

再生課題によるテスト効果

遠藤 正雄

Testing Effect with Free Recall Tests

Masao ENDO

Abstract

Testing improves retention. Initial recall tests between encoding and retrieval increase the performance of the final test. This phenomenon is called direct testing effect. Direct testing effect is interesting in the point that testing effect exists even when initial tests are given without feedback. Moreover, tests increase performance of later tests more than additional study of the encoding items. Some cognitive psychologists have studied the direct testing effect in the laboratory. The purpose of this paper is to review the study of direct testing effect with recall tests.

Key words : free recall, testing effect, initial test

自由再生、テスト効果、初期テスト

1. テストの間接的影響と直接的影響

教育場面では、テストが査定のために使われることが多い。多くの大学では、学習の修得度を測るため、期末試験が行われる。だが、テストには、学習効果を高める側面もある。週に1回あるいは学期に1回など定期的にテストを行うと、大学生は試験直前の集中した学習ではなく定期的に分散した学習をすることが知られている (Bangert-Drowns, Kulik and Kulik, 1991)¹⁾。加えて、テスト施行によって成績のフィードバックが与えられることになり、これが後の学習の指針となる (Fitch, Drucker, & Norton, 1951)²⁾。最終的なテストに先立つ初期テスト (initial test) を行うことが学習に良い影響を与える現象は、テスト効果 (testing effect) と呼ばれている (Carrier and Pashler, 1992)³⁾。

初期テストの結果をフィードバックすることで、今後の学習に対する動機づけが高まり、弱点を重点的に

補強する効率的な学習のプランニング効果が生まれる。これらの効果は、テストの間接的影響 (indirect effects) と呼ばれる。対して、テストをすること自体による記憶促進の効果が確認されている。学習の後と最終テストの間で一度あるいは数度テストを行った場合には、テストをしないときよりも最終テストの記憶成績は高くなる。つまり、保持期間におけるテスト(以下、「初期テスト」: initial test と呼ぶ)の試行は、学習材料の長期記憶に恩恵を与えるのである。この現象は直接的テスト効果 (testing effect) と呼ばれ、以前から着目されてきた (Gates, 1917⁴⁾); レビューとして Roediger & Karpicke, 2006 a⁵⁾); 認知心理学の実験室実験においても、テスト自体が将来の記憶に与える促進効果が確認されている。本論文は、自由再生課題によって検討された直接的テスト効果 (以下テスト効果) 研究の概観を述べ、展望を論ずることを目的とする。

受付 平成 19 年 4 月 27 日, 受理 平成 19 年 6 月 8 日
近畿福祉大学 〒679-2217 兵庫県神崎郡福崎町高岡 1966-5

2. テスト効果研究の端緒

厳密な実験デザインによるテスト効果が検討され始めたのは近年であるが、テスト効果研究は1910年代に端を発する。当時の2つの研究が注目すべきものとして挙げられる。Roediger & Karpicke (2006 a)⁶⁾に倣い、これらの研究を紹介する。

テスト効果の検証：Gates (1917) Gates (1917)⁷⁾は、認知的枠組みの大規模な調査を行い初期テストの効果を示した初めての研究である。Gates(1917)⁸⁾は、無意味綴りと散文(伝記)を符号化項目として用いた。実験参加者は小学校1年から8年生であった。彼らはまず、符号化項目を黙読した。次に、材料から目を背け、心の中で思い出すよう教示を受けた。そのとき、ちらりと符号化項目を見ること(glance)を許された。最後に符号化項目の再生テストが行われた。Gates (1917)⁹⁾は、この思い出しながら覚える部分(暗唱部分)の時間を操作し、最終的に行われた再生テスト成績への影響を検討した。その結果、どの学年でも暗唱時間の効果が見られた。ただし、散文では、暗唱部分の時間が長すぎると効果が小さくなることも示された。Gates (1917)¹⁰⁾の実験は「ちらりと見る」ことを許すため、フィードバックがある初期テストを行っている。また、現在の認知心理学の枠組みから考えると、実験デザインが曖昧であり、処理過程がはっきりとしない。しかし、想起によって、後の課題に大きな恩恵を与えることが客観的数値に示されたという意味で、重視されるべき研究である。

テスト効果の直接的影響：Spitzer (1939) Spitzer (1939)¹¹⁾は、フィードバックを行わないテスト効果研究の先駆けである。彼の実験は再認課題を用いていたが、再生課題を用いた研究にも大きな影響を与えたため、以下に紹介する。

Spitzer (1939)¹²⁾は、91の小学校の6年生、3605名を対象に調査を行った。符号化項目として散文を用い、実験参加者に覚えさせた。続く2ヶ月余りの間に、記銘文章に対する多肢選択テストを複数回行った。実験参加者は、初期テストまでの保持期間、初期テスト間の間隔日数、初期テスト回数によって、8つのグループに分けられ、比較された。

Spitzer (1938)¹³⁾の結果は、以下の3点を明らかにした。まず、テストをすることで忘却が抑制されることを示した。場合によっては、初期テストを行うことで、最終テストの再認成績は初期テストを上回ることもあった。次に、テストをすることで忘却が止まることが示された。初期テストの時期にかかわらず、初

期テスト実施によって忘却曲線がほぼ横ばいになった。最後に、初期テスト実施時期が遅いほど、最終テストに与える恩恵が小さくなることが明らかになった。初期テストを行うことで、初期テスト後の忘却は抑制される。だが、初期テストを行う時期が遅ければ、すでに忘却されている項目についてはその恩恵が与えられない。たとえば、符号化の直後と一週間後に初期テストを行ったグループの最終テスト(初期テストから63日後)は、符号化後21日後に初期テストを行ったグループの最終テスト(初期テストから63日後)よりも格段に成績がよい。初期テストが早ければ早いほど、最終テストの正再認率は大きいことを示していた。

初期テスト回数の統制がとれていないことなど、現在の研究の視点から見るとSpitzer(1939)¹⁴⁾の実験デザインにも不備がある。しかし、フィードバックを行わない初期テストによる最終テストへの直接的影響を示した点において、Spitzer(1939)¹⁵⁾は、テスト効果研究に大きな貢献をしたと考えられる。

3. テスト効果研究の推移

初期テスト群と初期テスト無し群の比較 通常の再生または再認課題の保持期間に初期テスト(initial test)を挿入して、最後のテストの成績にどのような効果が起こるかを調べることで、テスト効果の研究は進められた。テスト効果研究が始まった当初は、初期テストを受けた実験群と、初期テストを受けない統制群との比較が行われた。初期テストを受けた実験群は、初期テストを受けない統制群に比べ、最終テストの成績が高くなることが報告された(Gates, 1917¹⁶⁾; Spitzer, 1937¹⁷⁾)。この実験デザインでは、初期テスト実験群の実験参加者は、初期テスト内で実験参加者が想起した再生項目に接することになる。そこで、初期テストの有無の比較によって論じられたテスト効果は、単に符号化項目の反復呈示効果の表れであるとも考えられた。

初期テスト群と再学習群の比較 対して、近年の研究では、テスト効果を検討するために、再学習条件と初期テスト条件を比較している。再学習群は、初期テスト群と同様にあらかじめ記銘項目を学習する。その後、初期テスト群が初期テストを行う段階で、代わりにもう一度学習を行うのである。つまり、統制群である再学習群の刺激への接触は、実験群と同様もしくはそれ以上になる。それにもかかわらず、実験群は統制群(再学習群)よりも、最終テスト成績が高くなるのである。記憶成績のフィードバックがない場合にも、この効果は頑強に表れている。主に長期遅延条件におい

て、初期テスト群は再学習群よりも記憶成績が高い。再学習群と初期テスト群の比較研究では、様々な変数をからめた実験が行われ（反復回数：Roediger & Karpicke, 2006 b¹⁸）; Tulving, 1967¹⁹、保持期間：Roediger & Karpicke, 2006 b²⁰ など）、多くの状況で頑強なテスト効果が報告されている。なお、学習条件（再学習 / 初期テスト）と、初期テストから最終テストまでの期間を要因としたとき、交互作用が認められ、テスト効果は再び注目され始めた（Roediger & Karpicke, 2006 b²¹）; Wheeler, Ewers & Buonanno, 2003²²）。この注目すべき交互作用については後述する。

4. テスト効果に影響を及ぼす要因

初期テストの回数 初期テストを反復することは、遅延時間が長期の記憶課題に正の影響を与えるようである。Roediger & Karpicke (2006 b²³) 実験 2) は、初期テストあるいは再学習の反復効果を検討した。彼らの実験では、学習段階の操作および最終テスト時期の操作を行った。学習段階では、4 回学習群、3 回学習 1 回初期テスト群、1 回学習 3 回初期テスト群を設けた。テスト時期は、学習段階後 5 分後および 1 週間後に行った。その結果、テスト時期が 5 分後では、4 回再学習群、3 回学習 1 回初期テスト群、1 回学習 3 回初期テスト群の順に最終テスト成績が高かった。対して、テスト時期が 1 週間後では、1 回学習 3 回初期テスト群、3 回学習 1 回初期テスト群、4 回学習群の順に成績が高かった。反復学習は短期間後のテストに大きな効果を持ち、反復テストは長期間後のテストに恩恵を与えることが示された。

呈示時間と検索猶予時間 Roediger & Karpicke (2006 b)²⁴ および Tulving (1967)²⁵ は、単語項目の符号化の後、学習を複数回行う反復再学習条件と初期テストを複数回行う反復初期テスト条件での最終テスト成績の比較を行った。Tulving (1967)²⁶ では反復回数を増した場合にも、両条件における最終テスト成績に差がみられなかった。対して、Roediger & Karpicke (2006 b)²⁷ では、反復回数を増した場合、反復テスト条件が反復再学習条件よりも成績が高かった。Roediger & Karpicke (2006 a)²⁸ は、この原因として検索猶予時間の差を挙げている。Tulving (1967)²⁹ では、符号化項目 36 単語の 1 秒ペースで呈示し、再生を 36 秒で実験参加者に行わせた。対して、Roediger & Karpicke (2006 b)³⁰ は 40 単語を 3 秒ペースで呈示し 2 分の猶予時間で再生させた。直接の比較ではないが、呈示時間と検索猶予時間がテスト効果に影響を

与える可能性が示唆されている。

符号化項目の質 テスト効果研究の材料としては、単語 (Hogan & Kintsch, 1971³¹; Izawa, 1967³²; McDaniel & Masson, 1985³³; Tulving, 1967³⁴; Wheeler, et al., 2003³⁵) や、画像リスト (Wheeler & Roediger, 1992)³⁶ などが用いられた。それぞれの研究で、テスト効果は認められている。教育の文脈を考慮した実験として、散文を用いた研究もあり、いくつかの実験ではテスト効果が認められている (Spitzer, 1939³⁷; Roediger & Karpicke, 2006 b³⁸)。

保持期間 初期テストから最終テストまでの保持期間の効果を調べた研究がある (Roediger & Karpicke, 2006 b³⁹; 直後と一週間、Gates, 1929⁴⁰): 直後、一週間、1 年弱の比較など)。忘却が起こるため、初期テストから最終テストまでの保持期間が大きいほど成績は下がるが、その低下率は再学習条件ほど大きくない。この交互作用こそが、今日注目される直接的テスト効果である。

初期の研究では符号化時からテストまでの保持期間の効果を検証したものもある (Gates, 1917⁴¹; Spitzer, 1939⁴²)。概して、符号化時から初期テストまでの期間が短いほど、初期テストの効果は大きい。

実験参加者の構え Szpunar, McDerrott, and Roediger (印刷中)⁴³ は、実験参加者が、初期テスト後に最終テストがあることを知っているか否を操作し、最終テスト成績への影響を検討した。彼女らは、あらかじめ最終テストがあることを知らせた群、最終テスト実施について知らせなかった群、最終テストの存在を伝えず、かつ、強制的に覚えた単語を忘れるよう指示する群、初期テストを行わなかった群の 4 群での、最終再生テスト成績の比較を行った。最終テストの存在を知らなかった実験参加者でも、初期テストをしなかった実験参加者よりも最終テスト成績は高く、テスト効果はみられた。だが、最終テストのあることを知っているグループは、最終テストの存在を知らないグループよりも成績が高かった。Szpunar et al. (印刷中)⁴⁴ では、最終テストの存在を初期テスト時に知っていることで、テスト効果が増大することが示された。

5. テスト効果の説明理論

現在、テスト効果の説明理論として、精緻化理論、転移適切処理理論がある。

精緻化理論によると、初期テストの実施により、記銘痕跡の精緻化がなされ、検索ルートが増大する。つまり、初期テストは遅延に耐えうる記憶痕跡を作ると考えられる (Gardiner, Craik and Bleasdale, 1973)

⁴⁵⁾ Bjork (1994) は、労力を必要とする検索 (Effortful Retrieval) をすることで、精緻化が行われるため、困難な検索課題を行うことが初期テスト効果を上げると考えた。この説明として、Bjork and Bjork (1992)⁴⁶⁾ は、保持強度 (storage strength) と検索強度 (retrieval strength) を区別した。保持強度は、記憶痕跡の耐久性や学習の耐久性を反映している。検索強度は、記憶痕跡への瞬間的アクセス度を反映しており、検索流暢性の概念と同様のものである。検索強度は、痕跡が心に浮かぶことによって、どれだけ記憶が容易に表れるかということである。彼らのモデルでは、検索強度は保持強度と負の相関を仮定している。つまり、容易な検索 (高検索強度) は保持強度を高めない。一方、労力の要る検索実践は、保持強度を高め、耐久性のある長期学習を促進する。容易に思い出されるものは容易に覚えていられると考えがちだが、容易に思い出されるものは検索強度が高いだけであり、長期にわたって記憶に残るためには、労力を必要とした検索をすることによって保持強度を高めることが必要だと考えている。彼らの理論は、初期テストの処理過程を説明する有用な手段となりうる。「検索時の労力」の操作的定義を明確にし、データの蓄積による実証を行うことで、保持強度と検索強度の理論は補強されるであろう。

転移適切処理論によってもテスト効果を説明できる。転移適切処理では、先行学習中の符号化操作とマッチしたテストのプロセスが、記憶テストの成績に恩恵を与えていると考える (Morris, Bransford and Franks, 1977)⁴⁷⁾。符号化特定性原理 (Tulving & Thomson, 1973)⁴⁸⁾ と同様に、転移適切処理論は符号化と検索のプロセスの類似性による恩恵を強調している。テスト効果においても、初期テストの遂行によってなされた処理と、最終テストで行なわれる処理が類似しているため、その成績が増大すると転移適切処理論から解釈できる。

6. 展 望

最後に、テスト効果研究における問題点および展望を述べる。

まず、テスト効果が起こるための必要条件についてより詳細な検討が望まれる。再学習条件と初期テスト条件を比較したとき、テスト効果が起こらない場合 (Tulving, 1967)⁴⁹⁾、起こる場合 (Roediger & Karpicke, 2006 b⁵⁰⁾ など) の相違を生み出す要因が明らかになっていない。Tulving (1967)⁵¹⁾ は、初期テスト条件で再学習条件と同等の再生率を報告し、注目を浴びた。近年は初期テストの実施により、ある程度の

保持期間がある場合、再生の優位性が示されている。この差違を引き起こす要因として、回答猶予時間、呈示時間、符号化項目数などが原因として考えられているが、これらを直接検証する必要がある。

2つ目に、直接的テスト効果研究が教育場面にどのように貢献できるのか、具体的な提言をすべきである。直接的テスト効果は、フィードバックなしでも効果がある。しかし、通常フィードバックをした方が、初期テストの効果は高い (McDermott, 2006 a)⁵²⁾。「フィードバックがない場合でも効果がある」といった訴えかけでは、教育現場からは「効果がある方がいいのだから、フィードバック付きの学習をすべきではないか」と反駁され、教育への貢献は期待できないだろう。消極的アピールではなく、「ノートを開いて学習することが苦手な生徒であっても、例えば下校中に授業内容を思い出そうとするだけでも、学習効果の増大が期待される」などの具体的学習法を提案できれば、教育場に貢献することも可能となるだろう。

3つ目に、初期テストが及ぼす負の効果を追究すべきである。McDermott (2006 b)⁵³⁾ では、虚再生を引き起こしやすいリストで、初期テスト反復による虚再生増大効果が確認されている。テスト効果の効用と同時に初期テストによる弊害も検討し、前者を高め、後者を減らす初期テストの形態を見出すことが、教育への貢献につながるであろう。

なお、本論文では、初期テストや最終テストとして、自由再生課題のみのレビューを行った。再認識課題、手がかり再生課題を初期テストや最終テストとして用いた場合については稿を改めて論ずる必要があると考える。

引用文献

- 1) Bangert-Drowns, R.L., Kulik, J.A., Kulik, C.L.C. Effects of frequent classroom testing. *Journal of Educational Research*, 85, 89-99. 1991.
- 2) Fitch, M.L., Drucker, A.J., & Norton, J.A. Frequent testing as a motivating factor in large lecture courses. *Journal of Educational Psychology*, 42, 1-20. 1951
- 3) Carrier, M., & Pashler, H. The influence of retrieval on retention. *Memory & Cognition*, 20, 633-642. 1992.
- 4) Gates, A.I. Recitation as a factor in memorizing. *Archives of Psychology*, 6, 40. 1917.
- 5) The power of testing memory: Basic research and implications for educational practice. *Perspec-*

- tives on Psychological Science*, 1, 181-210. 2006a
 6) 5) に同じ
 7) ~ 10) 4) に同じ
 11) Spitzer, H.F. Studies in retention. *Journal of Educational Psychology*, 30, 641-656. 1939.
 12) ~ 15) 11) に同じ
 16) 4) に同じ
 17) 10) に同じ
 18) Roediger, H.L., Karpicke, J.D. Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Perspectives on Psychological Science*, 17, 249-255, 2006b
 19) Tulving, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175-184. 1967.
 20) ~ 21) 18) に同じ
 22) Wheeler, M.A., Ewers, M., Buonanno, J.F. Different rates of forgetting following study versus test trials. *Memory*, 11, 571-580. 2003.
 23) 24) 18) に同じ
 25) ~ 26) 19) に同じ
 27) 18) に同じ
 28) 5) に同じ
 29) 19) に同じ
 30) 18) に同じ
 31) Hogan, R.M., & Kintsch, W. Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562-567. 1971.
 32) Izawa, C. Function of test trials in paired-associate learning. *Journal of Experimental Psychology*, 75, 194-209.
 33) McDaniel, M.A., Masson, M.E.J. Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 371-385. 1985.
 34) 19) に同じ
 35) 22) に同じ
 36) Wheeler, M.A., & Roediger, H.L., III. Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240-245. 1992.
 37) 11) に同じ
 38) ~ 39) 18) に同じ
 40) ~ 41) 4) に同じ
 42) 11) に同じ
 43) Szpunar, K.K., McDermost, K.B., & Roediger, H.L. (In press). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition*. in press.
 44) 43) に同じ
 45) Gardiner, J.M., Craik, F.I.M., Bleasdale, F.A. Retrieval difficulty and subsequent recall. *Memory & Cognition*, 1, 213-216. 1973.
 46) Bjork, R.A., Bjork, E.L. A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum 1992.
 47) Morris, C.D., Bransford, J.D., Franks, J.J. Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519-533. 1977
 48) Tulving, E., Thomson, D.M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373. 1973.
 49) 19) に同じ
 50) 18) に同じ
 51) 19) に同じ
 52) McDermost, K.B., Chan, J.C.K. Effects of repetition on memory for pragmatic inferences. *Memory & Cognition*, 34, 1273-1284 2006a.
 53) McDermost, K.B. Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, 34, 261-267. 2006b.