Do students with learning disability benefit from adaptive learning software?

Summary

An important challenge of education is the tailoring of learning material and teaching methods to the specific demands of the group and individual. The Goldilocks project brought people from institutions of vocation and education training for people with learning disabilities from three countries together to investigate the potential benefit of the fact learning method *SlimStampen*. This scientifically informed learning software adapts to the individual student's learning rhythm and has been shown to lead to excellent results in elementary school classes and university courses. The project was successful in addressing practical concerns of administering *SlimStampen* to students with disabilities. The computer learning sessions were very well received by both teachers and students and emerged as a promising new classroom tool. However, results regarding the benefit of *SlimStampen* remained inconclusive.

Project report of the GOLD project (Goldilocks based learning overcomes learning disabilities). GOLD was part of the Leonardo da Vinci initiative of vocational education and training and funded by the European lifelong learning program (http://www.gold-leonardo.eu).

Individual learning variability

A major challenge of teaching is that all students are different. This is true for any training environment, but particularly apparent in schools for people with special educational and often physical needs, which may include the whole range of phenotypes that do not fit into mainstream education (e.g. a student with severe brain trauma and strong cognitive impairment or a student with dyskinesia and average intellect). There is a clear tradeoff between maximizing individual learning outcomes and including as many students as possible into the system, because the farther apart the students' abilities are the less efficient a joint class will be, leading to either boredom or despair for the students. However, the problem of individual variability between students is only partially reduced, but not solved by dividing the student population between institutions or classes. A primary goal of education improvement is therefore the research of techniques that might augment individual learning outcomes in a group setting.

Adaptive learning with *SlimStampen*

One such technique was developed by Dr. D. H. Van Rijn and graduate student M. Nijboer from the University of Groningen and is called *SlimStampen* (Dutch for clever cramming). SlimStampen is a scientifically based fact learning software environment that is based on models of human semantic (factual) memory. It adapts to the individual student's responses and calculates when a particular item should be presented again so that it is neither too easy nor too difficult to recall. In other words the retrieval effort should be *just right* for the particular student who is using the learning environment.

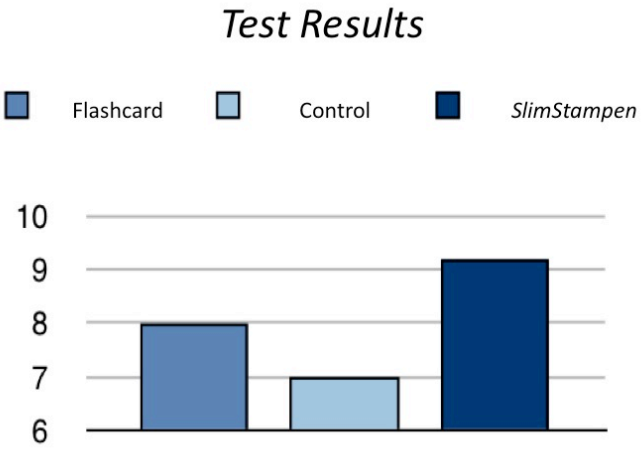

Figure 1. Test results of a Dutch high school French class with twenty students. Fifteen minutes of a normal lesson were given as autonomous learning period. In the Flashcard condition students were presented with groups of five words and proceeded to the next group when one list could be recalled correctly. Test results were obtained on a surprise examination the day after the learning session.

SlimStampen had already been shown to yield better learning outcomes than other learning software in Dutch school and university settings (see Figure 1; Van Rijn, 2010). However, it is still unknown how versatile the learning method is in practice, i.e. to what student groups and study items the benefit extends. To this end the

befittingly named Goldilocks (or simply GOLD) project was established in order to test the effectiveness of *SlimStampen* in a setting of vocational education and training for people with learning disabilities. In line with previous studies it was predicted that *SlimStampen* would be both more beneficial with regard to knowledge retention, as well as more motivating and fun to work with.

An easy and intuitive way to understand *SlimStampen* is to compare it to a tutor who helps a student by probing him/her with questions about the facts to be remembered. Many people have probably experienced this situation at one time or another in their lives. The tutor will try to adapt to the responses of the student and ask for items that the student struggles with or answers incorrectly quite frequently. Conversely the tutor will soon stop to check for items that the student can answer immediately. The tutor presumably knows by intuition at what time to probe for what fact, but there is actually a straightforward way to calculate the time when a fact should be retrieved again by considering the student's prior responses. This is how *SlimStampen* can substitute for the role of a personal tutor. This estimation process is based on two principles of fact learning, namely the importance of active retrieval of a fact from memory (*testing effect*) and the way in which those learning instances are spaced out over time (*spacing effect*).

Spacing effect and testing effect

The *spacing effect* states that spacing learning periods out over time is more effective than conducting one long learning session (Ebbinghaus, 1885; Figure 2). Importantly, this effect is simply a function of time and also holds within a single learning session. The *testing effect* states that retrieving information from memory increases the likelihood of future retrieval success more than perceiving the information from the outside and understanding it, e.g. by reading it (Karpike & Roediger, 2006).

Any decent learning method incorporates the testing and spacing effect in one form or another. The flashcard method is a well-known example. The facts are written on cards and put in a "starting box", meaning that they had never previously been rehearsed. Now, each time the learner successfully retrieves a fact the corresponding flashcard is put in a higher-level box. When a card reaches the last box it is declared as having been committed to long-term memory. However, no matter which box the card is in, when a fact cannot be recalled correctly the card has to be put back into the first box. This method is widely used and promises good results. It is very unlikely that any items reaching the last box will actually be forgotten any time soon. At the same time it incorporates individual differences by having a resetting mechanism after retrieval failures. However, this way of assuring good learning outcomes is extremely conservative, as many items would have probably already been retained at an earlier stage.

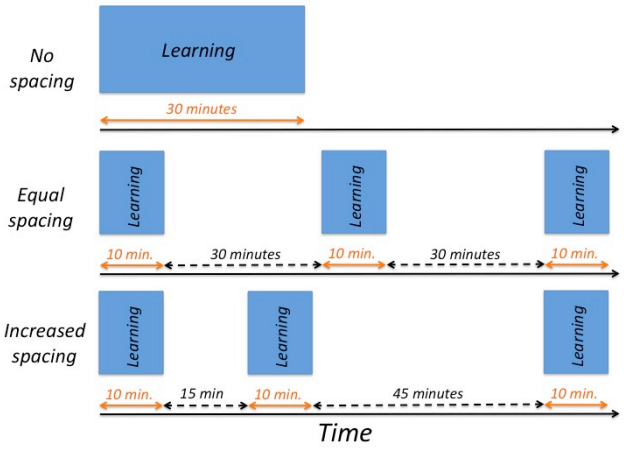

Figure 2. Three learning strategy schedules are shown with different spacing between sessions. Equally spaced learning intervals (mid-panel) are more beneficial than one long learning session (top panel). Further benefit comes from increasing time intervals from one session to the next (low panel). Note that the same principle applies within a single learning session for individual study items.

The improvement of *SlimStampen* or the advantage of having a personal tutor is that the times and number of rehearsals can be tailored for each item and student individually. A very difficult item in the flashcard system will be answered incorrectly many times, because the spacing between seeing an item is simply too great. In contrast, *SlimStampen* will soon "realize" that the initial spacing for this item is inadequate and re-present the item after a much shorter time, thus increasing the likelihood of successful retrieval and optimizing learning outcomes.

Memory activation and decay

Box 1

Figure 3 illustrates the activity of a certain fact in two distinct situations. The instance of a fact-retrieval is denoted by a peak, after which the activation slowly decays over time. The first peak, where the line starts, denotes the first time the item has been encountered and stored in memory. The blue, solid line shows the activity of a fact that is recalled four times in rapid succession (cramming). The red, dotted line on the other hand shows the activity of a fact that is retrieved from memory only three times, but with more time in-between retrievals (spacing). As can be seen, the activity of the blue fact decreases more rapidly after the last rehearsal than the activity of the red fact, in line with the spacing effect. Furthermore, the speed with which activity drops after each rehearsal decreases with each repetition and, importantly, depends on the activity level at the moment of retrieval. In other words, when a fact is still very fresh in memory, and activity is high, it is easy to recall, but it also decays away more quickly than if activity had been lower beforehand. This leads to the relative benefit of cramming knowledge immediately before an exam, but forgetting most of the facts shortly afterward.

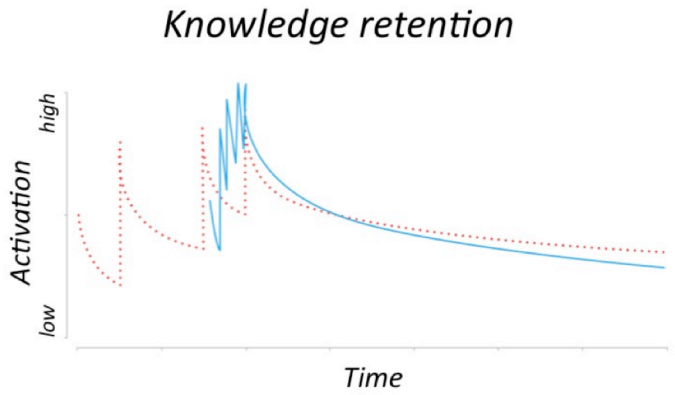

Figure 3. The red, dotted line shows the activation of a fact in memory that is recalled three times as seen by the peaks. The blue, solid line shows the activation of a fact with the same difficulty, but shorter intervals in-between retrievals and thus poorer long-term retention.

In Figure 3 the facts are assumed to be equally difficult to illustrate the effect of spacing and activity levels at the moment of retrieving a fact. A more realistic scenario is depicted in Figure 4, with one easy item (blue, solid) and one difficult item (red, dotted). As can be seen, the difficult item needs to be rehearsed five times to be roughly on the same level of activity as the easy item with only one rehearsal.

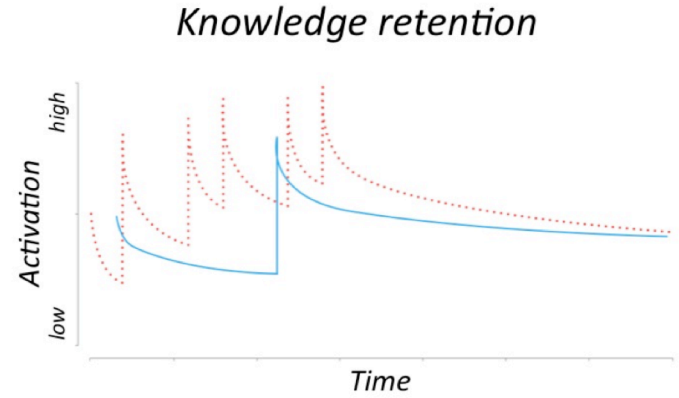

Figure 4. Activations of two facts in memory can be seen. The blue, solid line denotes an easy fact with one rehearsal, whereas the red, dotted line shows the activity of a difficult fact that has been rehearsed five times, resulting in a roughly equal activation in the end.

According to these memory models, the ease with which a fact can be recalled or the time needed to recall it is directly proportional to the time until the item should be retrieved again to yield optimal fact learning efficiency. That is why SlimStampen can calculate when an item should best be presented again to be a challenge for the learner, but not impossible to recall, optimizing the utility of the spacing effect. Similarly, it recognizes when an item is easily remembered and thus needs few rehearsals compared to more difficult items.

Method and study procedure

Box 2

Participants

A total of 110 participants took part in the study whose data was retained for analysis (see Table 1). Exclusion criteria were insufficient attendance of learning sessions (at least three were required), as well as unfinished exam sheets (less than half of the questions were answered). Otherwise, only data was excluded that was marked as ambiguous by the teacher who administered the experiment (e.g. when a student completed a session with another student's account, because the latter forgot to log out).

Procedure

In each group participants were randomly assigned to either studying with the *SlimStampen* or flashcard method. Teachers assisted students with logging on to the website that provided the study environment. Sometimes assistance was also required when putting in the students' answers. In such cases students were tested sequentially, otherwise testing was done in parallel on multiple workstations.

Material

Partners provided the learning material for each of their courses, which consisted of a mixture of multiple choice and essay questions for a first aid course in Orchardville, two lists of multiple choice questions for an IT course and a legislation course in CRPG, as well as three lists of essay questions regarding network administration in REA college. Some of the Orchardville question items also included pictures instead of or in addition to text.

Table 1

Overview of participants per condition and study group.

Group	**Date**	**SlimStampen**	**Flashcard**	**Total**
CRPG – ICT 1	23.07.2013	4	4	8
CRPG – ICT 2	26.11.2013	4	3	7
CRPG – legislation	13.09.2013	9	8	17
		17	15	32
Orchardville – first aid 1	05.07.2013	4	3	7
Orchardville – first aid 2	30.08.2013	4	4	8
Orchardville – first aid 3	11.10.2013	4	6	10
Orchardville – first aid 4	06.11.2013	6	5	11
		18	18	36
REA college – ICT list 1	07.11.2013	5	8	13
REA college – ICT list 2	07.11.2013	8	6	14
REA college – ICT list 3	07.11.2013	6	9	15
		19	23	42
		54	56	110

In detail, *SlimStampen* simulates the decay process of each fact in long-term memory according to theories of semantic memory (Pavlik & Anderson, 2005; Taatgen, 2009). According to these theories, the knowledge of a particular fact is expressed by how "active" that fact is. This activity value can be calculated depending on how often the fact has been retrieved, as well as when it has been retrieved. In Box 1 you can learn more about this knowledge retention.

Project GOLD

Three European institutions of vocational education and training for people with disability took part in the project: The Orchardville Society from Northern Ireland, CRPG from Portugal, and REA College from the Netherlands. The project was planned and conducted with the support of the University of Groningen. GOLD stands for *Goldilocks based learning overcomes learning disabilities* and is part of the Leonardo da Vinci initiative of vocational education and training and

funded by the European lifelong learning program. An account of the experimental design, participating students and courses, and an overview of the study's procedure can be found in box 2.

Hypotheses

It was hypothesized that the SlimStampen learning method would be superior to the flashcard method with respect to knowledge retention. In addition it was hypothesized that SlimStampen would be associated with less frustration during study periods in comparison to the flashcard system.

Findings

With respect to knowledge retention the current results indicate that there is no difference between learning methods (Figure 5). However, small sample sizes within groups and variability between participants, groups, teachers and locations naturally result in low statistical power and further data acquisition is certainly needed to draw strong conclusions (see box 3 for analyses).

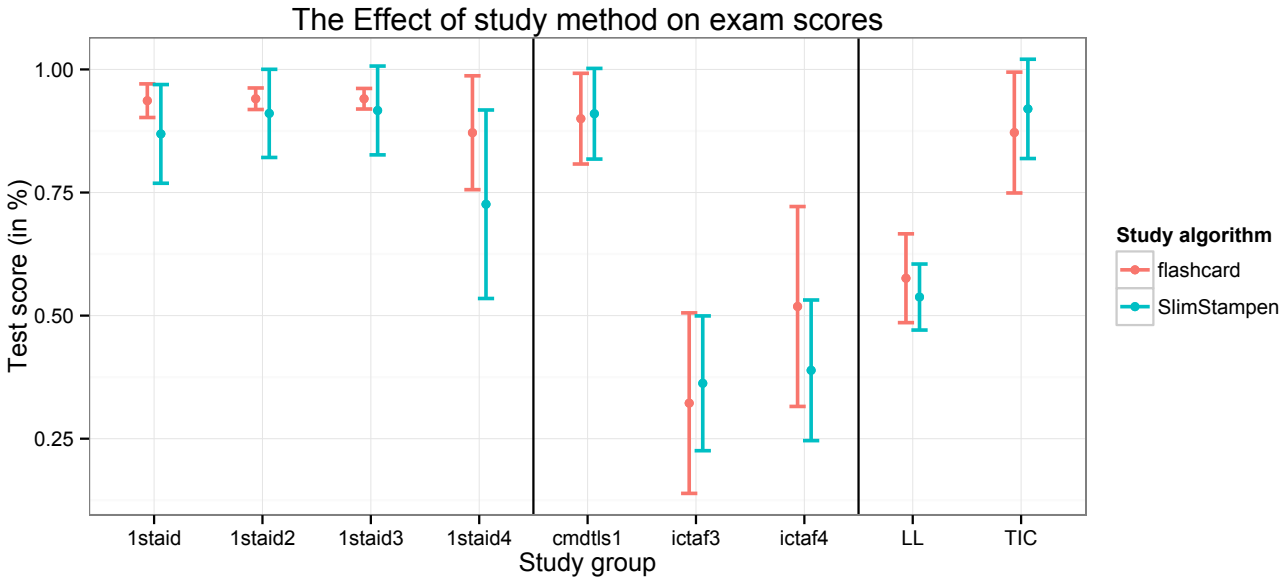

Figure 5. Final exam scores shown for each class and contrasting students who learned with the *SlimStampen* or flashcard method. Error bars represent 95% confidence intervals. That means that 95 out of 100 times, if the experiment were repeated, the average score would fall within the area of those bars. From left to right groups 1-4 belong to Orchardville society, groups 5-7 belong to REA College and groups 8-9 belong to CRPG.

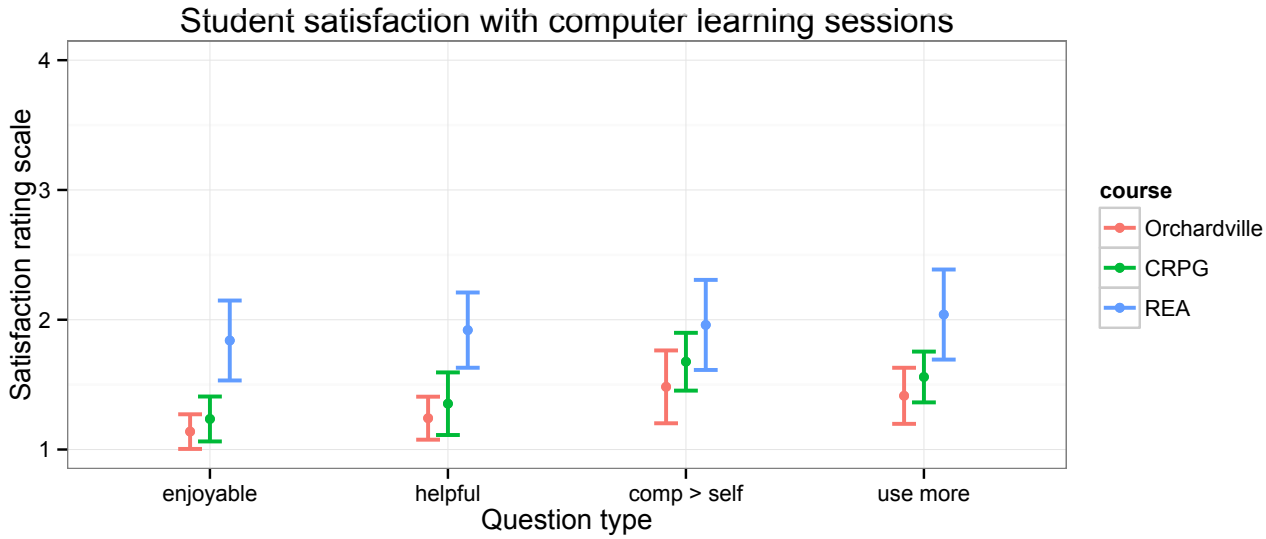

Figure 6. Learning satisfaction measured with a four point Likert scale (1 = completely agree, 4 = completely disagree) sorted by VET institution. Error bars denote 95% confidence intervals.

Statistical analyses

Box 4

Analysis

For statistical analysis the software R 2.15.2 was used (R core team, 2012), including the packages lme4 for mixed effects models (Bates, Maechler & Bolker, 2011) and ggplot2 for graphical representations (Wickham, 2009).

Results

A logistic mixed regression analysis of exam scores with study method as fixed effect and study group as random effect showed no significant effect of study algorithm (z = -0.577, p = 0.56). Including number of items seen and number of item rehearsals in the model showed no interaction between study method and number of items seen (z = -0.025, p = 0.98) or number of total rehearsed items (z = 0.165, p = 0.87). Furthermore there were no significant main effects of items seen (z = -0.505, p = 0.61) or items rehearsed (z = -0.576, p = 0.56). Overall the simple model is to be preferred over the more complex model (ΔBIC > 10). An overview of the results can be seen in Figure 5.

Furthermore, multiple-choice questions were included to make the testing sessions easier for the students, as typing answers on the keyboard would have been problematic for some. However, multiple-choice questions are fundamentally different from essay questions in that simple recognition of the answer is sufficient. This undermines the assumption made by *SlimStampen* that a correct answer is due to retrieval success (*testing effect*) and is expected to at least weaken its beneficial effect on retention.

Finally, the project was helpful in finding out what factors and potential issues need to be considered when designing a virtual learning environment for students with disabilities. That process, however, lead itself to a number of confounds and sources of noise in the data that undermine the strength of conclusions to be drawn. For example, the adaptations of the website to the students' needs changed the conditions between different classes, which probably by chance affected the SlimStampen group differently than the flashcard group (e.g. the screen resolution of some classes was so low that the response window of questions with pictures could sometimes not be seen without scrolling down, leading to a change in layout of that part of the website).

Unfortunately, testing learning satisfaction as a function of learning method was not possible, as questionnaires were completed anonymously. However, this has the advantage of students being unbiased in their evaluation of the computer learning sessions in general. As can be seen in Figure 6 students consistently rated learning satisfaction associated with studying with the computer as high, regardless of study method. The figure also shows that there is some variability between groups, with REA college students being somewhat less satisfied than students from the other facilities. This might well be due to inherent group differences, as well as differences in learning material. Students at REA College were probably the most capable in terms of general intelligence and reasoning ability, which is reflected in the students' comments about the learning sessions, which were also completed anonymously. For example, in some comments students reflect upon the usefulness of simply studying facts without thinking about why that is the correct answer.

Even though there is no quantitative and controlled assessment of student's evaluation of SlimStampen or the flashcard method, the overall impression by the teachers was that *SlimStampen* generally lead to less frustration during study sessions. Moreover, the computer sessions improved students' and teachers' satisfaction with the courses and lead to higher test performance and improved long-term knowledge retention as perceived by the teachers compared to previous years in which no computer learning sessions were administered. Consequently, it is not surprising that all partner institutions decided to continue to use the computer learning-environment in future classes and from now on only enable the *SlimStampen* learning method.

Conclusion

The GOLD project was not able to provide evidence for the usefulness of *SlimStampen* in a vocational education and training setting for people with learning disabilities. Nevertheless, it clearly showed that using electronic learning environments to help students to study facts increases learning satisfaction ('enjoyable') and is superior to normal classes ('use more') or when they have to study without the computer ('comp > self'). As the basic principles of *SlimStampen* are supported by scientific research and it has shown its merit in other student populations it is reasonable to assume that it might ultimately also benefit students with learning disabilities. This has to be borne out by future studies.

Authors

Steffen Bürgers, B.Sc., University of Groningen
Dr. Hedderik van Rijn, department of experimental psychology, University of Groningen
Enid Reichrath, M.Sc., toetsen meten & weten

References

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (Version 0.999999-2) [computer software]. Available from http://CRAN.R-project.org/package=lme4

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology.* Teachers College, Columbia University, given by Henry A. Ruger and Clara E. Bussenius (1913). Available on http://psychclassics.yorku.ca/Ebbinghaus/index.htm

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
ISBN 3-900051-07-0, URL http://www.R-project.org/.

Roediger, H.L., & Karpicke, J.D. (2006). Taking memory tests improves long-term retention. *Psycholgical Science, 17* (3), 249-255.

Taatgen, N.A. (2009). Kennisopslag, vergeten en geheugen. In R. Klarus & R.J. Simons (Eds.), Wat is goed onderwijs? Bijdragen uit de psychologie (pp. 33-46). Den Haag: Lemma.

Van Rijn, H. (2010). SlimStampen. Optimaal leren door kalibratie of kennis en vaardigheid. http://onderzoek.kennisnet.nl/onderzoeken-totaal/slimstampen

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New
York.