How to Promote Lasting Learning in Schools: Theoretical Approaches and an Agenda for Research

"Non vitae sed scholae discimus" ("Not for life but for school we learn"; Seneca, 65 B.C.E./2019). In his well-known complaint, the Roman philosopher Seneca criticizes the (philosophical) schools of his time for their, as he sees it, lack of relevance and utility of the learning content for real-life matters. While a similar criticism is also frequently voiced in contemporary normative debates about education, the relevance of school learning for life can be questioned from yet another perspective: How much do students actually retain from what they learn in school beyond the next class assignment? Is school instruction suitable for creating lasting knowledge and skills? And what can teachers do to foster learning that leads to lasting knowledge?

These questions are now timelier than ever. In the digital information society, just about any information seems to be accessible to everyone at the tap of a finger, and we rely on the Internet as a giant transactional memory system (Ward, 2013). The easy access to information creates the illusion that individual knowledge has become dispensable. Contrary to this view, the real-life importance of lasting knowledge cannot be overrated. Among other things, conceptual (declarative) knowledge is needed to make sense of new information (e.g., Bransford & Johnson, 1972), to critically evaluate information and compare conflicting claims (e.g., Richter & Maier, 2017), and to build mental models, that is the kind of mental representations needed to solve real-life problems and to interact with the world (e.g., Johnson-Laird, 1983). Procedural knowledge, such as cognitive skills, is obviously relevant for mastering all kinds of tasks. From a psychological perspective, conceptual and procedural knowledge are also the core components of *competences*, the acquisition of which is considered a major goal of education by many educational scientists and policy makers. Last but not least, knowledge is also required to acquire and retain new knowledge, preparing students for future learning (Bransford & Schwartz, 1999). In sum, there can be little doubt that building up knowledge, and lasting knowledge at that, is a central objective of formal education.

Considering the practical importance of lasting knowledge, we know surprisingly little about it. Empirical studies of long-term retention of educational content are rare. One of the few exceptions are Bahrick's studies of very long-term memory. For instance, Bahrick (1984) tested participants for their memory for Spanish (vocabulary, reading comprehension, grammar) that they had learned in high school or college. The time that had passed between the last class taken and the test varied from 1 to 50 years. One central finding was that the (collective) memory curves declined steeply in the first 3-6 years but reached an asymptote after that, meaning that memory remained stable except for a further decrease in old age possibly due to neurodegenerative effects. A second central finding was that the level of long-term retention depended on the amount of initial instruction, operationalized as the number of Spanish classes taken.

Thus, it seems that school learning can indeed create knowledge that lasts a lifetime (for the long-term retention of mathematics knowledge, see Bahrick & Hall, 1991; for a review of further studies, see Conway et al., 1992). Bahrick (1984) coined the term *permastore* for such very long-term memories. Moreover, the studies by Bahrick and colleagues show that the quantity of instruction makes a difference for the amount of knowledge that can be retained over years and decades. However, they cannot answer the question of whether specific instructional methods are better suited to create lasting knowledge than others. Next, we discuss approaches that aim to shed light on this question.

Theoretical Approaches to Foster Lasting Learning: Meaningful Learning and Desirable Difficulties

Basically, two broad lines of theory and associated research are devoted to lasting learning, the conditions that foster lasting knowledge and the underlying cognitive processes. The first, classical approach posits that *meaningful learning* is the key to creating lasting knowledge. In educational psychology, the work by Ausubel (1963) has laid the foundations for this approach by pointing out the pivotal role of prior knowledge for learning new information and the necessity that learners need to make sense of what they are learning, as opposed to pursuing rote learning, that is, simply memorizing information. Ausubel also pointed out that meaningful learning is an active process, implying that learners can employ cognitive strategies beyond rote rehearsal that can assist and benefit comprehension and learning. These basic ideas were subsequently taken up and developed further in various theoretical and instructional approaches such as generative and active learning (Fiorella & Mayer, 2016; Wittrock, 1974), and have become the mainstream view on learning in educational psychology and the educational sciences. The notion of meaningful learning also directly informs a variety of specific instructional methods, such as the use of advance organizers, concept mapping, or teaching and prompting learning strategies.

Many experiments have demonstrated unequivocally and for various types of information (e.g., verbal and pictorial information) that learned information is forgotten fast (often on a time scale of a few minutes) if it is not meaningfully processed, that is, comprehended well and integrated with prior knowledge (for examples of classical experiments, see Kintsch et al., 1990; Mandler & Ritchey, 1977). Moreover, meta-analytic reviews have shown, for a variety of broad instructional approaches and specific instructional methods, that meaningful learning can effectively improve learning in school settings. For instance, meta-analytic evidence is available for the effectiveness of learning strategy instruction (Donker et al., 2014), guided inquiry-based learning (Lazonder & Harmsen, 2016), reciprocal teaching (Rosenshine & Meister, 1994), concept mapping (Schroeder et al., 2018), or advance organizers (Stone, 1983) (see also the meta-meta-analysis by Hattie, 2009, Ch. 9 and 10, for further examples). It must be noted, though, that most of the available research has not specifically addressed the questions of whether and how instruction informed by the notion of meaningful learning fosters lasting learning.

The second major approach devoted to lasting learning has become known as *desirable difficulties* in learning (Bjork, 1994). Desirable difficulties are conditions in the instructional setting that make the learning process more difficult, both subjectively and objectively in terms of increasing cognitive load, but that foster long-term retention by stimulating cognitive processes conducive to lasting learning. Evidently, not all difficulties in learning are desirable. Most conditions that make learning more difficult merely impose extraneous cognitive load on the learner. For instance, badly designed learning materials, background noise, or presenting information that presupposes prior knowledge that students do not possess, do not qualify as desirable difficulties.

Applied memory research has identified various types of conditions that may, at least under certain circumstances, fulfill the criteria for desirable difficulties. These include having learners generate information themselves instead of presenting it (generation effect, McDaniel et al., 1988; Slamecka & Graf, 1978), varying the context between learning sessions instead of keeping it constant (Smith et al., 1978), or intentionally reducing the processing fluency during learning (Diemand-Yauman et al., 2011). Among the desirable difficulties that have attracted the most interest among researchers in educational psychology and the learning sciences are retrieval practice, spacing, and interleaving:

- *Retrieval practice* refers to the use of tests as learning opportunities, which are assumed to be more effective than restudying for long-term retention. This so-called testing effect has been demonstrated in many studies based on different materials, from simple verbal materials (Karpicke & Roediger, 2006) to contents of expository texts (e.g., Roediger & Karpicke, 2006), both in the laboratory and in classroom experiments (for meta-analyses, see Rowland, 2014; Yang et al., 2021). Compared to restudying the same materials, retrieval practice requires cognitive effort and potentially reveals knowledge gaps, which are likely to increase the difficulty of the technique from the learner's perspective. For example, students tend to believe that rereading is a more effective learning strategy than self-testing, although the evidence suggests the opposite (e.g., Kornell & Son, 2009).

- *Spacing* refers to the distribution of the total available learning time over several shorter units instead of massing it in one unit (or few longer units). Spacing is probably the best-researched desirable difficulty (for a meta-analytic review, see Cepeda et al., 2006). The effectiveness of spacing for long-term learning outcomes has been established for different time scales and for different contents such as verbal materials, cognitive skills, or motor skills. Importantly, positive evidence exists only for learning that involves repetition, such as spacing repeated presentations of the same information or spacing practice phases with identical or similar exercises. In such situations, spacing can benefit lasting learning, especially with longer (but not too long) intervals between learning sessions. Like retrieval practice, spaced learning is often experienced as more difficult than massed learning, as forgetting occurs between learning sessions and relearning proceeds less fluently after a delay. In line with this reasoning, students often (although not always) prefer massed over spaced learning (for a review, see Son & Simon, 2012).
- *Interleaving* refers to the sequence of exemplars in inductive learning settings, for instance, when concepts or principles are to be derived from a set of observations or case descriptions, or by practicing tasks that require different rules to solve them. In interleaved learning, the exemplars corresponding to different categories or rules are presented in an interleaved instead of a blocked fashion. Interleaved learning seems to play out its advantages if the discrimination between exemplars is crucial for learning, but positive evidence is available only for some types of materials, such as visual stimuli and mathematical tasks (for meta-analytic results, see Brunmair & Richter, 2019). Moreover, experiments that have specifically addressed the question of whether interleaving fosters lasting learning are still scarce, and the available findings are inconclusive. Studies have shown that learners perceive inductive learning with a blocked presentation of exemplars

as more successful than learning in an interleaved fashion, even if the actual learning outcomes follow the opposite pattern (e.g., Kornell & Bjork, 2008).

Promises and Challenges of Implementing Desirable Difficulties in Educational Contexts

The notion of desirable difficulties has gained considerable popularity in recent years, among researchers in cognitive and educational psychology and among educational practitioners. The increased interest may partly be due to the counterintuitive nature of the claim that making learning more difficult benefits learning. Another likely factor is the sustained effort of proponents of desirable difficulties to emphasize the relevance of desirable difficulties for education and to market the approach to educational practitioners (for examples, see the book by Brown et al., 2014; or the website https://www.learningscientists.org/). Most importantly, the approach of desirable difficulties is also backed up by strong experimental evidence, mostly from lab studies. For retrieval practice, spacing, and interleaving, meta-analyses exist that report at least medium overall effects of each desirable difficulty (Brunmair & Richter, 2019; Rowland, 2014; Yang et al., 2021), suggesting that these desirable difficulties have great potential for fostering lasting learning in educational contexts.

However, for both conceptual and empirical reasons, it also seems clear that simply transferring the learning principles into the classroom might not work in all cases. Empirically, the evidence for the effectiveness of desirable difficulties in the classroom is mixed and still scarce, especially for interleaving and spacing. For retrieval practice, many classroom studies underline the potential of practice tests for learning, but many of these studies were based on non-experimental designs or used a “no activity” control group (Yang et al., 2021). Moreover, many studies have focused on simple recall and recognition of information and neglected comprehension and transfer that are crucial for school learning. The available empirical evidence and questions regarding its scope and validity will be discussed in more depth in the contributions by Roelle et al. (this issue) for retrieval practice, by Ebersbach et al. (this issue) for spacing, and by Richter et al. (this issue) for interleaving.

Theoretically, explanatory accounts of desirable difficulties help us to understand the basic memory processes that are important in educational contexts as well, such as how retrieval practice contributes to the consolidation of memories (Karpicke & Roediger, 2008) or how contextual variability increases the chances that learned information is passively cued and retrieved at a later test (Smith et al., 1978). But the more complex types of learning that occur in the classroom require a broader theoretical perspective that includes how such basic processes interact with meaningful processing. For instance, existing accounts of retrieval practice fail to specify the role of initial comprehension of the learning materials, which might be a crucial precondition for the effectiveness of retrieval practice with complex materials such as expository texts (Rummer & Schweppe, 2022). Likewise, concepts from theories of meaningful processing are needed to explain how retrieval might foster not only memory of explicit information but also transfer (Pan & Rickard, 2018). With regard to spacing, the deficient processing account posits that distributed practice elicits deeper processing than massed practice (e.g., Hintzman, 1974). But to specify what "deeper processing" of a scientific explanation, a mathematical task, or other contents typically encountered in school means, researchers need to turn to theories of meaningful processing. To give a final example, to be effective in educational contexts, which usually require self-regulated learning to a considerable degree, desirable difficulties must stimulate learners to actively process information during learning. Therefore, instructional measures that guide and support learners to use relevant strategies in the right situations, for example, comparison strategies in interleaved mathematics learning (Ziegler & Stern, 2016), are likely to boost the effectiveness of desirable difficulties. Again, these questions will be discussed in more depth in the other articles in this special issue.

Conclusion: Towards a Theory of Lasting Learning in Educational Contexts

Returning to the questions raised in the beginning, some evidence exists that students can retain a considerable proportion of the knowledge acquired in school, provided that they received a sufficient dose of instruction. Thus, school instruction seems to be suitable, at least to some extent, to create lasting knowledge although more (and better controlled) longitudinal studies would be very useful for further clarifications. Regarding the question what teachers should do to foster lasting learning, the approaches of meaningful learning and desirable difficulties have both yielded principles and instructional methods that might contribute to this goal. However, a theoretical integration of the two complementary approaches seems to be an important objective for psychology and the learning sciences at large. Instead of holding “horse races” (Salomon, 2002) and asking which of the two approaches performs better as has been done in the past (e.g., Karpicke & Blunt, 2011; Rummer et al., 2017), we should move on to the question how ideas of meaningful learning and desirable difficulties can be fruitfully combined to create lasting knowledge. For example, research should examine to what extent the “deeper processing” that is supposedly instigated by desirable difficulties overlaps with or is distinct from those processes described in theories of meaningful learning. A related issue is to determine the roles of strategic and routine learning processes for the effectiveness of desirable difficulties. If strategic processes play a role, students could be taught relevant learning strategies to make the most of desirable difficulties. Likewise, instructional support might be effective in assisting students to use desirable difficulties constructively.

To examine these issues, experimental intervention studies in schools are needed that assess learning outcomes not only shortly after learning but also at delays of several weeks or longer. Such studies are ambitious and costly in terms of time and other resources and have rarely been conducted in the past. But the investment is likely to pay off. Longitudinal experimental studies in the classroom promise to significantly advance the theoretical understanding of lasting learning and how it can be effectively implemented in the classroom.

References

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. Grune & Stratton.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*(1), 1-29. https://doi.org/10.1037/0096-3445.113.1.1

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120*(1), 20–33. https://doi.org/10.1037/0096-3445.120.1.20

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717-726. https://doi.org/10.1016/S0022-5371(72)80006-9

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61-100. https://doi.org/10.3102/0091732X024001061

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Harvard University Press.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin, 145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Conway, M. A., Cohen, G., & Stanhope, N. (1992). Very long-term memory for knowledge acquired at school and university. *Applied Cognitive Psychology, 6*(6), 467-482. https://doi.org/10.1002/acp.2350060603

Diemand-Yauman, C., Oppenheimer, D. M., Vaughn, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. Cognition, 118, 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Donker, A. S., De Boer, H., Kostons, D., Van Ewijk, C. D., & van der Werf, M. P. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review, 11*, 1-26.

Ebersbach, M., Lachner, A., Scheiter, K. & Richter, T. (submitted). Using spacing to foster lasting learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. **[Platzhalter, muss am Ende aktualisiert werden]**

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717-741. https://doi.org/10.1007/s10648-015-9348-9

Hattie, J (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Lawrence Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966-968. https://doi.org/10.1126/science.115240

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*(2), 133-159. https://doi.org/10.1016/0749-596X(90)90069-C

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*(5), 493-501. https://doi.org/10.1080/09658210902832915

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681-718. https://doi.org/10.3102/0034654315627366

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 3*(4), 386. https://doi.org/10.1037/0278-7393.3.4.386

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27*(5), 521-536. https://doi.org/10.1016/0749-596X(88)90023-X

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*(7), 710–756. https://doi.org/10.1037/bul0000151

Richter, T., Nemeth, L., Berger, R., Borromeo Ferri, R., Hänze, M. & Lipowsky, F. (submitted). Using interleaving to foster inductive learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. **[Platzhalter, muss am Ende aktualisiert werden]**

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist, 52*(3), 148-166. https://doi.org/10.1080/00461520.2017.1322968

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255. https://doi.org/10.1126/science.1152408

Roelle, J., Schweppe, J., Endres, T., Lachner, A., von Aufschnaiter, C., Renkl, A., Eitel, A., Leutner, D., Rummer, R., Scheiter, K., & Vorholzer, A. (submitted). Combining retrieval practice and generative learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. **[Platzhalter]**

Rosenshine, B., & Meister, C. 1994. Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*, 479–530. https://doi.org/10.3102/00346543064004479

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rummer, R. & Schweppe, J. (2022). Komplexität und der Testungseffekt: Die mögliche Bedeutung der Verständnissicherung für den Nutzen von Abrufübung bei komplexem Lernmaterial [Complexity and the testing effect: The potential importance of securing understanding for the benefits of retrieval practice with complex learning materials]. *Unterrichtswissenschaft, 50,* 37-52. https://doi.org/10.1007/s42010-021-00137-4

Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? *Journal of Experimental Psychology: Applied, 23*(3), 293 – 300. https://doi.org/10.1037/xap0000134

Seneca, L. A. (2019). *Epistulae morales ad Lucilium, Liber XVII* [Moral letters to Lucilium, Book 17]. https://la.wikisource.org/wiki/Epistulae_morales_ad_Lucilium/Liber_XVII_-_XVIII (Original work published 65 B.C.E.)

Salomon, G. (2002). Technology and pedagogy: Why don't we see the promised revolution? *Educational Technology, 42*(2), 71-75. https://www.jstor.org/stable/44428740

Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review, 30*(2), 431-455. https://doi.org/10.1007/s10648-017-9403-9

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*(4), 342-353. https://doi.org/10.3758/BF03197465

Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review, 24*, 379–399. https://doi.org/10.1007/s10648-012-9206-y

Stone, C. L. (1983). A meta-analysis of advance organizer studies. *The Journal of Experimental Education, 51*(4), 194-199. https://doi.org/10.1080/00220973.1983.11011862

Ward, A. F. (2013). Supernormal: How the Internet is changing our memories and our minds. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 24*(4), 341—348. https://doi.org/10.1080/1047840X.2013.850148

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist, 11*(2), 87–95. https://doi.org/10.1080/00461520903433554

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309

Ziegler, E., & Stern, E. (2016). Consistent advantages of contrasted comparisons: Algebra learning under direct instruction. *Learning and Instruction, 41*, 41–51. https://doi.org/10.1016/j.learninstruc.2015.09.006