

テスト形式およびテスト形式の組み合わせが 記憶のテスト効果に及ぼす影響：再生テスト と再認テストを組み合わせた検討

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 / 大学院紀要 = Bulletin of
graduate studies

(巻 / Volume)

76

(開始ページ / Start Page)

23

(終了ページ / End Page)

34

(発行年 / Year)

2016-03-31

(URL)

<https://doi.org/10.15002/00012801>

テスト形式およびテスト形式の組み合わせが 記憶のテスト効果に及ぼす影響

—再生テストと再認テストを組み合わせた検討—

人文科学研究科 心理学専攻

博士後期課程3年 長 大 介

テストが果たす役割

テストと言うと学習者が学習内容をどの程度理解、定着しているかを測定するツールとして用いられることを連想するだろう。しかし、学習内容の定着の程度を測定する以外にも様々な役割がある(村山, 2006)。その中でも直観に反する役割としてテストによる学習内容の定着の促進が挙げられる。これは学習した内容を想起(テスト)することで、覚えるべき情報の保持が促進されるというものである。このテストによる保持の促進現象はテスト効果(testing effect)と呼ばれ、20世紀の初めには既に検討されている現象である(e.g., Bartlett, 1977; Gates, 1917; Spitzer, 1939)。一般的にテスト効果を扱った研究では記銘対象(本研究では単語を記銘対象とするので、本研究の記銘対象を学習項目と記し、区別する)を繰り返し学習すること(repeated study, restudying)よりも学習に加えてテストを行う、もしくはテストを繰り返すこと(repeated testing)が学習項目の保持を促進することを報告している。

テスト効果を検討する手続き

テスト効果を検討する研究で行われる手続きは3つのフェイズ(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)で構成されている。学習フェイズでは実験参加者に記銘対象に対する学習を求める。続く再学習/初期テストフェイズでは実験参加者に記銘対象についてもう一度学習を求める再学習、記銘対象についてテスト(以後、再学習/テストフェイズで実施するテストを初期テストと記す)に取り組むことを求める。最後に最終テストフェイズでは記銘対象についての(最終)テストに取り組むことを求める。最終テストフェイズを保持期間ごとに実施した研究もあるが(e.g., Carpenter, Pashler, Wixted, & Vul, 2008; Chan, 2010)、テスト効果を検討する研究では概ね3つのフェイズからなる手続きが用いられている。最終テストフェイズで実施するテストの記憶成績を比較した場合、記銘対象の学習を繰り返す再学習条件よりも記銘対象の学習とテストを行う、もしくはテストを繰り返すテスト条件の記憶成績が高くなることがわかっている。さらに最終テストフェイズで実施するテストの保持期間を操作することで、保持期間が短い場合には再学習条件ではテスト条件に比べて高い記憶成績になった。しかし、保持期間が長い場合にはテスト条件は再学習条件より高い記憶成績になる逆転現象がみられ、再学習とテストの保持に対する影響は保持期間によって異なることが明らかになった(e.g., Roediger & Karpicke, 2006b; Wheeler, Ewers, & Buonanno, 2003)。

現在までに多くの研究が行われ、テスト効果は単語や文章に固有の現象ではなく、記銘対象が第二言語(e.g., Kang, 2010)、視聴覚刺激(e.g., Butler & Roediger, 2007)、視空間刺激(e.g., Carpenter & Pashler, 2007)などを使用した場合でも見られることが示されてきた。これらの研究は実験室において様々な剰余変数を統制してテスト効果を検討してきた。このような剰余変数を統制する実験に対して、近年では教育場面での応用を視野に入れて実験室以外で一定の剰余変数がある状況でのテスト効果を検討した実験も行われ、剰余変数を統制した実験室実験と同じようにテスト効果が生じることが報告されている(e.g., Butler & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007)。以上の研究を総括すると記銘対象やテスト状況の違いにも関わらず、テスト効果が報告されていることからテスト効果の源泉が検索することそれ自体であることは明白である。

テスト効果を説明する理論

このテストによる保持の促進効果を説明するために様々な理論が用いられてきた(e.g., Roediger &

Karpicke, 2006a; 多鹿, 2008)。現在は3つの理論のいずれかを用いることで説明することが可能であるといえる。1つ目の理論は転移適切性処理 (Morris, Bransford, & Franks, 1977) である。転移適切性処理では学習時とテスト時の認知処理の類似の程度に注目し、2つの認知処理が類似している場合に記銘対象の検索が促進される。テスト条件では初期テストを最終テストと同じ形式で実施する。したがって、最終テストにおいて初期テストと同じ認知処理を実行することで、テスト条件は再学習条件よりも記銘対象を検索しやすくなると考えられている (Roediger & Karpicke, 2006b)。2つ目の理論は検索努力 (retrieval effort: Gardiner, Craik, & Bleasdale, 1973; Jacoby, 1978) である。検索努力とは認知資源の1つであり、記銘対象の想起に費やされる認知資源のことを指す。つまり、テスト条件では初期テストにおいて検索に対して検索努力が費やされることで、記憶痕跡が強化される。その結果、再学習条件に比べて記銘対象の保持が促進されると考えられている (e.g., Carpenter & DeLosh, 2005)。3つ目の理論が検索経路 (e.g., Bjork, 1975; McDaniel & Masson, 1985) である。検索経路によるテスト効果の説明では検索することによって検索経路 (= 記銘対象へのアクセスビリティ) の精緻化および検索経路の増大を仮定している。この検索経路の増大は検索時に使用される手がかりと関連している (e.g., Bartlett, 1977; McDaniel & Masson, 1985)。McDaniel & Masson (1985) はテスト効果の説明に対して検索経路を導入した。転移適切性処理では符号化時と検索の処理が一致している場合に保持が促進されることを予測する一方で、検索経路ではテストに用いる手がかりがテストによって異なる場合に保持が促進されるとした。検索経路は符号化の種類に対応しており、複数の検索利用可能性は、手がかりの数が多いほど機能するとし、初期テストと最終テストの手がかりが一致する場合に比べて一致しない場合にテスト効果が生じることを報告した。

転移適切性処理による説明の限界

テスト効果を説明する3つの理論のうち転移適切性処理を用いた説明は手続きと密接に関連し、初期テストと最終テストで同じテスト形式を採用することで、転移適切性処理による説明を可能としている。しかし、転移適切性処理を用いた説明にも限界があることが多くの研究によって示されている。Kang, McDerrott, & Roediger (2007) は初期テスト/再学習フェイズにおいて初期テストとして4つの選択肢からターゲット項目を選択する多肢選択テスト (multiple choice)、問題文に適切な記銘対象を補充する短答テスト (short answer) および再学習として記銘対象の読解 (read statement)、挿入課題 (filler/control) を実施し、最終テストでは多肢選択テストと短答テストを実施した。転移適切性処理に基づくなら最終テストが多肢選択テストの場合、初期テストで多肢選択テストを用いて回答した記銘対象の正答率は他の活動よりも高い正答率になり、短答テストを用いた場合には初期テストで短答テストを用いて回答した記銘対象の正答率が最も高くなることが予想された。しかし、得られた結果は興味深いものであり、最終テストのテスト形式に関わらず、初期テストでの多肢選択テストを用いて回答した記銘対象は短答テストを用いて回答した記銘対象よりも多く回答された。つまり、転移適切性処理による仮説は支持されなかった。このことから転移適切性処理は必ずしもテスト効果の説明として十分であるとは言えないことが指摘できる。

転移適切性処理を支持しない結果はテスト形式を操作した場合に見られる傾向がある (e.g., Carpenter & DeLosh, 2006; Glover, 1989, Hogan & Kintsch, 1971)。これらの先行研究は最終テストに関わらず、初期テストで自由再生や手がかり再生を行った場合に再認テストや多肢選択テストを行った場合に比べて高い記憶成績になることを報告した。このテスト形式による違いは再生と再認に含まれるプロセスの違いに由来する。Kintsch (1970) の再生の二段階説によると再生には探索過程と照合過程が含まれている¹。探索過程において実験参加者は実験者からの教示に従って、ある特定の時点 (学習時) で呈示された記銘対象を探索する。探索過程に続く照合過程では、探索過程によって抜粋された記銘対象候補に対して、特定の時点で呈示されたかどうかの判断を実行する。再生テストと再認テストは探索過程を含んでいるか否かが異なっており、検索努力はこの探索過程の実施に費やされる。したがって、初期テストで取り組んだテスト形式によって費やされた検

1 Kintsch (1970) のモデルでは再生と再認に含まれる探索過程の数の違いを仮定し、用いるテストによって両者に異なる効果を持つ変数が報告された (e.g., Kintsch, 1968)。しかし、再認テストにも体制化の効果が見られること (Mandler, 1980)、再認よりも再生の記憶成績が高くなること (Tulving & Thompson, 1973) などから、後に探索過程の数ではなく、照合の違いであるとする修正モデルが提案された (Anderson & Bower, 1974)。

索努力の量が異なることで、テスト形式による違いが得られたと考えられる。

検索努力と検索経路の関係性

上記の転移適切性処理の限界に対して本研究では検索経路と検索努力によるテスト効果の説明を試みる。Roediger & Karpicke (2006a) は、検索努力と検索経路について2つの理論は排反関係ではないことを指摘している。しかし、検索努力と検索経路から異なる結果を予想することができる。検索努力と検索経路から異なる結果が生み出す変数としてテスト形式の組み合わせが挙げられる。テスト効果を検討する実験では、学習とテストを1回ずつ実施する手続きを採用している (e.g., Roediger & Karpicke, 2006b)。これに対して学習とテストの組み合わせ、保持への影響を検討する手続として検索練習 (retrieval practice) がある。検索練習を扱った研究では、4つのセッションを1サイクルとし、学習とテストの組み合わせ方の保持への影響を検討し、1回のサイクルでの検討や複数のサイクルでの検討など多岐にわたる (e.g., 堀田・多鹿, 2008; Kang, 2010; Karpicke & Roediger, 2007; Roediger & Karpicke, 2006b)。いずれの研究においてもセッション中に同じ形式のテストが実施される。セッション中に異なるテスト形式のテストに取り組んだ場合、排反関係ではない検索努力と検索経路から異なる結果を予想することが可能となる。本研究ではこの点に注目し、初期テストにおいて同じテスト形式の繰り返した場合と異なるテスト形式を組み合わせる初期テストを行った場合に、どちらがテスト効果の生起に影響するのかを、テスト効果の説明に用いられている検索努力と検索経路に対応させて検討する。

検索努力と検索経路に対応させて検討するにあたっていくつかの問題を改善する必要がある。1つ目の問題とは、初期テストで複数のテストを行う場合、テストごとにテストに必要な時間が異なるという問題である。再生テストと再認テストではテストの実施時間が異なり、初期テストから最終テストまでの時間が剰余変数になる。本研究ではテスト形式として自由再生テストと再認テストを実施するが、標準的な再認テストでは、実験参加者に学習項目と妨害項目との弁別を求める諸否型再認テストが用いられている。再認テスト時には学習項目だけでなく実際には呈示されていない妨害項目も呈示しなければいけない。そのため、実験参加者のペースでテスト項目に対して判断を行っても、学習項目のみを検索する再生テストよりも必然的にテストに必要な時間が多くなってしまう。この問題に対して本研究では再認テストのテスト形式としてバッチ型再認テスト (Klatzky, 1980 箱田・中溝 1982) に変更する。バッチ型再認テストでは実験参加者に全ての学習項目と学習項目と同数の妨害項目を同時に呈示する。実験参加者は呈示されたテスト項目のうち学習項目を選び出すことを求める。テスト項目をすべて呈示することで、テスト時間の違いを統制することが可能になる。

目的および仮説

本研究ではテスト形式およびテスト形式の組み合わせがテスト効果の生起に及ぼす影響を検討し、テスト効果を説明する理論である検索努力と検索経路のどちらによって得られる結果の説明が可能であるかに言及することを目的とする。初期テストとして再生テスト、再認テストを繰り返す、もしくは組み合わせる実施し、最終テストとして保持期間ごとに自由再生テストを実施する。以下では保持期間ごとに検索努力と検索経路に対応する仮説に言及する。保持期間が短い場合に実施する直後テストではテスト形式による違いが得られる。特に学習を繰り返すことが想起を促進することが明らかになっている (e.g., Roediger & Karpicke, 2006b; Wheeler, Ewers, & Bunanno, 2003)。これは学習を繰り返すことで、学習項目の記憶痕跡が精緻化されるため、テストに取り組んだ場合よりも高い記憶成績になる。さらにテスト形式によって異なる記憶成績が得られることが予想される。特に初期テストにおいて再認テストを行った場合、初期テスト時に再生テストに比べて熟知性による判断が可能であるため、学習項目であると容易に判断することが可能である。また、再認テストの特徴として初期テスト時に学習項目をテスト項目として呈示するため、テスト項目を正しく学習項目と判断した場合、テスト項目の再学習が可能である。そのため、再認テストを繰り返す条件 (再認 - 再認条件) や再認テストを含む条件 (再認 - 再生, 再生 - 再認条件) は再生テストを繰り返す条件より高い記憶成績になると考えられる。直後テストに対して保持期間が長い場合に行う遅延テストでは検索努力と検索経路からそれぞれ異なる結果が予想される。検索努力がテスト効果の生起に対して重要である場合、再生テストは再認テストに比べて検索時により多くの検索努力を必要となること (e.g., Kang, et al., 2007)、検索努力は保持期間が長い場合に保持を促進することが明らかになっている (e.g., Roediger & Karpicke, 2006a, 2006b)。したがって、

初期テストにおいて再生テストを繰り返す条件（再生 - 再生条件）は、他のテスト条件（再認 - 再認, 再生 - 再認, 再認 - 再生条件）学習の繰り返しに比べて検索に対して多くの検索努力が費やされたため、高い記憶成績になる。一方で検索経路がテスト効果の生起に対して重要である場合、初期テストにおいて異なるテスト形式を用いてテストに取り組む条件（再生 - 再認, 再認 - 再生条件）ではテスト形式ごとに対応した認知モードや手がかりが生起するため、遅延テストにおいて同じテスト形式に取り組む条件（再生 - 再生, 再認 - 再認条件）よりも高い記憶成績になる。

実験 1

実験 1 では再学習 / テストフェイズでの学習および初期テストの実施回数を 1 回から 2 回に増やす。さらに初期テストとして取り組むテストの形式を操作することで、初期テストでのテスト形式の種類やテスト形式の組み合わせがテスト効果の生起に影響するかを検討する。そして、得られた結果について検索努力と検索経路のどちらを用いて説明が可能であるかに言及する。

方法

実験参加者 都内の大学に通う大学生および大学院生のうち実験参加の同意が得られた第一言語が日本語である 50 名が実験に参加した。いずれの実験参加者も矯正視力を含めて正常な視力を有していた。

実験計画 本研究の実験計画はテスト形式 5（学習 - 学習, 再生 - 再生, 再認 - 再認, 再認 - 再生, 再生 - 再認）×保持期間 2（直後, 遅延）からなる 2 要因混合計画であった。テスト形式を被験者間要因, 保持期間要因を被験者内要因とした。従属変数は最終テストにおける自由再生率および忘却量とした。忘却量は直後テストの自由再生率と遅延テストの自由再生率の差分とした。

実験刺激 実験で使用する学習項目は千原・辻村（1985）から清音カタカナ 3 文字名詞（e.g., キカイ, レタス）を対象とし、熟知価が高い項目（3.50-4.75）を 200 項目抜粋した。さらに抜粋した項目から熟知価および特定の 카테고리への偏りがないように 40 項目からなる刺激リストを 5 つ作成した。作成した 5 つの刺激リストのうち、1 つのリストを学習項目として使用する学習リスト, 2 つの刺激リストを初期テストにおける再認テストの妨害項目として呈示される妨害項目リストに割り振った。3 つの刺激セットの割り振りは実験参加者間でカウンターバランスをとった。

手続き 本研究は学習フェイズ, 再学習 / 初期テストフェイズ, 挿入課題フェイズ, 最終テストフェイズの順番で実施した。

学習フェイズ 学習フェイズでは実験参加者に項目の学習を求めた。学習項目の呈示にはパソコンを使用し、実験参加者にパソコンの画面上に映し出される単語について後でテストを行うため、可能な限り覚えることを求めたが、呈示される学習項目の数については説明しなかった。初めに注視点をパソコンの画面中央に 3 秒で呈示し、続いて学習項目をパソコンの画面上に 1 つずつ 6 秒間呈示した。学習項目がパソコンの画面上に呈示されてから 6 秒が経過したら次の学習項目が呈示された。学習項目の呈示順番はランダムで作成した。

再学習 / 初期テストフェイズ 再学習 / 初期テストフェイズは再学習 / 初期テスト要因の条件ごとに異なる方法で実施した。それぞれの条件において再学習試行もしくはテスト試行を 2 回実施し、いずれも 4 分間で実施した²。

学習 - 学習条件では学習フェイズで呈示された学習項目の再学習を 2 回行った。学習フェイズでは学習項目を個別に呈示したのに対して、学習 - 学習条件では学習フェイズで呈示された学習項目が印刷された用紙を用いて一斉に呈示した。実験参加者には用紙に印刷された項目は学習フェイズでパソコンの画面上に映し出された項目と同じだが、印刷された順番はパソコンの画面上に映し出された順番と異なっていることを説明し、再

2 Roediger & Thorpe (1978) は 50 個の単語をターゲットとする自由再生を行った。累積再生率の結果からグラフを作成したところ、テストを開始してから 4 分後に累積再生率が一定になることを明らかにした。本研究ではそれに倣ってテスト実施時間を 4 分とした。

度学習することを求めた。再学習試行は実験者の合図に合わせて開始し、4分間で行った。再学習試行を始めてから2分が経過したら残り時間をアナウンスした。1回目の再学習試行の終了後に2回目の再学習試行を行った。2回目の再学習試行は印刷されている順番が異なることを除き、1回目の再学習試行と同じ方法で実施した。再生・再生条件ではテスト試行として学習フェイズで呈示された学習項目をターゲットとする自由再生テストを2回行った。自由再生テストでは学習フェイズで呈示された学習項目を呈示された順番に関わらず、可能な限り思い出し、記録用紙に書き出すことを実験参加者に求めた。再生テストは実験者の合図に合わせて開始し、4分間で行った。テストを始めてから2分が経過したら残り時間をアナウンスした。1回目のテスト試行終了後、2回目のテスト試行を実施した。2回目のテスト試行では、1回目のテストで書き出した項目をもう一度書き込んでもよいこと、1回目の自由再生テストで書けなかった項目でも書き出してもよいことを新たに教示し、実験者の合図に合わせて実施した。再認・再認条件ではテスト試行として学習フェイズで呈示された学習項目をターゲットとする再認テストを2回行った。再認テストはバッチ形式で実施した。バッチ形式の再認テストでは学習項目と学習項目と同じ数の妨害項目が印刷されたテスト用紙を使用した。実験参加者にはテスト用紙に印刷された項目のうち学習フェイズで呈示されたことを思い出せる項目にチェックをつけることを求めた。再認テストは実験者の合図に合わせて開始し、4分間で行った。テストを始めてから2分が経過したら残り時間をアナウンスした。1回目のテスト試行終了後に2回目のテスト試行を実施した。2回目のテスト試行ではテスト項目が印刷されている順番と妨害項目が異なることを説明し、1回目のテスト試行と同じ方法で実施した。再認・再生条件では再認テストと自由再生テストをそれぞれ1回ずつ実施した。再認テストは再認・再認条件と同じくバッチ形式でテストを実施した。2回目のテスト試行での再生テストではテスト用紙に印刷されていたかどうかではなく、学習フェイズで呈示された学習項目を可能な限り思い出し、記録用紙に書き出すことを新たに実験参加者に説明し、再生・再生条件と同じ方法で実施した。再生・再認条件は再認・再生条件とはテストを実施する順番が異なっている事を除き、同じ方法で実施した。

挿入課題フェイズ 再学習/テストフェイズ終了後、挿入課題を実施した。実験参加者には一般的な知識の調査であるというカバーストーリーを与え、都道府県名の自由再生テストを5分間で実施した。

最終テストフェイズ 最終テストフェイズは保持期間要因の条件ごとに自由再生テストを実施した。直後テスト条件は挿入課題の直後にテストを実施し、実験参加者に学習フェイズで呈示された項目をできるだけたくさん思い出し、どんな順番でも良いので、記録用紙に書き出すことを求めた。直後テスト終了後、実験参加者の内観を求め、1週間後に別の調査があるため、もう一度実験室に来ることを説明した。遅延テスト条件は直後テストから1週間後に直後テストと同じ方法でテストを実施した。テスト終了後に実験参加者の内観報告を求めた。最後に実験のデブリーフィングを行い、実験参加に対してお礼を述べ、全ての実験を終了した。

結果

最終テストフェイズでの自由再生率を再学習/初期テスト要因に含まれる条件ごとに算出した (Figure 1)。

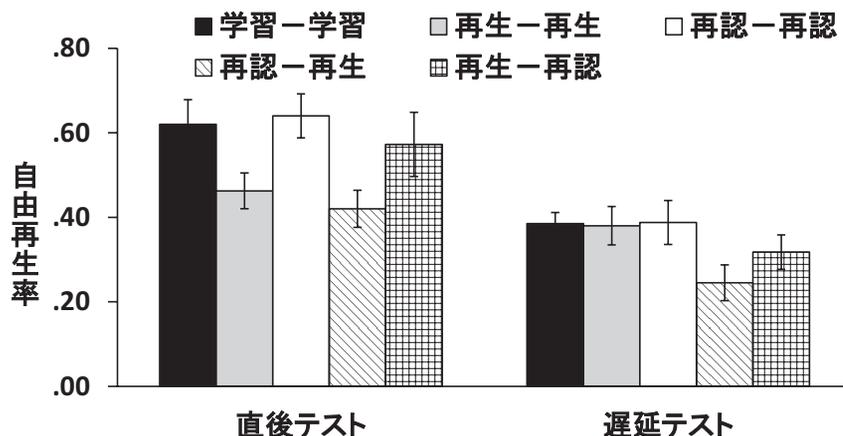


Figure 1. 保持期間別平均自由再生率

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2要因の分散分析を行ったところ、テスト形式要因の主効果は有意にならなかった ($F(4,45)=2.00$, ns , $\eta^2=.10$)。しかし、保持期間の主効果および交互作用は有意になった (それぞれ $F(1,45)=166.45$, $p<.01$, $\eta^2=.23$; $F(4,45)=4.53$, $p<.01$, $\eta^2=.02$)。交互作用が有意だったので、下位検定を行ったところ、直後テスト条件における再学習/初期テスト要因の単純主効果が有意になった ($F(4,45)=3.21$, $p<.01$, $\eta^2=.19$)。再学習/初期テスト要因の条件についてライアンの方法を用いて多重比較を行ったところ、再認-再認条件の自由再生率 ($M=.64$) は再認-再生条件の自由再生率 ($M=.42$) よりも、学習-学習条件の自由再生率 ($M=.62$) は再認-再生条件の自由再生率よりも高かった ($M=.42$)。また、テスト形式要因に含まれるいずれの条件において保持期間要因の単純主効果が有意になり、直後テストから遅延テストにかけての自由再生率の低下が見られた (再生-再生条件, $F(1,45)=5.82$, $p<.01$, $\eta^2=.08$; 再認-再認条件, $F(1,45)=54.58$, $p<.01$, $\eta^2=.39$; 再認-再生条件, $F(1,45)=26.21$, $p<.01$, $\eta^2=.31$; 再生-再認条件, $F(1,45)=55.66$, $p<.01$, $\eta^2=.31$; 学習-学習条件, $F(1,45)=47.27$, $p<.01$, $\eta^2=.25$)。自由再生率と同じく直後テストから遅延テストにかけての忘却量についても分析を行った。忘却量は直後テストの自由再生率と遅延テストの自由再生率の差として算出した。1要因分散分析の結果、再学習/初期テスト要因の主効果が有意になった ($F(4,45)=4.58$, $p<.01$, $\eta^2=.28$)。テスト形式要因の主効果が有意になったので、ライアンの方法を用いて多重比較を行ったところ、再生-再生条件の忘却量 ($M=.08$) は再認-再認条件 ($M=.25$)、再生-再認条件 ($M=.25$)、学習-学習条件 ($M=.23$) の忘却量よりも少なかった。

考 察

実験1では初期テストでのテスト形式の組み合わせがテスト効果の生起に与える影響を検討することを目的とし、再学習/テストフェイズで学習を繰り返す学習-学習条件、同じテスト形式のテストに取り組む再生-再生、再認-再認条件、異なる形式の異なるテストに取り組む再生-再認、再認-再生条件の自由再生率を比較した。その結果、Roediger & Karpicke (2006b) が学習を繰り返す場合よりもテストを行った場合に遅延テストにおいて自由再生率が高くなることを報告した結果とは異なり、テスト効果は見られなかった。本研究においてテスト効果が生じなかった原因として初期テストおよび直後テストにおける自由再生率の低さが挙げられる。遅延テストに先行するテストにおいて自由再生率が低い場合、その自由再生率を上限にして忘却量が生じる。そのため、学習-学習条件や再認-再認条件における自由再生率が高い場合、多くの忘却が生じるが、結果として再生-再生条件と同程度の自由再生率になったと考えられる。遠藤 (2007) は遅延テストにおける再学習条件とテスト条件の自由再生率を比較したところ、テスト効果が生じなかったことを報告し、その原因として自由再生率に床効果が生じたと考えた。また、Roediger & Karpicke (2006b) は、テストの遂行が難しい場合には再学習とテストの成績に違いが見られないとした。このことから本研究ではテスト時に自由再生の遂行難易度が高かったため、テスト効果が生じなかった可能性が指摘できる。

自由再生率でテストによる保持の効果が見られなかった一方で忘却量についてはテストの効果による忘却の低下効果が見られた。この結果は先行研究を支持するものである (e.g., Chan, 2010; Roediger & Karpicke, 2006b; Wheeler et al., 2003)。テストによる忘却の抑制に加えてテスト形式による違いが見られた。本研究では特に再生テストを繰り返した場合に直後テストから遅延テストにかけての忘却量は、学習や再認テストを繰り返した条件やテストを組み合わせた条件よりも少なかった。この学習の繰り返し、再認の繰り返しに対する再生の繰り返しによる忘却量の減少はテスト時に費やされる検索努力に由来する。再生の二段階説に基づくなら、再生では探索過程と照合過程を通して学習項目の判断を行う。その際、照合過程に比べて探索過程で多くの検索努力が費やされると考えられる。これに対して再認では照合過程のみで学習項目の判断が可能であるため、わずかな検索努力によって実行されると考えられる。その結果、初期テストにおいて費やされた検索努力によって忘却量が低下したと考えられる。

以上の結果は初期テストでのテスト形式の組み合わせと言うよりもむしろ、テスト形式、特に自由再生テストを行うことが保持を促進したことを示している。同時にこの結果が検索努力に由来している可能性を示唆している。しかし、同時に転移適切性処理による説明も可能であるため、その解釈には慎重を期する必要がある。

実験 2

実験 1 ではテストの影響が再生ではなく、忘却に見られ、検索努力によって忘却が減少した可能性が示唆された。この検索努力による保持の特徴として保持期間が長い場合に効果が表れることが挙げられる (e.g., Roediger & Karpicke, 2006a)。保持期間が短い場合に検索努力の効果が生じない場合でも保持期間を拡大することで、テスト効果が見られることがある (e.g., Runquist, 1983; Slamecka & Katsaiti, 1988)。また、実験 1 の忘却量にテストによる抑制効果が見られたことから保持期間を 2 週間に拡大することで、実験 1 よりも大きなテストによる抑制効果が生じ、検索努力による保持に対する促進効果が生じやすくなると考えられる。その結果、テスト形式やテスト形式の組み合わせのテスト効果の生起に対する影響が検討可能になる。

方法

実験参加者 都内の大学に通う大学生および大学院生のうち実験参加の同意が得られた第一言語が日本語である 50 名が実験に参加した。いずれの実験参加者も実験 1 には参加していなかった。

実験計画 実験 2 の実験計画はテスト形式 5 (学習 - 学習, 再生 - 再生, 再認 - 再認, 再認 - 再生, 再生 - 再認) × 保持期間 2 (直後, 遅延) からなる 2 要因混合計画であった。従属変数は最終テストでの自由再生率および忘却量とした。

実験刺激 実験 2 では実験 1 で使用した刺激と同じ刺激を使用した。

手続き 実験 2 では直後テストから遅延テストまでの保持期間を 2 週間とすることを除き、実験 1 と同じ手続きを用いて学習フェイズ, 再学習/テストフェイズ, 挿入課題を実施した。

結果

最終テストフェイズでの自由再生率を再学習/初期テスト要因に含まれる条件ごとに算出した (Figure 2)。

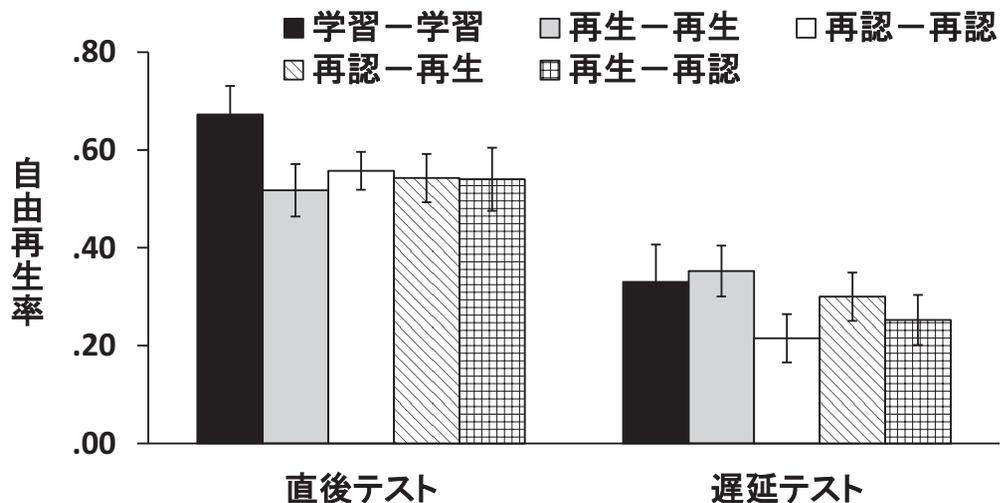


Figure 2. 保持期間別平均自由再生率

エラーバーは標準誤差を表す

最終テストフェイズでの学習項目の自由再生率に対して2要因の分散分析を行ったところ、テスト形式の主効果は有意ではなかった ($F(4,45)=0.75$, ns , $\eta^2=.03$)。保持期間要因の主効果およびテスト形式要因と保持期間要因の交互作用が有意になった (それぞれ $F(1,45)=303.62$, $p<.01$, $\eta^2=.38$; $F(4,45)=4.46$, $p<.01$, $\eta^2=.02$)。保持期間要因と再学習/初期テスト要因の交互作用が有意になったので、下位検定を行った。その結果、テスト形式要因のいずれの条件において保持期間要因の単純主効果が有意になり、直後テストから最終テストにかけて自由再生率が低下していた (学習-学習条件, $F(1,45)=93.34$, $p<.01$, $\eta^2=.41$; 再生-再生条件, $F(1,45)=21.66$, $p<.01$, $\eta^2=.21$; 再認-再認条件, $F(1,45)=93.34$, $p<.01$, $\eta^2=.62$; 再認-再生条件, $F(1,45)=46.79$, $p<.01$, $\eta^2=.40$; 再生-再認条件, $F(1,45)=68.08$, $p<.01$, $\eta^2=.40$)。しかし、直後テストおよび遅延テストにおいてテスト形式要因の単純主効果はいずれも有意にならなかった (それぞれ $F(4,90)=1.19$, ns , $\eta^2=.10$; $F(4,90)=1.07$, ns , $\eta^2=.08$)。

実験参加者の内観報告において学習項目を用いてストーリーを作成して覚える方法が多く報告された。このことから再学習試行において学習項目を一斉呈示したことで、学習項目の中から体制化しやすい学習項目を選択的に覚えていた可能性が指摘できる。そこで再学習/初期テストフェイズでの学習項目の呈示方法を一斉呈示から学習フェイズと同じ個別呈示に変更して呈示方法の影響を検討した。最終テストでの自由再生率に対し呈示方法要因 (一斉呈示/個別呈示) \times 保持期間 (直後/遅延) からなる2要因分散分析を行ったところ、保持期間要因の主効果が有意になり、直後テストから遅延テストにかけて自由再生率の低下が見られた ($F(1,18)=92.50$, $p<.01$, $\eta^2=.38$)。しかし、呈示方法要因の主効果および交互作用は有意にならなかった (それぞれ $F(1,18)=0.16$, ns , $\eta^2=.00$; $F(1,18)=0.01$, ns , $\eta^2=.00$)。以上の結果は、実験参加者が再学習試行時に体制化しやすい学習項目を選択的に符号化するという方略の促進効果が見られないことを明らかにした。そのため、学習-学習条件の自由再生率が呈示方法に由来するものではないことを示した。

自由再生率に続いて忘却量に対して1要因分散分析を行ったところ、テスト形式要因の主効果が有意になった ($F(4,45)=4.46$, $p<.01$, $\eta^2=.28$)。テスト形式要因の主効果が有意になったので、ライアンの方法を用いて多重比較を行ったところ、再生-再生条件の忘却量 ($M=.16$) は学習-学習条件 ($M=.34$)、再認-再認条件 ($M=.34$) の忘却量よりも低かった。

以上の結果をまとめると、保持期間を2週間に延長した場合でもテスト効果は認められなかったが、忘却量については初期テストとして再生を繰り返し行った場合に他の条件に比べて忘却量が少なかった。これらの結果は実験1の結果を再現し、テスト形式およびテスト形式の組み合わせについて検索努力による部分的な説明が可能であるといえる。

考 察

実験2では遅延テストの保持期間を1週間から2週間に変更してテスト形式およびテスト形式の組み合わせの影響を再検討した。保持期間を拡大することで、検索努力の効果を検出しやすくなると仮定した。しかし、実験2の結果は実験1の結果と同じく学習項目の保持というよりも忘却に影響したことを示した。

保持期間を拡大したことで、検索努力の効果が検出しやすくなったにも関わらず、テストによる保持の促進効果が見られなかった原因として自由再生率の低さと共に直後テストと遅延テストを被験者内要因として操作したことが挙げられる。つまり、まず再学習/テストフェイズにおいて学習項目の学習を繰り返すことで、学習項目の記憶痕跡が精緻化される。続く直後テストにおいて実験参加者は自身の学習状態について判断する際に熟知性や検索の流暢性を指標にする (e.g., Jacoby, Bjork, & Kelly, 1994)。つまり、学習を繰り返したことで、熟知性が高まるため、他の条件よりも多くの項目を再生することが可能となる。その結果、再生できた学習項目に対してテスト効果が生じ、遅延テストにかけての忘却が抑制されたと考えられる。またもう1つの影響が転移適切性処理である。直後テストと遅延テストを同じテスト形式で行った場合、同じ検索方略を用いることで再生が容易に行われた可能性がある。遅延テスト終了後に直後テストと遅延テストで使用した検索方略を尋ねたところ、多くの実験参加者が同じ検索方略を使用していることを報告した。このことから同じ検索方略を用いること自体が再生に影響していたといえる。以上のことから実験2では再学習/テストフェイズ

で学習を繰り返した場合、テストを行った場合に比べて学習項目の記憶痕跡が精緻化され、直後テストで再生されやすくなった。さらに遅延テストで同じ検索方略を用いることで、直後テストで再生した項目の再生が再生されやすくなったと考えられる。

実験 3

実験 1 および実験 2 では遅延テストにおいてテスト効果は見られなかった。これらの結果は、再学習 / テストフェイズで学習項目の再学習を行った実験参加者が直後テストと遅延テストに取り組んだことで、直後テストで再生された学習項目に対してテスト効果が生じた可能性が指摘できる。実験 3 では保持期間要因を被験者間要因として操作し、遅延テストのみ実施した。分析時には実験 1 の直後テストを用いてテスト形式およびテスト形式の組み合わせの影響を再検討する。

実験参加者 都内の大学に通う大学生および大学院生のうち実験参加の同意が得られた第一言語が日本語である 50 名が実験に参加した。いずれの実験参加者も実験 1, 2 には参加していなかった。

実験計画 本研究の実験計画はテスト形式 5 (学習 - 学習, 再生 - 再生, 再認 - 再認, 再認 - 再生, 再生 - 再認) からなる 1 要因被験者間計画であった。従属変数は遅延テストの自由再生率とした。

実験刺激 実験 3 は実験 1 および 2 で使用した刺激と同じ刺激を使用した。

手続き 実験 3 は最終テストフェイズにおいて遅延テストのみを実施することを除き、実験 1 と同じ方法で実施した。

結果

実験 3 では最終テスト時の自由再生率に対して実験 1 の直後条件を追加したテスト形式 5 (学習 - 学習, 再生 - 再生, 再認 - 再認, 再認 - 再生, 再生 - 再認) × 保持期間 2 (直後: 実験 1, 遅延: 実験 2) からなる二要因分散分析を実施した。最終テストフェイズでの学習項目の自由再生率を保持期間要因ごとに算出した (Figure 3)。

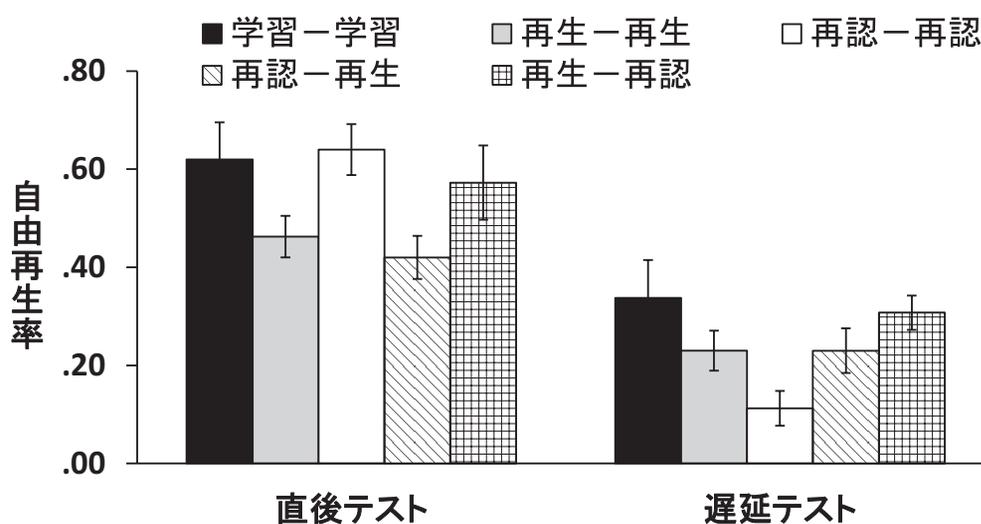


Figure 3. 保持期間別平均自由再生率

エラーバーは標準誤差を表す

最終テストフェイズでの自由再生率について分散分析を行ったところ、保持期間の主効果、テスト形式の主効果および保持期間とテスト形式の交互作用がそれぞれ有意になった (それぞれ, $F(1,90) = 75.42$, $p < .01$, $\eta^2 = .40$; $F(4,90) = 2.80$, $p < .05$, $\eta^2 = .05$; $F(4,90) = 2.91$, $p < .05$, $\eta^2 = .06$)。保持期間要因とテスト形式要因の交

相互作用が有意になったので、単純主効果の検定を行ったところ、テスト形式要因に含まれるいずれの条件においても保持期間要因の単純主効果が有意になり、直後テストから遅延テストにかけての自由再生率の低下が見られた（学習 - 学習条件, $F(1,90)=13.42, p<.05, \eta^2=.71$; 再生 - 再生条件, $F(1,90)=9.09, p<.05, \eta^2=.04$; 再認 - 再認条件, $F(1,90)=11.81, p<.01, \eta^2=.46$; 再認 - 再生条件, $F(1,90)=6.07, p<.05, \eta^2=.03$; 再生 - 再認条件, $F(1,90)=46.79, p<.01, \eta^2=.24$ ）。また、保持期間要因のそれぞれの条件においてテスト形式要因の単純主効果が有意になった（それぞれ $F(4,90)=3.18, p<.05, \eta^2=.06$; $F(4,90)=2.56, p<.05, \eta^2=.06$ ）。直後テスト条件では再認 - 再認条件の自由再生率 ($M=.64$) と再認 - 再生条件の自由再生率 ($M=.42$) の間に有意傾向が見られ、再認 - 再認条件の自由再生率は再認 - 再生条件の自由再生率よりも高かった。また、遅延テストでは学習 - 学習条件の自由再生率 ($M=.33$) は再認 - 再認条件の自由再生率 ($M=.23$) よりも高かった。

考 察

実験3では実験1および2においてテスト効果が生じないことが保持期間を被験者内要因として繰り返しテストを行った可能性を考慮して被験者間要因として操作して再検討を行った。その結果、自由再生率にテスト効果は生じないことが明らかになった。

実験1, 2に見られるテストに対する学習の優位性は直後テストを行ったことによって生じたテスト効果ではなく、実験参加者が用いた符号化方略である体制化に由来すると考えられる。再学習/テストフェイズにおける学習試行では学習フェイズと異なる順番で印刷された学習項目を学習する。つまり、学習フェイズおよび学習試行を含めて学習項目を3回学習するが、それぞれの学習においていずれも初頭部に配置される学習項目は異なっている。そのため、学習ごとに体制化されやすい初頭部の学習項目が異なることで、より多くの学習項目が長期記憶に変換されたと考えられる。その結果、テスト効果が生じなかったと考えられる。

総 合 考 察

本研究ではテスト効果に対するテスト形式およびテスト形式の組み合わせの影響を検討し、検索努力と検索経路のどちらの理論によって説明できるかに言及することを目的とした。テスト効果の生起に対して検索努力が重要であるなら、初期テストで再生を繰り返した場合に遅延テストにおいて自由再生率が他の条件に比べて高くなる。一方で検索経路が重要であるなら、初期テストで再生テストと再認テストを組み合わせた場合に遅延テストにおいて自由再生率が他の条件に比べて高くなることが予想された。また、直後テストから遅延テストにかけての忘却量は自由再生率とは逆のパターンを示すことが考えられた。本研究の結果は、テスト形式およびテスト形式の組み合わせはテスト効果の生起に影響しないことを示している。一連の実験においてテスト効果が生じなかった原因として再生率の低さを指摘した。しかし、初期テストにおいて再生テストを繰り返した場合、記憶高進 (hypermnnesia) やレミニセンス (reminiscence) が生じていた。記憶高進とはテストを繰り返し行った場合にフィードバックがないにもかかわらず、記憶成績が上昇する現象である (e.g., 林・太田, 2002)。これに対してレミニセンスとはテストを繰り返し行った場合、以前に行ったテストでは思い出せなかった項目を新たに思い出す現象である (e.g., 林・藤岡・本多, 2008)。前者がテストを行うごとに成績が上昇するのに対し、後者は想起した内容が変化するものである。本研究の再生 - 再生条件では実験参加者のうち5割の実験参加者に記憶高進、7割の実験参加者にレミニセンスが見られた。このことは1度でも検索された項目は実際には直後テストで再生された項目以外にもあることを示している。初期テストにおいて学習項目を正しく検索することによって後の検索が促進されることが報告されている (e.g., Butler, 2010)。また、初期テストにおける学習項目の検索成功回数が多いほど、保持が促進されていることも報告されている (e.g., van den Broek, Segers, Takashima, & Verhoeven, 2013)。しかし、本研究の結果と併せることで、初期テストで学習項目を正しく検索できることは必ずしも保持に貢献するわけではないと考えられる。そのため、今後はどの程度検索が成功することで、テスト効果の生起に影響するかを検討していくことが望ましい。

自由再生率においてテスト効果が生じなかった一方で、忘却量ではテストによる抑制効果が見られた。

Roediger & Karpicke (2007b) も本研究と同様に学習を繰り返す場合に比べてテストを繰り返した場合に忘却が少なくなることを報告し、本研究の結果はこれを追認するものである。テストをすることの忘却への影響は、テスト形式によって異なり、本研究では初期テスト時に再生テストを行った場合に直後テストから遅延テストにかけての忘却量が少なかった。このことから忘却量の結果は検索努力によって説明が可能であることを示唆している。しかし、忘却量は直後テストと遅延テストでの再生率を用いて算出しているため、直後テストで再生できた学習項目が少ない場合、項目あたりに費やされる検索努力の量が多くなり、忘却しにくくなる。これに対して直後テストで再生率が高い場合、項目あたりに費やされる検索努力の量が少なくなり、忘却しやすくなる可能性がある。そのため、本研究でテスト効果が得られなかった原因として挙げている直後テストにおける再生率の低さを改善することが今後の大きな課題になるといえる。

引用文献

- Anderson, J. R. & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition*, 2, 406-412.
- Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 719.
- 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.
- Butler, A.C. (2010). Repeated Testing Produces Superior Transfer of Learning Relative to Repeated Studying.. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 36, 1118-1133.
- Butler, A., Karpicke, J.D., & Roediger, H.L. (2007). The Effect of Type and Timing of Feedback on Learning from Multiple-Choice Tests. *Journal of Experimental Psychology: Applied*, 13, 273-281.
- Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514-527.
- 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.
- Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14, 474-478.
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36, 438-448.
- Chan, J. C. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, 18, 49-57.
- 千原 孝司・辻村 祐子 (1985). 清音 3 音節名詞について : 40 カテゴリー・500 語の熟知価. 滋賀大学教育学部紀要, 35, 75-99.
- 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.
- Gardiner, J. M., Craik, F. I. M., & Bleasdale, R. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, 1, 219-216.
- Gates, A.I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 6.
- Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392.
- 林 美都子・藤岡 真也・本田 真大 (2008). 記憶高進現象における意識の影響 - 検索意図と代理検索回路仮説の観点から -. 心理学研究, 79, 317-324.
- 林 美都子・太田 信夫 (2002). 記憶高進研究の近年の動向. 筑波大学心理学研究, 24, 59-73.

- 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.
- 堀田 千絵・多鹿 秀継 (2008). テストが学習材料の長期の記憶成績に及ぼす正の効果と負の効果. 日本教育心理学会総会発表論文集, *50*, 290.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of verbal learning and verbal behavior*, *17*, 649-667.
- Kang, S. H. K.(2010) Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*. *38*, 1009-1017.
- Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528-558.
- Karpicke, J.D., & Roediger, H.L.(2007).Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151-162.
- Kintsch, W. (1968).Recognition and free recall of organized lists. *Journal of Experimental Psychology*, *78*, 481-487.
- Kintsch, W. (1970). Models for free recall and recognition. *Models of human memory*, 331-373.
- Klatzky, R. L.(1980). *Human memory: Structures and Processes Second Edition*. San Fransisco & London: Freeman and Company. (クラツキー, R.L. 箱田 裕司・中溝 幸夫(共訳) (1982). 記憶のしくみ II - 認知心理学的アプローチ - サイエンス社)
- Klatzky, R. L.(1982). *Human memory: Structures and Processes*. San Fransisco & London: Freeman and Company.(クラツキー, R.L. 箱田 裕司・中溝 幸夫(共訳). 記憶のしくみ II - 認知心理学的アプローチ - サイエンス社)
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, *87*, 252-271.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371-385.
- 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.
- Roediger, H. L., & Karpicke, J. D.(2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181-210.
- Roediger, H.L, & Karpicke, J, D.(2006b). Enhanced Learning Taking Memory Test Improves Long-Term Retention. *Psychological Science*. *17*, 249-255.
- Roediger, H. L., & Thorpe, L. A.(1978). The role of recall time in producing hypermnesia. *Memory & Cognition*, *6*, 296-305.
- Runquist, W. N.(1983). Some effects of remembering on forgetting. *Memory & Cognition*, *11*, 641-650.
- Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 716-727.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641-657.
- 多鹿 秀継 (2008). テストが学習材料の長期の記憶成績に及ぼす影響. 神戸親和女子大学大学院研究紀要, *4*, 57-65.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*, 352-373.
- Van den Broek, G. S., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, *22*, 803-812.
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571-580.