Investigating whether Information is Retained for Longer if Taught by Teacher-Led or Student-Led Lessons.

Sharif Khan-Bennett

March 2020

School of Physics and Astronomy
University of Birmingham
United Kingdom

1 Introduction

A model commonly used to describe the effectiveness of different teaching methods at memorably imparting information is the National training laboratory's learning pyramid. It states that a learner will remember varying amounts of information, depending on the way it was imparted. For example, it states that a learner will only remember 10% of what they are lectured about, but they will remember 90% of what they teach to others Letrud (2012). This is presented in figure 1.1

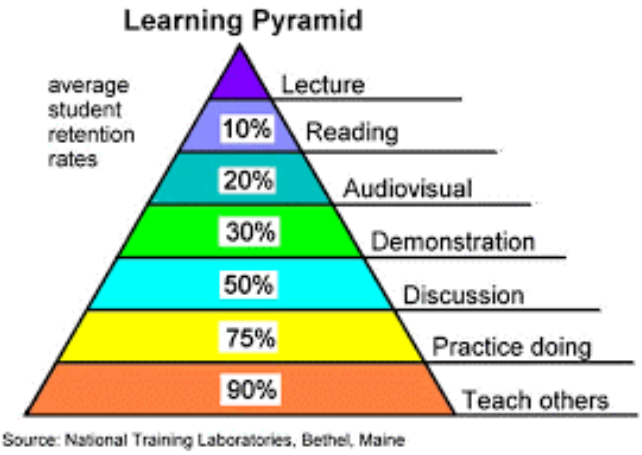

Figure 1.1: NTL's Learning pyramid Letrud (2012)

Despite its wide usage, it has faced a great deal of criticism. Its sceptics point out its apparently arbitrarily chosen percentage brackets for each teaching method. They also question the methods used in obtaining these statistics, stating that there is no indication as to how much time had passed between learning and testing. i.e., how long it had taken for a student to forget 90% of what they were lectured about. Finally, they suggest that although the teaching methods had been represented as discrete and separate experiences in the pyramid, in reality, a learner utilises a combination of these methods when learning, implying that separating them is misrepresentative. For example, a learner may be exposed to reading, audiovisual and demonstrative forms of teaching while in a lecture, blurring the lines between the top four levels on the pyramid. It is for these reasons that some people demand that NTL redact this pyramid Letrud (2012).

Despite its criticism, it is still commonly suggested that some methods of teaching are more effective than others at conveying memorable information. For example, Jerry J. Wellington and Gren Ireson pointed out the advantage of "meaningful learning" over "rote learning", the former being non-arbitrary, substantive and non-verbatim, such as a physics practical or teaching others. The latter being "purely arbitrary associations" such as learning by reading or by being lectured. They suggest that if the learner lacks the relevant prior knowledge, the rote method is insufficient as it results in its recipient only being able to internalize a "random series of words" Wellington and Ireson (2008). They point out some of the negative effects which rote learning may encourage in what they call "right-answerism" where students would regurgitate random facts to satisfy the teacher as opposed to cementing their knowledge. They say: "'I hear and I forget, I see and I remember, I do and I understand."

Although Wellington and Ireson may not have applied such specific and precise rates of retention for different teaching methods, they still allude to a similar hierarchy as suggested by NTL in which students remember more of what they do for themselves and less of what they are told. This investigation sets out to compare the retention rates of 19 A-Level students from Bournville College when learning from a range of teaching methods. These methods including passive experiences such as lecturing, and active experiences including practical work and teaching others. The drop in test score after one week will be used as a proxy for retention rate, as described in the next section. This data will be used to test whether lecturing is

indeed less memorable than other methods and whether there is truth in the relative effectiveness of each teaching method suggested by the learning pyramid.

2 Method

The teaching methods that are investigated in this report are lectures, practicals and students presentations. To measure the retention rate associated with a particular method, a series of tests will be used to measure the average rate in depreciation of marks from when the lesson took place. The first tests would be completed immediately after a teaching method had been applied and a second set would take place the following week, testing the same material with a set of similar questions. A greater average drop in marks would indicate a shorter retention time and a less effective teaching method.

As the case with all social experiments, it is expected that there would be a large variation in the results due to the randomness associated with human behaviour. To mitigate some of these effects, a set of controls are introduced to try and ensure consistency when testing each teaching method. Firstly, there must be a constant time interval between the first and second test such that the students have an equal amount of time to forget (or remember) each lesson. To ensure that memory is the only factor which affects any changes in results each test must be of similar difficulty. This can be very subjective but will be attempted by ensuring each paper is worth an equal number of marks from standardized questions from previous papers. Each lesson will also have to be a similar difficulty so that results are not affected by some lessons being more difficult. Again, this can be subjective but will be controlled by covering equal amounts in the A-Level syllabus and with help with my teacher supervisor. Additionally, topics which have been taught recently beforehand should be avoided so that the results aren't skewed in favour towards a certain method due to the material having been recently covered.

Ideally, after the first lesson, the students shouldn't continue to study the material before the second test otherwise it would defeat the point of the test being a memory exercise. Enforcing this in the classroom would be impractical however, so there are some variables which should be mentioned which were not controlled. For example, it was not known if students continued to study material after the first lesson. Additionally, it might be expected that there is a greater measured retention in the A2 class over the AS class due to there being a naturally higher amount of pressure from exams causing students in one class to be naturally more engaged than in the other.

The content that each teaching method used was selected by ensuring that it had not been taught in a long time. For this reason, the lessons were delivered as revision sessions where the content and delivery methods are outlined in table 1.

Class	Delivery Method	Content
AS	Class Practical	Hooke's Law
AS	Lecture	Young's Modulus
A2	Lecture	Materials
A2	Class Presentations	Fields

Table 1: Methods used and content covered in each lesson

It may already be apparent that there is an overlap in content in the AS classes, that being the similarity between Hooke's Law and Young's Modulus. The reason for this is that it was impractical to jump about the syllabus as the AS students were only just being intro-

duced to materials. It could be done with the A2 students, however, as they were at the stage of revising previously learned topics. This may result in the test results from the AS Lecture on Young's Modulus being disproportionately larger, as discussed in the next section.

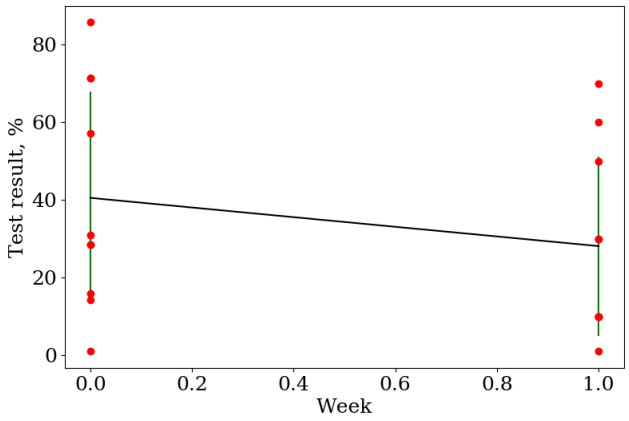

Figure 3.2: Results from week one and week two testing retention from an AS lecture on Young's Modulus.

3 Results

Figures 3.1, 3.2, 3.3 and 3.4 show the spread of test results immediately after the relevant lesson, and from the following week. The test scores are shown as red dots with the average drop as a black line, connecting the standard deviation of results on each week as green lines.

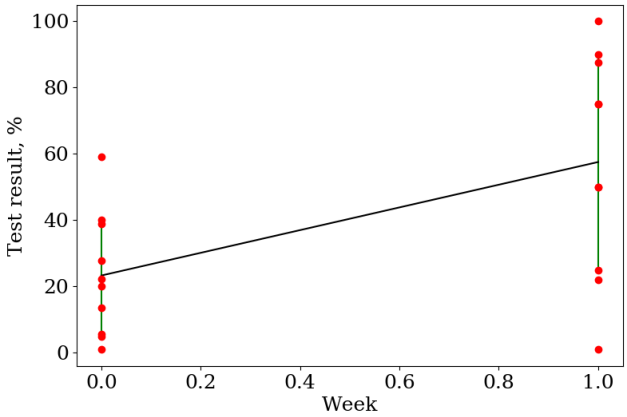

Figure 3.1: Results from week one and week two testing retention from an AS practical on Hooke's Law.

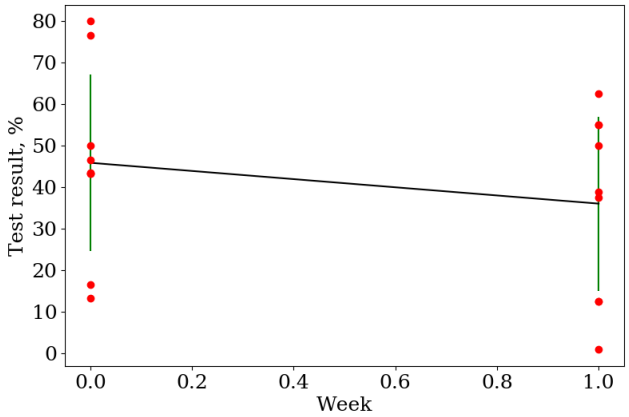

Figure 3.3: Results from week one and week two testing retention from an A2 lecture on Materials.

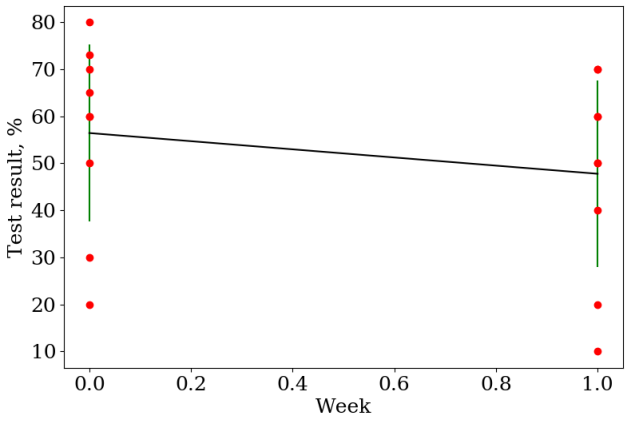

Figure 3.4: Results from week one and week two testing retention from A2 class presentations on Fields.

It can be seen that, with one exception, all the test scores diminished as expected. The plots which fall at a slower rate, such as figure 3.4 indicate a longer retention time, and a more memorable initial lesson. Figure 3.1 displays the average mark increasing, unlike in all the other cases. This could be due to the AS students never having studied materials before. It is presumed that in the following week, although they may not have covered Hooke's law again, they continued studying materials, in accordance with the a-level syllabus, familiarising them with the ideas and equations surrounding Hooke's law. This could explain why the average test result had doubled by the second week.

These results were used to test whether there was any truth to be found in the NTL learning pyramid. By looking at the test results from the AS class and the A2 class separately from figures 3.1 through to 3.4, the retention rates of different methods of teaching could be compared to see whether they fit into model proposed by the NTL.

The AS practical saw an increase in average mark by a factor of 2.468 ± 1.356, whereas the results from the lecture were multiplied by a factor of 0.694 ± 0.568, indicating that the practical was 3.66 ± 0.962 times more effective than the lecture as a method of delivering information. This is considerably less than the factor of 7.5 as suggested by the NTL. Whether it can be treated as evidence to reject the learning pyramid hypothesis will be found using hypothesis testing with a significance level of 0.05% (a z value of 1.64). The z value of this result is found using the following formula

$$z = \frac{\tilde{x} - \mu}{\sigma\sqrt{n}} \tag{3.1}$$

where $\tilde{x}$ is the factor which is being tested, in this case, 7.5, μ is the factor suggested by this investigation, σ is the associated error of μ and n is the number of test results used to find this factor, in this case, the number of student in the AS class, 10. This results in a z factor of 1.25, less than the 1.64 requirement for a 0.05% significance level. Consequently, the data is not sufficient to reject the claims of the learning pyramid.

The same method was applied to the A2 class in comparing the lecture-style with the method of teaching others where the learning pyramid suggests that the latter should be 9 times more effective at maintaining test results. It was found that the A2 lecture experienced a drop in average mark by 0.786 ± 0.457 and the class presentation method experienced a drop in mark given by the factor 0.846 ± 0.353, indicating that the class presentation method was only 1.07 ± 0.405 times more effective as a delivery method than a lecturing-style. This gives a z value of 6.627, much larger than the 1.67 threshold. Therefore it can be said that the sample did not behave according to the predictions of the learning pyramid.

By modelling the retention curve as a decaying exponential

$$R = e^{-at}, \tag{3.2}$$

where R is the amount of knowledge retained, given by the second test result divided by the first test result, and t is time in weeks, the retention curve for each student's change in result can be found. Cases where a was negative, indicating that a test result had improved were removed from the data set as it was assumed that one of the control variables aforementioned had been violated, making the memory test redundant. In each case, a mean value of a was found along with its standard deviation to plot retention curves with errors for each teaching method. This is shown in figures 3.5, 3.6, 3.7, 3.8 where the data has been extrapolated past the first week.

The value of a for the first set of test results was found to be -0.596 ± 0.720 resulting in figure 3.5

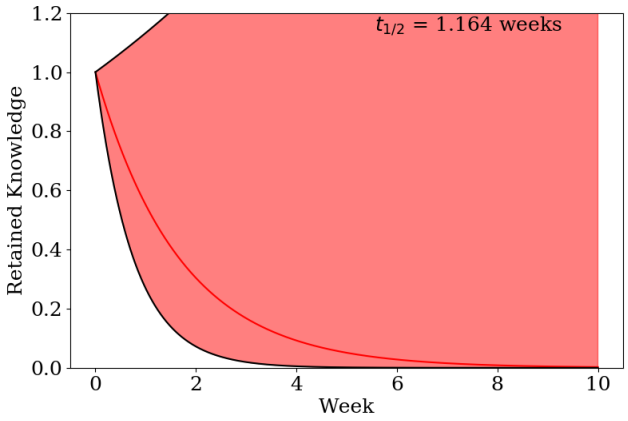

Figure 3.5: Mean retention curve from an AS practical and its error as given by the shaded area.

The value of a for the second set of test results was found to be -0.836 ± 0.808 resulting in figure 3.6

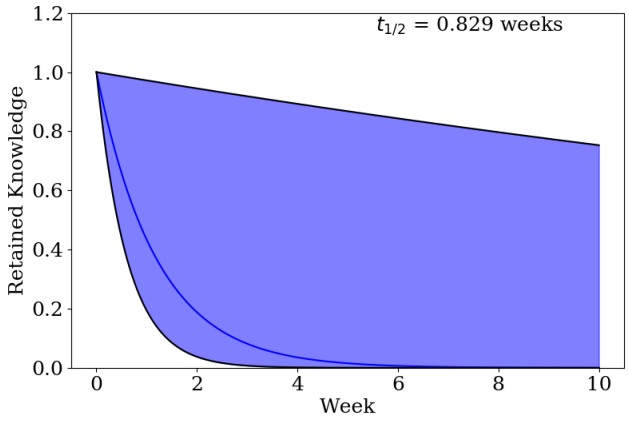

Figure 3.6: Mean retention curve from an AS lecture and its error as given by the shaded area.

The value of a for the third set of test results was found to be -0.938 ± 0.836 resulting in figure 3.7

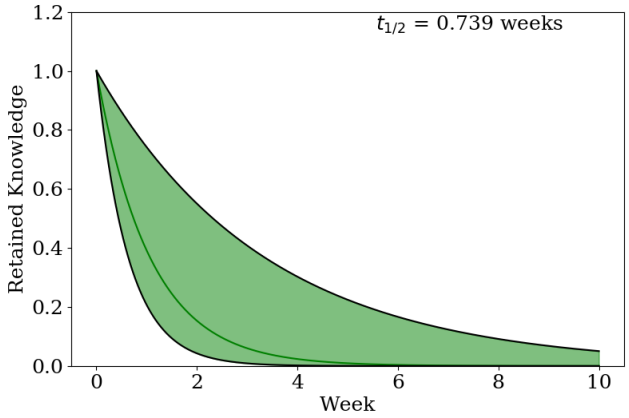

Figure 3.7: Mean retention curve from an A2 lecture and its error as given by the shaded area.

The value of a for the fourth set of test results was found to be -0.716 ± 0.639 resulting in figure 3.8

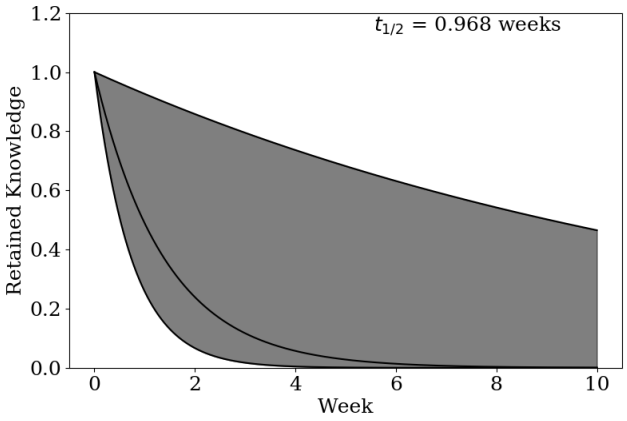

Figure 3.8: Mean retention curve from A2 class presentations and its error as given by the shaded area.

It can be seen from the figures above that the lecture-style method of teaching had the worst retention times of the three methods investigated, as expected. This can be shown quantitatively by comparing the half-lives of retention in each case. The practical produced a half-life of, 1.164 weeks, the highest of the four, the class presentation method produced a half-life of 0.968 weeks and the lectures produced half-lives of 0.829 and 0.739 weeks. This appears to indicate that despite what had been suggested by the NTL learning pyramid, that practical work may, in fact, be more effective as a teaching method than teaching others. However, the results in this experiment are limited due to data collection being cut short as a result of unforeseen circumstances. This leads to the very large errors associated with the value of a in the decay curves. There is such a large error associated with the AS practical that the upper limit of expected retention appears to diverge, despite cases of improved test results being removed. Consequently, the results in this investigation should be treated with scepticism as a result of the lack of repeated testing which would have taken place in subsequent weeks, populating figures 3.1 through to 3.4 with results past the one week mark.

4 Conclusion

This investigation set out to show how effective different methods of teaching are by comparing the retention times of each as a result of the drop in test scores over one week. It was shown that students perform better when asked to recall knowledge acquired from active methods of learning, specifically practical and class presentation lessons. It was shown that although the lecture-style of teaching was least effective, that the sample did not conform to the strict brackets suggested by the learning pyramid, putting into question its validity. It would not be possible to use this data to reconstruct a pyramid of the relative effectiveness of each method due to the very large errors associated with each result. Even if it was the case, the pyramid would not be universal, rather only represent the sample in question. This investigation also does not aim to suggest that any one method of teaching should be used exclusively instead of another, as suggested by Letrud (2012). Rather, these methods should be used in unison to provide diversity in a dynamic classroom.

Future experiments may want to continue measuring the test results past the one week mark, to get a better representation of the retention curve, and test whether a decaying exponential is a suitable model for this, they may also want to investigate on larger samples to obtain more reliable data. Unfortunately, this was not possible in this case due to the data collection period being cut short by COVID-19.

References

Kåre Letrud. A rebuttal of ntl institute's learning pyramid. *Education*, 133:117–124, 01 2012.

J.J. Wellington and G. Ireson. *Science Learning, Science Teaching*. Routledge, 2008. ISBN 9780415433938. URL https://books.google.co.uk/books?id=0KRCXAlyCBQC.