and about 30% with standard methods.

PERCENTAGE CORRECT

METHOD	0-9%	10-19%	20-29%	30-39%	40-49%	50-59%	60-69%	70-79%	80-89%	90-99%	100%
"Standard" Method	**2**	**2**	**2**	**2**		**1**	**1**			**1**	**1**
Folded List							**1**	**3**	**2**	**1**	**5**

When using their "Standard Method", participants reported growing boredom over time with the learning task. One participant noted that when using the "Standard Method", she felt that it was ineffective and that she was learning nothing, and so she became bored and frustrated, and began paying attention to distractions and began snacking. Many were unfocused and multi-tasking (with, for example, the television on or Facebook open) when they were trying to learn with their standard method. Finally, at least one participant applied the folded list method incorrectly.

Discussion and Conclusions.

Based on the numerical results, it appears that the folded list method was more effective. However, participants identified biases and problems. First, there was an "order effect". The task became more boring over time, and so the "Standard Method", which all used second, was applied with less interest and focus than the folded list. Second, the learning tasks were not equally challenging. Some letters are similar to other letters, and some letters are similar to those in the Greek, Hebrew, or other alphabets that some participants already knew. Third, it was clear that the choice of alternative method was important—for this task, flash cards were as effective (or more effective) than folded lists, though one flash-card user noted that **shapes** were reproduced more accurately from her folded-list studying. *(The only time I, personally, still use flashcards is to learn faces and foreign languages.)*

If the goal of the experiment was to evaluate folded lists as learning tools, the experiment had too many uncontrolled variables and biases to answer the question. However, as an example of the potential value and pitfalls of trying new methods of teaching or learning, it worked well. No one in the class felt that this test provided broad evidence about the usefulness of folded lists as a learning tool. However, many or most felt that for this task, in **this** situation, folded lists worked better than most of the alternatives they chose.

Class members recommended a better learning task in a redesigned experiment: using an alphabet that was entirely unfamiliar. For example, the Korean alphabet would have been unfamiliar to all class members. Second, they recommended that the order effect be corrected, by alternating time with each method, or perhaps assigning half of the participants to begin with one method and the other half to begin with the other method. Content that was less boring or more relevant to them would have helped to alleviate the boredom. It was noted that the folded list method made it more difficult to be **un**focused, because the exercise requires visual attention while sketching the Cyrillic letters when looking at the letters sounds and (alternating) writing the sounds when looking at the Cyrillic letters. The class concluded that designing an experiment like this in their classroom, using class participation, and then analyzing it (just as we did) to identify biases and flaws, and then doing it again (which we did not have time for) was something they would consider trying in their classroom.

This metacognition experiment applied in a class.

This is an experiment that could be modified for use with a class as they learn and practice a new method. The learning task could be content needed in the class, which would at least help students believe that the learning was useful. Bias due to the difficulty of the material could be controlled by having each method applied on content that was equivalent in difficulty, and with each specific set of content studied by half of the class with one method and half of the class in the alternative method. Order effects could be controlled by assigning half of the class to

start with one method and half with the other method. Would students lose valuable study time if they compared two methods of studying? Not if both have evidence of effectiveness, or if one is a method that students are **already** choosing and the other has evidence of effectiveness. Drawing to learn and retrieval practice, two of the core parts of folded lists, both have research-based evidence of effectiveness for learning. Students would likely be no worse off, and potentially would gain experience and information about a new study method by doing this experiment.

Am I recommending that teachers use their students as guinea pigs for anything they want to learn? Not at all! Our class had a lengthy discussion about whether doing metacognition experiments with their students would be ethical and appropriate. The full discussion was complicated, but the general view at the end of our discussion of the ethics was that teachers often try new methods of teaching or ways to have their students study or learn. Having them do a metacognition experiment is just one way to do it. The recommendation was that if you, as a teacher, are interested in having students try a new or better learning tool, then consider doing so in a way that gives you and your students useful information. Our class discussed whether or not they would even talk with students about experimental design, sources of error and bias, and the need to be cautious about conclusions, and then have the students help design the experiment (or have the students design the experiment fully, with coaching from the teacher). Teachers could talk with students about ways to practice and keep testing a method themselves to see if, when, and how well it might work. As teachers, of course we have to be careful and ethical—no one would enjoy having a student to tell a parent "my teacher is experimenting on me" followed by a valid complaint to a principal or superintendent—your students and your supervisors need to know and understand what you are doing **in advance**. We all need appropriate review and approval for exercises in our classes. If it is clear to you, your class, and your supervisors that your goal is solely to help your students understand and improve their own learning and studying, then this could be a useful classroom assessment activity.

Overall, classroom assessment-motivated experiments on one's own teaching and on one's students' learning have two goals: (A) to test approaches that might improve teaching, and (B) to help students understand their own learning as they experience different ways of studying.

The experiment that my class did had some problems, but it achieved most of our goals. It was (A) a test of an approach that might improve teaching and learning. It did (B) help participants think about learning as they experienced two different ways of studying. In addition, (C) it stimulated a good discussion of biases and flaws in experimental design, and students were very interested in the discussion. Completely unplanned was (D) a discussion of the ethical issues.

This short, simple experiment provided an interesting conclusion: the folded list worked better for this task done in this way. Logical conclusions have been common when my freshman students do metacognition experiments, and, equally, my students are often not very confident, just based on the first set of results, that the conclusion was valid. In others, we (and my students) have retained skepticism, just as we should. As a teacher, I care about retesting with unbiased methods, so as a teacher I would be likely to try the experiment again after modifying the methods to solve some of the problems we noticed. I would choose important material to study. I would try to make sure that the level of difficulty was similar for both. I might adjust the practice time, and I would have students record how interested and focused they were as they applied the two different methods.