

初期テスト時の学習項目と妨害項目の類似性がテスト効果の生起に与える影響II：異なるテスト形式の初期テストと最終テストを用いた検討

CHO, Daisuke / 長, 大介

(出版者 / Publisher)

法政大学大学院

(雑誌名 / Journal or Publication Title)

大学院紀要 = Bulletin of graduate studies

(巻 / Volume)

81

(開始ページ / Start Page)

37

(終了ページ / End Page)

47

(発行年 / Year)

2018-10-31

(URL)

<https://doi.org/10.15002/00021339>

初期テスト時の学習項目と妨害項目の類似性がテスト効果の 生起に与える影響 II

—異なるテスト形式の初期テストと最終テストを用いた検討—

人文科学研究科 心理学専攻
博士後期課程 3年 長 大介

The influence of similarity between target and distractor at recognition test on the testing effect II

-Does testing effect occur when initial test and final test are different format? -

Daisuke CHO (*Major in Psychology, Graduate school of Humanities, Hosei University*)

学習項目をテストすることは学習項目を繰り返し学習することよりも学習項目の検索を促進する。この現象はテスト効果と呼ばれ、近年の記憶研究において盛んに研究されている。本研究ではテスト効果の説明に用いられるエピソード文脈説に基づいて、初期テストにおける学習項目と妨害項目の類似性が、テスト効果や学習項目の検索に及ぼす影響を検討した。実験参加者は初めに学習項目を学習し、続いて学習項目の再学習もしくは初期テストとして学習項目と妨害項目を対呈示する二肢強制選択テストに取り組んだ。最後に実験参加者は5分後、もしくは1週間後に最終テストとして自由再生テストに取り組んだ。その結果、いずれの保持期間後に行った最終テストではテスト効果および学習項目の検索に対する類似性の影響は認められなかった。この結果から初期テストと最終テストのテスト形式が異なる場合には、学習項目と妨害項目の類似性は学習項目の検索に影響しないといえる。しかしながら、本研究では初期テストにおいて実験参加者は学習時の処理文脈を適切に再現していなかったために生じた可能性がある。そのため、実験参加者が初期テストにおいて学習時の処理文脈を再現し、学習項目を検索することを促す要因を明らかにしていただくことが今後の課題といえる。

キーワード：テスト効果 (testing effect), エピソード文脈説 (episodic context account), 自由再生テスト (free recall test)

導入

これまでに行われてきた記憶研究が報告してきた結果は、いずれも記憶を良くすることに貢献してきた。たとえば、学習項目を覚えるときに、学習項目に共通する特徴に注目して、学習項目をまとめる体制化、学習項目に新たな情報を追加して覚えやすくする精緻化、学習項目のある特徴（意味的、形態的、音韻的特徴）に注目して覚えやすくする処理水準効果 (Craig & Lockhart, 1972; Craig & Tulving, 1975) などが挙げられる。多くの研究は、実験参加者がこれらの認知処理を行った学習項目は、これらの認知処理を行っていない学習項目よりもテスト時に多く検索できることを報告してきた。このように学習項目を良く覚えるには、学習項目に対してどのような認知処理を行ったかが重要である。しかしながら、時には学習項目をどのように覚えたかよりも学習項目を思い出すこと、すなわち、検索（テスト）することが学習項目の検索にとって重要であることはあまり知られていない。なぜなら、テストとは多くの場面において測定するためのツールとして利用されているか

らである。特に教育場面では、学習者が何をどれだけ身につけているかを測定し、成績を評価するために使われている（村山, 2006）。

このようなテストが学習項目の検索を促進する現象はテスト効果（testing effect）と呼ばれ、20世紀の初めには報告され（e.g., Abbot, 1909; Gates, 1917）、近年では盛んに研究が行われ、現在では検索練習（retrieval practice; Bjork, 1988）、テスト強化学習（test-enhanced learning; Roediger & Karpicke, 2006a）、検索による学習（retrieval-based learning; Karpicke, 2012）などの名称がつけられている¹。

テスト効果の手続き

テスト効果研究において用いられている実験の手続きは概ね3つのフェイズで構成されている。第1フェイズである学習フェイズでは、実験参加者は実験者からテスト教示（“後でテストを行うので、できるだけたくさん覚えてください”）を与えられ、学習項目の学習に取り組む。続く第2フェイズである再学習/初期テストフェイズでは、実験参加者は学習項目をもう一度学習する再学習（restudy, study-study）および/もしくは学習項目の初期テスト（initial test, intervening test）に取り組む。最後に第3フェイズである最終テストフェイズでは、実験参加者に学習フェイズで学習した学習項目についてのテスト（以後、初期テストと区別するために最終テストと表記）に取り組む。このような手続きを用いることによって、多くの研究が、初期テストにおいて検索した学習項目は再学習した学習項目よりも最終テストにおいて検索されやすくなることを報告してきた（Carpetner, 2009; Carpenter & DeLosh, 2006）。また、文章刺激を実験で用いた場合には、文章ごとに学習（読解）と初期テストもしくは再学習を行う手続きが用いられることもある（e.g., Roediger & Karpicke, 2006a）。

テスト効果の説明理論

現在までにテスト効果を説明する際には、初期テストと最終テストのテスト形式の一致によって説明する転移適切処理（Morris, Bransford, & Franks, 1977; Roediger & Karpicke, 2006a）、初期テストにおける検索経路の増加と記憶痕跡の強化による説明（Bjork, 1975; Bjork & Bjork, 1992; McDaniel & Masson, 1985）、学習項目と意味的に関連した媒介語の活性化によって説明する精緻検索説（elaborative retrieval account; Carpenter, 2009; Carpenter & DeLosh, 2005）、初期テストにおける検索の難しさと学習項目の検索に対する認知的努力の関連によって説明する検索努力説（retrieval effort, Pyc & Rawson, 2009）などを挙げるができる（Karpicke, 2017; Roediger & Karpicke, 2006b）。テスト効果の結果を説明する際、これらの理論は背反関係ではなく、複数の理論を用いて結果を説明することができる。特に転移適切処理による説明は初期テストと最終テストのテスト形式が一致している限り有効である。

これらの理論を使うことによってテスト効果を説明できるが、テスト効果は初期テストにおいてどのようなテスト形式を用いたかの影響を受ける（Rowland, 2014）。これらの理論について Karpicke, Lehman, & Aue (2014) は、それぞれの理論について問題点を指摘し、テスト効果の説明として初期テストにおける学習時の処理文脈の再現に注目したエピソード文脈説（episodic context account）を提案した。エピソード文脈説では4つの仮定によって、テスト効果を説明している。1つ目の仮定とは学習項目の文脈表象の緩やかな変化である。学習項目の文脈表象は時間経過にしたがって緩やかに変化する（Mensink & Raaijmakers, 1989）。しかし、取り組む課題や目的に決定的な変化がある場合には急激に変化する（Lehman & Malmberg, 2013）。すなわち、学習フェイズでの学習項目の学習後に行う再学習では、学習フェイズと同様に学習項目を学習するため、学習項目の文脈表象は時間経過にしたがって変化する。一方で、初期テストでは、実験参加者が取り組む課題は学習項目の符号化から検索に変わるため、再学習に比べて文脈表象がより大きく変化すると考えられる。2つ目の仮定とは処理文脈の再現である。すなわち、初期テストにおいて学習項目の検索手がかりとして学習時の処理文脈の再現および利用である。検索手がかりとして学習時の処理文脈を再現した場合、再現した処理文脈の持つ特徴と

¹ テスト効果研究は、テストの実施の有無を比較した研究（Chan & McDermott, 2007）、テストに対する統制条件として再学習を操作した研究（Roediger & Karpicke, 2006a）、学習とテストの組み合わせを比較した研究（Karpicke, 2012）に分けることができる。テスト効果の名称の違いは手続きの違いを反映しているものの、名称の使い分けに関しては明確な基準を見いだすことはできない。そのため、本稿ではより包括的な名称であるテスト効果を使用する。

一致する特徴を持つ単語は学習項目候補のまとまりである探索セット (search set) に追加される (Raaijmakers & Shiffrin, 1981). そのため, 学習項目を検索する際, 特定の特徴を持つ探索セットに検索を制限することが可能になる. 3 つ目の仮定とは文脈表象の更新である. 学習においてフェイズにおいて学習項目を学習した場合, 学習項目の文脈表象には学習時の特徴が追加される (A). 続く, 初期テストにおいて学習時の処理文脈 (A) を再現して, 学習項目を正しく検索した場合, 学習項目の文脈表象 (A) に初期テストにおいて正しく検索された処理文脈の特徴 (B) が追加され, 学習時と初期テスト時の処理文脈の特徴を含む文脈表象 (AB) に更新される (Lohnas & Kahana, 2014). 4 つ目の仮定とは再現された文脈表象による検索の制限である. 最終テストにおいて, 検索手がかりとして学習項目の処理文脈 (AB) を再現することによって, 探索セットに含まれる単語のうち, 学習時の処理文脈 (A) と初期テスト時の処理文脈 (B) の両方を含む単語のみに検索を制限することにより, 学習時の処理文脈 (A) のみを持つ単語よりも効率的に検索することが可能になる.

Karpicke et al. (2014) によって提案されたエピソード文脈説に基づくことによって再学習と初期テストの違いに加えて, 初期テストにおけるテスト形式の比較 (再認<再生, Glover, 1989; Hogan & Kintsch, 1971), 取り組む課題の比較 (再学習<生成=手がかり再生, Karpicke & Zaromb, 2010; 再学習<リスト弁別課題, Whiffen & Karpicke, 2017), 手がかり語と学習項目の連想強度の比較 (再学習<弱連想<高連想; Carpenter, 2009) などを説明することが可能になる (e.g., Karpicke, 2017; Karpicke et al., 2014). これまでのテスト効果の説明に用いられてきた理論は初期テストのテスト形式に依存しやすいのに対し, エピソード文脈説は学習時の処理文脈の再現というテスト形式に依存しない説明であることから, これまでの説明理論とは異なり, 包括的な説明理論であるといえよう.

学習時の処理文脈を再現する影響の検討

Karpicke et al. (2014) によって提案されたエピソード文脈説に注目して長 (印刷中) は, 初期テストとして従来の研究では検索の恩恵が生じにくいとされてきた再認テストを用いて, エピソード文脈説の説明可能性を検討した. エピソード文脈説に基づくなら, 再認テストは再生テストよりも学習項目の検索時に学習時の処理文脈を再現する必要性が低いため, 検索の恩恵が少ないといえる. そのため, 初期テストにおいて再認テストを用いた際に, 学習時の処理文脈を再現する必要性を操作することによって, テスト効果に対する処理文脈を再現する影響を検討した.

実験は従来のテスト効果研究で用いられている手続きに基づいて行われた. はじめに学習フェイズでは実験参加者は学習項目を学習し, 続く再学習/初期テストフェイズでは初期テストとして学習項目と妨害項目を用いた二肢強制選択テストを行い, 学習項目と妨害項目が類似している高類似初期テスト (リスク-リスト), 学習項目と妨害項目が類似していない低類似初期テスト (リスク-キカイ) および学習項目の再学習に取り組んだ. 最後に最終テストフェイズでは実験参加者は2つの異なる保持期間 (5 分間, 1 週間) ごとに最終テストに取り組んだ. その結果, 最終テストにおいてテスト効果が認められた. また, テスト効果は類似性によって異なり, 初期テストにおいて妨害項目と類似していた学習項目のみに認められた. この類似性によるテスト効果の違いが, 初期テストにおいて学習時の処理文脈の再現したことに由来するかを評価するため, 最終テストでは学習項目に加えて, 初期テストにおいて呈示された妨害項目もターゲットとして呈示した. なぜなら, 学習時の処理文脈を再現しながら再認テストに取り組み, テスト項目として呈示された妨害項目もまた学習項目と同様に符号化されるからである (Jacoby, Shimizu, Daniels, & Rhodes, 2005). すなわち, 学習項目と妨害項目が対呈示される二肢強制選択テストでは, 初期テストにおいて学習項目を正しく検索した場合には妨害項目もまた学習項目の処理文脈の一部として見なすことができる. したがって, 学習項目の検索によって妨害項目を含む文脈表象に更新されると予測された. 妨害項目をターゲットとするテストの結果は仮説を支持するものであり, 学習項目と類似した妨害項目は学習項目と類似していない妨害項目よりも検索された.

以上の結果から, 初期テストにおいて学習項目と妨害項目が類似している場合には, 実験参加者は学習時の処理文脈を再現して再認判断を行っていたと考えられる. したがって, エピソード文脈説が仮定する, 学習時の処理文脈の再現がテスト効果の生起に対して重要な役割を果たしている可能性が示唆された. しかしながら, 長 (印刷中) が報告した結果については留意すべき点がある.

先行研究の問題点と本研究の目的

長（印刷中）は最終テストにおいて類似性によるテスト効果の違いを報告した。しかし、類似性による記憶成績の違いおよび低類似初期テスト条件において、テスト効果を見いだすことができなかった。したがって、エピソード文脈説は部分的に支持されたにとどまる。この低類似初期テスト条件に認められた結果は最終テストの実施方法に由来するといえる。

長（印刷中）では、最終テストとして二肢強制選択テストを実施したが、学習項目と妨害項目の類似性は操作しなかった。そのため、エピソード文脈説の4つ目の仮定（最終テストにおいて、検索手がかりとして処理文脈を再現し、利用すること）を十分に満たせなかった可能性がある。すなわち、最終テストにおいて、低類似初期テスト条件では検索手がかりとして処理文脈を再現して再判断を行なった試行と処理文脈を再現せずに再判断を行った試行が混在していたため、テスト効果に対する処理文脈を再現することの影響が平均化されていたといえる。

この最終テストにおける処理文脈の再現もまた重要な要因である。たとえば、Halamish & Bjork (2011) は、最終テストにおいて難易度の異なるテストとして手がかり再生テスト（低難度）と自由再生テスト（高難度）を行ったところ、自由再生テストにおいてのみテスト効果が生じたことを報告した。このことから、学習項目と新項目の類似性を操作することなく最終テストを行ったため、処理文脈を再現するか否かを十分に統制できていなかったといえる。以上のことから最終テストにおいて処理文脈を再現することなく、学習項目の検索が可能であったため、最終テストにおいて低類似初期テスト条件にテスト効果が生じなかった、および記憶成績に類似性の違いが認められなかったと考えられる。

最終テストにおける処理文脈の再現の影響を統制するために、本研究では最終テストにおいても処理文脈を再現する必要性の高い自由再生テストを行うことによって、改めてテスト効果に対する初期テストにおける学習項目と妨害項目の類似性を操作し、初期テストにおいて学習時の処理文脈を再現する影響を検討し、テスト効果の生起において学習時の処理文脈が果たしている役割を明らかにする。

仮説

これまでのテスト効果研究に基づくなら、初期テストにおいてテストした学習項目は再学習した学習項目よりも最終テストにおいて検索されやすい。さらに初期テストにおける学習項目と妨害項目が類似している場合、学習項目と妨害項目が類似していない場合に比べて学習時の処理文脈を再現して再判断を行いやすいため、妨害項目と類似している学習項目には類似していない学習項目よりも強いテスト効果が生じることが予測される。さらにテスト効果に対する再学習と初期テストの影響は保持期間によって異なり、最終テストまでの保持期間が短い場合には再学習した学習項目は初期テストにおいて検索した学習項目よりも良く検索される。一方で、最終テストまでの保持期間が長い場合には初期テストにおいて検索した学習項目は再学習した学習項目よりも検索される (Roediegr & Karpicke, 2006a; Toppino & Cohen, 2009; Wheeler, Ewers, & Buonanno, 2003)。エピソード文脈説に基づくなら、初期テストにおける学習項目と妨害項目の類似性は学習時の処理文脈の再現に影響し、最終テストでは類似性によって異なる強さのテスト効果が生じる。学習項目と妨害項目が類似している場合には学習項目と妨害項目が類似していない場合に比べて、最終テストである自由再生テストにおいて学習項目候補となる単語が多くなる。そのため、記銘項目の検索が困難になるため、最終テストにおいても学習時や初期テスト時の処理文脈を再現して検索手がかりとして積極的に利用することが予測される。

まとめると、保持期間が 5 min の直後テストでは再学習した学習項目が初期テストした学習項目よりも検索される（再学習 > 低類似初期テスト = 高類似初期テスト）。保持期間が 1week の遅延テストではテスト効果が生じることに加えて、妨害項目と類似している学習項目が最もよく検索される（再学習 < 低類似初期テスト < 高類似初期テスト）。

方法

実験参加者 都内の私立大学に通う大学生および大学院生のうち第一言語が日本語である 24 名が参加した。

いずれも矯正視力を含めて正常な視力を有していた。実験参加者には実験への参加は自由意志に基づき、実験に参加しなくても実験参加者自身に不利益が生じないこと、一度実験への参加を同意しても実験の途中で参加を辞退でき、辞退しても不利益が生じないことを口頭および書面で説明した。

実験計画 保持期間（直後テスト、遅延テスト：間）と学習方法（低類似初期テスト、高類似初期テスト、再学習：内）を用いた2要因混合計画だった。従属変数として初期テストにおける学習項目に対する旧判断であるヒット率および、最終テストにおける学習項目の自由再生率を用いた。

実験刺激 実験刺激として天野・近藤（2003）から単語親密度が中程度（7段階評定の5.0以上）のカタカナ3文字の単語、280語を選定し、単語プールを作成した。単語プールから15語を抜粋し、学習項目として呈示する学習リスト（e.g., リスク, メロン, グラフ）を作成した。続いて学習リストに含まれる特定の単語と同じ文字を2文字含む単語、15語を抜粋し、初期テストにおいて妨害項目として呈示する単語で構成される高類似リスト（e.g., リスト, サロン, グラム）を作成した。さらに学習項目リストに含まれる特定の単語と同じ文字を含まない単語、15語を抜粋し、初期テストにおいて妨害項目として呈示する単語で構成された低類似リスト（e.g., キカイ, ヘルツ, コンマ）を作成し、それぞれの刺激セットに含まれる単語を組み合わせた刺激セットを作成した。この手順を用いて6つの刺激セットを作成した。作成された刺激セットを3つ組み合わせた2つの刺激リストを作成した。実験では刺激リストに含まれる3つの学習リストに学習方法要因の条件を割り振った。また、低類似初期テスト条件と高類似初期テスト条件に割り振られた場合には、低類似リストと高類似リストのサブリストと組み合わせて使用した。すなわち、低類似テスト条件では学習リストと低類似リストのサブリストを組み合わせて使用し（リスク-キカイ）、高類似初期テスト条件では学習リストと高類似リストのサブリストを組み合わせて使用した（リスク-リスト）。残りの10語のうち6語は初頭効果および新近効果用のバッファとして学習フェイズの初頭部と新近部にそれぞれ3語を割り当てた。残りの4語は練習試行における妨害項目として使用した。

手続き

実験は学習フェイズ、再学習/初期テストフェイズ、最終テストフェイズの順番で行った。また、学習フェイズと再学習/初期テストフェイズ終了後に挿入課題を行った。

学習フェイズ 学習フェイズにおいて実験参加者は45項目の学習を行った。実験参加者には呈示される項目について後でテストを行うことを説明し、できるだけたくさんの学習項目を覚えることを求めた。学習項目はディスプレイの中央に1つずつ、5000msで呈示した。学習項目を呈示してから5000msが経過したら、間を置かずに次の学習項目が呈示されることを説明した。実験参加者に学習フェイズでの取り組み方について不明な点がないことを確認し、実験参加者の任意のタイミングで学習項目の学習を開始した。

挿入課題 I 学習フェイズが終了した直後に挿入課題を行なった。実験参加者には都道府県名の自由再生テストを用いた一般的な知識に関する調査を行なっている、というカバーストーリーを説明し、挿入課題への取り組みを求めた。実験参加者には5分間で都道府県名と県庁所在地名をできるだけ思い出し、書き出すことを求めた。

再学習/初期テストフェイズ 挿入課題が終了した直後に再学習/初期テストフェイズを行った。再学習/初期テストフェイズでは学習項目をもう一度学習する再学習試行、初期テストとして学習項目と類似した妨害項目を用いて二肢強制選択テストに取り組む高類似初期テスト試行、学習項目と類似していない妨害項目を用いて二肢強制選択テストに取り組む低類似初期テスト試行をそれぞれ15試行ずつ行った（合計45試行）。

はじめに実験参加者には課題として学習フェイズで呈示された学習項目を再学習、もしくはテストすることを説明した。さらに取り組む課題は試行ごとに異なり、どちらの課題に取り組むかはディスプレイ上に呈示される指示文にしたがって取り組むことを併せて説明した。続いて再学習と初期テストの取り組み方について説明した。再学習試行でははじめに学習項目の再学習を求める指示文（「もう一度覚えてください」）がディスプレイ中央に2000msで呈示された。続いて学習項目がディスプレイ中央に1つ、3000msで呈示された。実験参加者には学習項目がディスプレイに映し出されている間にもう一度覚えることを求めた。学習項目の呈示から3000msが経過したら、ディスプレイ上から学習項目が消え、ブランク画面に切り替わった。ブランク画面は

2000ms で呈示され、実験参加者にはブランク画面中にも直前に映し出された学習項目を覚え続けることを求めた。ブランク画面終了後、次の試行で取り組む課題の指示文が呈示された。初期テスト試行では再学習試行とは異なる指示文（「どちらの単語が今日の実験のはじめに映し出されましたか？」）がディスプレイの中央に2000ms で呈示された。続いてテスト項目として学習項目と妨害項目がディスプレイの中央に左右に並んで3000ms で呈示された。実験参加者にはテスト項目が呈示されている間にそれぞれのテスト項目を注意深く見てから、学習フェイズで呈示されたことを思い出せるテスト項目を決めるように求めた。テスト項目の呈示から3000ms 後、ディスプレイ上からテスト項目が消え、ブランク画面に切り替わった。ブランク画面は2000ms で呈示され、実験参加者にはブランク画面中に学習フェイズで呈示されたことを思い出せる項目を口頭で回答することを求めた。ブランク画面の呈示から2000ms 後に次の試行の指示文が呈示された。

再学習/初期テストフェイズの教示後に課題の取り組み方に慣れることを目的として練習試行を行った。練習試行では初頭、新近バッファを使用し、再学習試行、低類似初期テスト試行、高類似初期テスト試行をそれぞれ2試行ずつ行った。練習試行での課題に取り組む順番および練習試行で使用するテスト項目は全ての実験参加者で同じだった。練習試行終了後、実験参加者の取り組み方が教示と異なる場合には、実験者が実験参加者の課題の取り組み方について教示とは異なる点を指摘し、再度取り組み方を説明した。本試行は実験参加者が課題の取り組み方を理解していることを確認できてから開始した。実験者は、低類似初期テスト試行と高類似初期テスト試行中に実験参加者が報告する回答を記録用紙に記入した。

本試行では同じ条件の試行が3回以上連続しないこと、低類似初期テスト試行と高類似初期テスト試行では学習項目が同じ位置に3回以上連続しないことを条件として呈示順番および呈示位置を決定した。

挿入課題 II 再学習/初期テストフェイズ終了後に挿入課題 I と同じ方法で挿入課題 II を実施した。実験参加者には挿入課題 II では挿入課題 I で書き出した都道府県や県庁所在地の名前を含めて書き出すことを求めた。

最終テストフェイズ 本研究では保持期間ごとに最終テストとして自由再生テストを実施した。

直後テスト条件では挿入課題 II の直後に最終テストを実施した。はじめに実験参加者には自由再生テストの取り組み方（「これから行うテストでは、今日の実験の初めに映し出された単語を思い出し、どんな順番でも構わないので、できるだけたくさん書き出してください」とテスト実施時間（「テストは4分間で行います。テストを始めてから2分が過ぎたら、残り時間が2分であることをアナウンスします。残り時間に注意してテストに取り組んでください」）の教示を与えた。実験参加者に最終テストの取り組み方について不明な点がないことを確認した後に実験者の合図に合わせてテストを開始した。テスト終了後、これまでの実験の取り組み方についてアンケート形式の内観報告への回答を求め、実験のデブリーフィングを行なった。一方、遅延テスト条件では挿入課題 II の直後に学習フェイズと再学習/初期テストフェイズの取り組み方についてアンケート形式による内観報告を求めた。内観報告後、実験参加者には保持期間中に学習項目の想起を防ぐために、最終テストのカバーストーリーとして1週間後に実施する質問紙調査への参加を求めた。挿入課題 II から1週間後に直後テスト条件と同じく最終テストとして自由再生テストを行った。遅延テスト条件の実験参加者には、直後テスト条件の実験参加者に与えた自由再生テストの取り組み方について教示の一部を変更した教示（「これから行うテストでは、1週間前の実験の初めに映し出された単語を思い出し、どんな順番でも構わないので、できるだけたくさん書き出してください」）を与え、直後テスト条件と同じ方法で最終テストを行った。遅延テスト終了後、最終テストの取り組み方、記銘項目の想起の有無や想起頻度についてアンケート形式による内観報告を求めた。

結 果

初期テスト

はじめに初期テストとして行なった二肢強制選択テストにおける学習項目のヒット率を、再学習を除く条件ごとに算出し、保持期間要因および学習方法要因ごとに Figure 1 にまとめた。初期テストにおいて実験参加者はほぼ全ての学習項目を正確に再認した ($M=.90$)。また、妨害項目に対する旧判断である誤警報率も低かった ($M=.10$)。

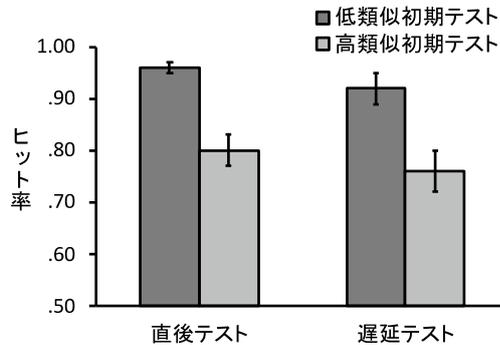


Figure 1. 初期テストにおける学習項目の平均ヒット率。
エラーバーは標準誤差を表す。

初期テストにおける学習項目のヒット率に対して保持期間（直後テスト，遅延テスト）と再学習条件を除いた学習方法（低類似初期テスト，高類似初期テスト）の2要因分散分析を行った。その結果，保持期間要因の主効果 ($F(1,22)=0.66, ns, \eta_p^2=.02$) および交互作用 ($F(1,22)=0.21, ns, \eta_p^2=.009$) はいずれも有意ではなかった。一方，学習方法要因の主効果は有意になり，低類似テスト条件のヒット率 ($M=.94$) は高類似初期テスト条件のヒット率 ($M=.86$) よりも有意に高かった ($F(1,22)=12.59, p<.01, \eta_p^2=.36$)。

最終テスト

テスト効果に対する初期テストにおいて学習時の処理文脈の再現する影響を検討するため，初頭効果および新近効果用のバッファを除いた学習項目の再生率を分析した。最終テストにおける学習項目の自由再生率を算出し，学習方法および保持期間ごとに Figure 2 にまとめた。

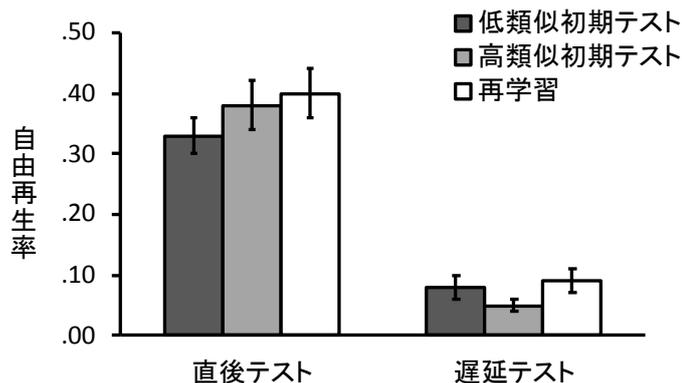


Figure 2. 最終テストにおける学習項目の平均自由再生率。
エラーバーは標準誤差を表す。

最終テストにおける学習項目の自由再生率に対して保持期間（直後テスト，遅延テスト）と学習方法（低類似初期テスト，高類似初期テスト，再学習）の2要因分散分析を行った。その結果，保持期間要因の主効果が有意になり，直後テスト条件の自由再生率 ($M=.37$) は遅延テスト条件の自由再生率 ($M=.09$) よりも有意に高かった ($F(1,22)=87.39, p<.01, \eta_p^2=.79$)。一方，学習方法要因の主効果 ($F(2,44)=0.68, ns, \eta_p^2=.03$) および交互作用 ($F(2,44)=0.93, ns, \eta_p^2=.04$) はいずれも有意にならなかった。

考 察

本研究では、長（印刷中）と同様に初期テストにおける学習項目と妨害項目の類似性を操作することによって、テスト効果に対する学習時の処理文脈を再現することの影響を検討した。また、初期テストのテスト形式の転移の観点から、最終テストにおいて学習項目をターゲットとする自由再生テストを行った。その結果、保持期間における自由再生率の違いは認められた（直後テスト>遅延テスト）が、学習方法の違いは認められなかった（再学習≒低類似初期テスト≒高類似初期テスト）。したがって、初期テストにおける学習項目と妨害項目の類似性がテスト効果および学習項目の検索を促進する仮説は支持されず、初期テストと最終テストのテスト形式が異なる場合には、再認テストにおける学習項目と妨害項目の類似性はテスト効果および学習項目の検索を促進するほどは影響しないといえる。

これまでの研究に基づくなら、初期テストとして再認テストを用いた場合には自由再生テストほど最終テストにおいてテスト効果および学習項目の検索に影響しないことが報告されてきた（e.g., Carpenter & DeLosh, 2005; Glover, 1989）。本研究の結果は先行研究を追認するものであり、初期テストにおいて再生テストと同じく処理文脈の再現を伴う再認テストを用いたとしても、学習項目にテスト効果や検索による恩恵をもたらすわけではないことを示唆している。したがって、初期テストとして再認テストを用いた場合、初期テストにおける検索の恩恵は限られたものであるといえよう。

本研究ではテスト効果および学習項目の検索に対する類似性の影響が認められなかった。しかしながら、以下2点から、本研究ではテスト効果や学習項目の検索に対する学習項目と妨害項目の類似性の影響は十分な検討が行われているとはいえない。1点目は初期テストにおいて学習時の処理文脈を再現することなく学習項目を検索していた可能性である。学習項目の再認判断には異なる2つの検索過程があり、記憶の意図的な利用として学習エピソードの詳細を想起する回想過程と、記憶の自動的な利用として知覚的流暢性を基礎とする熟知性である（e.g., 藤田, 1999; Jacoby, 1991）。このうち、学習時の処理文脈を再現することは回想過程に相当する。本研究では初期テストにおける学習項目と妨害項目の類似性によって、再認判断の手がかりとして学習時の処理文脈を再現することを想定した。初期テストにおける学習項目と妨害項目との類似性は学習項目の検索を阻害するが（e.g., Roediger & McDermott, 1995）、類似性によって検索の難易度が高い場合には学習時の処理文脈を再現して判断することによって、学習項目を正しく検索することが可能になると予測された。しかしながら、初期テストにおいて学習項目の記憶成績に類似性の影響が認められ、妨害項目と類似している学習項目は類似していない学習項目ほど正しく検索されなかった（ $86 < .94$ ）。初期テストの結果から本研究の高類似初期テストでは学習時の処理文脈を十分に再現することなく再認判断を行っていたといえよう。その結果、学習項目の文脈表象も更新されず、最終テストにおいて学習項目の検索手がかりとして機能しなかったため、テスト効果の生起に影響を及ぼさなかったと考えられる。

2点目は最終テストにおいて床効果が生じた可能性である。本研究も遠藤（2007）と同様に最終テストまでの保持期間（5分、1週間）を操作し、保持期間ごとに最終テストを行ったところ、遅延テストにおいてテスト効果が生じなかった。この結果について遠藤（2007）は遅延テストにおいて床効果が生じた可能性に言及した。本研究の最終テストのうち、直後テストでは一定数の学習項目が検索されたが（ $M=.37$ ）、遅延テストではわずかな学習項目だけが検索された（ $M=.09$ ）。したがって、本研究の遅延テストにおいても遠藤（2007）と同様に床効果が生じていたと考えられる。その結果、自由再生率にテスト効果が認められなかったといえる。また、遅延テストにおける床効果は1点目に挙げた処理文脈が適切に再現されていなかったことによって生じている可能性がある。

このように本研究では、初期テストにおける学習項目と妨害項目の類似性にも関わらず、学習時の処理文脈を適切に再現することなく再認判断を行ったことで、テスト効果および学習項目に対する影響は確認できなかった。そのため、今後の課題として実験参加者の学習時の処理文脈を再現することを促す方法を用いた検討が必要である。初期テストにおいて、実験参加者が学習時の処理文脈を再現する必要性のある課題として2つの課題を挙げることができる。学習項目の検索に学習時の処理文脈を再現することが伴う課題としてソースモニタリング課題（Johnson, Hashtroudi, & Lindsay, 1993）が挙げられる。ソースモニタリング課題では、実験参加者

は初めに複数の条件下 (e.g., 視覚呈示, 聴覚呈示) で学習項目を学習する. 続くテストでは, 実験参加者は呈示されるテスト項目がどの条件下で学習されたか (視覚呈示, 聴覚呈示, 呈示されていない) を判断する. ソースモニタリング課題に加えて情報源制限検索 (source-constrained retrieval: Jacoby et al., 2005; Karpicke et al., 2014) もまた再認テスト時に学習時の処理文脈の再現が伴う. 情報源制限検索では, 実験参加者は初めに複数の方向付け課題に取り組む. 続くテストでは, 実験参加者はテスト項目が特定の方向づけ課題で呈示された項目か否かの判断を求められる. 一般的な再認テストでは, 学習時のエピソード情報は必ずしも利用されるわけではない. 対して, ソースモニタリング課題や情報源制限検索では, テスト項目に対する判断に学習時のエピソード情報の利用が伴う. したがって, 初期テストにおいて, 学習時の処理文脈の再現を伴う再認テストを用いることによって, 実験参加者は二肢強制選択テストよりも初期テストにおいて単独呈示されたテスト項目に対して学習時の処理文脈を再現して再認判断に取り組むことが可能になるといえる (e.g., Karpicke et al., 2014).

初期テストにおいて実験参加者に学習時の処理文脈の再現を伴う課題への取り組みを求めることに加えて, 学習フェイズの実施方法を操作することによっても, 初期テストにおいて学習時の処理文脈を再現しやすくすることは可能である. 本研究では, 学習フェイズにおいて学習項目を1回しか学習していないため, 初期テストにおいて学習時の処理文脈の再現にわずかなエピソード情報しか利用できなかったと考えられる. たとえば, Roediger & Karpicke (2006a) は, 最終テストまでの保持期間が短い場合にはテスト回数よりも学習回数が学習項目の検索を促進することを報告した (e.g., Coppen, Verkoijen, & Rikers, 2011). この直後テストに認められる学習回数の優位性は学習項目を繰り返し学習することによって, 学習項目の熟知性 (知覚的流暢性) に加えて, 回想 (学習エピソード) が豊富になったことに由来するといえる.

この学習項目の呈示回数と熟知性, 回想は密接に関連している. たとえば, Jacoby, Jones, & Dolan (1998) は, 実験参加者に主観的な記憶経験の違いについて報告を求める Remember/Know 手続き (Gardiner, 1988; Tulving, 1985) を用いて, 回想と熟知性に対する学習項目の呈示回数 (1回, 2回, 3回) の影響を検討した. この Remember / Know 手続きでは実験参加者はテスト項目に対する再認判断に加えて, 再認した項目が学習時に呈示された際の詳細な情報を想起できるか (Remember 判断), 再認した項目が呈示されたことはわかるが, 詳細な情報は想起できないか (Know 判断) の判断を求める (詳しくは藤田, 1999). 学習項目の呈示回数を操作した結果, Jacoby et al. (1998) は学習項目の呈示回数の増加にしたがって学習項目の記憶成績と同時に, Remember 判断率も増加することを報告した.

さらに続く実験では, 除外テスト (exclusion test) を実施した. 除外テストとは, 学習項目の検索過程に含まれる回想と熟知性のうち, 再認判断に対する熟知性の影響を測定することが可能なテストである. Jacoby らは学習フェイズにおいて実験参加者に学習項目を視覚呈示 (read) と聴覚呈示 (heard) した. 続く除外テストでは, 実験参加者に呈示されるテスト項目が聴覚呈示された単語として再認できる場合には「yes」と判断し, 聴覚呈示された単語として再認できない場合には「no」と答えること, すなわち, 視覚呈示された単語に対しても「no」と答えることを求めた. 除外テストにおいて, 実験参加者が視覚呈示された単語に対して「no」と判断した場合には「視覚呈示された」というエピソードの回想に基づいて「聴覚呈示されていない単語」と判断し, 「yes」と判断した場合には「視覚呈示された」というエピソードは回想できないが, 「単語が呈示された」という熟知性に基づいて判断していると考えられた. Jacoby らが行なった除外テストでは, 呈示回数の増加にしたがって, 視覚呈示された単語に対して「yes」と判断する割合 (≒FA 率) が低下することを報告した². 学習項目に対する Remember 判断率の増加と視覚呈示された単語に対する「yes」判断率 (≒FA 率) の低下からも, 学習項目を呈示する回数を増加させることによって, 実験参加者は学習項目を検索する際に回想を利用しやすくなるといえよう (e.g., Yonelinas, 2002).

これまでの研究は初期テストとして再認テストは再生テストほどテスト効果や学習項目の検索を促すことがないことを報告してきた (e.g., Carpenter & DeLosh, 2006; Glover, 1989; Hogan & Kintsch, 1971). 一方で, エピソ

² Jacoby et al. (1998) では, テスト項目が呈示されてから 700ms 以内に判断を行う Deadline 条件とテスト項目が呈示されてから 1200ms 後に判断を行う Wait 条件を操作した. その結果, 学習項目の呈示回数と再認判断の制限時間の交互作用が有意になり, Deadline 条件では呈示回数の違いが認められなかったが, Wait 条件では呈示回数の増加にしたがって, 視覚呈示された項目に対する「yes」判断率が低下した.

ード文脈説に基づくことで、再認テストでも再生テストと同様に、初期テストにおいて学習時の処理文脈を再現して学習項目を検索することによって、再認テストでもテスト効果や学習項目の検索が促進されることが報告されている (e.g., 長, 印刷中; Karpicke et al., 2014). このことから、初期テストのテスト形式に関わらず、初期テスト時に学習時の処理文脈を再現する必要のある状況を作ることが今後の課題として挙げることができる。このような観点から検討することによって、再認テストの特性に合わせた方略の利用に対して有益な知見をもたらしていくことが可能であるといえる。

引用文献

- Abbot, E.E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs*, *11*, 159-177.
- 天野 成昭・近藤 公久(2003). NTT データベースシリーズ日本語の語彙特性第 1 期 CD-ROM 版 三省堂
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomenon. In R. L. Solso(Eds), *Information processing and cognition: The Loyol Symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.
- Bjork, R.A.(1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P.E. Morris, & R. N. Sykers (Eds.) *Practical aspects of memory: Current research and issues* (pp. 396-401). New York, Wiley.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From learning processes to cognitive processes: Essays in honor of William K. Estes*, *2*, 35-67.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1563-1569.
- Carpenter, S. K., DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619-636.
- Carpenter, S.K., & DeLosh, E.L. (2006) . Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268-276.
- Chan, J. C. K., & McDermott, K. B.(2007). The testing effect in Recognition Memory: A Dual Process Account. *Journal of Experimental Psychology*. *33*, 431-437
- 長 大介 (印刷中). 再認テスト時の学習項目と妨害項目の類似性がテスト効果の生起に与える影響. 認知心理学研究.
- Coppens, L. C., Verkoefen, P. P., & Rikers, R. M. (2011). *Learning Adinkra symbols: The effect of testing*. *Journal of Cognitive Psychology*, *23*, 351-357.
- Craik, F. I. M., & Tulving. (1975). Depth of Processing and the Retention of Words in Episodic Memory. *Journal of Experimental Psychology*, *104*, 268-294.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671-684.
- 遠藤 正雄(2007). テスト効果に及ぼす遅延期間の影響. 近畿福祉大学紀要, *8*, 179-182.
- 藤田 哲也(1999). 潜在記憶の測定法. 心理学評論, *42*, 107-125.
- Gardiner, J. M. (1988). Functional aspect of recollective experience. *Memory & Cognition*, *16*, 309-313.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*, 1-104.
- Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392-399.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 801-812.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562-567.
- Jacoby, L. L. (1991). A Process Dissociation Framework: Separating Automatic from Intentional Uses of Memory. *Journal of Memory & Language*, *30*, 513-541.
- Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgement and exclusion errors. *Psychonomic bulletin & review*, *5*, 705-709.
- Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychonomic Bulletin & Review*, *12*, 852-857.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin*, *114*, 3-28.
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, *21*, 157-163.
- Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J.H. Byrne(Ed.), *Learning and memory: A comprehensive reference* (pp. 487-514). Oxford, UK: Elsevier
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of learning and motivation*, *61*, 237-284.
- Karpicke, J.D., & Zaromb, F.M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*, 227-239.
- Lehman, M., & MalMBERG, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *12*, 155-189.
- Lohnas, L. J., & Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology=Learning, Memory, and Cognition*, *40*, 755-764.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371-385.
- Mensink, G. J. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*, 172-186.
- 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*, 519-533.
- 村山 航(2006). テストへの適応. 教育心理学研究, *54*, 265-279.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory ?. *Journal of Memory and Language*, *60*, 437-447.

- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological review*, 88, 93-134.
- Roediger, H.L., & Karpicke, J. D.(2006a). Enhanced Learning Taking Memory Test 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*, 1, 181-210.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, 21, 803-814.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: a metaanalytic review of the testing effect. *Psychological Bulletin*, 140, 1432-1463
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental psychology*, 56, 252-257.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, 26, 1-12.
- Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240-245.
- Whiffen, J. W., & Karpicke, J. D.(2017). The Role of Episodic context in Retrieval Practice Effect. *Journal of experimental psychology, Learning, memory, and cognition*, 43, 1036-1046.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46, 441-517.