Does Overt memory testing lead to better learning than Covert memory testing?

Tara Soltani

Handledare: Fredrik Jönsson
VETENSKAPLIG UNDERSÖKNING 15 HP, VT 2011
STOCKHOLMS UNIVERSITET
PSYKOLOGISKA INSTITUTIONEN
DOES OVERT MEMORY TESTING LEAD TO BETTER LEARNING THAN COVERT MEMORY TESTING?

Tara Soltani

The testing effect is described as an improvement of memory performance at a final test as a result of previous test sessions. Several studies reported that initial testing during learning provides a better memory representation at a delayed retention interval. The aim of present study is to assess whether an Overt or a Covert retrieval procedure leads to better final recall. Participants were at initial testing instructed to retrieve and write down the words (Overt group), retrieve it without writing down (Covert group) or only restudy the words (SSS group). Memory was tested after 15 minutes and again after 1 week. The results indicated a significant main effect of learning group and retrieval interval, where the Overt group performed reliably better than remaining groups. The idea behind the testing effect may work as a guideline of how the testing procedure can be conformed into educational contexts, which advocates for more consistent use of testing in classroom for optimal learning.

Cognitive researchers have demonstrated that learning occurs during testing and that repeated testing can, under some circumstances, improve memory retrieval at a later memory recall more than an equal amount of repeated studying does (e.g., Butler, 2010; Carpenter & Pashler, 2007; Carpenter, Pashler & Cepeda, 2009; Karpicke & Roediger, 2008; Leeming, 2002; Roediger & Karpicke, 2006; Rohrer & Pashler, 2010; Wheeler, Ewers & Buonanno, 2003). It is well known that elaborative study is beneficial for subsequent recall. Learning occurs when people are studying and encoding a given material, whereas testing is considered as a relatively natural way of measuring the subject’s learning during the study period. Activities requiring students to practice retrieval are used less frequently in teaching. The basic idea of the testing effect is that testing itself enhances the amount of learning of the material compared to elaborative studying. Present study focuses on investigating the difference between these two study and testing strategies.

Theories of the testing effect

Additional Exposure and Overlearning.

One idea behind the testing effect suggests that a test provides additional exposures to a material and these extra exposures lead to the testing effect (Thompson, Wenger & Bartling, 1978). Many experiments have compared a condition where students studied the material and took a delayed final test with another condition where students studied the material, took an initial test and a delayed final test. The latter showed better performance on the final test (Roediger & Karpicke, 2006). Other experiments have represented material for study in the control condition and nevertheless obtained a strong testing effect. The usual restudy control is given a greater exposure to the

I would like to thank my supervisor Fredrik Jönsson for all the help and support. He developed the idea and the structure of this work.
material since subjects are re-exposed to the material they could produce on the test before, which suggests that other processes as well are responsible for the effect. Some researchers have argued that the act of testing may reflect an “overlearning” of materials practiced earlier on a test (e.g., Thompson et al., 1978), suggesting that overlearning provides a better delayed retention rather than the process of retrieval itself. This explanation can be problematic, explaining why studying promotes better short-term retention whereas repeated testing produces better long-term retention (e.g., Roediger & Karpicke, 2006). This implies that repeated studying results in overlearning on immediate tests but is not reinterpreted into a better long-term retention since most testing conditions produces better recall on delayed tests than repeated studying conditions. It is also suggested that retrieval processes such as more elaboration and effortful encoding engaged in a test are responsible for increased learning. On the other hand investigators may indicate that some items are easier to remember than other items for different reasons and the easy items are recalled both on initial test and again on the final test, Thus, it creates an illusion that the test itself causes better learning (Roediger & Karpicke, 2006), because the easy items receive additional practice through the testing and are better recalled later.

**Effortful Retrieval and Desirable Difficulties.**

A second theory, suggests that some aspects of the retrieval process must be a factor behind the testing effect since overlearning and additional exposure could not explain all of it. A variety of retrieval hypothesis has been advanced and many writers have claimed that retrieval effort explains some parts of the testing effect (e.g., Gardiner, Craik, & Bleasdale, 1973). Tests that require recall of a material provide clearer testing effect than other methods such as recognition. One possible reason is that test recalls need more effort and deeper level of processing than recognition tests do (Bjork, 1975). Relatedly, increasing the spacing of initial tests provide better retention in due to enhanced retrieval effort, which could reflect greater results on delayed tests (Karpicke & Roediger, 2006; Whitten & Bjork, 1977). Many studies demonstrate the positive effect of retrieval effort of delayed retention, for example Gardiner et al., (1973) tested students on general knowledge and measured how long time it took them to answer the questions. At the end of the lecture, students were given a final free recall test on the answers. This indicated that the longer time it took students to produce an answer the higher was the chance that they would recall that answer on the final free recall test. Bjork (1994, 1999) argued that teachers should concentrate on creating desirable difficulties in order to increase student’s learning. Delayed feedback and spaced practice are techniques that are responsible for forming a desirable difficulty. However, testing makes up a desirable difficulty compared to studying (Roediger & Karpicke, 2006). People are not often engaged in difficult learning activities even the fact that such activities may enhance learning, for example Baddeley and Longman (1978) demonstrated an experiment where they used either massed or spaced practice conditions as they trained a number of subjects typing on keyboard. Subjects preferred to a greater extend the massed practice condition even though spaced practice appeared to be better than massed practice. The testing effect is one example of how conditions that are more effortful and slower at initial learning are also able to produce stronger long-term retention relative to conditions that contain rapid gains since this provide poorer long-term retention. Students may choose not to retest themselves during learning because repeated testing is perceived as much more effortful and demanding than repeated studying. Teachers as well may decide not to give students many tests during classes (Baddeley & Longman, 1978).
Transfer Appropriate Processing.

Finally, a third theory of the testing effect is the transfer-appropriate processing account. To minimize forgetting the processes that are required at retrieval attempts should achieve perfect match between encoding and prior learning. Thus, the same processes that are involved at encoding should also be present at the retrieval opportunity (e.g., Fisher & Craik, 1977; Morris, Bransford & Franks, 1977). Taking a test allows students to practice and engage in retrieval processes during learning, leading to practicing the same processes needed to improve subsequent retrieval. This permits a greater retention than engaging in encoding activities like repeated reading (Roediger & Karpicke, 2006). The concept of transfer-appropriate processing offers an explanation for the testing effect, which means that testing is beneficial for learning since testing leads students to engage in retrieval processes that transfer information in long-term memory to later contexts and situations (Roediger & Karpicke, 2006).

Previous research and the importance of the testing effect

Tests are commonly used as an assessment to measure students learning ability in educational contexts. Previous findings suggest that taking tests are more preferable than studying and the material is more likely to be successfully recalled in the future. Tests can be utilized as instruments to promote learning and retention, as demonstrated by Roediger and Karpicke (2006). According to them, testing is not only for measuring knowledge it may even have the capacity to increase the amount of knowledge we possess. They assume that testing is a powerful way of improving the memory of a material even when no feedback is given to the subjects. This phenomenon describing enhanced memory performance after a test is made is called the testing effect and although the concept has been studied before by cognitive psychologists and although it has been the subject of many experimental studies over the years it is, according to Roediger and Karpicke (2006), neither appreciated nor used in many educational contexts. Studies show that the testing effect is strongest for longer time intervals, but often absent at shorter time intervals. One reason for this is that the memory naturally declines the longer time passes between encoding and retrieval, regardless of performing a test or not (Roediger & Karpicke 2006).

Many educators refer to that testing in school should be minimized so that more time will be spent on lectures (Roediger & Karpicke, 2006). The most basic courses in the universities have few tests in due to midterm and final exams. Repeated test on a material may have a greater impact on future memory ability than devoting additional time restudying the material. Karpicke and Roediger (2008) argued that introducing frequent testing in the classroom might enhance educational achievement at all education levels. They let college students learn a list of 40 Swahili-English word pairs. They manipulated whether the word pairs remained in the list or were dropped out immediately after they were recalled. Students began by studying the list of Swahili-English words both in the period of studying and testing, (e.g., iktisadi - ?). There were a total of four conditions (ST, SnT, STn and SnTn; S = studying, T = testing). The first condition (ST) was characterized by the fact that both restudying and retesting of all the words were completed. In the second condition (SnT) subjects studied the whole list and once a word pair was correctly recalled it was dropped from further study. After the first round it was only the words that subjects had not remembered in previous test period that appeared in the studying period. In the third condition (STn), correctly recalled words were also dropped from further testing but were studied in each follow up study period. In the final condition (SnTn), recalled word pairs were dropped out in both the
In addition, Carpenter et al., (2009) used the same concept to assess 8th Grade Student’s retention of U.S history facts. A sheet of paper containing 30 questions was given to each student (e.g., Who assassinated president Abraham Lincoln?). The answer was provided for 15 of these (e.g., John Wilkes Booth) but not provided for the remaining 15. For students whose answer was provided, denoted as study item, were asked to read both question and answer. For students whose answer was not provided, denoted as test/study item, were instructed to read the question and write down the answer. Students then received a review sheet with 15 items randomly assigned to be reviewed through restudying (study), 15 items randomly assigned to be reviewed through testing (test/study) and last 15 items randomly assigned to No Review. The 30 items that showed up on the review groups (study and test/study) occurred in a different order for each student and they were asked to complete the review and then receive the answer sheet, which was created to contain the same 30 items that appeared on their review. The 30 items now appeared in alphabetical order, which was intended to increase the probability that students would read each item again. They were then instructed to read the questions and answer all 30 items. All students received one review session and one final test session. Facts were reviewed after one week for the Immediate Review Group, 16 weeks for the Delayed Review group or not reviewed at all. All students received a test nine month later after they completed the review. The final test contained the same 15 study items and 15 test/study items that had showed up on the review along with 15 items that had not appeared on the review. They were informed to write down an answer to each of the 45 questions. Results showed that facts reviewed through testing were significantly beneficial to 8th grade students’ retention of facts from U.S History than facts reviewed through restudying and twice as well as those given no review. Review by testing after a 16 week time interval appeared to be the best retention for facts and these results support the notion that testing can increase long term retention of a given material.

Moreover Leeming (2002) demonstrated that multiple testing during courses lead to students continually studying the course material rather than just having an intense learning period just before the exam. Leeming had a class studying introduction in psychology twice a week whom he tested each time during the first 10-15 minutes of each class. Thus, a typical class would have time for 24 to 24 exams per semester. From another section of Introductory Psychology, students had only three exams during the same semester. The retention test covered the chapters from the textbook and contained randomly short essay, “fill in the blank” and multiple-choice questions. Students in each class completed a questionnaire where they indicated whether they agreed, disagreed or were unsure concerning “one-exam-a-day” procedure. Responses showed a clear positive evaluation where students reported that the method forced them keep up with the material, thus led to more studying. However results regarding student’s performance revealed that exam scores were higher for students that were given daily exams than students in traditional classes. The number of high grades was increased and
the number of low grades was decreased in all classes with repeated testing procedure. Leeming (2002) suggested that the higher scores of the repeated testing class were a result of better initial learning rather than some effect of memory processes. Finally students reported that the “one-exam-a-day” increased studying and learning.

The difference between Overt and Covert testing

Almost all previous studies investigating the testing effect have used an Overt retrieval procedure during testing in the learning phase. More specifically, subjects explicitly retrieve the items by writing down their responses, alternatively saying them out loud. Covert retrieval on the other hand is less used in studies of the testing effect (e.g., Carpenter et al., 2006; Carpenter & Pashler, 2007; Carpenter, Pashler, Wixted & Vul, 2008). This procedure shares much with the Overt retrieval procedure because the students are explicitly instructed to test memory, but they are not asked or allowed to output their responses in any way (e.g., by writing it down or saying it aloud). The main aim of the present experiment is to directly compare the strength of the testing effect as a function of an Overt or a Covert retrieval procedure. A study only condition is included to see if we replicate the testing effect. No previous research has made this comparison between these testing procedures. Will a Covert retrieval procedure lead to as strong testing effect as the more commonly used Overt procedure?

Although no direct comparison has been made between the two procedures, some few studies have used a Covert procedure (e.g., Carpenter, Pashler & Vul, 2006; Carpenter & Pashler, 2007; Carpenter et al., 2008). For example, Carpenter and Pashler (2007) examined whether Covert testing can enhance learning of visuospatial information in maps. The main goal of the study was to assess whether the testing effect could be obtained with a visuospatial task that did not require an Overt verbal response. Participants were instructed to learn two maps containing 12 features each, such as rivers, buildings and roads. One of the maps was learned through additional studying (study condition) and the other map was learned through testing with feedback (test/study condition). Participants learned the first map through one method (e.g., test/study) and then learned the second map through the other method (e.g., study). There were a total of four conditions, where Map A was presented for either study or test/study followed by Map B, which was also presented for either study or test/study. Participants were randomly assigned to one of the four conditions. They were informed that they would learn two maps by using both the study procedure and the test/study procedure they practiced. All participants learned both of the maps, first one through completing a test followed by feedback (test/study) and the last one through spending additional study time (study). During test/study, participants studied the map with 12 features and were then asked to figure out which feature was missing out of 12 and to form a mental image of it and its location. After they had formed the image they saw the complete map and this procedure was repeated with a different feature until all 12 had been tested. During study, participants studied the map with all 12 features and were then given opportunity to restudy the complete map again. After a visual attention task during 30 minutes all participants were instructed to draw both maps they had been shown. Participants map drawing appeared to be significantly better for maps when tested covertly unlike maps learned through studying. Testing produced enhancement of map learning. According to Carpenter and Pashler (2007) this study was the first demonstration where the testing effect could be found through visuospatial task and did not require any Overt verbal response in contrast to other studies with the same topic.
that requires typing, writing or speaking a response loud which shows that the testing effect is not limited.

Further study with similar basis, using Covert retrieval was presented by Carpenter et al., (2006). In their experiment they investigated what types of learning benefit from a cued recall test. At first subjects were presented with 40 weakly related cue-target pairs in a study sessions and afterwards, they were given an additional chance to learn each word pair. This was performed through either restudying the pairs (A+B), denoted as study trials or a cued recall test (A→?), denoted as test/study trials and then followed by a presentation of the correct pair (A+B). More exact, subjects were exposed to 40 word pairs, one at a time. Subjects completed a test/study trial on 20 of the pairs where they had to covertly retrieve the target (A→?). The target then appeared and both items were shown. For remaining 20 pairs, subjects were given additional chance to view the cue and the target, i.e. study trial. The following day, subjects returned to complete the experiment. Testing revealed a benefit at final testing compared to restudying. Consequently the subjects in the test/study trials produced higher final test accuracy. Later on, Carpenter et al., (2008) investigated whether memory tests enhance learning and reduce forgetting more than additional study opportunities do, using Covert retrieval attempts. Subjects learned obscure facts or Swahili-English word pairs by either receiving additional study chance (study) or completing a test (test/study). They were tested after 5 min or up to 42 days. In all experiments testing increased recall more than studying did. However the rate of forgetting was consistently lower for the test/study group than for the study group. Since this study revealed a reduction in forgetting in due to testing, results are in line with prior studies, all showing support for testing (e.g., Karpicke & Roediger, 2008; Leeming, 2002; Roediger & Karpicke, 2006; Thompson et al., 1978; Whitten & Bjork, 1977). Further, a probable reason why the act of tests may be beneficial could be the fact that they reveal which items have been learned and which ones that have not been learned and require more studying. More correctly, when subjects try to recall an item and fail, they may find a new way of encoding the material in subsequent study opportunity (Carpenter et al., 2008).

Aims

In sum, most testing effect experiments have required the participants to combine study with overt memory testing during an initial learning session. The first and main aim of the present experiment is to investigate whether a Covert testing procedure leads to a testing effect comparable to that of an Overt testing procedure. This will be investigated in two (related) ways, first in terms of absolute performance levels at two different retention intervals (15 min and 1 week), and secondly in terms of the degree of forgetting across a week. The second aim is to replicate the testing effect, and for that reason a study only condition was also included as a control task. With support of previous research, repeated testing is expected to increase learning more than repeated studying does since participants are given opportunity to actively retrieve the words. To clarify, both the Overt and Covert groups are expected to demonstrate a lower degree of forgetting than the study only group.
Method

Participants
A total of 60 participants (42 females and 18 males), with a mean age of 24.27 (SD = 3.53) years were recruited to this study through public information at Stockholm University. The participants were given either a movie voucher or course credit as compensation for their participation. The participants were randomly assigned into one of three groups, Overt, Covert or a repeated study group, denoted as SSS.

Materials
The experiment was run on PC computers with E-prime 2.0 software (Psychological Software Tools, Pittsburgh PA). A total of 40 word pairs were used, namely 20 Swedish-Swedish word pairs taken from Shaps, Johansson and Nilsson (1974) and 20 word pairs from a study by Nelson and Dunlosky (1994). Nelson and Dunlosky (1994) used Swahili-English word pairs, which were translated into Swahili-Swedish word pairs for this study. Swahili is often used because it has the same standards as English alphabet and also relatively few subjects have been exposed to Swahili vocabulary (Nelson & Dunlosky, 1994).

Procedure
The experiment took place in a laboratory located at Stockholm University in the Department of Psychology. Participants were tested individually in front of a computer where they were exposed to 20 Swahili-Swedish word pairs and 20 Swedish Association word pairs. Before the experiment started participant were asked to sign a consent where they were informed about the purpose of the experiment and their right to drop out at any time during the procedure.

Learning Phase
In the first part of the experiment, participants began with a practice session before the actual task, in which two word pairs were shown consisting of a word in Swahili and the translation in Swedish, for example “malkia – drottning”, as well as two mathematical filler activities were shown as a distraction moment, for example, $12 \times 60 = 720$. Number 1 on the keyboard indicated “True” answer and 0 indicated “False” answer. Participants chose one of the two options. This distraction element was a repeating element between each study session and each test session in learning phase. After the practice session participants were exposed to all 40 word pairs and were informed to attempt to memorize all of the items (consisting of Swahili and Association words) that appeared each in a time, followed by each other. This procedure was repeated totally three rounds (3 study and 3 test sessions) for the two testing groups (Overt and Covert), or just 3 study sessions for the third repeated study group (SSS).

After one round of exposure consisting of all 40 words in the study session, two of the three groups were asked to either retrieve the words overtly by writing down the equivalent words on the keyboard or retrieve the words covertly by an attempt to only rehearse them in test session. The third group did only restudy the word pairs in 3 study sessions without being tested in the learning phase unlike the other groups, more exact they saw all the 40 word pairs three times after each other. The study list was presented at a rate of one word every 6 seconds in the study session and they were given 10 seconds to either retrieve the words and visualize them (covertly) or retrieve the words by writing them down (overtly) in the test session (depending on group belonging).
before the computer program automatically proceeded on to the next word. However for the repeated study group (SSS) all words were constantly running without any recall attempt. After each study session all participants performed the 45 seconds verifying multiplication task as filler activity. All 40 word pairs were both studied and later on tested.

**Filler Activity**

In the second part of the experiment, participants were instructed to perform a working memory test containing numerous mathematical tasks during approximately 15 minutes while they also had to study letters. The mathematical figures were combined with a letter for instance (8/2) + 6? while letters such as H, F, Q appeared. This was meant to serve as a distraction moment between the first part and the third part of the experiment however this mathematical task will not be analyzed in present study.

**Short Recall Test**

Just before the end of the experiment’s first part all groups were tested on all 40 items after a short interval of 15 minutes. Each word pair appeared for 6 seconds and participants in all groups received the instructions to write down the Swedish translation equivalent or the association equivalent before the computer program automatically continued to the next word, e.g. “Ndoo -?” or “Stearin - ?”. Participants were thanked and reminded about the follow-up next week.

**Final recall test**

Participants were then asked to return after one week to complete the last part of the experiment, also denoted as delayed final recall test. It consisted of all of the 40 word pairs exposed earlier in the learning phase. Like previous study all groups were shown a word and were then asked to write down the corresponding word. The final recall test was identical to the short interval recall test. All groups were given the same amount of time and the same word pairs were tested.

**Design**

A 2 x 3 mixed-factorial design was used with Retention Interval, Groups and Memory performance as variables. Retention Interval (Short vs. Long) was manipulated within-subjects since same participants attended both occasions and group (Overt, Covert and SSS) was manipulated between-subjects since there were three separate groups. The design contained of totally three groups and the difference between the testing groups (Overt and Covert) was in the execution of the testing procedure during learning phase. During test sessions, the Overt group was instructed to overtly write down their response on the keyboard after each “exposure-round” (all 40 pairs), whereas the Covert group was instructed to covertly retrieve the responses by explicitly rehearse the corresponding word. The repeated study group (SSS) on the other hand did only restudy the material in 3 study sessions without any use of Overt or Covert method to retrieve the words since they were not included in any test session during learning phase. This group was used as an attempt to optimize the differences between study and test strategies. All groups were treated in the exact same way as they were exposed to the same word pairs, same distraction elements and the same amount of time. To see if the result was affected by having considered the three groups differently, participants in all groups were tested before leaving (short recall test) and were reminded to return after one week to complete the experiment’s last part in form of a final recall test.
Overt | S → T | S → T | S | Filler Activity | Short Recall Test | Long Recall Test (Final Testing)
Covert | S → T | S → T | S
SSS | S → T | S | S

Figure 1. The figure illustrates number of study and test periods in every group. Each box represents a study and a test session. Overt indicates the group who were asked to write down the words, the Covert group was asked to rehearse the same word whereas the SSS group only completed three rounds of study sessions. There were a total of 3 study sessions, 3 test sessions, a filler activity task, a short recall test and a final recall test for all conditions. Participants in all groups performed a mathematical distracter task during 45 seconds after each study period and before each test period.

Results

In order to identify any effect a 2 x 3 mixed ANOVA was used with the between groups variable Learning Group (Overt, Covert and SSS) and the within groups variable Retention Interval (Short/Long recall) as the independent variables. Memory performance in the final recall tests was the dependent variable. The alpha level was set to .05 and the effect size is denoted with partial eta squared ($\eta^2_p$).

There was a significant main effect of Learning Group, $F_{2, 57} = 3.42, p < .05, \eta^2_p = .11$. This proves that memory performance depends on learning group and that group belonging is an important factor in the number of correctly recalled words. Tukey post hoc tests showed that the Overt group ($M = .75, SD = .18$) performed reliably better than both the SSS ($M = .60, SD = .18$) and Covert ($M = .72, SD = .18$) groups ($ps < .05$), whereas the latter two did not differ significantly. Means and standard deviations for performance at short respectively long recall for all groups are displayed in Table 1.

There was also a main effect of Retention Interval, $F_{2, 57} = 211.21, p < .01, \eta^2_p = .79$. As expected people remember less after a longer retention interval ($M = .59, SD = .03$) compared to after a shorter interval ($M = .80, SD = .02$). There was a learning Group x Retention Interval interaction $F_{2, 57} = 4.08, p < .05, \eta^2_p = .13$, which implies that the effect the type of retention interval had on memory performance after a week depended on learning group and the other way around as shown in Table 1.

Table 1. Means and standard deviations for all groups at short and long retention intervals.

<table>
<thead>
<tr>
<th>Group</th>
<th>Retention Interval</th>
<th>Short</th>
<th>Mean</th>
<th>Long</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overt</td>
<td></td>
<td></td>
<td>.82</td>
<td>.67</td>
<td>(.19)</td>
</tr>
<tr>
<td>Covert</td>
<td></td>
<td></td>
<td>.83</td>
<td>.60</td>
<td>(.15)</td>
</tr>
<tr>
<td>SSS</td>
<td></td>
<td></td>
<td>.72</td>
<td>.49</td>
<td>(.21)</td>
</tr>
</tbody>
</table>

As evident from Table 1, and the Group x Retention Interval interaction, the Overt and Covert group does not differ significantly at the short recall test, 15 min after learning. However, they do seem to differ at the 1 week test (see Table 1). More specifically, the interaction implies that people forget more in the Covert than Overt group, and they also seem to forget more in the SSS group. In a further analysis, we calculated a forgetting
score for each group, where memory recall of 15 min (short interval) was subtracted from memory recall after 1 week (long interval). This gives negative values that indicate how much has been forgotten over a week, see figure 1. The forgetting score was then entered into a one way independent ANOVA with learning group as between groups variable, $F_{2, 57} = 4.09, p < .05, \eta^2_p = .13$, which proved to be significant. Tukeys post hoc tests showed that forgetting was significantly lower in the Overt group ($M = -.15, SD = .10$) than both Covert group ($M = -.24, SD = .12$) and the SSS group ($M = -.23, SD = .10$), $p < .05$. The latter two did not differ significantly. To conclude, a Covert retrieval procedure leads to the same amount of forgetting over a week as in the study only condition, and both are inferior to the Overt group.

![Figure 1. Proportions of mean difference score in forgetting. Memory recall of 15 min was subtracted from memory recall after 1 week. The bars represent standard errors of the means.](image)

Discussion

The first aim of the present experiment was to investigate whether a Covert testing procedure leads to a testing effect comparable to that of an Overt testing procedure, where the Overt group was instructed to overtly retrieve the equivalent of the Swahili/Association word (through writing) whereas the Covert group was instructed to covertly retrieve them (without writing down). The second aim of this experiment was to replicate the testing effect and to tell whether repeated testing could increase memory in form of learning compared to repeated study, where a study group was added, SSS. With two different retention intervals (15 min and 1 week) we were able to tell whether we could see an additional sustainability of the word pairs as a result of testing and studying. Results showed a difference between the study group and the testing groups where the Overt group performed significantly better than remaining groups, as well as produced less forgetting. But no difference could be seen between the two latter groups. Results also showed that performance was markedly reduced after one week of delaying.

Study aimed to examine whether testing could promote greater memory performance than repeated studying due to more retrieval attempts, leading to further elaboration.
Results showed the fact that testing was more beneficial than studying. These results are in line with a number of prior research speculations about the testing effect (e.g., Carpenter & Pashler, 2007; Carpenter et al., 2009; Karpicke & Roediger, 2006, 2008; Leeming, 2002; Whitten & Bjork, 1977). This study also managed to show a significant lower rate of forgetting in the Overt group due to testing, similar to previous studies that have reported reduction in forgetting (e.g., Roediger & Karpicke, 2006, Thomson et al., 1978). A possible reason might have been the opportunity for subjects to continuously retest their knowledge, leading to elaborative processing of the material unlike the Covert group and repeated studying group. A speculation is that one elaborates further when one is allowed to write down the responses during testing in contrast to not writing down. The effect could be explained on the basis of effortful retrieval (Bjork, 1975), since recall tests need more effort and deeper level of processing than other forms of studying.

This study is consistent with the idea of transfer-appropriate processing, suggesting that memory performance gives the best outcome when processes engaged in during encoding match those engaged in during retrieval (Morris et al., 1977). According to this theory, the Overt testing group ought to perform best since the design of the delayed final testing was similar to what participants in this group performed during learning, thereby leading to the same processes involved at both encoding and retrieval. This was however proven to be true in results of present study, which showed support for the repeated testing group who was instructed to overtly produce the words in both learning phase as well as in final testing. The transfer appropriate processing theory may be an explanation for why the Covert and SSS group did not perform at the same level as the Overt group. The procedure they engaged in did not require matching of encoding and retrieval and the final testing procedure differed from what they had performed during learning phase. These two groups are not consistent with the idea and should therefore underachieve according to the theory (Morris et al., 1977). A replication of the transfer-appropriate processing could be implemented to be able to see whether the procedure at encoding can affect later retrieval. In due to restrictions of this study it remains as a target for future research.

Since there was no significant difference between the Covert group and the group that studied repeatedly (SSS) in present study, it is possible that both Covert retrieval and studying lead to the same amount of elaboration. This is in contrast to what previous research has shown as they used the same Covert retrieval procedure. The test/study conditions appeared to give significantly better results than all study conditions (e.g., Carpenter et al., 2006; Carpenter & Pashler, 2007; Carpenter et al., 2008). All studies acquired a study and a test/study trial, where the testing procedure consisted of explicitly rehearsal of the material. A possible explanation of why it could not be replicated in present study may be in due to that both these strategies required equal processing. The two testing groups might not have differed from each other as much as in studies mentioned earlier because of a variety of factors such as design, degree of difficulty or the procedure. These facts could have reduced the likelihood of detecting significant differences between the two groups. An amount of researches have been applied to the testing effect where they have compared study with test conditions. However, no one have yet made this comparison between an Overt and a Covert retrieval procedure, which is of interest because it compares two different test conditions with each other, where both have proven to promote positive impact on learning (e.g., Carpender & Pashler, 2007; Carpenter et al., 2009; Karpicke & Roediger,
In addition, both testing groups were in present study exposed as many times for the complete word pairs during learning phase, i.e., “malkia – drottning”. On the other hand, participants in the Overt group were “re-exposed” to the words since they were given opportunity to produce the requested word during all test sessions by actually writing them down, but only in cases when the words were correctly recalled. More exactly, each time they could produce the correct Swahili or Association word they were provided with an extra exposure of the complete word pair, which might have facilitated retrieval attempts at later testing. The re-exposure might have led to the difference between the Overt group and remaining groups. Another related reason why repeated testing produced better overall recall than studying may be that testing can tell which items have been learned and which ones need further studying. In subsequent studies they might find new ways of encoding the material (Carpenter et al., 2008).

It is worth addressing that all groups were asked to perform a short recall test (to write down the responses) of the words just before the end of the experiment's first part, which could have affected the outcome of the study. Consequently this could have lead to elaboration of the material, which resulted in a decreasing gap between the groups at final testing a week later. If this short recall test would be removed it could result in an even clearer differences among the groups. One weakness of the experimental design was the inability to control what participants have done during the time that had passed between the experiment's two parts. There may be a chance that some participants had rehearsed some words during the week or friends that participated in the experiment had been communicating about the content even though they were informed to avoid it. In addition it may also have been due to individual differences in terms of learning ability, previous knowledge of language, perceptual speed, motivation and interests.

Present study provides a clear improvement regarding the possibility of testing. We now know that in order to write down our responses, specifically in this experiment we do contribute to an extra form of learning, which is valuable information. The idea of the testing effect provides a guideline of how the testing procedure can be conformed into educational contexts. This advocates for more consistent use of multiple testing in the classroom for optimal learning (Roediger & Karpicke, 2006). Further studies such as Leeming (2002) demonstrated that repeated testing under classes encourage students to continually study the course material long before the exam date instead of devoting an intense studying time just before. Testing can be seen as a resource to enhance student’s learning ability and thus promote higher grades. Furthermore, present study was restricted to the learning of word pairs, therefore it would be interesting to operate an additional experiment with use of another type of material such as reading longer texts, which is rather more similar to the material being studied at higher educational levels.

To conclude, the present study reveals a testing effect as a result of acquiring repeated testing and study groups. We were able to attain a positive effect of testing, where the role of testing appeared to be the most durable method in terms of recalling as many correct words as possible and hence provided the lowest rate of forgetting, in particular the act of writing down the responses, showed to be superior to any other method. This supports the fact that recall attempts lead to improved memory over time. The role of the testing effect can be applied to many areas such as educational contexts and all types of material to be learned. As a fundamental starting point this concept may work as a
more effective study technique and thus enhance the overall learning for all individuals.

References


