Desirable Difficulties and

Studying in the Region of Proximal Learning

Janet Metcalfe

Columbia University

Please address correspondence to:

Janet Metcalfe,

Department of Psychology,

401B Schermerhorn Hall,

Columbia University,

New York, New York, 10027

jm348@columbia.edu

212-854-7971 212-865-0166

Introduction

The Desirable Difficulties perspective (Bjork & Bjork, 2006, in press; Bjork, 1988, l992; Bjork & Linn, 2006) and the Region of Proximal Learning framework (e.g., Kornell & Metcalfe, 2006; Metcalfe, 2002, 2009; Metcalfe & Kornell, 2005) appear, at least on the surface, to be at odds. The Desirable Difficulties perspective says that individuals should make things hard on themselves (but in a good way). Instructors should not spare the learners, but instead they should challenge them and make learning difficult. People should embrace difficulties, because it is through those difficulties that long term learning occurs. The term 'desirable' is added to difficulties, but even so the message is that learning should be a challenge. The Region of Proximal Learning framework proposes that learning is optimized by having the learner study materials that are not very difficult, given the individual's current state of learning. Indeed they should be the easiest possible as-yet-unlearned items-- just beyond what the learner has already fully mastered, but not much beyond. Too much difficulty, within this framework, is seen as maladaptive and potentially disheartening.

Both of these views are intuitive. The desirable difficulties perspective capitalizes on the adage ‘no pain, no gain,' and on the pervasive work ethic emphasizing the moral value of hard work and diligence that many cultures claim for their own. The Region of Proximal Learning framework resonates with the developmental philosophies of Piaget (1952) and Vygotsky (1987), including the idea that there is a zone of proximal development that is just beyond what the child is capable of easily on his or her own, in which he or she could learn via scaffolding. It also conforms to the transitional learning stage in Atkinson's (l972) Markov model , in which items that are not yet permanently learned, but rather are in a state of being almost learned, are those on which the learner should concentrate. It takes seriously the so-called 'labor in vain' effect of Nelson and Leonesio (l988) whereby people may spend a great deal of time selectively working on the most difficult items, but this extra time and effort is without payoff . It results in no memory benefit. In the world of education, the Region of Proximal Learning framework resonates to the notion of 'just right' books for children in which their own reading level is assessed and then they are given books very close to their own skill level. Allowing children to read at their own level is thought to promote fluency and faster, not slower, progress. It is also thought to make reading more enjoyable and less frustrating. The adage that comes to mind for the Region of Proximal Learning framework is that it is sensible to first take the 'low hanging fruit.'

The apparent conflict between the Desirable Difficulties and the Region of Proximal Learning views does not do justice to either framework, however. Both are more nuanced than such a simplistic overview would suggest. The goal of this chapter is a clarification of where the two frameworks are in agreement, and where there are real differences. One issue, at the heart of both frameworks, is the extent to which the learner is actively involved in his or her own learning. This sense of engagement may be the most critical marker of both whether the person is operating within their own Region of Proximal Learning, and also of whether the amount of difficulty they are experiencing is the desirable amount. Both frameworks, it is argued, seek a level of challenge at which learners remain engaged: not so easy that they are bored, but not so difficult that they are overwhelmed.

There are four domains that Bjork and his colleagues have specified as ways in which learning should be made difficult. The Desirable Difficulties framework recommends (1) that the learner should engage in retrieval practice (or self-generation or testing), (2) that feedback should be reduced, (3) that practice should be spaced, and (4) that study of different topics should be interleaved. The Region of Proximal Learning framework agrees on two of these and disagrees on the other two. The two recommendations on which the Region of Proximal Learning and the Desirable Difficulties framework are in agreement are those referring to retrieval practice (i.e., self-generation or testing) and spacing of practice. Although the conclusions on these two are the same, the reasoning leading to the convergent conclusions is different. The recommendations on which the two frameworks differ are feedback—which the Region of Proximal Learning framework advocates but the Desirable Difficulties framework abjures, and interleaving--which the Desirable Difficulties framework for the most part advocates, but about which the Region of Proximal Learning framework has reservations. Here, I shall discuss each of these four recommendations and review empirical findings that bear on each.

Before turning to the specifics of these four recommendations, it is worth noting that the Desirable Difficulties framework and the Region of Proximal Learning focus on different cencerns. The Desirable Difficulties framework is process oriented. The underlying assumption is that making the cognitive processing in which the individual engages as difficult as possible is key to improving memory. Difficult encoding or retrieval is thought to etch in retrieval routes, making the items memorable. Demonstrations, such as that of Benjamin, Bjork, and Schwartz (1998) in which the items that people answered with the most difficulty were later found to be most memorable, illustrate this point. In contrast, the Region of Proximal Learning framework has focused more on the content of the materials, and the strategies that people use to ensure that easily-learned materials—but those that will not be remembered without at least some additional effort -- are not, inadvertently, overlooked. Of course, content and process are intertwined, and so the processing needed for simpler materials may, itself, be easier. Items that require more extensive and effortful processing are, themselves, more difficult items. Furthermore, both of these frameworks point to a middle ground. The term *desirable*, is key in the desirable difficulties framework—*very* difficult is not desirable. And the Region of Proximal Learning framework does not say that the learner should study the easiest items. Those items will have been learned already, and hence are not in the person's Region of Proximal Learning. Any item that is not yet learned is not *very* easy.

Retrieval practice

Both the Desirable Difficulties perspective and the Region of Proximal Learning framework agree that the kind of focused attention that is rallied in the service of retrieval practice (Bjork, 1988), gives rise to excellent learning results. The Region of Proximal Learning framework is entirely pragmatic on how to study: any process or strategy that helps learning is embraced. The Desirable Difficulties framework sees retrieval attempts as a difficulty of just the right degree to be desirable. Hence, there is complete agreement on this first of the domains emphasized in the Bjork (2004) and Bjork and Linn (2006) framework.

The memory enhancement seen with self-generation, the benefits attributable to retrieval practice, and the learning enhancement observed with testing appear to be different faces of the same diamond. They all involve the engaged effort of the individual to actively come up with the answer to a question. Retrieval practice is the more generic term, but it is plausible that both generation effects and testing effects benefit memory as a result of a common mechanism. Whether the reason for the memory enhancement is that retrieval practice entails a particular optimal amount of difficulty, or because the person accords preferential treatment and perhaps attentional resources to materials and events in which the self is actively engaged is unknown. Whatever the cause, the memory enhancing effects of generating are robust.

The generation effect was first detailed by Slamecka and Graf (l978), in a paper that compared memory for words that were generated by the participants themselves to words when they were simply presented and that had to be read. Memory performance in the generate condition was superior--leading to the idea that self-generation was a memory tonic. A meta-analysis of studies on the generation effect in the nearly three decades since this original paper was conducted by Bertsch, Pesta, Wiscott, and McDaniel (2007). They investigated 86 studies which included 445 measures and a total or 17,711 participants. The generation superiority resulted whether recognition, cued recall or free recall was the criterion measure, whether learning was intentional or incidental, whether the design was between participants or within participants, and whether the lists were blocked or random. Generation resulted in better memory for both older adults and younger adults. It resulted in better memory when the materials were numbers or words, though the beneficial effect was smaller when the materials were nonwords, and, alone among manipulations, anagrams showed no effect of generation. There was a beneficial effect of generating regardless of whether the generate rule was rhyme, association, category inclusion, sentence completion, calculation, synonym generation, word fragment completion, antonym generation or letter rearrangement, and whether the participants were asked to generate a whole target or only part of a target. The generation effect occurred when the lists were short, medium, or long. It obtained at all test delays. And, finally, it occurred when the generation task was easy, of moderate difficulty and when it was difficult. Indeed, difficulty, seemed to have no impact.

Thus, if what self generation is doing is inducing difficulty, it is a desirable difficulty. But it is not clear that the benefits of self-generated retrieval are due to the difficulty rather than some other aspect of the task, especially since the difficulty of the generation task itself had no effect. There is no dispute, however, about the robustness of the data.

In an interesting variation on the theme of generation effects, Kornell, Hays and Bjork (2009) showed that a memory facilitation of the correct answer resulted from attempted retrieval even when the person generated a wrong answer. There was, of course, a caveat in this study, and that was that the person needed to be given feedback. The benefits of feedback will be detailed shortly. In this case, it was a necessity.

There is now abundant evidence that test taking, like generation, results in memory benefits (Butler & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel & Fisher, 1991; McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006 a, and b). Although the testing effect applies not only to recall tests but to recognition tests (though not so robustly), larger effects of testing are seen in recall testing, probably because people more reliably generate in that situation. Although there has been little deep theoretical understanding or even speculation concerning the reasons for either the generation effect, or the testing effect, it seems likely that both have something to do with the fact that the self is actively involved, rather than because of some optimal degree of difficulty.

Feedback

The second way in which learning can be made difficult in what is thought to be a desirable way, within the Desirable Difficulties framework, is by reducing feedback (Bjork & Linn, 2006). Unlike the use of generation or testing, in which there was agreement between the Desirable Difficulties and the Region of Proximal Learning framework, in the case of feedback, the two frameworks differ. Providing feedback, especially for incorrect responses, but also for correct responses that may not yet be fully mastered and which the person may not realize they have gotten correct, is recommended in the Region of Proximal Learning framework. Reducing feedback in such cases is seen as a lost opportunity for learning-- not desirable.

There are several cases, in the motor learning literature, where not getting (or reducing) a certain kind of feedback, called knowledge of results or KR benefited performance (Winstein & Schmidt, 1990; Wulf & Schmidt, 1989). It is not clear what to make of these results. KR consists of seeing a line graph of one's own movement together with a graph of the correct movement, immediately after having tried to enact the correct movement. It is not entirely clear whether KR is like correct/incorrect feedback (which has no beneficial effect in verbal memory tasks) or corrective feedback (which has consistently shown benefits to memory), or, indeed, whether it is simply a distraction. It is also not clear whether results about motor movements generalize to verbal tasks, regardless of the type of feedback. Despite these findings in the motor learning literature indicating that reducing feedback may sometimes help, I have been unable to find any cases where decreasing or eliminating corrective feedback helps verbal learning.

There is, by contrast, a large literature on the beneficial effects of feedback (Anderson, Kulhavy & Andre, 1971; Butler, Karpicke, & Roediger, 2007, 2008; Butler & Roediger, 2008; Kulik & Kulik, 1988; Lhyle and Kulhavy, 1987; Metcalfe & Kornell, 2007; Metcalfe, Kornell, & Son, 2007; Pashler, Cepeda, Wixted, & Rohrer, 2005). When memory for the correct answer is compared on what were previously incorrect responses that have been given corrective feedback as compared to those that have been given no feedback, the former are always better than the latter. For example, Metcalfe and Finn (under review) investigated the correction of incorrect responses to general information questions. Final performance without feedback was on the floor at around 4%. With corrective feedback, however, performance increased, sometimes as much as by 75%. In Butler et al.'s (2008) experiments, with items not given corrective feedback, later test performance was .41 in one experiment and .47 in another (reflecting the fact that many of the initial answers were initially correct, of course). When participants were given feedback, performance increased to .87 and .83 respectively.

Pashler et al. (2005) have shown that feedback needs to be corrective to be effective: the correct answer needs to be given, not just a statement of whether the answer was correct or not. Rarely is simply telling the person that they were wrong sufficient to allow them to correct their error. Nor is telling them that they were right enough to make them more right the next time. The one exception to this general rule occurs when people do not know that they were correct. Although there is usually a very high correlation between responses on one test and responses on the next, in the case of low confidence correct responses, people often fail to produce the correct response on the subsequent trial, unless they are told that it was correct on the first test (Butler et al., 2008). This is a special case in which mere correct/incorrect feedback has an informative effect. Providing correct/incorrect feedback followed, not by the correct answer itself alone, but by the chance to make a second guess on a multiple choice test (that included the correct answer of course), also resulted in some improvement in performance (Metcalfe & Finn, under review). Usually, though, when people are wrong they are wrong because they do not know the answer. Unless told what the answer is, they flounder.

Although there is broad consensus *that* feedback needs to be given to enhance memory, *when* the feedback needs to be given is more controversial. There are a number of cases in which corrective feedback that is given at a delay is more effective at improving memory than is feedback given immediately. For example, Guzman-Muñoz and Johnson (2007) showed, in a study investigating memory for geographical representations, that delayed feedback resulted in a more laborious acquisition, but better eventual retention, than did immediate feedback, on a test of the location of individual places. This entire pattern fits with the Desirable Difficulties framework very nicely. However, the delayed feedback was also more informative than the immediate feedback, in this study, because it involved seeing an entire map (including the relations among to-be-learned places) rather than just the location of individual places. As the authors themselves suggested, the result might have obtained not because of the delayed feedback, *per se*, but rather because the beneficial configural information was more salient in the delayed than in the immediate feedback condition.

Schooler and Anderson (1990) also found an advantage for delayed feedback in a study on learning of LISP. In their study, learners who were given immediate feedback went through the original learning trials more quickly than students given delayed feedback. Those given immediate feedback were slower on the criterion test problems, however, and made more errors. The authors suggested that the detrimental effect of the immediate feedback, in the situation they studied, resulted because the feedback competed for the working memory resources that were needed to accomplish the task itself, resulting in a disruption of compilation, within the ACT* framework.

However, a number of experiments (see, Butler et al., 2007, Kulhavy, 1977; Kulhavy & Anderson, l972) indicate that even disregarding the benefit that the delayed feedback may provide in allowing a coherent overview of the to-be-learned materials, or the disruptive effect that immediate feedback might have because it distracts from the task at hand, delaying feedback is often beneficial. As Kulhavy and Anderson (l972) noted, studies conducted in the laboratory have tended to show a delayed-feedback advantage. However, many of these studies controlled total time. They thereby forced the last presentation of the correct answer to be closer to the test in the delayed than in the immediate feedback condition. Clearly an advantage due to a shorter retention interval will ensue. However, even controlling for the lag to test, advantages for delayed feedback are sometimes found. Controlling for lag to test, Metcalfe, Kornell and Finn (2009) found that delayed feedback resulted in better vocabulary learning with grade school children. The delayed feedback advantage was not significant with college students, however. This difference in the results might have come about because children are different than adults, of course. However, the overall level of performance was also higher for the grade school children than for the adults (the materials, of course, were different). It is possible that whether immediate or delayed feedback is most effective depends on the person, the person's age, on the level of recall and/or the number of errors. These possibilities needs to be explored further, since the determination of when it is optimal to give feedback may be a factor having a considerable impact on memory. These results, overall, suggest that introducing a delay in feedback may, at least under some circumstances, be a desirable difficulty.

Finally, the Region of Proximal Learning model points to the fact that feedback is particularly important for certain items that are *almost learned*. These are the items that are supposedly in the person's Region of Proximal Learning, and are also items that Butterfield and Mangels (2003) have called 'metacognitive mismatch' items. These are the items on which the person thought they were right, but in fact were wrong (high confidence errors), or on which they thought they were wrong but in fact were right (low confidence corrects). Butterfield and Metcalfe (2001, 2006), Butterfield and Mangels (2003), Fazio and Marsh (2009), and Metcalfe and Finn (under review) have all shown that these high confidence errors are very easily corrected--indeed they are hypercorrected provided that the participant is given corrective feedback. A small amount of feedback converts a highly confident error into a well-entrenched correct response. Similarly, Butler et al. (2007) have shown that the other kind of metacognitive mismatch, the low confident correct responses, will often show up wrong on a subsequent test. This tendency, though, is easily converted to stable correct responding when those initially low-confident correct responses are bolstered by feedback.

Spacing

Both the Desirable Difficulties and the Region of Proximal Learning framework agree on the benefit of spaced practice over massed practice. Interestingly, though, both frameworks also agree that there are some particular conditions under which massed practice is indicated.

The Landauer and Bjork (l978) model of expanding retrieval practice recommends that at the beginning of learning, repetitions of a to-be learned pair should be close together in time, to ensure that when the cue is presented alone, the target will be correctly retrieved. Then the spacing can be increased somewhat, as long as the target continues to be correctly retrieved. As learning becomes more solid, the spacing can be further increased. The limiting factor in this model is that, unless feedback is given, the answers participants generate need to be correct for memory to be strengthened. The reason the Desirable Difficulties framework advocates spacing (but with the caveat that retrieval must be correct) is that spaced retrieval is more difficult than massed retrieval, and so it helps memory performance more. Expanding retrieval practice, rather than simply maximizing spacing, is, thus, advocated (Landauer & Bjork, 1978; Schacter, Rich, & Stamp, 1985; Storm, Bjork, & Storm, in press, c.f., Balota, Duchek, Sergent-Marshall & Roediger, 2006; Carpenter & DeLosh, 2005)

The Region of Proximal Learning framework also advocates spaced practice, in general, except when full learning or full encoding has not occurred on the first practice. Then the Region of Proximal Learning framework proposes that the learner should stay with the current item until encoding is complete, rather than flitting to other items. The concern in this model, and the reason that continued study on a single item is sometimes advocated, is with learning/encoding/comprehension at the time of study. The spacing predictions in the Region of Proximal Learning framework result from its stop rule which determines when the person should perseverate and when he or she should stop studying one item and turn to another. The rule in the model is that the person should turn to another study item when the learning on the first item is no longer showing progress. This is given by the learner's judgment of the Rate Of Learning (jROL). When this rate approaches zero, learning on the current item is no longer productive, and the learner should turn to other items. This lack of learning may come about because the item being studied has already been sufficiently learned (or in the case of very difficult items because the item is intractable and no progress is in evidence). Such a stop rule fits well with the idea that some items can be quickly and easily learned, or, at least learned to the point that no further efforts will be beneficial at time t. Other items take more time to reach a learning asymptote. Indeed, a pervasive finding in the literature is that people study items with low judgments of learning (JOLs, i.e., the subjectively difficult items) longer than items with high JOLs (the subjectively easy items, see Son & Metcalfe, 2000, for a summary of the literature on this pervasive correlation, and see Dunlosky & Metcalfe, 2009, for discussion of these issues). In the case that the person's jROL has reached an asymptote and no further learning is occurring, the person should turn to other items that may be more yielding, rather than continuing to study an item that has already been learned to the maximum possible at that moment and for which no further learning gain can be expected, at least immediately. This stop rule results in the spacing prediction.

Recently, Metcalfe and Jacobs (2010) have proposed an analogy between the strategy that results from the stop rule in the Region of Proximal Learning framework and animal foraging. The analogy may be useful in allowing us to think about massed and spaced practice. The idea is that a person studying an item is like a hummingbird or a bee, say, taking nectar from a flower. The person should stay until they have extracted as much usable memory information out of the item as possible; the bird or bee should stay until it has extracted as much nectar out of the flower as possible--only then turning to another item or another flower. Once the source has been depleted it is a good idea to look elsewhere since there will be no further (immediate) gain from staying on an item or flower whose nectar is depleted. But the nectar will replenish, and it will be advantageous to come back to that flower then.

How long does it take to learn or understand the items (or to take the nectar offering)? That will, of course, depend upon the items, and their difficulty or complexity (or the richness of the source). Whether spacing or massing (which is continuing to stay with the current item) is more advantageous, depends on where the learner is in the process of nectar extraction. In experimental situations this will depend upon the presentation rate. For verbal paired associate learning, people may not have encoded what they need to form a deep and useful encoding that will help their later memory within 1/2 s, especially if the materials are relatively difficult. But most likely they will have done so by the time they have studied it for 5 seconds, depending upon the materials, of course. If there are mediators, in the materials, but they are difficult to discern (as in the a Spanish-English pair such as of "vodevil-music hall," say), then massing practice at fast presentation rates may allow learners the time needed to pick up on the mediators that they need to comprehend the meaning. Under such conditions massed practice may produce an advantage over spaced practice. Metcalfe and Kornell (2005) found that massing was more effective than spacing with medium difficulty Spanish-English pairs such as that given above, but only if the presentation rate was very short. Otherwise--with either easier materials, or with a slower presentation rate--spacing resulted in superior memory. Notably, Nancy Waugh (1970), in one of the first systematic studies of the spacing effect, reported the same presentation rate interaction: spacing was advantageous only at slow presentation rates.

The Desirable Difficulties perspective, as captured in the Landauer and Bjork (l978) expanding retrieval practice model, recommends massing practice of poorly learned items so that they will be successfully retrieved. Only then should they be spaced. The Region of Proximal Learning framework also recommends massing under some specific conditions, where initial learning or encoding has not been completed during the first presentation.

Interleaving

The last of the difficulties thought to be desirable is interleaving. The Desirable Difficulties framework recommendation is that, like spacing, interleaving may make study more difficult for the learner initially but that difficulty will produce better long term learning. Note, however, that Bjork (2006) in his Psychology Learning and Teaching, keynote address noted that "desirable difficulties are desirable because responding to them successfully engages processes that support learning, comprehension, and remembering." Despite the fact that interleaving is usually cited as a desirable difficulty, if comprehension is impaired by interleaving, interleaving may be undesirable. Interleaving operates at a superordinate level of analyses rather than at the item level, and so is slightly different from spacing. A coherent structure of some sort is assumed to be fragmented by interleaving. If there were multiple members of the same category, for example, and several different categories to be learned, the question is whether it is better to stay within one category, and complete study on , say, all of the category members before turning to the members of another category (blocking practice) or to switch and alternate among categories (or interleave). Similarly, interleaving could apply to stories and narratives. Is it better to complete one story before turning to the next and the one after that (massing practice of each), or to read a bit of one, then turn to the next and the one after that, reading bits of each, before coming back to the first one?

The answer given by the Region of Proximal Learning framework is that it depends upon whether the second and third category exemplar (or story event), are providing new information, and hence promoting learning, or whether they are redundant with the exemplars that have just been presented, and therefore offer no immediate opportunity for learning. The same rules apply as applied for whether massed or spaced practice would be more advantageous.

An early example of learning that was either blocked (massed) or interleaved was given by Kurt and Hovland (l956) using category exemplars that were geometric patterns that varied on four dichotomous properties. The participants had to learn the categories, and these categories were determined by rules. After each exemplar was presented participants were told the category name, and from these they were to infer the rules. The results showed that this rule-determined categorization was better--though only marginally --when the exemplars from the same concept were presented in succession (blocked), rather than being interleaved. Presumably blocking allowed the participants to infer the complex rule that determined category membership across exemplars. Having alternate categories interleaved was a distraction. A similar modern experiment showed the same results, with rule-driven artificial categories (Garcia, Kornell & Bjork, 2009).

Along similar lines, Gagne (l950) showed that learning was faster and fewer errors were made when three complex stimuli that were highly similar to one another in a list of 12 items (composed of four groups of such highly confusable items), but that had to be discriminated, were grouped together rather than being interleaved. The grouping allowed the confusions between similar items to become more evident, and the learning to proceed at a more rapid pace. In these studies, blocking rather than interleaving, was advantageous.

Blocking rather than interleaving was also advantageous in narrative learning. In such cases, interleaving may be harmful because it interferes with rather than supports deep comprehension. Mandler and Deforest (l979) presented adults and children with two stories either in a blocked fashion, or interleaved. All of their participants were able to recall the stories in their canonical form, even if they had been presented in the interleaved form. They had difficulty maintaining or remembering the interleaved order --suggesting that the canonical order was more natural, and providing support for the idea that a grammar-like story schema is used in text comprehension. When changes were made in the presentation order of such a canonical structure, even with textual markers to indicate the correct placement of the displaced events, recall errors resulted (Mandler &

Goodman, l982). This finding, again, points to the idea that a story structure may be part of the basis of text understanding, and that interleaving may hurt comprehension. Finally, when Mandler (l978) directly compared both the quantity and quality of recall in the same two stories that had either been presented in their canonical form, or which had been interleaved with one another, it was found that interleaving hurt memory. Presumably the availability of the story schema, and the co-ordination of that schema with the incoming information in a coherent way, was important for the deep comprehension that facilitated later memory. The memory advantage to presenting the stories, all of a piece in their canonical form, rather than interleaving them, was greater for children participants than for adults.

In contrast to the above findings, some studies have shown that interleaving has beneficial effects, even in induction situations. Kornell and Bjork (2008) expected interleaving to hamper induction by obscuring the commonalities or structure that define a concept or category. They presented participants multiple paintings by different artists along with the artist's name. The paintings of a single artist were presented either consecutively (massed) or interleaved with other artists' paintings. When asked which artist painted each of a series of new paintings the authors found that performance was better in the interleaved than in the massed condition. Older adults showed the same result as did younger adults (Kornell, Castel, Eich, & Bjork, in press).

These findings are surprising. They run counter to the authors' own intuitions that, in this situation, massing should have been better for induction than interleaving. The kind of inferential processes needed in induction, intuitively, seemed to require that the exemplars be directly compared to one another, as could only occur with massed presentation. Comprehension of the category structure would seem to depend upon it. Even more interesting than the authors' mistaken intuitions is the fact that the participants themselves, in both experiments, thought that their own performance was better in the massed condition. They thought this even after having completed the experiment! It would, seem, then, that whether it is better to mass practice or to interleave is a question that is better left to empirical research than to intuition. Our logical arguments, and our intuitions, both prospectively and retrospectively, as experimenters and even as participants, may provide the wrong answers and, in this situation, are not to be trusted.

Conclusion

The Region of Proximal Learning Framework proposes that there is a level of difficulty that is desirable. It is a level that is just beyond what the person has mastered well. At this particular level of difficulty there is a window of opportunity for learning. When a person studies materials that are in this range, especially if the study itself involves generation and self testing, and especially if the presentation is spaced, a small investment of time and effort can pay off in large learning gains. These gains can be amplified with feedback. Furthermore, it is suggested that when the person studies in this way, and sees the rewards of these efforts, it may set up a pattern of success in which learning itself becomes intrinsically rewarding and pleasurable. The proposal, here, is that both the Region of Proximal Learning framework and the Desirable Difficulties perspective, despite using different terminology, agree that this region--a region that is challenging (i.e., desirably difficult) but not too challenging-- is the 'just right' level of difficulty. Directing effective learning methods at this particular level of difficulty results in labor with gain rather than labor in vain.

References

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology, 62*, 148-156.

Atkinson, R. C. (l972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96*, 124-129.

Balota, H. P., Ducheck, J. M., Sergent-Marshall, S. D., & Roediger, H. L. (2006). Does expanding retrieval produce benefits over equal interval spacing? Exploration of spacing effects in healthy aging and early Alzheimer's disease. *Psychology and Aging*, 21, 19-31.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index, *Journal of Experimental Psychology: General, 127*, 55 - 68.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*, 201-210.

Bjork, E. L., & Bjork, R. A. (in press). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society.* New York: Worth Publishers.

Bjork, R.A. (l988). Retrieval practice and the maintenance of knowledge. In M.M. Gruneberg & P.E. Morris (Eds.) *Practical Aspects of Memory: Current Research and Issues, Vol 1: Memory in Everyday Life*, (pp. 396-410), New York, NY: John Wiley & Sons.

Bjork, R. A, (1992). Interference and forgetting, In L.R. Squire (Ed.) *Encyclopedia of learning and memory* (pp. 283-288), New York: McMillan.

Bjork, R. A. (2006)

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L-G. Nilsson and N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 109-133). Psychology Press: Hove and New York.

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science: Introducing desirable difficulties. *Association of Psychological Science Observer, 19,* 29.

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*, 273-281.

Butler, A. C., Karpicke, J. D., and Roediger, H. L., III (2008). Correcting a meta-cognitive error: Feedback enhances retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918-928.

Butler, A. C., & Roediger, H. L., III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527.

Butler, A. C., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*, 604-616.

Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research. 17*, 793-817.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1491-1494.

Butterfield, B. & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1*, 1556-1623.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619-636.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage Publications, Inc.

Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin and Review, 16*, 88-92.

Gagne, R. (l950). The effect of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology, 40*, 61-73.

Garcia, M. A., Kornell, N., & Bjork, R. A. (2009). Difficult rule-based category learning benefits from massed practice. *Poster presented at the 50th Annual Meeting of the Psychonomic Society,* Boston, MA, Nov. 19, 2009.

Guzman-Muñoz, F. J., & Johnson, A. (2007). Error feedback and the acquisition of geographical representations. *Applied Cognitive* Psychology, 22, 979-995.

Kornell, N. & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585-592.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (in press). Spacing as the friend of both memory and induction in younger and older adults. *Psychology and Aging.*

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989-998.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the Region of Proximal Learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 609-622.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*, 211-232.

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63*, 505-512.

Kulik, J. A, & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79-97.

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology, 51*, 239-243.

Landauer, T. K. & Bjork, R. A. (l978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, and R. N. Sykes (Eds.), *Practical aspects of memory,* pp. 625-632. London: Academic Press.

Lhyle, K. G., & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology, 79*, 320-322.

Mandler, J. M. (l978) A code in the node: The use of a story schema in retrieval. *Discourse Processes. 1*, 14-35.

Mandler, J. M. & DeForest, M. (l979). Is there more than one way to recall a story? *Child Development, 50*, 886-889.

Mandler, J. M., & Goodman, M. S. (l982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior, 21*, 507-523.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19,* 494-513.

McDaniel, M. A., & Fisher, R. P. (1991). Test and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192-201.

McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review, 14*, 200–206.

Metcalfe, J. (2002). Is study time allocated selectively to a Region of Proximal Learning? *Journal of Experimental Psychology: General*, *131*, 349-363.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Direction in Psychological Science*, *18*, 159-163. .

Metcalfe, J. & Finn, B. (under review). Hypercorrection of High Confidence Errors: Did they Know it All Along?

Metcalfe, J. & Jacobs, W.J. (2010). People's study time allocation and its relation to animal foraging. *Behavioral Processes, 83,* 213-221

Metcalfe, J., & Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *Journal of Memory and Language, 52*, 463-477.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors and feedback. *Psychonomic Bulletin and Review, 14*, 225-229.

Metcalfe, J., Kornell, N. & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37,* 1077-1087.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based program to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology, 19*, 743-768.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676-686.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3-8.

Piaget, J. (l952). *The origins of intelligence in children.* New York: International Universities Press.

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249-255.

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*. 181-210.

Schacter, D. L., Rich, S. A., & Stamp, M. S. (1985). Remediation of memory disorders: Experimental evaluation of the spaced retrieval technique. *Journal of Clinical and Experimental Neuropsychology, 7,* 79-96.

Schooler, L. J., & Anderson, J. R. (1990). The disruptive potential of immediate feedback. *Proceedings of the 12 Annual Conference of the Cognitive Science Society*, *12,* 702- 708.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592-604.

Son, L. K. & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204-221.

Storm, B. C., Bjork, R. A., & Storm, J. C. (in press). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*.

Vygotsky, L. S. (l987) *The collected works of L. S. Vygotsky, Volume 1. Problems of general psychology including the volume thinking and speech.* R. W. Rieber & A. S. Carton (Eds.), New York: Plenum Press.

Waugh, N. C. (l970). On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior, 16*, 465-478.

Winstein, C. J., & Schmidt, R. A. (l990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 677-691.

Wulf, G. & Schmidt, R. A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 748-757.