

**JUILLET 2017 – VOL. 7 N° 2**

## **LES DIFFICULTÉS DÉSIRABLES : DES OBSTACLES POUR FAVORISER L'APPRENTISSAGE**

Michaël LÉVESQUE-DION<sup>1</sup>\*

<sup>1</sup> École de psychologie, Université Laval, Québec, Canada

\* [michael.levesque-dion.1@ulaval.ca](mailto:michael.levesque-dion.1@ulaval.ca)

### **Pour citer l'article**

Lévesque-Dion, M. (2017). Les difficultés désirables: Des obstacles pour favoriser l'apprentissage. *Psycause : Revue scientifique étudiante de l'École de psychologie de l'Université Laval*, 7(2), 18-25.

# LES DIFFICULTÉS DÉSI RABLES : DES OBSTACLES POUR FAVORISER L'APPRENTISSAGE

Michaël LÉVESQUE-DION, étudiant au Baccalauréat en psychologie  
École de psychologie, Université Laval, Québec, QC, Canada

## Résumé

Le concept de « difficulté » est à l'origine de nombreuses fausses croyances en ce qui concerne l'apprentissage. En particulier, il s'avère que certaines méthodes qui semblent initialement faire obstacle à l'apprentissage augmentent en fait la rétention à long terme de l'information. Ces « difficultés désirables » prennent plusieurs formes : l'espacement, l'interfoliage (traduction libre) et le testing sont toutes des façons contre-intuitives de favoriser l'apprentissage. La présente revue de littérature permet de voir dans quelle mesure les difficultés désirables peuvent être utiles et comment elles peuvent être employées dans un contexte académique.

**Mots-clés :** apprentissage, mémoire, difficultés désirables, espacement, interfoliage, effet de test

## Abstract

The concept of “difficulty” lies at the root of many misconceptions concerning learning and memory. More precisely, certain techniques which initially seem to impede learning actually increase long term retention. These “desirable difficulties” take many forms: spacing, interleaving and testing are all counter-intuitive ways to foster learning. This literature review shows how desirable difficulties can be useful and how they may be implemented in an academic context.

**Keywords :** learning, memory, desirable difficulties, spacing, interleaving, testing effect

## Introduction

Le concept de « difficulté » est à l'origine de nombreuses fausses croyances en ce qui concerne l'apprentissage. D'un côté, plusieurs pensent à tort qu'il suffit de faire plus d'efforts pour mieux retenir, sans égard aux méthodes employées. D'autres pensent plutôt qu'il faut éviter autant que possible les erreurs et les obstacles en cours d'apprentissage. Or, ces deux approches reposent sur une conception erronée de l'apprentissage humain et mènent directement à

l'utilisation de méthodes d'apprentissage inefficaces. En fait, certaines méthodes d'apprentissage qui font obstacle à la performance immédiate peuvent augmenter la rétention de l'information à long terme (Bjork & Bjork, 2011). Ces « difficultés désirables » ont des effets contre-intuitifs, ce qui explique peut-être pourquoi elles sont peu connues et relativement peu utilisées dans des contextes pédagogiques (Bjork & Bjork, 2011; Roediger & Karpicke, 2006). Dans l'immédiat,

ces méthodes semblent avoir un effet délétère sur l'apprentissage : la performance est faible ou progresse lentement, ce qui suggère que l'individu n'apprend pas. Or, lorsque la performance est testée après un délai, ces méthodes s'avèrent bénéfiques pour la rétention à long terme : les résultats sont supérieurs à ceux obtenus avec des méthodes « standard » couramment utilisées (Bjork & Bjork, 2011). Bjork & Bjork (2011) proposent un principe général pour expliquer ce phénomène : plus une information est difficile à récupérer (p. ex., parce qu'elle a été étudiée il y a longtemps), plus sa trace en mémoire sera renforcée lorsqu'elle sera récupérée. Si une information est facile à récupérer dans l'immédiat, sa trace en mémoire sera peu renforcée, et il sera plus difficile de la récupérer après un délai.

## Objectif

Les mécanismes spécifiques qui sous-tendent l'effet des difficultés désirables ne font pas l'objet d'un consensus (Rowland, 2014), et il est tout à fait envisageable que différentes formes de difficultés désirables exercent leurs effets bénéfiques via des mécanismes différents (Soderstrom & Bjork, 2015). La présente revue de littérature se concentrera plutôt sur les aspects pratiques associés aux difficultés désirables dans le but de répondre à la question « quels sont les principaux types de difficultés désirables et quelles peuvent être leurs applications dans le domaine académique ? » En ce sens, la présente revue s'adresse particulièrement aux étudiants ou aux éducateurs qui voudraient bonifier leurs techniques d'étude ou d'enseignement.

## Une variété de difficultés désirables

### L'espacement et l'interfoliage

Plutôt que d'étudier un sujet à une seule occasion, l'espacement (spacing ou étude distribuée) consiste à distribuer l'étude d'un sujet sur plusieurs séances séparées dans le temps. Une méta-

analyse récente des résultats de plusieurs dizaines d'études sur ce sujet montre qu'à temps d'étude égal, il est préférable de distribuer ses efforts : la performance à un test subséquent sera généralement supérieure (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006). Sommairement, ce phénomène montre l'importance de l'oubli pour l'apprentissage : lorsqu'il y a un délai entre les expositions au matériel, il faut le réapprendre (Bjork & Bjork, 2011). Ces efforts additionnels se traduisent par une trace plus solide en mémoire.

Une autre forme de difficulté désirable proche de l'espacement consiste à alterner entre divers sujets au cours d'une même période d'étude, plutôt que d'étudier chaque sujet l'un après l'autre, par blocs. Cette technique appelée interleaving, ou interfoliage en français (traduction libre), constitue en quelque sorte une forme d'espacement plus « actif » : l'espace entre les séances d'étude sur un sujet est rempli par l'étude d'autres sujets. Une telle méthode a notamment été utilisée en laboratoire pour améliorer l'apprentissage de formules mathématiques. Rohrer et Taylor (2007) ont assigné aléatoirement leurs participants à l'un de deux groupes : le groupe « Mixte », qui devait résoudre des problèmes mathématiques interfoliés, et le groupe « Par blocs », qui étudiait un type de problème à la fois. Pendant la période de pratique, la performance était meilleure chez les participants du groupe « Par blocs » (89 % vs 60 % des problèmes résolus correctement). Effectivement, il est plus facile de résoudre plusieurs problèmes similaires d'affilée, requérant tous la même méthode, que de devoir changer de méthode à chaque fois qu'un nouveau problème se présente (Meiran, 1996). Cependant, une semaine après la dernière période de pratique, la performance était largement supérieure chez les participants du groupe « Mixte », avec une performance de 63 % en moyenne, alors que les participants du groupe « Par blocs » avaient seulement une performance de 20 %. Face aux problèmes interfoliés, les participants étaient forcés de récupérer la méthode appropriée en

mémoire à chaque nouveau problème, ce qui aurait renforcé la trace de ces informations en mémoire. Les participants du groupe « Par blocs », qui n'avaient pas à faire cet effort à chaque essai, n'ont pas connu les mêmes bénéfices à long terme (Rohrer & Taylor, 2007).

Cela étant, il faut rappeler le fait que la performance immédiate, pendant la pratique, était meilleure lorsque les solutions aux problèmes mathématiques étaient étudiées une par une. L'aisance qui découle de l'utilisation de la méthode « par blocs », en comparaison avec les difficultés associées à l'interfoliage, introduit la possibilité que ses utilisateurs pensent à tort qu'il s'agit d'une méthode d'apprentissage efficace. Kornell et Bjork (2008) ont investigué cette hypothèse à l'aide d'une tâche où les participants devaient associer des peintures à leur auteur. Avant de prendre part à la tâche, les participants devaient compléter une phase d'apprentissage où des peintures, accompagnées du nom de leurs auteurs, leur étaient présentées une à la fois pendant 3 secondes sur un écran d'ordinateur. Encore une fois, l'interfoliage de peintures issues de différents artistes pendant la pratique était comparé au regroupement de peintures en blocs du même artiste, et l'étude comportait à la fois un volet intra-sujet et un volet inter-sujet, avec deux échantillons différents. Dans le volet intra-sujet, pendant la phase de pratique, les participants étudiaient tantôt les œuvres de 6 artistes en alternance, et tantôt les œuvres de 6 autres artistes, regroupées par artiste (avec le contrebalancement approprié entre les sujets). Dans le volet inter-sujet, les participants étudiaient les œuvres des 12 artistes selon l'une ou l'autre des méthodes. Lors de la phase de test, quelques minutes après la phase de pratique, de nouvelles peintures de chaque artiste étaient présentées en ordre aléatoire, et les participants devaient rapporter le nom de l'artiste. Tel qu'attendu, la performance au test était largement supérieure dans la condition où les œuvres des différents auteurs étaient mélangées pendant la pratique.

Les participants dans le volet intra-sujet, qui avaient pu essayer les deux techniques, devaient aussi répondre à la question suivante : « Laquelle des deux techniques est la plus efficace, selon vous ? ». Soixante-trois pour cent des participants ont jugé la technique « par blocs » plus efficace, et un 15% additionnel a jugé que les deux techniques étaient aussi efficaces l'une que l'autre. La plupart des participants semblaient donc croire que l'étude « par blocs » était supérieure, probablement parce que cette technique tend à produire une impression de facilité. En effet, après avoir observé 10 peintures du même artiste, les particularités propres à cet artiste sont aisément disponibles en mémoire.

Les effets de l'interfoliage ont été reproduits dans divers contextes : notamment, l'apprentissage de la prononciation de mots français (Carpenter & Mueller, 2013), et l'apprentissage de diagnostics basés sur des données d'électrocardiogramme (Hatala, Brooks & Norman, 2003). Étant donné que l'interfoliage implique nécessairement l'espacement des différents items à l'étude, il est possible de penser que les bénéfices associés à l'interfoliage soient dus uniquement à l'espacement. En contrôlant la durée écoulée entre la présentation des items (l'espacement) pour qu'elle soit équivalente dans une condition « Mixte » et dans une condition « Par blocs », Taylor et Rohrer (2010) ont montré que l'effet de l'interfoliage dépasse le simple effet de l'espacement. Leurs résultats suggèrent que l'interfoliage peut notamment améliorer la performance en augmentant la capacité des utilisateurs à discriminer entre les différents items. Une hypothèse alternative serait que l'interfoliage amplifie l'oubli dû à l'espacement en augmentant l'interférence en mémoire. Le fait d'étudier un nouveau sujet, par exemple, rend l'information sur les autres sujets plus difficile à récupérer (Wixted, 2005), ce qui augmente la solidité de leur trace mnésique lorsqu'ils sont finalement récupérés (Bjork & Bjork, 2011).

Les études présentées jusqu'à maintenant concernant l'interfoliage présentent une faiblesse méthodologique: tous les participants, peu importe la technique utilisée lors de la pratique, complétaient une évaluation finale interfoliée, où l'ordre des problèmes ou des items était aléatoire. Ainsi, il est possible que la performance des participants ayant utilisé l'interfoliage lors de la pratique soit meilleure simplement parce qu'ils étaient plus familiers avec la forme de l'évaluation, un peu comme dans le phénomène de traitement approprié au transfert (Morris, Bransford & Franks, 1977). Une telle limite ne pourrait être adressée de façon satisfaisante qu'avec une nouvelle étude dans laquelle les deux types d'évaluation seraient employés: la moitié des participants complèteraient une évaluation finale « Par blocs » et l'autre moitié une évaluation finale interfoliée. En attendant, dans la « vraie vie », les problèmes à résoudre ne se présentent pas un « type » à la fois (Bjork & Bjork, 2011).

### L'effet de test et la rétroaction

L'effet de test (testing effect) réfère à la différence de performance observée, lors d'une évaluation, entre des individus ayant étudié deux fois le matériel susceptible d'être évalué et des individus ayant étudié une fois le matériel, puis complété une forme d'évaluation de pratique. Le fait de compléter une évaluation est une difficulté désirable puisqu'il est plus facile de relire ou de réviser des informations que de tenter de les récupérer en mémoire, et que cet effort de récupération supplémentaire donne lieu à un apprentissage plus durable.

Dans une étude de Roediger et Karpicke (2006), des étudiants au baccalauréat devaient compléter deux conditions expérimentales, contrebalancées entre les participants: « étude + étude » et « étude + test ». Une séance d'étude consistait simplement à lire et relire un texte, pendant 7 minutes. Une séance de test consistait quant à elle à tenter de rappeler autant d'informations du texte que possible, par écrit, pendant 7 minutes. Les participants

avaient donc deux textes à lire, un qu'ils étudiaient deux fois et un autre qu'ils étudiaient une fois, puis sur lequel ils se testaient. Les participants étaient ensuite séparés aléatoirement en trois groupes de quarante participants chacun, qui complétaient l'évaluation finale respectivement cinq minutes, deux jours ou une semaine après la dernière séance d'étude ou de test. Après cinq minutes, la performance était supérieure pour le matériel qui avait été étudié deux fois de suite. Cependant, après un délai de deux jours, la performance était beaucoup plus grande pour le matériel sur lequel les étudiants s'étaient testés, et les bénéfices persistaient avec un délai d'une semaine. Il s'agit encore une fois d'un effet de difficulté désirable: même si la performance est moins bonne dans l'immédiat, elle est meilleure à plus long terme, ce qui indique que l'apprentissage est en réalité plus solide.

Dans une seconde expérience, Roediger et Karpicke (2006) ont cherché à étendre la portée de leurs résultats en élaborant trois nouvelles conditions expérimentales dans un devis inter-sujets. Les participants étaient assignés aléatoirement à (a) quatre périodes d'études subséquentes (ÉÉÉÉ), (b) trois périodes d'étude suivies d'une période de test (ÉÉÉÉT), ou (c) une période d'étude suivie de trois périodes de test (ÉTTT). À l'évaluation finale, 5 minutes après la dernière séance d'étude ou de test, la performance était plus élevée pour les participants ayant complété plus de séances de test: la performance pour le groupe ÉÉÉÉ était supérieure à celle du groupe ÉÉÉÉT, dont la performance était supérieure à celle du groupe ÉTTT. Cependant, chez les participants évalués une semaine plus tard, le patron inverse pouvait être observé: la condition ÉTTT était supérieure à ÉÉÉÉT, qui elle était largement supérieure à la condition ÉÉÉÉ. Il semble donc qu'une seule séance de test était suffisante pour augmenter la performance à long terme. Autre fait intéressant, lorsque les participants étaient questionnés, après la dernière séance de pratique (d'étude ou de test), à savoir à quel point

ils pensaient bien se rappeler du texte une semaine plus tard, les participants les plus confiants étaient ceux dans la condition ÉÉÉÉ. Encore une fois, une seule séance de test était suffisante pour diminuer la perception des participants qu'ils seraient capables de bien rappeler le passage une semaine plus tard (il n'y avait pas de différence entre ÉÉÉT et ÉTTT sur ce point). Le fait de se tester (ici, tenter de rappeler les informations étudiées) permettrait vraisemblablement à l'individu de se rendre compte des trous dans son apprentissage, et conséquemment d'émettre des prédictions plus réalistes. Le fait de relire plusieurs fois un texte, au contraire, serait susceptible de produire une impression de maîtrise qui encourage l'individu à surévaluer son apprentissage réel et conséquemment, à sous-performer au moment de l'évaluation.

Concernant l'effet de test de façon plus générale, une méta-analyse de Rowland (2014) rapporte une taille d'effet modérée ( $g^1 = 0,50$ ). Par ailleurs, l'effet est plus grand lorsque l'évaluation finale se fait sous forme de rappel indicé ( $g = 0,57$ ) ou de rappel libre ( $g = 0,49$ ) plutôt que d'un test de reconnaissance ( $g = 0,31$ ), lorsque le délai entre la pratique et l'évaluation finale est supérieur ou égal à un jour ( $g = 0,69$ ) plutôt qu'inférieur à un jour ( $g = 0,41$ ), et lorsqu'une rétroaction est fournie lors de la pratique ( $g = 0,73$ ), comparativement à une absence de rétroaction ( $g = 0,39$ ). Outre leurs bénéfices directs en termes d'apprentissage, les évaluations (peu importe leur forme) sont une excellente façon d'identifier les défauts dans sa mémoire (Bjork & Bjork, 2011). En effet, il peut être difficile d'avoir une idée exacte de sa maîtrise d'un sujet simplement en se posant la question « est-ce que je connais bien ce sujet? », surtout si l'on vient tout juste de réviser. L'information est alors facilement récupérable, et il est difficile d'évaluer la « solidité » réelle de la trace mnésique (Bjork & Bjork, 2011). Ce qui est plus productif, c'est de se poser des questions sur le sujet en question. Il devient alors évident s'il est bien maîtrisé ou non.

Le fait de se tester (p. ex., se poser des questions sur un matériel à assimiler) implique nécessairement de faire des erreurs. Malgré les résultats observés dans les études précédentes, il y a donc lieu de se demander si les évaluations améliorent la rétention d'informations correctes, mais aussi d'informations incorrectes (p. ex., comme dans le phénomène des flashbulb memories; Talarico & Rubin, 2003). En cohérence avec les résultats de Rowland (2014), plusieurs études suggèrent une méthode facile pour éviter que des erreurs ne s'incrument en mémoire : fournir une rétroaction. Kornell, Hays et Bjork (2009) ont investigué plus précisément l'effet des rétroactions à l'aide d'une tâche d'association de mots faiblement liés sémantiquement. Dans la phase d'apprentissage, les participants devaient apprendre des paires de mots faiblement liés sémantiquement (p. ex., étang - grenouille), puis reproduire les associations lors d'une évaluation subséquente (p. ex. étang - \_\_\_\_\_). L'étude comportait deux conditions : dans la première, les participants devaient étudier l'association de mots présentée à l'écran pendant 13 secondes. Dans l'autre condition, seul le premier mot était présenté pendant les 8 premières secondes de l'essai, et les participants devaient tenter de deviner le second mot. Puis, dans les 5 secondes suivantes, les deux mots étaient présentés simultanément à l'écran. Après avoir retiré les paires de mots pour lesquelles les participants avaient réussi à deviner le deuxième mot (c'est-à-dire, en conservant seulement les essais où les participants avaient fait une erreur lors de la pratique), la performance était significativement plus élevée dans cette condition. Ainsi, après s'être trompés de réponse en tentant de deviner le second mot, la présentation d'une rétroaction a permis aux participants de mieux performer que si la bonne réponse leur avait été présentée du premier coup.

Ces résultats ont seulement été reproduits au niveau conceptuel par Richland, Kornell et Kao (2009), à l'aide d'une tâche différente : les

1 N.B. : Le  $g$  de Hedges est une mesure de taille d'effet similaire au  $d$  de Cohen qui contrôle pour la différence de taille entre les échantillons des deux conditions expérimentales comparées.

participants devaient lire un texte, puis répondre à une dizaine de questions sur son contenu. La moitié des participants avait deux minutes pour compléter un pré-test qui contenait cinq des 10 questions de l'évaluation finale, suivies de huit minutes pour lire le texte. L'autre moitié des participants n'avait pas de pré-test et disposait simplement de 10 minutes pour lire le texte. Après avoir retiré, pour chaque participant individuellement, les questions auxquelles le participant avait eu la réponse correcte au pré-test (peu fréquent), les résultats ont montré que la performance était largement supérieure pour les items testés préalablement. Ainsi, malgré la production initiale d'une réponse incorrecte, le fait d'avoir une rétroaction indirecte (la lecture du texte) a permis aux participants d'améliorer leur performance par la suite.

## Portée et applications

Divers mécanismes ont été proposés pour expliquer la façon dont les difficultés désirables peuvent améliorer la rétention à long terme de l'information, mais il n'y a pas de consensus à l'heure actuelle (Rowland, 2014). L'incertitude quant aux mécanismes à l'origine des effets des difficultés désirables ne devrait toutefois pas être vu comme un obstacle à leur utilisation. En effet, même si les mécanismes ne sont pas tous compris, de nombreuses études ont montré leurs effets positifs. Des recherches récentes ont d'ailleurs montré l'efficacité de techniques basées sur les difficultés désirables dans un contexte académique : l'espacement (Seabrook, Brown & Solity, 2005), l'interfoliage (Rohrer, Dedrick et Stershic, 2015) et les tests de pratique (Bjork, Little et Storm, 2012; Roediger, Agarwal, McDaniel & McDermott, 2011) ont toutes été utilisées avec succès dans de vraies salles de classe. D'autres recherches ont montré que le fait de se tester soi-même peut aussi avoir des bénéfices substantiels, p.ex. à l'aide de flashcards (Dobson & Linderholm, 2015; Karpicke, 2009).

Plus concrètement, comment ces techniques peuvent-elles être utilisées, dans la classe et lors de l'étude? En ce qui concerne la technique de testing, simplement poser des questions en classe (en invitant les étudiants à y répondre dans leur tête ou en levant la main), effectuer des retours rapides sur la matière des cours précédents, ou fournir aux étudiants des exercices de pratique et de la rétroaction à la suite des examens, sont autant de méthodes susceptibles d'être efficaces (Brown, Roediger & McDaniel, 2014). Pour l'étude, cacher la réponse à une question et tenter de la rappeler, ou faire une séance de questions avec un collègue étudiant sont deux exemples simples et faciles à utiliser (Brown, Roediger & McDaniel, 2014). Concernant l'espacement, il s'agit d'être assidu : étudier quelques heures chaque semaine plutôt que passer la nuit à étudier avant un examen (Brown, Roediger & McDaniel, 2014). Les deux techniques peuvent être efficaces à court terme, mais la seconde est beaucoup moins susceptible de produire un apprentissage durable (Bjork et Bjork, 2011). Pour utiliser l'interfoliage, alterner entre les sujets et entre les cours au cours d'une même période d'étude, plutôt que de se concentrer sur un seul sujet pourrait être un moyen appliqué d'utiliser cette technique. En classe et dans les évaluations de pratique, présenter des problèmes ou des questions «pêle-mêle» plutôt que «par thème». Une dernière chose très importante à considérer : il faut cultiver des attentes réalistes. Les difficultés désirables peuvent sembler contre-productives à court terme (Bjork & Bjork, 2011) : il faut donc savoir à quoi s'attendre et être prêt à persévérer pour en tirer les bénéfices.

## Conclusion

Traditionnellement, les examens sont vus comme une façon de vérifier l'apprentissage qui a été fait. La recherche sur les difficultés désirables montre que ce sont aussi des occasions d'apprentissage. Pour favoriser la rétention à long terme de

l'information, il n'est toutefois pas nécessaire d'attendre à l'examen: les enseignants peuvent utiliser diverses techniques d'évaluation en classe ou fournir à leurs élèves des moyens de s'autoévaluer, et les étudiants peuvent se tester de diverses façons lors de leurs séances d'étude. L'espacement, l'interfoliage, et toute autre technique qui demande de faire des efforts pour récupérer l'information sont aussi susceptibles de produire des gains durables et d'augmenter la rétention de l'information au-delà de l'examen final. Des recherches plus poussées seront nécessaires afin d'identifier clairement les mécanismes qui sous-tendent les effets des difficultés désirables et de définir des paramètres optimaux pour encadrer leur utilisation. D'ici là, l'idée générale demeure pertinente: pour apprendre, il faut se donner plus d'occasions de récupérer l'information, même si c'est difficile. Sur le moment, les difficultés désirables peuvent donner l'impression à l'apprenant qu'il fait du surplace... mais s'il persévère, ses efforts seront récompensés.

### Lectures suggérées

- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, 64, 417-444.
- Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.

### Liste de références

- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth Publishers.
- Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, 3(3), 165-170. doi: 10.1016/j.jar-mac.2014.03.002
- Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.
- Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & cognition*, 41(5), 671-682. doi: 10.3758/s13421-012-0291-4
- 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. doi: 10.1037/0033-2909.132.3.354
- Dobson, J. L., & Linderholm, T. (2015). Self-testing promotes superior retention of anatomy and physiology information. *Advances in Health Sciences Education*, 20(1), 149-161. doi: 10.1007/s10459-014-9514-8
- Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education*, 8(1), 17-26. doi: 10.1023/A:1022687404380
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469-486. doi: 10.1037/a0017341.
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories is spacing the "enemy of induction"?. *Psychological science*, 19(6), 585-592. doi: 10.1111/j.1467-9280.2008.02127.x

- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989. doi: 10.1037/a0015729
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1423. doi: 10.1037/0278-7393.22.6.1423
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519-533. doi: 10.1016/S0022-5371(77)80016-9
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15(3), 243-257. doi: 10.1037/a0016496
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological science*, 17(3), 249-255. doi: 10.1111/j.1467-9280.2006.01693.x
- Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17(4), 382. doi: 10.1037/a0026252
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6), 481-498. doi: 10.1007/s11251-007-9015-8
- Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, 107(3), 900. doi: 10.1037/edu0000001
- 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. doi: 10.1037/a0037559
- Seabrook, R., Brown, G. D., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, 19(1), 107-122. doi: 10.1002/acp.1066
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176-199. doi: 10.1177/1745691615569000
- Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, 14(5), 455-461. doi: 10.1111/1467-9280.02453
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24(6), 837-848. doi: 10.1002/acp.1598
- Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science*, 14(1), 6-9. doi: 10.1111/j.0963-7214.2005.00324.x