Student study-behavior and knowledge retention

Bram van Vliet
University of Twente
The Netherlands
g.j.b.vanvliet@student.utwente.nl

ABSTRACT

This research aims to determine if it is possible to reliably distinguish study-behaviour with and without spacing behaviour and based on that, determine to which degree study-behaviour influences knowledge retention. The study-behavior of 29 students in a digital learning environment was analysed and split into three groups based on the degree of spacing behaviour. 17 Of these students gave permission to examine their exam results. These students were then requested to participate in post-exam retention tests. The data that was collected was indicative of long-term knowledge retention improvements in the students that applied spaced learning strategies, proving this requires further studies with a larger sample size.

Keywords

Study behaviour,study-behaviour Knowledge retention, Living Textbook, Website traffic analysis, Spaced learning

1. INTRODUCTION

1.1 Opening

The purpose of education is to provide people with knowledge and skills. Most educational institutions gauge the success of their students through testing. Although testing is a very effective method of determining the current knowledge of a person, a single test does not examine the quality of the acquired knowledge, not its longevity. For a single test, a student can study in two days, pass with flying colours and then forget all content within a week. This behaviour cannot be detected with a single test at the end of a study period. However, studies have shown that multiple smaller tests promote spaced learning[5] which has been theorised, but not proven, to promote long-term retention[6]. In the current study, we therefore examined the relation between spaced learning and long-term knowledge retention.

1.2 State of the art

Much research has been done into the factors that affect studying, such as the effect of immediate feedback, cumulative assessment, and spaced learning.

32th Twente Student Conference on IT Jan. 31st, 2020, Enschede, The Netherlands.

1.2.1 Spaced learning

Spaced learning is the concept of spreading study time over a longer period with long intervals between training sessions. This is opposed by cramming, clustered or massed learning, such as only studying in the two days before an exam. Smolen, Zhang and Bryne provided a review on the current research into spaced learning[6]. They concluded that spaced learning leads to more robust memory formation than massed learning. The topic of retention was briefly mentioned in the context of verbal learning. They found that greater spacing worked well for longer retention intervals, however, shorter intervals between training sessions worked better on the short term retention.

1.2.2 Immediate feedback

Dihoff et al.[4] discussed five different methods of examination and providing feedback and evaluated the effects of the different methods on response recall. They did this by making the students take one of five different exams, each with varying degrees of feedback and in different formats. They then asked the participants to come back two weeks later, asking them to take the exam again, this time simply in a written answer format and asked them to identify the answer they gave last time. They found that providing immediate feedback on the exam, with the use of an IF AT form[1], caused students to have the best recall of the answer they gave last time and have the highest(>85%) score on the second exam.

1.2.3 Cumulative assessment

"Cumulative Assessment is a method that combines repetitive testing, repetition of content, compensation across tests, and feedback between tests, in order to stimulate students to study" -Cecilio-Fernande et al., p.2 [3]

Kerdijk et al. [5] applied this principle to a group of 78 students who were randomly assigned to a control group or a cumulative testing group. They noticed a slight increase in performance on questions regarding the latter part of the course in the student group that applied spaced learning. They did not find an improvement in the short term memory retention, but hypothesised that spacing and cumulative assessment would have a positive effect on long-term knowledge retention.

1.3 Research aims

From the above review of the literature, it is clear that spaced learning in the form of lecturing influences the duration of knowledge retention and as mentioned in section 1.2.3, spaced learning could improve long term memory

[1] An IF AT is a form on which you can scratch out a little metallic film for each option(A, B, C, etc.). If it is the correct answer, a little icon is revealed. This provides instant feedback on the correctness of an answer.

retention. In the research of Kerdijk et al.[5], a measure for spacing was not required, as the groups were predetermined. However, not every study can benefit from or has the option to apply cumulative assessment in the way they did. Additionally, by doing so, the research already influences the way people study. In order to reason about the effect of spacing behaviour, a method for determining what that behaviour is, outside of pre-selection, is required. This leaves us with the following two research questions:

1. Can two types of study-behaviour (spaced and clustered) reliably be distinguished?
2. To what extent does study-behaviour impact learning and knowledge retention?

Regarding research question one, the hypothesis was, based on preliminary data and study-behaviour analysis, that it is posible to reliably distinguish two types of study-behaviour.
Regarding the second research question, the hypothesis was, based on related works, that study-behaviour does influence learning and knowledge retention.

2. METHODS

Data about study-behaviour has been collected through the Living Textbook (*see section 2.2.1*) with the students' consent. During the module, the students had six tasks. These tasks were small, single-person projects about various topics. At the end of the students' module, they had one big exam (*section 2.2.3*). After this exam, the students were given two more tests, as part of this research, to inspect the knowledge retention(*section 2.2.4*).

2.1 Subjects

The group of students consists of 42 master students, following the GIMA[2] course, a distance learning course on geographical information management. They use digital learning tools, among which the Living Textbook, in addition to regular textbooks. These students were chosen based on their usage of the Living Textbook(*see section 2.2.1*). Of these students, 35 consented to be tracked within the Living Textbook. Furthermore, 17 of the students consented to the viewing of their exam performance, 13 students took the first retention test and eleven took the second retention test. Under the students who participated, 2x20 euros was given by method of a lottery as an incentive to participate.

2.2 Tools

2.2.1 The Living Textbook

The Living Textbook is a project to modernise education[2]. It provides a concept map(*Figure 1*) rather than a book or PDF, which allows users to see the relations between concepts, rather than having pages which follow each other and refer to one another. The Living Textbook also allows links to external sources, such as research papers or web articles. It can be easily updated and, most relevant for this research, it offers the option to allow tracking.

The website saves which pages are visited, which links are clicked and by whom. This is currently being used to determine if the so-called learning paths that teachers can create within the system are being used by the students. This data allows drawing conclusions about students' study-behaviour; subjects like when they start studying, how many concepts they browse through, how much time is spent on each of these concepts and how frequently they visit the concepts which will appear on their exam.

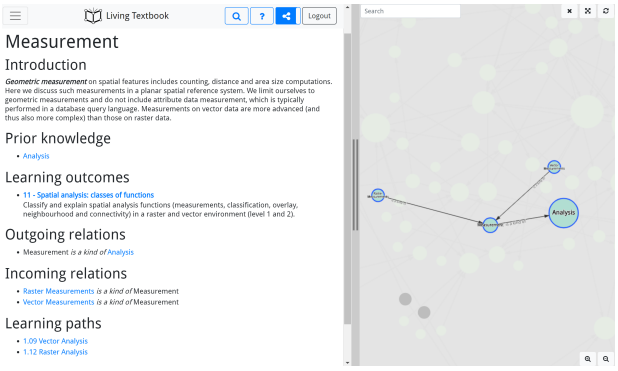


Figure 1: Sample page from the Living Textbook. Contents of the current concept on the left and the concept map, with related concepts highlighted, on the right

Many of the studies discussed in section 1.2 focus on lab testing or class activities. However, the Living Textbook provides a unique opportunity, as it allows us to analyse the behaviour of students when they are not being observed and have control over how they study. Therefore the data from the Living Textbook is ecologically valid and might provide a more reliable view into the actual study-behaviour when compared to a lab setup.

2.2.2 Student grouping protocol

The data that was collected about the web traffic by the Living Textbook was then converted into one graph per student which displayed the hours spent per day. Graphs like Figure 2a, 2b, and Figure 2c. This data did have one limitation, namely that it was not possible to determine when someone left a page. The only data available was when they opened which page, so if they opened two pages after each other, it was simple to calculate how long they had been on the first page, but one could only estimate how long they had been on the second. Because of this, a clear indicator was used when this happened. This is displayed as the hours estimated bar. It is 0.3 hours for each time that it was not possible to calculate how long a user was on a page. 0.3 was chosen so that it did not change up the scale of the images, but it is clearly visible.

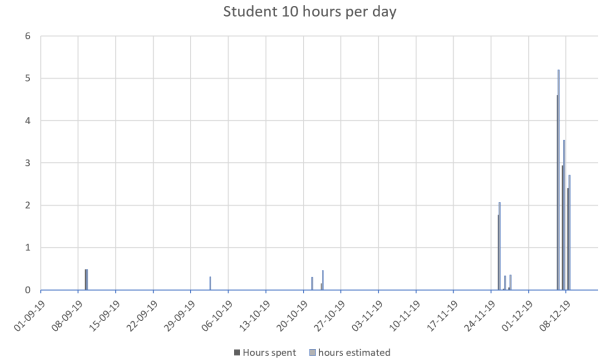
This data was then ordered manually into one of three groups:

- Assignment students
- Clustered learners
- Spaced learners

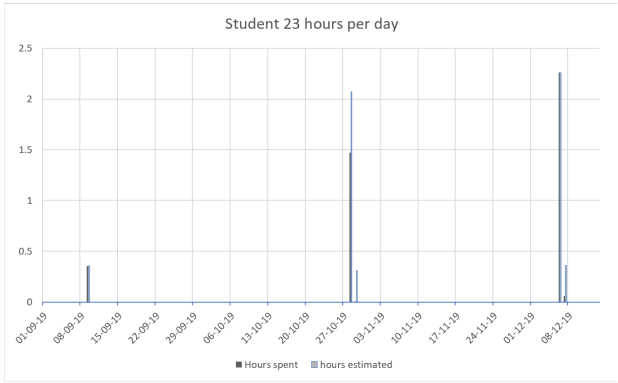
The criteria based on which this grouping was performed will be discussed in section 3.1

To measure the accuracy of the grouping criteria, a fellow computer science student was asked to perform the grouping, based on the criteria. This happened in two instances, as the first criteria proved too imprecise. For both instances, different computer science students were asked to preform the grouping as to avoid bias.

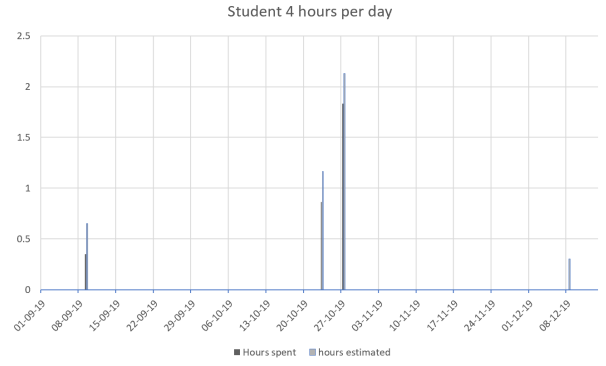
[2]GIMA stands for Geographical Information Management and Applications



(a) Sample data of a spaced student



(b) Sample data of a clustered student



(c) Sample data of an assignment focused student

Figure 2: sample data of different student types. Hours estimated is set to 0.3 for each visit of unknown or unreasonable time(>4 hours) This way it was recognisable that there was some activity, but it was unknown how long that activity was.

2.2.3 Exam

At the end of the module, the students were tested through a single exam. This exam was divided into several sections, relating to the different subjects. Questions 35, 36, 39, 40 and 41 were related to the knowledge that students could only obtain from the Living Textbook, so these were the only relevant questions for this research, as for only these questions, it was known when and how much the students studied.

2.2.4 Tests

To measure long-term retention, a subset of the questions from the exam was posed to the students 1 and 4 weeks after their initial exam. The questions on the first test were slightly adapted questions from the exam which were about the Living Textbook's content. They were about the same topic as the original questions but had little things altered, such as an OR that became an AND. For the second retention test, we went back to the original exam questions in order to test retention. This allowed us to link the answers to the retention test questions directly to the exam questions and determine how much was forgotten.

2.3 Procedure

When users first login in the Living Textbook, they are asked if their tracking data may be stored for academic purposes; this is a setting they can alter at any time.

After the exam in early December, the students had a mandatory module evaluation session, during which they were introduced to this research and requested to fill in a signup sheet if they wished to participate in the research. They were requested to not study for the retention tests and were informed of the 2x20 euros that would be raffled at the end.

Those that signed up were emailed one week later with a link to the first retention test and requested to take it without communicating the answers with fellow students. Thirteen students took this test

Then, four weeks after the exam, just after the Christmas holiday, the students were emailed again with the request to take the second retention test. Eleven students took this test.

3. RESULTS

3.1 Student grouping

For this research 3 groups of students were distinguished: Assignment learners, clustered learners and spaced learners. These were sorted into these groups by the following criteria:

On the 29^{th} of October, the students had an assignment due for which they had to use the Living Textbook. Students that only displayed activity around that moment and no significant activity afterwards, are labelled as assignment learners. Significant activity here means anything beyond half an hour on a single day. So if there is a large cluster of activity around the 29^{th} and before the test there are about 20 minutes of activity, that student is classified as an assignment learner.

Clustered learners, or crammers, are students that don't apply spaced learning and thus end up doing all their studying in a couple of days. Specifically, they spend a lot of time just before the exam, but not a lot before that.

Finally, spaced learning is a method by which people study multiple times over a longer period, people that have multiple moments of study, spread out over the time-frame are spaced learners. This could mean studying a little bit every week or every couple of weeks, but it could also mean

studying a lot during a certain week and refreshing that knowledge every other week. To determine clustered or spaced learning, the activity around the assignment may be left out of consideration.

These criteria arose from analysing errors that were found in grouping with the original criteria:
On the 29^{th} of October, the students had an assignment due. Students that primarily displayed activity before that moment are labelled as assignment learners. Clustered learners or crammers are students that don't apply spaced learning and thus end up doing all their studying in a couple of days. If a student is only active on two days during the module and then becomes very active in the last week before the test, they are a clustered learner. Spaced learning is a method by which people study multiple times over a longer period, people that have multiple moments of study, spread out over the time-frame are spaced learners. This could mean studying a little bit every week or every two weeks, but it could also mean studying a lot in a certain week and refreshing this knowledge every week or every couple of weeks.

The first issue was that the definition for cramming was not elaborate enough, thus it was altered to become: *Clustered learners, or crammers, are students that do not apply spaced learning and thus end up doing all their studying in a couple of days. Specifically, they spend a lot of time just before the exam, but not a lot before that.*

The second issue is that the assignment made it hard to differentiate between crammers and spacers, because crammers did not have just one spike of activity, but also had activity around the assignment moment. Therefore the criteria were *altered even further: On the 29^{th} of October, the students had an assignment due. students that only displayed activity around that moment and no significant activity afterwards, are labelled as assignment focused students. Significant activity here means anything beyond half an hour on a single day. So if there is a large cluster of activity around the 29^{th} and before the test, there are about 20 minutes of activity, that person is an assignment learner.*

After this clustering was applied, 29 students with useful results were left. Six students were filtered out because they had no notable activity within the Living Textbook. Asking a computer science student to perform the same clustering resulted in a clustering which was 93% similar(as can be seen in table 1 and in the confusion matrix in table 2), which is good by machine learning standards. The confusion came mainly from students that had two or three study moments, including the assignment, which led to uncertainty as to whether that should count as spaced learning or not. When these cases were re-examined, the cases were labeled as spaced learning in the case of three study moments and as clustered learning in the case of two moments. This was based on the line *"To determine clustered or spaced learning, the activity around the assignment may be left out of consideration."* as this means that a student who studied two times including the assignment moment, was a crammer and the student that had three study moments was not a crammer and was therefore a spaced learner.

3.2 Study-behaviour

The graphs that were described in section 2.2.1 and shown in Figure 2a, 2b, and 2c, were grouped according to the criteria that were described in Section 3.1. The division was as follows:

student group	actual grouping	control grouping
assignment learners	17	16
clustered learners	5	4
spaced learners	7	9
total	29	29

Table 1: Division of students across the three categories

	actual values		
	Assignment	Clustered	Spaced
Assignment	16	0	0
Clustered	0	4	0
Spaced	1	1	7

Table 2: Grouping confusion matrix

- Assignment focused: 17 students
- Clustered learners: 5 students
- Spaced learners: 7 students

Of the assignment focused students, 10 gave permission to view their exam results and 7 took part in the retention tests. Of the clustered learners, one gave permission to view their exam results. This student also participated in the retention tests. And finally, among the spaced learners, four gave permission to view their exam results and three took part in the retention tests.

3.3 Behaviour and exam results

The results of the unpaired t-test between the exam scores of the assignment learners and the spaced learners, as described in table 4, show no statistically significant difference between the two groups with the two-tailed P value of 0.3877. The mean score of all students was 11.82, with a standard deviation of 4.76. The only score we had for clustered learners was a 19, which is one and a half standard deviation above the average.

Student group	average	standard deviation
Assignment learners	12.1	4.63
Clustered learners	19	0
Spaced learners	9.75	2.28
Overall	11.82	4.76

Table 3: Exam scores per student group

Two tailed P value	0.3877
$\mu_1 - \mu_2$	2.35
t	0.8964
df	12
SEM	2.622

Table 4: unpaired t-test of the exam scores between assignment learners and spaced learners

3.4 Behaviour and test results

Despite what was discussed in section 3.2 and 3.3, Table 5 shows large differences between the assignment focused students and the spaced learners. This means that on average, the assignment focused students forgot more than the spaced learners. An unpaired t-test, as described in table 6 does not prove this difference to be statistically significant with a P value of 0.3105. This is due to a high variance in both groups. Most assignment focused students scored significantly lower on the first and second

Student group	average	standard deviation
Assignment learners	-1.4	4.67
Clustered learners	3	0
Spaced learners	2.33	2.49
Overall	0.33	4.24

Table 5: Difference of score between exam and second retention test per student group

Two tailed P value	0.3877
$\mu_1 - \mu_2$	2.35
t	0.8964
df	12
SEM	2.622

Table 6: unpaired t-test of the difference in score for the exam and the second retention test between assignment learners and spaced learners

retention test than they did on their exam, which indicates a decrease in knowledge retention.

4. DISCUSSION

The aim of this research was to determine the posibility of distinguishing two types of study-behaviour reliably and to determine to what extent study-behaviour influences learning and knowledge retention. Grouping criteria were developed which led to a 93% accurate grouping. With the use of t-tests, this research was unable to determine statistically significant influence on learning and on retention.

The grouping criteria that were developed were sufficient, as only two students were misclassified. These misclassifications were mainly due to the fact that the relevant students barely did any studying. This issue would not be as prevalent if the Living Textbook was the primary study tool, as students would not have a choice but to study from the Living Textbook. After closer inspection and a more strict interpretation of the criteria, these misclassifications were easily rectified.

The division of students over the three groups makes sense, as the students were not explicitly told to study from the Living Textbook, thus most presumed the activity they had had during the assignment was sufficient for the test. Aside from that, properly applying spaced learning takes effort and dedication, which not many students want to go through for a small part of an exam.

Finally, on the topic of student activity, it makes sense that there was, overall, not that much activity. This was due to the fact that the Living Textbook was only one of many learning tools that the students had available to them, not the primary one, as the presumption was at the beginning of this research.

The average scores per study-behaviour group, as described in Figure 3, show that the clustered learners scored highest by far. This was largely due to the fact that there was only one student in that group that allowed us to view their test results. The expected results, based on the works of Smolen et al.[6] would have been: spaced learners as the highest scoring group, followed closely by the clustered learners and finally, the assignment focused students. This was the expected result because assignment focused learning is essentially clustered learning, just further in advance. The reason the averages turned out as they did, was because of statistical outliers. These would have been filtered out, were it not for the fact that that would have left us with a sample size of 8, 0 and 3 students in the assignment, clustered and spacing groups respectively.

The spaced learners had much better results on the first and second retention tests on average. This would affirm the hypothesis that spaced learning improves long-term knowledge retention. However, due to the low student numbers, this claim cannot be supported by statistics and requires further examination.

The aim of the retention tests was to use questions from the exam which had been slightly altered(AND turned into OR, minor changes in phrasing). However, it became clear after the first test, that the students did not read the questions well enough, and these questions were read wrong as a result. These changes were reverted for the second retention test. Mostly due to this reason, this research focuses on the second retention test, rather than on the first one.

Both the assignment and the spaced learner groups had people in them whose scores went up. This can be explained by the testing effect[1]. This effect states that students learn simply from taking a test. Taking a test requires students to refresh their knowledge of the subject of the test, which leads to an improvement to their knowledge retention. So future research should be careful in planning the retention tests.

The largest limitation of this research was the number of subjects, which leads to the absence of statistical significance. The solution to this would be to re-execute this research with a larger student group. The source code for the graph generation can be found at `https://github.com/Kyanite3221/studybehaviorAnalysisLivingTextbook/` and may be used for research purposes. This code produces graphs with similar information as the graphs in this paper, however, they do not look as nice.

Due to the low amount of students, it was possible to manually group the students based on a written set of criteria. For larger student groups, this becomes a formidable task and the use of clustering algorithms might be required.

5. CONCLUSION

Over the course of this research, it has become clear that it is possible to distinguish at least two types of study-behaviour based on a set of written criteria, which require adaptations based on the curriculum and a preliminary data analysis. These criteria can be simple yet precise enough that someone without any knowledge about the course can preform the grouping with a 93% accuracy when compared to the person who developed the criteria.

This research was not sufficient to prove anything statistically significant about the effects of study-behaviour on knowledge retention outside of some suggestive data which affirms the current claims about the effects of spaced learning on long-term knowledge retention[5, 6]. The data did suggest that spaced learners maintained their gained knowledge better than those that did not apply spaced learning principles. To affirm this claim, however, more research, with a larger and more specific group of students is required.

To conclude, the Living Textbook provides an excellent platform to perform an in-depth analysis of students' self study-behaviour, which may be grouped based on the degree of spacing or lack thereof with the use of a set of written criteria. This grouping can then be used for further research into knowledge retention.

6. REFERENCES

[1] P. Agarwal, J. Karpicke, S. Kang, H. Roediger III, and K. McDermott. Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22(7):861–876, 2008. cited By 144.

[2] P. Augustijn, R. Lemmens, M. Verkroost, S. Ronzhin, and N. Walsh. The living textbook: Towards a new way of teaching geo-science. In *AGILE 2018 : Proceedings of the 21st AGILE Conference on Geo-information science Lund, Geospatial technologies for all*. Association of Geographic Information Laboratories for Europe (AGILE), 2018.

[3] D. Cecilio-Fernandes, M. Nagtegaal, G. Noordzij, and R. Tio. Cumulative assessment: Does it improve students' knowledge acquisition and retention? *Scientia Medica*, 28(4), 2018. cited By 0.

[4] R. Dihoff, G. Brosvic, and M. Epstein. The role of feedback during academic testing: The delay retention effect revisited. *Psychological Record*, 53(4):533–548, 2003. cited By 57.

[5] W. Kerdijk, J. Cohen-Schotanus, B. Mulder, F. Muntinghe, and R. Tio. Cumulative versus end-of-course assessment: Effects on self-study time and test performance. *Medical Education*, 49(7):709–716, 2015. cited By 15.

[6] P. Smolen, Y. Zhang, and J. H. Byrne. The right time to learn: Mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience*, 17(2):77–88, 2016. Cited By :43.