An Improved Design for In-Class Review

By E. Jane Maxwell, Lisa McDonnell, and Carl E. Wieman

We present the theory and implementation of a review strategy based on testing rather than lecturing. We also show the results of a beginning-of-course review using the format of a two-stage examination, in which students complete a set of questions individually, then again as a group. This format offers several benefits compared with the typical lecture review: (a) students engage with the review topics much more deeply and more accurately gauge their own preparation; (b) students receive immediate, corrective feedback from their peers and clarify their understanding through discussion during the group stage; and (c) the instructor receives detailed information on students' background understanding that can be used to tailor instruction. These proposed benefits are supported by the improved performance of groups during the second stage and by student opinions collected by survey several days after the review activity. The two-stage review therefore serves to both diagnose and remediate deficiencies in background understanding, leaving students and instructors better prepared for the course.

A substantial amount of class time is spent reviewing material from previous courses or the previous class meeting. It is common for instructors to give review lectures that can occupy some hours at the beginning of a term and/or 5–10 minutes of review at the start of each individual class. The intention in these review sessions is generally to focus students' attention, clarify their understanding of "previously learned" concepts, and prime them to connect this prior knowledge to the new topics or problems that will follow.

At the University of British Columbia, we had trained observers (Wieman, Perkins, & Gilbert, 2010) watching the attention of students during classes. They found that students' attention largely switched away from the course material during this lecture review. Not only did they get little from the review, but they also required additional time to reengage when new material was introduced.

Although we were surprised to discover that this time-honored practice was not very effective, once confronted with the data, it was easy to understand why. There is a well-established result from cognitive psychology that familiarity with a topic makes people erroneously believe they understand/have learned it (Willingham, 2003). A less-studied result, but one that can be confirmed at many faculty meetings, is that a person will quickly become bored and disengaged if lectured on something they believe they already know. The combination of these two effects means that few students who have previously heard about the topic being reviewed will pay attention and benefit. Those students who have not heard about the topic are very unlikely to learn much from such a brief review, and thus it will also be of little value to them.

Guided by the cognitive psychology literature on the benefits of testing for retention, self-evaluation, and learning (Bjork, 1994; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006), we tried an alternative strategy. We replaced all review lecturing with *problems* that covered the review topics. The students solved these problems in class and responded using clickers. Working on the problems activates thinking about the relevant material and forces students to test their understanding. If they get a question wrong, and often even if they don't, they are then primed to think more deeply about the follow-up discussion and ask questions to understand better. Also, if there are test questions that everyone in the class answers correctly, that is immediately obvious and the instructor can move on, leaving more time for topics where many have difficulty.

When we and others tried this approach, it seemed to work better—the students were more engaged and the instructor was better informed. However, we wondered about alternatives that might be even more effective, taking advantage of the targeted and timely feedback that can be provided to a student by a classmate during collaborative learning. This sort of feedback is difficult or impossible for an instructor to do in a large course but seemed particularly well suited to review, where a student might only need a brief reminder of some terminology or result they had forgotten to solidify their understanding. On a small scale, where one is spending a few minutes reviewing a previous class, peer instruction (Crouch & Mazur, 2001) seemed like a reasonable solution. In that method, students click in answers and then, if a significant fraction have the question wrong, students discuss with their neighbors and revote.

In considering the more substantial review needed at the start of a course that relies on material covered in previous courses, we decided to try a more extensive option, specifically a "two-stage review." This uses the format of a two-stage exam (Cortright, Collins, Rodenbaugh, & DiCarlo, 2003), in which students answer a series of review problems individually, turn in their answers, and then redo the same problems in a group of three to five students.

Implementation

We first tested this approach in a third-year course in analytical chemistry offered at the University of British Columbia. The class consisted of ~90 chemistry majors and honors students, and ~25 students enrolled in the Bachelor of Medical Lab Science (BMLS) program. All students had completed a common prerequisite course in analytical chemistry, but the emphasis of that course can vary from term to term. In addition, students enrolled in the BMLS program are not specifically required to take a physics course on electricity and magnetism, despite the fact that a basic knowledge of electrostatics and circuits is required to understand the operation of many analytical instruments.

This variation in students' background knowledge was a major concern. As such, the two-stage review activity seemed a particularly appropriate means of addressing the following four goals:

1. Capture a snapshot of students' understanding of key concepts that could be used to tailor lectures and activities.
2. Communicate to the students our expectations for their background understanding and provide students with immediate feedback on their level of preparation.
3. Stimulate interactions between the chemistry and BMLS cohorts.
4. Engage students in active participation on the first day of class, which previously had been relatively unproductive.

The instructor and other members of the course team (teaching assistants, etc.) met several weeks before the first class to identify the topics and key concepts to include in the review activity. This list of topics was then used to develop a set of 18 multiple-choice questions (see the appendix, available at www.nsta.org/college/connections.aspx). The questions were targeted at a "quiz" level rather than "final exam" level, in that they did not ask the students to integrate multiple concepts or present the concepts in a novel context. Arriving at a correct solution generally required students to define key terms (four questions), make simple predictions, or identify true/false statements based on key chemical or physical concepts (14 questions) that were directly addressed in the prerequisite courses. Students were expected to know these concepts and terms, which were to be used and expanded upon in the course.

On the first day of class, the instructor spent ~40 minutes (of the 80-minute class period) introducing the course syllabus and delivering a brief introductory lecture—all of the activities that typically took place in the first class. The instructor then introduced the review activity, with a great deal of emphasis on the fact that it was *not* graded, but an opportunity for review intended to benefit the students and the subsequent instruction.

To ensure mixing of the two cohorts, students were preassigned to groups of five, with one BMLS major and four chemists per group. Students reorganized into their groups before the start of the activity, guided by a group assignment sheet and seating chart projected at the front of the room. In contexts where creating heterogeneous groups is not a priority, this process could be simplified.

Each student received a question sheet and a standard Scantron multiple-choice answer sheet to start the review. Students spent 15 minutes (half of the time remaining in the class period) working individually to answer as many of the questions as possible. Once the Scantron sheets were collected, each group received an Immediate Feedback Assessment Technique (IF-AT) multiple-choice form (www.epsteineducation.com). IF-AT forms are multiple-choice answer sheets with a top layer of scratch-away material, similar to a scratch lottery ticket. For each question, the student group scratches away the top layer of their chosen answer (A–E). If their answer is correct, the revealed area will contain a star. If no star is visible, the group continues to try until they scratch off the correct answer. Scoring is typically allocated as 4 points for a correct answer on the first attempt (only one square scratched), 2 points for the second attempt, 1 point for the third attempt, or 0 points if more than three attempts are required. This creates an incentive for groups to discuss the question until they reach a new consensus. In our case the activity was not graded, but we did explain the scoring system to the class, and some groups chose to tally their "score" regardless. Groups worked for 15 minutes to complete the second stage of the review activity.

Although we had been concerned about the potential reaction to giving a test-like activity on the first day of class, we found that the students' attitudes were generally positive and enthusiastic. There was minimal grumbling, and students worked diligently both individually and in groups. During the second stage, the groups were engaged in animated discussion and appeared invested in finding the correct answer (as evidenced by frequent celebratory outbursts when a scratched answer revealed a star). A few individual students required prompting from an instructor or teaching assistant to sit so they could interact with their group, but this is not unexpected given their lack of experience with activities of this type.

Because students had only 15 minutes to answer 18 questions, they were reminded that there was no penalty for not finishing, but the majority of individuals and groups completed the questions. Of the six (out of 23) groups who did not finish on time, two groups voluntarily stayed late to finish, indicating how engaged students became during this process and how seriously they took the exercise.

Analysis of the individual and group answers provided us with some very valuable insight into our students' background understanding of important prerequisite topics. Figure 1 shows the individual and group scores.

The questions can be roughly grouped into three categories. In the first category, the majority of individuals chose the correct response, and the group scores approached 100% (e.g., Questions 1–6, 10, 14 in Figure 1). This material needed no further review; individuals who were initially incorrect generally received corrective feedback from their group, and the one or two groups that did not choose the correct answer on the first attempt realized their error from the IF-AT card.

In the second category, individual responses varied, but most groups chose the correct answer (e.g., Questions 8, 11, 15, 16). These questions we flagged as requiring some additional review. We did this by providing students with remedial tutorials via the course website.

The third category comprised questions for which less than 60% of the groups chose the correct answer

FIGURE 1

Individual and group scores (percentage answering correctly) for 17 questions of the two-stage review (one question was excluded from the analysis because of ambiguous wording). Group scores were calculated as the percentage of groups choosing the correct response on their first answer attempt, as indicated by the number of squares revealed on the Immediate Feedback Assessment Technique card. The full text of the multiple-choice questions is provided in the appendix (available at www.nsta.org/college/connections.aspx).

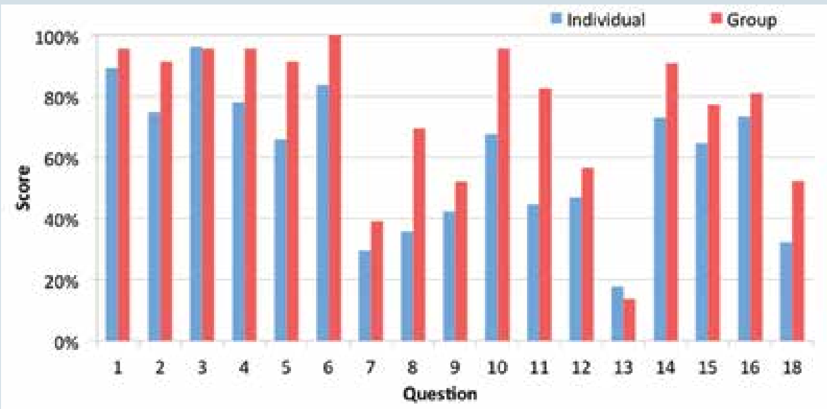
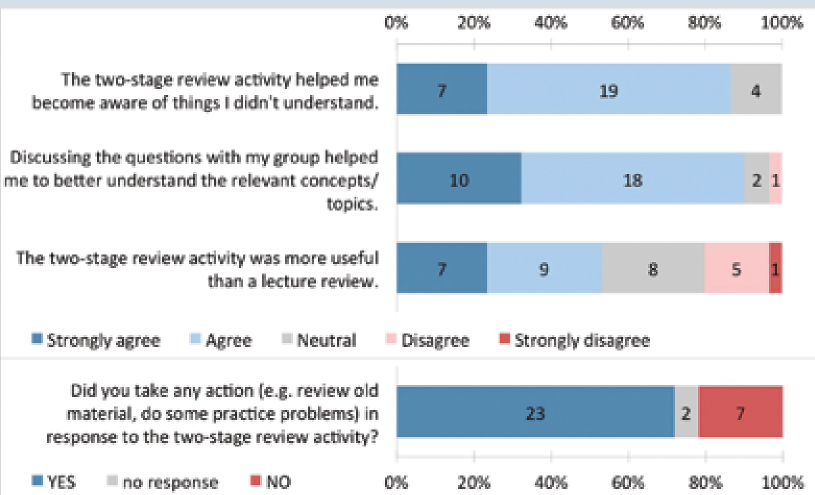


FIGURE 2

Student survey responses regarding the two-stage review activity. The number on each bar indicates the number of individuals choosing that response. Yes/No responses to the last question were coded from students' open-ended responses to the question shown.



on their first attempt (e.g. Questions 7, 9, 12, 13, 17). When we examined these questions more closely, we found that the most commonly chosen incorrect answers revealed major misconceptions, some of which were known to us and others that we had not previously realized. This category was particularly important in indicating where to invest additional time in the course.

We posted on the course website an answer key that included explanations of both the correct answers and commonly chosen incorrect answers. In future implementations, we plan to provide individual feedback to students on the basis of their answer choices.

We also used the same two-stage review format in a second-year genetics course. It worked much the same as for the analytical chemistry course. The genetics students were asked to complete a brief online survey four days after the review, and 32 of the 50 students did so. The survey consisted of three limited-choice questions (Figure 2) and three open-response questions. As shown in Figure 2, the majority of students agreed with our assessment of the benefits of the two-stage review format. Particularly significant is the fact that 72% of the respondents indicated that they had taken some action (e.g., reviewed old material, done some practice problems) in response to the two-stage review activity.

In response to an open-ended question asking what they liked most about the activity (Table 1), students frequently mentioned the group discussion, the opportunity to gauge their preparedness for the course, and the opportunity to clarify their understanding. For example, one student responded:

> I liked that the activity was put under test conditions at first, to realize how much I personally know. Then the group portion allowed me to understand the material better by talking in groups and teaching and learning the material to our group members.

A final open-ended question asked students what they liked least about the review activity. There were few negative responses, with no complaint standing out as common.

Discussion

Although we have not compared what students learn in the sort of active review presented here with the more traditional review lecture, we have considerable indirect evidence of benefits. First, there is a qualitative difference in the level of student attention and engagement, and as noted from the survey data, most students subsequently went back to review material covered on the review test. Second, as noted, it is far more informative for the instructor (and students) as to what students do and do not know and therefore enables more efficient use of class time. The instructor of the previously mentioned genetics course estimates that the results of the two-stage review allowed her to reduce the time spent on review topics by two full class sessions. Third, on all but one of the questions of the two-stage review, the groups did better than the individuals, and on many questions essentially all the groups got the correct answer. We know from recent research on two-stage exams (Gilley & Clarkston, 2014) that when tested later, nearly all individuals learn the answers chosen by their groups during a two-stage exam. Finally, there is an overwhelming body of research showing that active learning methods achieve more learning than traditional lecture (Freeman et al., 2014); that result should apply to this situation. (Similar benefits for similar reasons would likely apply to replacing the common end-of-class summary with an active exercise; for example, students briefly listing and sharing the three to four main points covered in class that day.)

We also have several anecdotal reports of other instructors who have switched from review lectures to these two-stage reviews. They are universally enthusiastic.

TABLE 1

Repeated themes from the students' responses to open-ended survey questions.

What did you like most about the two-stage review activity?	*N*
Discussing/working with group	20
Gauge ability/feedback/review	11
Clarify/improve/correct understanding	5
Scratch cards	4
Doing individual stage first	3

Conclusions

Managing the variation in students' background understanding is one of the greatest challenges for successful teaching. In an ideal scenario, review should aid in both diagnosing and remediating deficiencies in preparation and understanding. We have presented a strategy for active review by testing plus follow-up peer discussion that draws on work in cognitive psychology to provide a more effective review. This strategy provides both students and instructors with an accurate gauge of student preparation. Students receive immediate feedback and reinforcement from the collaborative process of peer discussion and teaching. ■

Acknowledgments

This work was supported by the University of British Columbia (UBC) through the Carl Wieman Science Education Initiative. We thank Dr. Dan Bizzotto and Chad Atkins for their *help in implementing the activity. We acknowledge many useful discussions with Sarah Gilbert and the UBC Science Teaching and Learning Fellows.*

References

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 182–205). Cambridge, MA: MIT Press.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education, 27*(3), 102–108.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69,* 970–977.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, USA, 111,* 8410–8415.

Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching, 43,* 83–91.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 19*(5865), 966–968.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Wieman, C., Perkins, K., & Gilbert, S. (2010). Transforming science education at large research universities: A case study in progress. *Change: The Magazine of Higher Learning, 42*(2), 6–14.

Willingham, D. T. (2003). Why students think they understand—when they don't. *American Educator, 27*(4), 38–41.

E. Jane Maxwell (ejmaxwell@chem.ubc.ca) *is a science teaching and learning fellow in the Department of Chemistry and at the Carl Wieman Science Education Initiative at the University of British Columbia, Vancouver, BC, Canada.* ***Lisa McDonnell*** *is a science teaching and learning fellow in the Department of Zoology and at the Carl Wieman Science Education Initiative at the University of British Columbia.* ***Carl E. Wieman*** *is a professor in the Department of Physics and Graduate School of Education at Stanford University in Stanford, California.*