Overview: Experiential Learning Resources
Prepared by Dean’s Fellow, Lian Hai Guang for Yale-NUS Centre for Teaching and Learning

0. Quick Notes

A quick summary of useful tips and pointers extracted from the Focus Article.

1. Focus Article


A brief 7-page paper covering how to think about active learning and studies about one specific form of it. This draws out implications of how active learning can improve even the most basics aspects of a learner’s learning experience—note-taking and revision.

2. Supplementary Materials


A more extensive albeit slightly backdated study of different variants of active learning. What is of value in this paper is the section on how students reacted to the various forms of active learning implemented during the study.

3. Extra Materials

A list of links to extra resources if you are interested in learning more about Active Learning.
Active learning refers to the idea of engaging with one’s learning materials in a deliberate and effortful way, in contrast to simply absorbing them inertly or passively. It is debatable as to what counts as active or passive, which brings up a genuine question. Another question corollary to this is asking how different types of active learning compare with each other in terms of their effectiveness e.g. aiding content retention.

In psychology, active learning can be framed in terms of generation advantages or generation effect. The idea is that learners perform better when made to produce or generate (in part or in full) assigned learning materials, rather than just reading through them. Studies have shown that there is a difference in performance between students who have gone through a generative (active) process of going through material versus merely reading through it; the difference in performance (%) between the two groups measures the magnitude of the generation effect.

The article claims that most studies on the generation effect emphasise a particular aspect of their analysis and end up twisting the full picture; explaining why individual studies on the topic have been inconsistent with each other. The authors instead suggest, and employ, the use of a statistical method called meta-analysis. This involves combining the results of 86 studies on the generation effect, and averaging them out to obtain a “bottom-line” sense of its magnitude. They also categorised these studies in order to compare between different types of effects.

A table summarising the various generation effects from different generative learning strategies is included on page 73, accompanied with explanations to understand the table.

Revelations from the study can be potentially applied to improving learners’ memory performance in two ways: (1) how notes are taken (2) how revision is done. Two experiments, one by Van Blerkom et al, and another by Foos, Moora and Tkacz, show that learners who produce information while working through material perform better than learners who do not.

The article also notes that learners have a general difficulty with discerning between important and less important information; notes taken by learners typically contain a lower proportion of important information. It suggests that even if complete notes are handed out to them, the problem will still remain because learners will still unlikely be able to tell what is crucial from what is inconsequential. Additionally, this encourages a more inert form of learning.

The article concludes with the recommendation that making students responsible for producing their own notes, and also having a more generative way of doing revision will better their scores. This is in consideration that help will be rendered externally to support them in identifying what is critical from what is not. Student who use more generative (active) study techniques are also likely to learn more.
The idea that learning is more effective when one spends effort on the material instead of using more passive types of review is well recognized in the field of education (e.g., Thomas & Rhower, 1986). What is often less clear, however, is exactly what is means by “effortful” or “active” versus “passive” learning, and what types of active learning, if any, are more effective educational tools than others. In the field of psychology, one way these concepts are described and tested is as an experimental finding called the generation effect or generation advantage. In its simplest form, the generation effect is the finding that, in most cases, information is more likely to be remembered when it was studied under conditions that required learners to produce (or generate) some or all of the material themselves, versus reading what others had prepared. In this chapter, we will summarize the research on the generation effect and how strong an impact it has on memory performance, describe what conditions make it more or less effective (experimentally), and offer some suggestions for how these experimental designs might be adapted to college classroom experiences.

In experimental research on the memory advantage of generating material participants use a particular researcher-provided rule to produce information that is later tested in one of several ways. These rules vary widely and include things like creating list of words that are the opposite of (or that rhyme with) provided words, completing the last word in a sentence, or even the completion of addition or multiplication problems. These generated words or numbers are then tested using different types of measures including recognition (e.g., multiple-choice) or some type of recall (e.g., essay) format. The percent correct on the test for these participants is compared to the percent correct of a similar group who read prepared lists of words (or numbers) but produced none of the information themselves. The difference between these percentages is the size of the generation effect (Slamecka & Graf, 1978).

As happens in any scientific field, individual experiments conducted on the same topic often find different results. This inconsistency makes trying to apply those results to real-life circumstances difficult, as one experimenter’s results sometimes appear to conflict with others’. This has happened in work on the generation effect as well, as researchers use different materials, study slightly different aspects of the same phenomenon, and use different experimental designs and different groups of participants (who are diverse themselves in many ways). Underlying all this variability is the true size of the benefit of generating material over reading it. The apparent change in how effective generating is over reading is largely because each of these individual studies pulls out one aspect of the effect for
analysis, distorting the complete picture. One way to get more of a “bottom-line” feeling about any particular effect is to use a statistical method called a meta-analysis. This technique converts all the results of individual experiments into a common measure (called an effect size) that summarizes the seemingly disparate experiments. Effect size measures give a more practically useful estimate of how much generation of information boosts memory performance over reading only, and it can also be used to compare groups of results that are different based on a category of interest (e.g., the memory benefits of solving addition problems versus multiplication problems).

In 2007, we conducted a meta-analysis with our colleagues Richard Wiscott and Michael A. McDaniel. In it, we combined the results of 86 studies on the generation effect. In addition to summarizing these studies to determine (on average) how beneficial the generation of material is across the board, we also carved these studies up into categories. This allowed us to investigate what kinds of effects were larger or smaller—what types of study, testing or individual person characteristics made producing your own answers more beneficial for memory performance over reading them as compared to others. Before we get to the results of our study, we pause here for some brief statistical commentary. First, remember that people’s scores on all the different types of memory tests fall on a continuum, representing very poor to excellent performance—regardless of the kind of studying they do. Although the best estimate of a score in the group is the average one, individuals will be scattered over the whole scale. This means that while the average score for generators might be higher than for average readers, some generators will score lower than some readers, and some readers will score better than some generators. So while we can make recommendations to improve memory performance that are likely to work for many (or even most) students, we cannot promise they will work for everyone (see Lee & Kalyuga, this volume, for context on this point). The second comment has to do with interpreting the size of the generation effect itself, or the differences among generation benefits depending on how they are produced. For example, we found in our meta-analysis that producing an entire word improved memory performance more than generating only part of that word (such as adding the final letter). But how useful or practical is this difference? How many generators in these two conditions did better than those who only read?

Although guidelines exist that allow determinations of whether these average differences are small, medium or large, those descriptions may not be meaningful for application in a practical context. For the following discussion, we will provide the statistical effect sizes that we found in our meta-analysis (these are represented as something called Cohen’s $d$, which is translated as how many standard deviation units one average score is from another), but we will also describe what that effect size means in terms of how much the memory scores of the individual people taking the tests in the different conditions overlap. For example, across all the studies we evaluated, the average size of the generation effect was $d = .40$. In psychology, this is referred to as a medium-sized effect. Not very helpful, we know. However, a little more useful is the translation of this effect size to the percent of overlap in the distribution of scores—about 66% of people who generated information had better test scores than the average score of those who only read the information. Thus, in general, the larger the effect size is, the less the reading and generating distributions overlap and the more generators who score better than the average reader. To summarize some of the most potentially useful information we found in our meta-analysis:
<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect size (d)</th>
<th>% of generators scoring better than average reader score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition test</td>
<td>.46</td>
<td>68</td>
</tr>
<tr>
<td>Free recall test</td>
<td>.32</td>
<td>63</td>
</tr>
<tr>
<td>Older Adults tested (65+)</td>
<td>.50</td>
<td>69</td>
</tr>
<tr>
<td>Younger adults tested</td>
<td>.41</td>
<td>66</td>
</tr>
<tr>
<td>Rhyming rule used to generate words</td>
<td>.46</td>
<td>68</td>
</tr>
<tr>
<td>Sentence completion used to generate words</td>
<td>.60</td>
<td>73</td>
</tr>
<tr>
<td>Synonym rule used to generate words</td>
<td>.41</td>
<td>66</td>
</tr>
<tr>
<td>Letter rearrangement used to generate words</td>
<td>.37</td>
<td>64</td>
</tr>
<tr>
<td>25 or fewer words generated or read</td>
<td>.60</td>
<td>73</td>
</tr>
<tr>
<td>More than 1 day retention interval between generating/reading and testing</td>
<td>.64</td>
<td>74</td>
</tr>
</tbody>
</table>

Based on these results, we can expect certain types of test preparation to be (on average) more effective than others. For example, the effect is stronger for multiple-choice types of tests as compared to essay testing, and is very effective among older adult learners as well. Of the different kinds of generation rules used, completing sentences appears to lead to the biggest advantage over reading only, but even relatively minor types of generation (like switching two underlined letters) improves test scores over reading alone. Finally, the effect is strong for small blocks of information, and to leads to relatively long-term retention of the material.

These results apply to participants in the kinds of experiments that are typically conducted in psychology. These experiments usually follow the same format, where participants are given a list of words to either read over or to create themselves using a particular rule. Does this same generation effect principle work in more real life situations where students are attempting to learn lecture or textbook material? We believe the answer to this is yes, although investigations of the usefulness of generating information in more realistic situations is less common than laboratory-based research (DeWinstanley, 1995; Peynircioğlu & Mungan, 1993). Considering potential applications of the generation effect to improve memory performance, two have been found to be practically useful (although we will mention a possible third later on). The first of these has to do with how students take lecture notes in the first place (or how they take notes from their textbooks) The second has to do with how students study material from lecture notes they have taken, or from a text source they have read. In reviewing the experiments done in these situations, there are indications that having students produce at least part of the information for which they will later be held responsible leads to better test scores both when used as a study technique (King, 1992) and also when used as a note-taking strategy (Russell, Caris, Harris, & Hendricson, 1983). This research has also indicated some potential limitations and ways that the technique should be supplemented to make it more effective.

On their own, students chose a wide variety of study techniques, but outlining information, highlighting notes or text material, summarizing notes/text, and writing practice questions are among the most common. Research on whether these techniques produce more learning (i.e., higher test scores) than
more passive forms of studying (such as reading over notes, text, or study guides produced by an instructor) tends to agree that students who produce information have higher average test scores than those who only re-read information they are given (e.g., Paris, Wixson & Palincsar, 1986; Van Blerkom, Van Blerkom, & Bertsch, 2006). One of the most commonly investigated study techniques is writing practice questions. In these experiments, students are asked either to write questions in advance of reading information with which they are familiar in order to help them organize what they read (e.g., perhaps asking “What are some of the causes of the Civil War?” before they start reading a passage on it for an American History class), or in cases of unfamiliar material, to create multiple-choice or essay type questions based on lecture or text material after they have read it. These questions can then be used by the student later in preparation for the test. In experimental examinations of the effectiveness of these study techniques, students’ later test scores are compared to those students who either did not see practice questions at all, or those who read over practice questions and answers created by the class instructor. For example, Van Blerkom et al. (2006) compared test scores covering a passage on locus of control among four groups of students: the first read the text passage and copied it, the second read and highlighted, the third read and took notes, and the fourth read and wrote potential test questions and answers in the margins as they read. Two results were clear: students who created practice questions on their own had the highest scores on test questions based on that information and students are generally not good at targeting what is considered to be important information from notes or texts. In the Van Blerkom et al. study, an average of only 40% of the ideas about locus of control that were on the test were also targeted by the students writing practice questions. In another example, Foos, Mora and Tkacz (1994, Experiment 2) found that students who created their own practice questions got 86% of similar content test questions correct, compared to students who were provided with practice questions (who got 72% of similar test questions correct). The problem was that students creating their own questions only targeted about 20% of the important concepts they were asked to learn from the passage of text.

This difficulty with discerning more important from less important information appears to be a general problem, particularly among poorer academically performing students. This issue is not only with what students choose to study but also in the notes they take from lectures or text sources in the first place, as students’ lecture notes typically contain a low percentage of important ideas. King (1992) found that only about 15% of important ideas were recorded in students’ notes, which unfortunately is a not an unusually low number. Clearly, if students are not taking adequate notes, their test scores will suffer. Kobayashi (2005) combined the results of 57 studies comparing the learning outcomes of students who did or did not take notes and found an average effect size of $d = .22$ (this translates to only slightly more than half of note takers with better outcomes over the average non-note taker). Basically, this means that the average students’ note taking skills are so poor that it is not surprising they do not do well on tests. Although the generation effect appears to be viable in classroom settings, it has empirically highlighted an important problem about which most class instructors were already aware: many students fail to tell the difference between more and less important information.

Although many different interventions exist to help students learn effective study techniques, if they are not working with adequate information in the first place, no amount of study intervention is likely to produce a large effect. In today’s technology-assisted learning environments, some instructors are attempting to solve the problem of poor note taking by providing students with either hard copy or electronic versions of their lecture notes. This practice is problematic in two ways. First, as previously indicated, even when students have a complete record of the information they are supposed to learn (e.g., a textbook chapter), many of them are woefully inadequate at parsing the core ideas or most important points from more shallow details when they study. Second, providing students with a
complete set of lecture notes enables a more passive type of initial learning of the material. Having a complete set of lecture notes does appear to result in higher test scores for students when compared to those of un-assisted student note-takers: as the un-assisted students are likely only recording a low percent of relevant information, this result is not surprising. The work on the generation effect, however, would predict higher scores (over longer retention intervals) for students who only receive general outlines of lecture or text information, but have to complete the full information themselves. This pattern has been found by several different groups of researchers (e.g., Barnett, 2003). Katayama and Crooks (2003) gave their participants a text passage to read along with either a complete or partial set of notes about the content of the passage. The partial condition provided participants with column and row headers to help guide and organize the notes the students took. After time to read the passage and study either the complete notes or the ones they had taken (in the partial condition), the participants took a test that consisted of several types of questions. Not only did students in the partial condition perform better on the test overall (particularly for application types of questions), but their higher scores were more consistent over a one week retention interval. Russell et al (1983) compared multiple-choice test scores of students based on three types of note taking: one group was given comprehensive notes about the lecture content; one group partial notes which contained an outline of the lecture material, tables and figures; and a skeleton notes group which received only a brief outline of the lecture information. The skeleton notes groups had higher test scores than the other two groups on both immediate and delayed (2 or 4 week) testing.

**Educational Implications**

In reviewing what educational practices are predicted by research on the generation effect to lead to higher test scores, we have seen evidence that this basic empirical finding does transfer to more practical situations. Making students responsible for the production of information both during note taking and also during study does appear to improve test scores, with the caveat that many students appear to need some external structure (e.g., outlines) during note taking to ensure that they are recording appropriate information. What other areas of education might be improved through the use of these generation principles? Although research evidence is lacking on the following suggestions, they proceed directly from the same principles above, which have provided evidence of effective transfer. The first suggestion has to do with a particular tendency we have noticed among our own students and has been the topic of discussion with other faculty as well: we find that students often want answers to questions in class that they could figure out for themselves. It seems to be either questions about information they have already received (perhaps in an earlier lecture), or questions about examples or applications of principles or theories they have learned. In many cases, these kinds of application questions are ones that students could answer on their own with a little cognitive effort (possibly instructor-guided). While it is certainly easier and more efficient for the class instructor to answer these questions (sometimes repeatedly) and move on to the next block of material, doing so may hinder students’ learning. Clearly, the traditional Socratic method of using questions and answers to lead individuals to self-discovery of knowledge is not an efficient way to teach a classroom of 50 Introduction to Psychology students. However, helping them attempt to figure out the answers to some of their own questions either by having them look back through their notes or by asking questions that get them to think about how a theory explains behavior are likely to produce better learning than providing complete answers.

In addition to attempting to generate the answers to their own classroom questions, students who modify preferred study techniques to make them more generation-based are also likely to learn more. For example, making and using notecards is a common way students study course material. In preparing
their cards, they often copy terms or definitions from their notes or textbook onto the cards, putting the answer on the back. Although this method is likely to benefit from the copying process compared to re-reading the original information, it is not particularly generative in that no manipulation of the information occurs. Teachers commonly tell students to put the information into their own words, which is a more generative process, but one with which our students often struggle. They seem to have difficulty translating technical information into language with which they are familiar. The first author of this chapter has found success in helping students to make this conceptual leap by telling them to try to limit the number of words that can be written on any card (say, to perhaps half of the source original ones). For example, instead of copying the complete definition of schema as “an expectation about objects, events, others’ or our own behavior based on past experiences,” students might read this definition in their notes and write “we think things happen because they have before”, or “expectations based on something’s history”, or even “expectation, things, people, based on the past” (full sentences are not necessary). Limiting word counts forces students to re-word or paraphrase more effectively. Is this more work than copying the original definition? Of course it is. Is it likely to lead to better memory for the concept of a schema over time? Yes.

Reading text and copying that information onto notecards is less work than creating study questions and answers. Reading over complete lecture slides received from an instructor or downloaded from a website is easier than taking one’s own notes in class. Answering a students’ question in class is easier than leading them through the process of coming up with that answer on their own. At some point, we need to value the work that goes into effective learning instead of trying to find the easiest path to knowledge (see Clark and Bjork, this volume, for an extended discussion on desirable difficulties).

References


Active Learning Resources
Prepared by Dean's Fellow, Lian Hai Guang for Yale-NUS Centre for Teaching and Learning

Recommended materials are marked with ‘★★’

Extra Reading Materials on Active Learning:


From Cornell University’s Centre for Teaching Excellence:


University of Illinois (Urbana Champaign) Centre for Innovation in Teaching & Learning:


University of Waterloo’s Centre for Teaching Excellence: