Mastery learning in a large first year physics class

Paul Francis, Cristina Figl and Craig Savage, Centre for Learning and Teaching in the Physical Sciences, The Australian National University, Australia
paul.francis@anu.edu.au, cristina.figl@anu.edu.au, Craig.Savage@anu.edu.au

Abstract: *In 2009 we tried an experiment in our large core first year physics course: we introduced mastery learning. The basic idea behind mastery learning is that any student can learn anything well, but that it takes some students much longer than others. We should therefore let students proceed through a course at different speeds, while insisting that they totally master each section of the course before moving on.*

The students have to get over 80% in each homework assignment before they are allowed to take the next one. They are, however, allowed to take different versions of each assignment multiple times until they reach this threshold. At the end, the weaker students would have covered less content than the strong ones, but they should have fully understood whatever they did. In the laboratory component, the students were assessed in each experiment against a set of lab mastery goals. The students could pass the lab component only if they have mastered each of these goals at least once.

Did it work? Logistically it worked very well, somewhat to our surprise. There were a number of striking unexpected benefits: students did more work, complained less about the workload, asked for help more often, and showed an improved ability to solve questions first time around. Gains in student conceptual understanding were much improved, but this may be due to other innovations introduced in the course. Examination performance, however, did not improve, even on the most basic material. Students could do the problems when given unlimited time and assistance from peers, but not in exam conditions.

Introduction

The motivation for this experiment came out of several frustrating observations. Marking exams was a depressing activity each year, as we realised how many of our students still had major problems with the most basic material. The lecturers of follow-on 2nd and 3rd year courses were reporting similar things: students seemed to get through several years of physics while still having major problems with foundational material. In the laboratories, for example, very basic skills like producing good graphs, handling uncertainties or describing the experimental method in adequate detail were still rarely practiced even after three years of explicit instruction. A related problem was reported by our tutors: students seemed rarely interested in the detailed feedback they were provided with. Students seemed to regard feedback on last week's assignment as "ancient history" – they would seldom even read it. As a result, they would make the same mistakes over and over again.

Perhaps, we thought, the problem lies in the very nature of a conventional lecture course. We typically proceed from topic to topic at a pre-determined pace. Most students do not fully master one topic before we go on to the next, or we would have to award everyone 100%, which is frowned on by university administrators. So a mediocre student would develop enough of an incomplete or shallow understanding in each topic to get a pass mark, but might go through their entire degree without having *fully* understood anything.

What could we do about this? The educational literature suggested one possible solution: mastery learning (Block, 1971). The basic idea (Carroll, 1963) was that any student can master any topic, but that it takes some students more time than others (typically by a factor of three). As formulated by Bloom (1968), mastery learning consists of two components. Firstly, the course material is divided into small sequential components (perhaps chapters in a textbook), and a student is not allowed to proceed from one component to the next until they have demonstrated mastery of the first component, typically by getting a mark of over 80% in some test. Secondly, students are given as much time as they need to master each component, and are also helped with personalised assistance. This personalised assistance comes partially from the teacher, partially from collaborative peer groups, and partially from the availability of a wide range of instructional materials. A closely related technique is the "Personalised System of Instruction" developed by Keller (1968). The main difference here is that the personalised help is provided by "proctors", who are typically higher year students.

There is a very substantial body of research, mostly dating from the 1970s, which claims that mastery learning leads to enormous educational gains (for reviews, see Guskey & Gates, 1986 and Kulik, Kulik & Bangert-Drowns, 1990). Curiously, however, this technique seemed to fade from both academic attention and use from the 1980s onwards (Cracolice & Roth, 1996; Davis & Sorrell, 1995; Eyre, 2007) , despite studies which continued to show its effectiveness. The reasons for this may have been political, but the literature demonstrates a number of possible pedagogical problems with this technique, the worst of which was student procrastination. Almost all users of these techniques have reported that many students respond to the lack of deadlines with almost infinite procrastination (Eyre, 2007 and many references therein).

Despite this, there is some evidence that the widespread use of computers is making mastery-based techniques easier to implement, and they may be slowly coming back into fashion (eg. Pear & Crone-Todd, 1999).

How did we implement mastery learning in our course?

Our reading of the literature suggested that mastery learning was worth a go. But how could we implement it in a large, traditionally timetabled first year class? Our class, PHYS1101, had an enrolment in 2009 of 230 students. The first half of the course covered Newtonian mechanics, while the second half covered electricity and magnetism. We actually tried two different implementations of mastery learning: one in our lectures (and associated homework), and a different one in our weekly laboratories. Human ethics approval was obtained for research in this class.

Lecture Implementation

We broke the lecture material into "levels", each of which corresponded to a chapter in our textbook (Chabay & Sherwood, 2007: level 1 covered vectors and momentum, level 2 covered force and impulse, etc). For each level, we created two mostly on-line tests, an "a" test and a "b" test. Once the students felt that they understood the material in one level, they would take the "a" test. If they got a score of at least 8/10 in this test, they were considered to have mastered that level, and could proceed to the next level. If not, they would receive automated detailed feedback on any wrong answers, and could seek help from tutors, peers, and if necessary at special drop-in tutorials. They then attempted the "b" test. Hopefully this time around, they would achieve 8/10, but if not, they got a third chance to master the level – they would write-up on paper all the questions they got wrong and hand it to their tutor for marking. If the tutor considered their write-up satisfactory, they were considered to have mastered the level, but if not, the tutor could set more work as required. To minimise procrastination, students were required to complete a minimum of one test per week, which might be an "a" test, a "b" test or a write-up. In lectures, we would cover whatever material the bulk of the students were up to, or any material which, based on scores on different test questions, many of the students were having trouble with. Students who were ahead of most of their peers would need to rely on the textbook. Each half of the course was treated separately – there were separate levels for mechanics and for electricity and magnetism. In each half of the course, there were seven "standard levels" plus three more "advanced levels" – more difficult questions requiring the synthesis of knowledge from across all the standard level material.

To minimise plagiarism, all questions in the levels were randomly generated, so no two students were asked the same set of questions, and different random numbers were used in individual problems so even when two students were asked the same questions, the answers were different. In addition, we set the students an exam which was worth only a token amount of marks. The exam only tested material from the levels they had mastered: different versions of the exam were given to different students depending on how much they had mastered. The exam results were used only to make sure that they had really done the levels themselves.

Laboratories Implementation

During the semester, each student had to complete eight labs, each three hours in duration. The labs were not changed from previous years, but a new marking scheme was brought in. Seven mastery goals were defined (preparation, record keeping, describing the method, showing the raw data, data presentation, uncertainties, analysis). Demonstrators would assess each student's performance in each lab against each of these goals on a scale of 0, 0.5 or 1 marks. The sum of these marks would constitute the student's laboratory grade. In addition, each student had to get a mark of at least 0.5 in each goal at least once during the course, or they would get no marks at all for the lab component. Half way through the course, extra help was offered to all the students who had failed to ever get a mark for one or more of the goals.

How did it work?

Lectures

Almost immediately, two things became apparent. Firstly, a small group of around 5 break-away students started moving rapidly through the levels, with some of them mastering a half a term's worth of levels within two weeks of the course starting. These students then spent a lot of time working on the advanced levels and continued to come to class throughout the semester.

At the other extreme, examination of the logs produced by our learning management system (WebAssign) revealed that around 20% of the class did the "1a" test almost as soon as it was posted (without waiting for it to be covered in lectures), got most of it wrong and then immediately took the "1b" test. They got similarly low scores second time around – typically making the same conceptual errors. When we became aware of this, we explicitly warned the class that the "a" and "b" tests covered the same material, and that they needed to pay attention to the feedback.

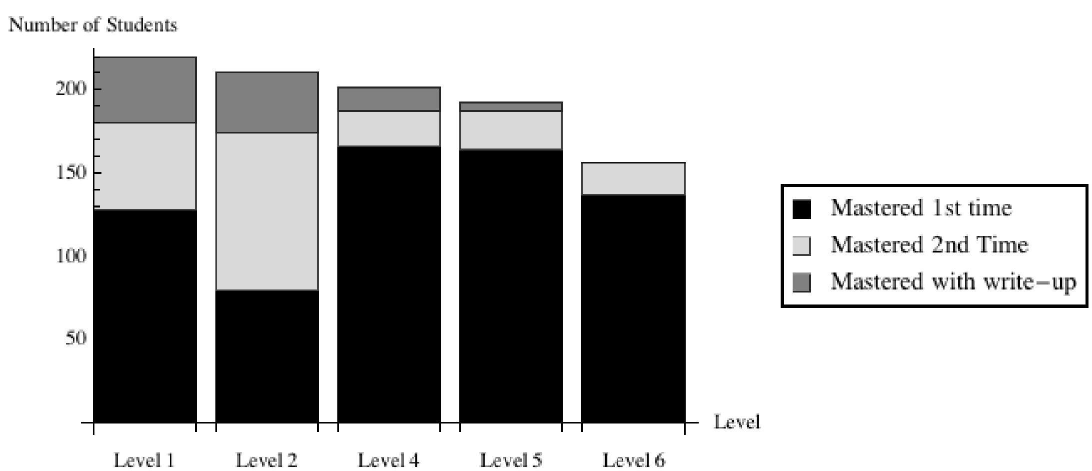

Figure 1: Change in the numbers of students mastering each mechanics level at the first, second or third try as the semester proceeded. Only purely computer-marked levels are shown. Most of the drop in numbers between Level 1 and Level 2 is due to students changing their enrolment.

Another surprise came from a detailed study of the wrong answers – it became clear that about 30-40% of incorrect answers were due to carelessness and misreading the question, rather than any lack of physics understanding. In a traditional course, this sort of error rate would lower their grade but not cause them to fail (particularly because most of the marks are typically for working rather than for the answer), but the mastery approach put a much greater premium on getting things right first time. As the semester progressed, however, students became much more likely to master a level first time, as can be seen in Figure 1. We assume that they had learned to self-check their answers, and to make sure they understood something before submitting – all valuable skills.

How rapidly did students progress through the levels? Most did the minimum of one test per week, and generally mastered each level first time. A few (< 5%) went much faster, but an increasing number were left behind as they failed to master one level first time. This progress can be seen in Figure 2. We were surprised that so few students took advantage of the flexibility of the system to cover material at a different pace: the sharp diagonal take-off line indicates that most students treated this just like a weekly homework assignment, albeit with an unusually high pass mark. It is clear, however, that procrastination was not a major problem.

The only major logistical problem we encountered was with levels that required tutor marking. Students tended to avoid these levels (such as level 3 in Fig 2), and rush on with more advanced computer-marked levels, planning to come back to the human-marked levels later (we had thought of banning this, but it was necessary for students to be able to move on so that they had something to do while waiting for marks to come back). Some students delayed coming back to these levels too long.

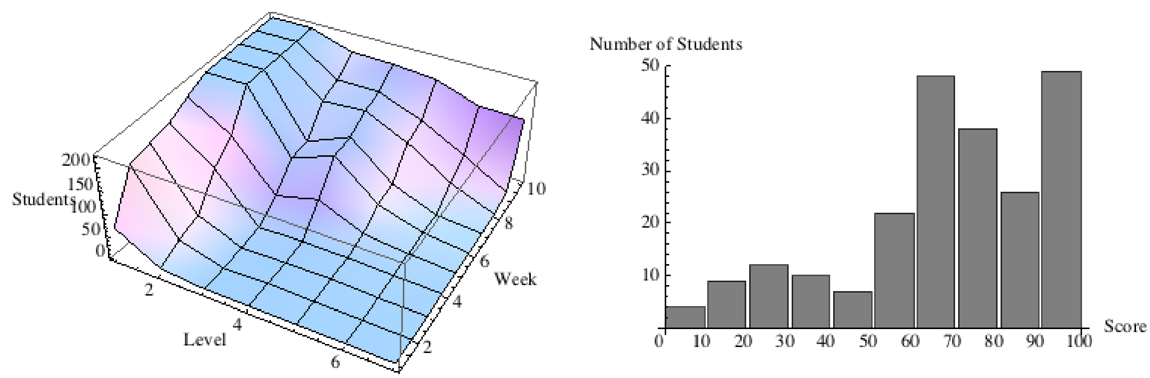

Figure 2: The number of students who had mastered a particular Mechanics level, as a function of time (left panel) and the final grade distribution, on a scale of 0-100% (right panel).

To our great relief, the overall final grade distribution (Figure 2) was acceptable. The variable number of levels mastered did give a reasonable mark distribution. The two peaks in the distribution correspond to people who did and did not attempt the advanced level problems. Unfortunately, the drop-out rate rose from 12% to 18%. We ascribe this to the relatively unforgiving nature of the mastery approach: weak students were forced to confront their weakness early on when they failed to master the basic levels repeatedly, and had no hope that they could claw the marks back in a final exam. Higher drop-out rates seem to be a generic feature of mastery-based courses (Cracolice & Roth, 1996).

Did the breadth of student knowledge suffer because of the mastery approach? Around 70% of students attempted each of the standard levels and had hence covered all the course material, at least at some level (Figure 2). The remaining 30% may therefore have missed doing any problems on the material in the later chapters. These students are, however, the ones with persistent major conceptual problems with the more basic material, so it is not clear how much they would have gained from an attempt at the more advanced material.

Special drop-in tutes were held to give extra support to those who were stuck on the earlier levels. These were attended by around 20% of the students at least once, but the students who attended were predominantly those who were already getting good marks, and wanted even better ones, rather that the genuinely struggling students. We asked some of the weaker students why they did not use these opportunities to get extra assistance, and they typically reported that timetable clashes or the hassle of getting to campus were the barriers.

The rate of coverage of topics changed considerably. An examination of scores in various early-level questions showed that a substantial number of students had major conceptual problems with very basic concepts, such as Newton's third law and the nature of vectors. A great deal of class time was spent on these basics, leading to complaints from students who were working on more advanced levels that they were not being supported in lectures. Nonetheless, an analysis of the exam answers of even the better students showed that most errors concerned these basic conceptual materials, rather than the more advanced material.

Student feedback was generally mildly favourable. But there was one striking change. A typical student did 50% more homework this year than last, as measured by the number of questions attempted, but we had almost no complaints about the amount of homework this year. Last year, however, we were bombarded with constant and vocal complaints about the quantity of homework. Many students acknowledged that this course was very hard work, but they didn't seem to complain about it. Perhaps the mastery approach gave students a feeling that they were more in control, or made the link between effort and grades more explicit?

Laboratories

Feedback from the demonstrators was that the new marking scheme made the marking much faster and easier. The students apparently didn't rush through the experiments as much but had the courage to stop and say "no, instead of finishing the whole lab badly I would rather concentrate on this bit and do it properly". The former marking schemes, which were different for every lab, gave the students the impression that they were marked on their results more than on their approach. This led to the students rushing through the labs without focusing on important details. Overall, lab marks were substantially lower under the new system.

As Figure 3 shows, students made rapid progress on most goals. Average marks improved rapidly over the first few labs, before plateauing at around 85%. We were pleased that students seemed to continue to use their newly acquired skills, rather than demonstrating them once and then reverting to their previous ways. For the more difficult skills, a slow and steady gain was seen. Anecdotally, this was a major improvement on the previous year, though this is hard to quantify as no baseline comparison data was available.

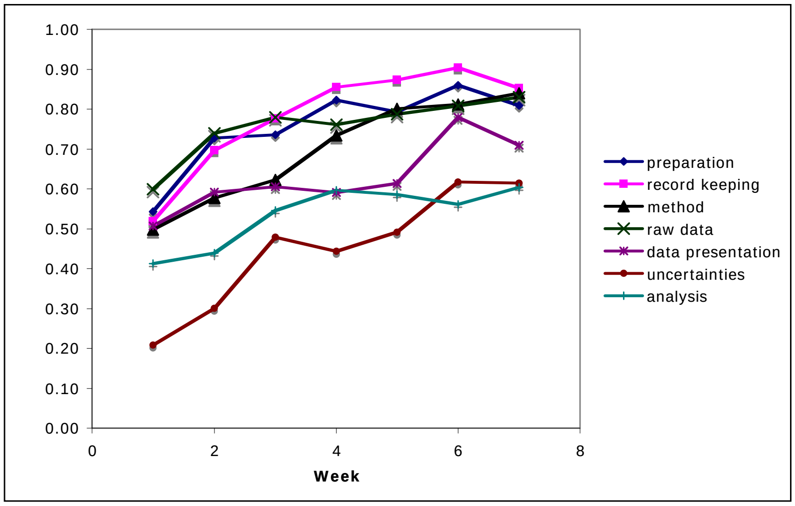

Figure 3: The average mark for each laboratory mastery goal, as the course progressed. A mark of 0.5 at least once against each goal is needed to pass the lab component.

The mastery approach made one other major difference – the instructors were actually approached by many of the weaker students, who had realised that they were failing to achieve some goals and were requesting help to understand what they were doing wrong: something very rarely seen before. A substantial number of students reported that the labs were “too hard’’ and that they couldn't go to any of the offered drop in tutorials (but complained only a few days before the due dates).

Did the students really master the basics?

Did we succeed in our primary goal of forcing students to really master the basics? As measured by getting high scores in the levels and by the lab marking scheme, the answer is yes. But it is possible that students could be using some shallow-learning strategy to achieve 80% scores in the levels, without deep understanding. To test this, we had two instruments. Firstly, we did a pre- and post-test of their conceptual physics knowledge, using the Force Concept Inventory (Hestenes, Wells & Swackhamer, 1992), a widely used and well validated test. The average normalised gain (gain in score divided by maximum possible gain in score) increased from 0.26 to 0.4 – a substantial gain, and one that brings this course well above the levels reached by most traditionally taught physics courses (Crouch & Mazur, 2001). A gain of 0.4 is however typical of courses which, like ours, use peer instruction in lectures. The course did not in general concentrate on FCI topics more than the previous year. In one case (Newton’s third law), it became obvious from student scores in the early levels that most of the class had major misconceptions, and so two lectures were devoted to this topic (and indeed very large gains were seen in the relevant questions), but most of the other topics we identified as problematic and hence concentrated on in lectures (such as vectors, and uncertainties) are not covered by the FCI.

Our second instrument was the exam. We re-used questions that had been used in the previous year’s exam, to allow a direct comparison of student performance. Based on student performance in the levels, we were expecting a great improvement in exam scores. But no significant difference was found. Indeed, the success rates were identical to within 1%. Many students who had done very well in the levels did very badly in exam questions covering the same material. The typical mistakes made by students were very similar to those seen in previous years – typically blind quoting of erroneously chosen equations, or attempts to solve problems using the wrong physical principles.

How could the students do so well in a homework situation, and so badly in an (open book) exam situation? To answer this question, we interviewed a sample of 16 students who had done well in the homework levels, but particularly badly in the exam. During the interview, they were given similar problems to solve on a whiteboard, and were asked about what strategies they employed to do well in the homework. The good news is that, with only two exceptions, they were capable of solving the problems, given time and an occasional hint. Three common themes emerged. Firstly, they almost all blamed special circumstances on exam day (mostly illness, staying up late the night before and family problems). Secondly, many of them mentioned that, as one student put it, when describing the exam questions, “I thought they’d be similar to the questions in the levels”. The questions in the exam covered exactly the same physics as the homework level questions, but throughout the course we’d deliberately tried to make all questions “context rich” (Heller & Hollabaugh, 1992), so that students needed to make an effort to work out which physical principle to apply in each situation. Clearly this was difficult in an exam situation, and easier in a homework situation. When asked how they managed to work out which physical principle to use when doing the homework levels, they stated that either they relied on discussion with friends, or they sat and thought about it for a long time, trying out various options in a trial-and-error way.

A few students also commented that the lack of a high-stakes exam caused them to focus, while doing the assignments, on getting the right number in the box, rather than learning the material. As one student put it, “we’re in the mindset that we just do a level and then we don’t have to know that material ever again.”

Conclusions

The mastery approach had a number of unexpected benefits, such as students doing more work with less complaint, students spontaneously asking for help, and students improving their ability to get questions right first time. Before starting this experiment, we were terrified that the logistical details of having a class all proceeding through the material at different paces would go terribly wrong – thankfully it all worked pretty smoothly. But did we succeed in our core goal of forcing students to really master the basics? Here the evidence is mixed. By several indicators, student performance significantly increased. But the exam data suggest that we made no change. We seem to have taught the students to solve problems collaboratively, and in situations where they have a lot of time to think. But in exam situations, these techniques no longer work.

References

Block, J. H. (1971). *Mastery Learning, Theory and Practice*. New York: Holt, Rinehart and Winston.

Bloom, B. S. (1968). Learning for Mastery, *UCLA-CSEIP Evaluation Comment*, 1, No. 2

Carroll, J. B. (1963). A model of school learning, *Teachers College Record, 64*, 723-733

Chabay, R. W. & Sherwood, B. A. (2007). *Matter and Interactions*. Hoboken: Wiley

Cracolice, M.S., & Roth, S.M. (1996). Keller’s “old” personalised system of instruction: a “new” solution for today’s college chemistry students. *The Chemical Educator, 1*, 1-18, retrieved Jun 18, 2009 from http://chemeducator.org/bibs/0001001/00010004.htm

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

Davis, D., & Sorrell, J. (1995). *Mastery learning in public schools.* Educational Psychology Interactive. Valdosta, GA: Valdosta State University. Retrieved June 18, 2009 from http://teach.valdosta.edu/whuitt/files/mastlear.html

Eyre, H. L. (2007). *Keller's Personalised System of Instruction: was it a fleeting fancy or is there a revival on the horizon?* Behavior Analyst Today, Summer 2007, retrieved June 18, 2009 from http://findarticles.com/p/articles/mi_6884/is_3_8/ai_n28461022/

Heller, P. & Hollabaugh, M. (1992), Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *American Journal of Physics,* 60, 637-644

Hestenes, D., Wells, M. and Swackhamer, G. (1992). Force Concept Inventory, *The Physics Teacher, 30*, 141-158

Guskey, T. & Gates, S. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. *Educational Leadership, 43*(8), 73-80

Keller, F. S. (1968). Good-bye teacher. *Journal of Applied Behavior Analysis, 1*, 79-89.

Kulik, C. C., Kulik, J. A. & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, *60*, 265-299.

Pear, J. J. & Crone-Todd, D. E. (1999). Personalised system of instruction in cyberspace. *Journal of Applied Behavior Analysis, 32*, 205-209.