

テストが学習材料の長期の記憶成績に及ぼす影響

The Influence on testing long-term retention of learning materials

多 鹿 秀 継*

Tajika Hidetsugu

本研究の目的は、記憶テストの実施が学習材料の長期の記憶成績に及ぼす影響を、認知心理学の実験研究に基づいて文献展望したものである。最初に、教育測定・評価の分野におけるテストの意味を明確にした。ついで、当該のテーマに関する基礎的な実験研究は古くからなされていることを説明し、再生テストないしは再認テストを学習時に反復して実施することが、単に学習材料を学習することよりも後の最終テストで促進効果を生じるというテスト効果を明確にした。最後に、テスト効果を説明するいくつかの仮説を概観した。

キーワード：テスト効果、再生テスト、再認テスト、検索仮説、転移適切性処理仮説

1 本研究の背景と目的

今日、学校教育における測定・評価のテストをめぐる様々な問題、例えば学力テストに見られる学力低下の問題等は、測定・評価の領域を専門とする教育心理学者だけでなく、教授・学習過程の研究領域を専門とし、学校における授業過程を「授業目標—授業過程—子どもの測定・評価」の一連の教授・学習過程と見なし、測定・評価を効果的な授業過程を構築するための学習環境の1つとして理解する教育心理学者を含めて、適切な取り扱いの求められる喫緊の問題として議論がかまびすしい状況にある(山森・荘島, 2006)。因みに、2007年の4月に、国語と算数・数学に関する全国学力・学習状況調査(全国学力テスト)が、小学6年生と中学3年生を対象に文部科学省により実施された。そして、10月下旬には、その結果が公表された。全国学力テストは、小学6年生で約114万人、中学3年生で約108万人の児童・生徒を対象とした大規模のものであった。

本論文では、測定・評価の領域で取り上げられてきたテストを、認知心理学における記憶の基礎的な研究成果に基づいて取り上げるものである。記憶の基礎的な研究から、テストを実施することが後の長期の記憶の保持を高めるということが知られている。本論文の目的は、テストが学習材料の長期の保持を高めるとする主だった研究をreviewし、テストが学習材料の長期の記憶成績の促進効果を説明するい

くつかの仮説を吟味することである。

ところで、授業過程を通して子どもが獲得した知識を、様々な方法に則ってテストする場合、テストは大別すると以下の4つの意味において実施されるといえるだろう。

1つは、テスト本来の目的である教育測定・評価と結びついたものである。つまり、テストを実施することは、まず学習した内容の定着の程度を確かめることにある(多鹿, 1989)。この場合、テストは授業で学習した内容の理解や記憶を吟味することであり、学習した内容の発展課題あるいは応用課題として、転移可能性を測定することでもある。

よく知られているように、学習内容の吟味方法としてのテストには、客観テストと論文テストがある。認知心理学の用語を敷衍すると、客観テストは再認テストが中心であり、真偽法(○×式)や多枝選択法がよく知られている。他方、論文テストは、記銘した内容を記述する単純な再生テストから、単に記銘した内容の再現に限定されず、記銘した様々な内容や先行知識を組み合わせ推論する形式のテストまで幅が広い。再生テストは、解答が一義的に決定される場合は客観テストと呼んでよい。

教育測定・評価に関わる教育心理学者は、評価の資料としてテストを分析し開発することに焦点を当て、テストの信頼性や妥当性を高めることに基本的な関心があるといえる。それ故、児童・生徒が学習した内容の何を知っているかを測定するためのアセ

* 本学大学院教育学専攻教授

メントとしてテストを工夫すること、これは確かにテスト実施のために必要なものである。

テストを実施することの第2の意味は、第1のテスト実施の意味合いの結果に基づき、学習者に学習内容の理解度や進捗の程度をフィードバックすることで、学習への動機づけを高め、学習への取り組みを促すことにある(橋本、1956; 村山、2006)。

テストを実施することの第3の意味は第2の意味に類似するが、学習者自身が自分の学習結果をモニターすることによって自己評価し、次の学習活動へどのように取り組むかのプラン作成を支援するためにテストを実施することを目指したものである。第3は、ある意味で児童・生徒のメタ認知の活性化を促す取り組みといえる。

テストを実施することの第4の意味は、テストを実施することによって、学習内容の保持をより堅固なものにしかつ保持を高めることに関するものである。テストを実施することあるいは受けることにより、学習材料の後の保持を高めることを、一般にテスト効果(testing effect)と呼ぶ。テスト効果に関する研究成果は、古典的な研究においても既知の内容であった(Abbot, 1909; Gates, 1917; 橋本, 1956, 1959; Jones, 1923-1924; Spitzer, 1939; なお、これらの文献の内、橋本と Spitzer 以外は Roediger & Karpicke (2006a) による)。認知心理学の記憶研究と教育との関連を取り上げた Psychonomic Society 学会の最近のシンポジウム(2005年11月)において、授業場面における基礎研究の成果が議論され(McDaniel, Roediger, & McDermott, 2007)、テスト効果に関する研究が新たに報告され始めた(Roediger & Karpicke, 2006a)。

本論文では、テストを実施することに関する第4の意味に焦点を当て、教師がテストを行うことにより、児童・生徒の学習した内容に関する長期の記憶成績が上昇することを、認知心理学における記憶研究の主だった文献を展望することによって明確にし、その説明概念を吟味するものである。即ち、本論文では、テストを実施することが獲得した知識を単に測定してその程度を確認するだけでなく、児童・生徒の知識の性質やアクセスの可能性を変えること、換言すると、学習のみの条件に比べて知識の長期の保持を促進することを明確にすること、並びに得られた結果を説明するいくつかの仮説を取り上げて吟味するものである。

再学習時間とテスト時間を等価にした場合、ある

学習材料をテストすることは、当該の学習材料を何度も再学習するよりも、それらの学習材料の長期の記憶において促進効果をもつ(McDaniel et al., 2007)。このように、学習よりもテストを繰り返すことにより、学習内容の貯蔵期間をおいた場合、後の最終テストの保持成績が学習のみを繰り返す条件群よりも促進されることを、上述したようにテストを実施することによる長期の保持の促進効果、あるいは単にテスト効果と呼んでいる。

テスト効果は、認知心理学における記憶研究では比較的良好に知られている効果であるといえる。教育心理学の領域では、テスト効果よりもむしろ第1や第2の意味でのテスト研究が盛んである。しかしながら、テスト効果の詳細を吟味することは、認知心理学における記憶研究だけでなく、教授・学習過程を吟味する教育心理学にとっても、今日議論されている教育測定・評価の問題に対して、貴重な示唆を与えるものといえる。

本論文では、それ故、テストがもつ様々な課題、例えば、テスト形式、テスト項目の選択、テスト得点の扱い、テストの信頼性や妥当性の問題等の課題を吟味するものではない。また、教室場面における教科のテスト事態や動機づけに言及するものでもない。本論文では、テストを実施することが学習した内容の保持にどのような影響を及ぼすのかを、記憶の実験研究の成果に基づいて主だった文献から展望し、学習内容の長期の保持にとって、テストすることによる促進効果を明らかにすることに焦点を当てたものである。

勿論、テストを実施した場合でもテスト効果が生じない研究もいくつか存在する。テスト効果が認められない研究の1つとしてテストによる干渉効果があり、他に負の暗示効果が認められる。

テストによる干渉効果とは、学習材料を再生することで他の学習材料の再生を損なうことを意味する。例えば、部分リスト手がかり効果(part-list cueing effect or part-set cueing effect)として知られている現象は、テストによる干渉効果の一例として認められることができる(Slamecka, 1968)。通常、学習したリストの一部を構成する項目をテスト時に提示して残りの項目の再生を求めることは、学習項目の一部を手がかりとして提示せずに学習した項目をすべて再生するように求める場合に比べて、符号化時に提示されたばらばらの各項目を1つにまとめる体制化方略と呼ばれる符号化処理方略を利用して記憶す

ることが一般的に知られていること (Tulving, 1962) から、与えられる一部の項目が体制化の一部を構成しているため、残りの項目も再生されやすくなることが予想される。

しかしながら、Slamecka (1968) の1つの実験では、記銘材料としてカテゴリに属する事例 (単語) が使用されたが、リスト内の様々なカテゴリに含まれる単語の中から、再生時に1つないしは2つの単語を手がかりとして提示すると、そのような手がかり単語が全く与えられない場合に比べて、残りの単語の再生が悪くなった。Slamecka (1968) の実験結果はその後拡張され、学習リスト内の与えられる手がかり項目の数が多くなると、同じリスト内の残りの項目の再生の割合が低くなることが報告されている (Rundus, 1973)。

このような研究に基づき、Roediger (1978) は10カテゴリ各5項目からなる項目リストを実験参加者に提示した。テスト時に、すべてのカテゴリから項目を再生するように教示した。手がかりとして与えられるカテゴリ数が各群によって異なった。その結果、手がかりの多い群ほど、多くの項目を再生した。しかしながら、カテゴリ手がかりが多く与えられた群ほど、リストの一部であった他のカテゴリの再生が悪かった。

また、検索誘導性忘却 (retrieval-induced forgetting) として知られる検索に特有の効果もテストによる干渉効果として位置づけることができる (Anderson, Bjork, & Bjork, 1994; 文献展望に関しては、Anderson, 2003)。検索誘導性忘却の現象は、多様な学習事態やテスト方法のもとで得られており、比較的一般的な忘却現象であるとして知られている。検索誘導性忘却の基本的な実験手続きは以下の通りである。学習時に複数のカテゴリ名と事例とが対になった材料を学習する (例えば、「赤いーれんが」、「赤いートマト」、「家具ーつくえ」・・・)。学習後の練習段階で、学習時に提示されたカテゴリの半数のうち、当該のカテゴリと事例の対の半数に対して、カテゴリ名を手がかりとして反応項の事例の単語完成課題を行う (「赤いーれ□□」)。ついで、実験と無関連な課題を行った後、学習時のすべてのカテゴリ名が手がかりとして提示され、それと対になっていた事例を再生するように求められる (「赤いー□□□」、「赤いー□□□□」、「家具ー□□□□」・・・)。その結果、練習段階で提示されていなかった残り半数のカテゴリ名と対にされた事例の再生成績 (「つ

くえ」) を基準としたとき、練習段階で提示された半数のカテゴリ名と対になった単語完成課題の事例 (「れんが」) の再生は、基準 (「つくえ」) の再生よりも促進される。しかしながら、練習段階で単語完成課題として練習していない他の事例 (「トマト」) の再生は、基準の単語 (「つくえ」) の再生よりも悪いものであった。

このような検索誘導性忘却では、学習時に提示された項目対の反応項は、練習時に同じ刺激項をもつ反応項をテストされることにより、最終テスト時に負の効果をもつことを示している。

テストによる抑制効果に加えて、負のテスト効果として負の暗示効果を指摘することができる。負の暗示効果とは、児童・生徒がテストから学習する間違った情報を正解とする信念に言及するものである (Remmers & Remmers, 1926)。Roediger and Marsh (2005) は、テスト効果を測定する手続きに基づいて、多枝選択テストを使って負の暗示効果を吟味した。彼らの問題の設定は、負の暗示効果が正のテスト効果を圧倒するほどの強い負のテスト効果を示すかどうかであった。18の短文を読ませ、これらの短文と読んでいない18の他の短文の項目を混ぜた多枝選択テストを実施した。多枝選択テストの項目は、正解1つに対して、2、3、5つのdistractorで構成された。多枝選択テストの結果、読んだ短文の項目を再認する方がよかった。また、distractorが増えるとエラーも増えた。このテストの後に、最終手がかり再生が実施された。このテストでも、多枝選択テストで示されていない情報についての質問もした。その結果、学習時に読んだ短文に関してテスト効果が見られ、テストされた項目はテストされない項目よりもよく再生された。学習していない短文に対しても、同様にテストされていない項目よりもテストされた項目の方がよかった。しかし、同時に負の暗示効果も見られ、選択枝の数が多きほど、再生の成績が悪かった。

2 再生と再認によるテスト効果

長期の保持におけるテスト効果を見た古典的な研究例として知られる Gates (1917) の研究を、Roediger and Karpicke (2006a) に従ってまとめると (183 ページ)、Gates (1917) は学習材料として無意味音節と文章を実験参加者として学童期の1年から8年の児童・生徒に読ませ、その後、記銘した材

料を再生するように教示した。この実験では、記録した学習材料を再生する一連の作業を暗唱(recitation)と呼び、暗唱の時間を操作した。即ち、この実験ではトータルの学習時間は一定であり、学習時間の0、20、40、60、80、90%を暗唱としての自己テストに費やすための時間の割合として操作した。この直後再生では提示順に再生させ、3ないし4時間後に再度再生テストを実施した。これを遅延再生と呼ぶ。実験の結果、無意味音節材料では1年生を除いて暗唱の効果がみられ、暗唱に時間をかけるほど再生率が高くなった。また、1年生を使用していない文章材料でも、すべての学年で暗唱の効果がみられ、暗唱に時間をかけるほど再生率が高くなった。また、このことは、忘れた材料の再学習を含む暗唱による再生行為は、学習を促進することを示している。

また、Spitzer (1939) は、多数の小学6年生に文章を学習させ、直後テストの後63日の遅延の範囲で、様々なスケジュールに沿って多枝選択テストを実施した。その結果、各々の遅延後の最初にテストされた得点を結ぶと、63日までの遅延期間で典型的な忘却曲線が見られた。しかし、各遅延後の最初のテストの後に遅延期間を設けて再テストした場合、それらの得点は、同じ遅延後の最初のテスト得点に比べて、高いものであった。

更に、Tulving (1967) は36項目からなるターゲットの単語を、3種の学習やテストの作業サイクルからなる条件群の1つに割り当てられた実験参加者に記録させた。1つは標準群と呼ばれ、学習-テスト-学習-テストの4試行の作業サイクルを通して1サイクルを構成し、6サイクル24試行(それ故、12回の学習試行と12回のテスト試行からなる)を実施する条件群であった。また、反復学習群と呼ばれる群は、1サイクルを学習-学習-学習-テストで構成し、18回の学習試行と6回のテスト試行からなる条件群であった。更に、反復テスト群と呼ばれる群は、1サイクルを学習-テスト-テストで構成する条件群であり、6回の学習試行と18回のテスト試行の機会をもつ条件群であった。学習時間とテスト時間を等価にしたので、全体の24試行に要する時間は条件群間で異ならなかった。即ち、学習時間は1項目1秒で提示され、再生時間も36項目分として36秒間の口頭再生が求められた。

Tulving (1967) の再生結果は、各サイクルの最

終再生の結果(各サイクルの4つ目の試行である再生テストの結果)が比較された。実験の予想として、もし、学習試行が項目の保持にとって重要であれば、反復学習群、標準群、反復テスト群の順に再生成績はよくなるであろうし、テスト効果が見られるなら、反復テスト群、標準群、反復学習群の順に再生がよくなるであろう。通常は、前者の予想が一般的であるといえる。

Tulving (1967) の実験の結果、3群の学習曲線は類似したものであった。最終の24試行目のテストで、反復学習群と標準群では20項目を再生し、反復テスト群では18.5項目を再生した。反復テスト群における他の2群に対するテスト効果は見られなかった。しかしながら、反復テスト群は、24試行のうちわずかに6試行のみを学習しただけであるにもかかわらず、18試行の学習を行った反復学習群と同等の再生数を生み出したことは、ある意味でテスト効果と言えるかもしれない。

本来の研究の目的はテスト効果を吟味することではないが、Hogan and Kintsch (1971) の実験2では、実験参加者に40項目からなる単語リストを、1項目当たり2秒提示で、4回学習させる学習-学習-学習-学習の反復学習群と、1回の学習後3回の自由再生テストを繰り返す学習-テスト-テスト-テストの反復テスト群における48時間後の再生テストと再認テストの結果を比較した。その結果、自由再生テストでは、反復学習群の再生数は6.1項目で、反復テスト群の再生数は8.2項目であり、反復テスト群のテスト効果が認められた。他方、再認テストでは、反復学習群の再認数は32.8項目で、反復テスト群の再生数は27.7項目であり、テスト効果は認められなかった。なお、再認テストの結果は、48時間後に再生と再認の両テストを受けた条件と再認テストのみを受けた条件群の平均値である。

Thompson, Wenger, and Bartling (1978) はHogan and Kintsch (1971) の研究を発展させた。実験2では、実験参加者に40項目からなる単語リストを、4回学習させる学習-学習-学習-学習の反復学習群と、1回の学習後3回の自由再生テストを繰り返す学習-テスト-テスト-テストの反復テスト群に加えて、一度リストを学習-テストし、再生に失敗した項目のみを再学習してすべての項目をテストされる学習-テスト(忘却項目の再提示)-テスト(忘却項目の再提示)-テスト(忘却項目の再提示)の反復再提示群に振り分けて実験を行った。

その結果、20 分後の最終再生テストでは反復テスト群が最も悪い成績を示した。しかしながら、48 時間後の最終再生テストでは、反復テスト群と反復学習群の成績に違いはなく、反復再提示群が最もよかった。

Tajika (1986) は、実験 1 において、小柳・石川・大久保・石井 (1960) の熟知価表から選択した 80 項目の高熟知価からなる三音節名詞を用いて、半数は学習項目とし、残り半数は再認テスト用 distractor 項目として、学習-テスト-テスト-テスト条件 (以下では 1-3 群) と学習-学習-学習-テスト (以下では 3-1 群) の 4 試行の作業を通して 1 サイクルの実験を大学生に実施した。学習では、1 項目 2 秒提示で 40 項目を継時提示した。項目間隔は 1 秒であった。学習と学習、学習とテストの間隔は 30 秒であった。テストは 3 種類で構成され、自由再生、三音節名詞の頭文字 1 字を手がかりとして与え、残りの 2 文字を再生させる手がかり再生、及び再認テストであった。なお、再認テストでは、選択した 80 項目の残りの 40 項目を含めた再認表からターゲット項目を選択するように教示した。テスト時間は、学習時間とテスト時間を等価にするために 120 秒であった。テストは、各群の第 4 試行目のテストを直後テストとし、直後テストの 1 時間後に最終テストを実施した。実験結果は、1-3 群の 3 回目の最終テスト得点と、3-1 群の 1 回目だけの最終テスト得点、及び両群の 1 時間後の遅延テスト結果を比較した。

Tajika (1986) の実験 1 の結果、1-3 群の直後テストの再生率は .28 で遅延テストの再生率は .29 であった。他方、3-1 群の直後テストの再生率は .41 で遅延テストの再生率は .31 であった。また、1-3 群の直後テストの手がかり再生率は .70 で遅延テストの手がかり再生率は .62 であった。また、3-1 群の直後テストの手がかり再生率は .92 で遅延テストの手がかり再生率は .82 であった。更に、1-3 群の直後テストの再認率は .68 で遅延テストの再生率は .63 であった。他方、3-1 群の直後テストの再認率は .85 で遅延テストの再認率は .74 であった。

各条件群の忘却率を「(直後テスト得点の割合 - 遅延テスト得点の割合) ÷ 直後テスト得点の割合」によって算出したところ、再生テストの場合、1-3 群で -.05、3-1 群で .24 であった。文字手がかり再生テストの場合、1-3 群で .07、3-1 群

で .12 であり、再認テストの場合、1-3 群で .11、3-1 群で .11 であった。このことから、Tajika (1986) においても、わずか 1 回しかリスト項目を学習していないにも関わらず、再生テストに関して再生率は低いが、忘却率を基準に結果を判断すると、テスト効果が認められたといえるだろう。

Tajika (1986) の実験 2 では、対連合学習として、小柳他 (1960) の三音節名詞を実験 1 と同一の基準から 24 対の項目対を構成し学習させた。また、同じく同一の基準から再認用の distractor 項目を選択した。再生テストは刺激項を手がかりとして提示して反応項を求めた。また、再認テストでは、学習した 12 対と刺激項は学習項目で反応項は distractor 項目である 12 対からなる再認表を作成し、正解の対を選択させた。1 対の提示時間は 3 秒であり、対間の時間は 1 秒、テスト時間は 96 秒であった。実験 1 と同様に、1-3 群も 3-1 群も、ともにテストは再生ないし再認で統一された。実験の結果、1-3 群の直後テストの再生率は .69 で遅延テストの再生率は .59 であった。他方、3-1 群の直後テストの再生率は .98 で遅延テストの再生率は .62 であった。また、1-3 群の直後テストの再認率は .85 で遅延テストの再認率は .81 であった。他方、3-1 群の直後テストの再認率は .99 で遅延テストの再認率は .92 であった。

各条件群の忘却率を実験 1 と同様に算出したところ、再生テストの場合、1-3 群で .14、3-1 群で .37 であった。再認テストの場合、1-3 群で .04、3-1 群で .05 であった。このことは、実験 1 と同様に、再生テストでは、1-3 群は 3-1 群に比べて忘却率が低く、テスト効果が見られたといえる。なお、再認に関しては、課題が容易であり、成績の天井効果が認められた。

また、Allen, Mahler, and Estes (1969) は実験参加者に対連合学習を 5 試行ないしは 10 試行行い、テストなし、1 回のテスト、あるいは 5 回のテストを実施した。1 日後に、刺激項のみを提示して反応項を回答させる最終の手がかり再生テストを実施したところ、テストを受けない条件では 5 試行の学習よりも 10 試行の学習を行う方が若干優れていたが、5 回のテストを受けた条件群では学習の試行数の違いの差異は見られなかった。

更に、生成効果の研究として知られている Jacoby (1978) では、多くの対連合学習 (例えば「はな-チューリップ」) を行った後、対を再学習さ

せる条件群（再学習群：「はな－チューリップ」）か、学習時に提示した対の反応項の一部を断片文字にした課題を与え、対の刺激項に対する反応項を生成するように教示される条件群（生成群：「はな－チ□－□ッ□」）が設定された。再学習か生成テストで提示される対は、学習直後に提示されるか 20 対の干渉対の学習後に提示される遅延提示に操作された。実験の終了時に、学習対の刺激項を手がかりにして反応項を求める最終の再生テスト（「はな－□□□□□□」）を実施した。その結果、対の反応項を生成する条件群が再学習群よりよい成績を示した。更に、直後のテストよりも遅延のテストを実施した方が、最終の再生テストでよい成績を得た。

これら上記の研究以外にも、おびただしい数の様々な研究がテスト効果を報告している（最近の文献に関しては、McDaniel et al. (2007) 及び Roediger & Karpicke (2006a) を参照のこと）。以下では、最近の研究を1つだけ紹介しておこう。

Karpicke and Roediger (2007) は、Tulving (1967) の研究に鼓舞され、Tulving (1967) の研究方法を改善することによって、テスト効果を見出した。Karpicke and Roediger (2007) の実験1では、提示項目数は 40 項目とし、提示時間は 3 秒であった。Tulving (1967) では 1 秒の口頭再生であったので、学習者が再生する時間を十分に確保した。試行数は 20 試行とし、5 回のサイクルで最終テストの比較を行った。また、Tulving (1967) では上述のように反復テスト群はテストを繰り返すため、サイクルの最後のテストでは学習直後の短期記憶の効果を得られないと考えられ、他の条件群に対して不利であると理解された。そのため、短期記憶と長期記憶の効果を分離した。更に、1 週間後の遅延最終再生テストを実施した。条件群は Tulving (1967) と同一の 3 条件群から構成され、標準群（学習－テスト－学習－テスト）、反復学習群（学習－学習－学習－テスト）、及び反復テスト群（学習－テスト－テスト－テスト）であった。

Karpicke and Roediger (2007) の実験1の結果、1 週間後の遅延最終再生テストで、反復テスト群の再生が反復学習群に比べてよい成績を示し、テスト効果が認められた。また、標準群が最もよい遂行結果であった。

彼らの実験2は、どのような種類の反復練習が長期記憶の保持の効果をもつのかを吟味した。標準群に加えて、学習－学習－テスト－テスト（反復群）、

並びに標準群と同じ学習－テストの試行数をもつが、試行の内容で異なる2つの条件群を設定した。即ち、標準群と同じ学習－テストの試行順序であるが、先行のテストで再生された項目は次の学習で除かれ、テストではすべての項目を再生するように求められる群（学習除外群）、先行のテストで再生された項目は次の学習で除かれかつテストでも学習した項目のみを再生するように求められる群（学習・テスト除外群）の4群の比較をした。それ故、実験2では、反復学習群や反復テスト群は構成されなかった。

実験2の結果、学習・テスト除外群は最も早く学習を達成したが、1 週間後の遅延最終再生テストでは最も悪かった。以前に再生された項目を反復学習する標準群は学習除外群に比べて保持の促進は見られなかった。他方で、標準群と学習除外群は反復群よりもよい成績であった。

3 テスト効果を説明する仮説

ここでは、Dempster (1996) 並びに Roediger and Karpicke (2006a) を参考にして、テスト効果を説明する以下の3つの仮説を簡潔に紹介しよう。

3-1 過剰学習仮説

過剰学習仮説によると、テストを繰り返し実施することが、単に学習を繰り返し実施してテストすることよりも効果的である理由は、学習者がテスト中に学習材料に繰り返し触れることによる、換言すると、学習材料の一部を過剰学習したことによるというものである。Thompson et al. (1978) の研究からも理解できるように、テストは学習した材料の一部に追加的に触れることになり、この余分に触れることがテスト効果を生み出すといえる。

テスト効果の実験は、一般に項目を学習し遅延の最終テストを受ける反復学習群と、項目を学習してテストされ、その後に遅延の最終テストを受ける反復テスト群を比較する。それ故、この過剰学習による説明の背景には、反復テスト群は、反復学習群に比べて何度もテストを繰り返すことにより、項目へより多くの時間をかけているという全学習項目への接触時間とテスト効果との交絡が見られる。しかしながら、学習時間とテスト時間を等しくした研究においてもテスト効果は見られており (Karpicke & Roediger, 2007)、過剰学習説はテスト効果を適切に説明することはできないといえる。

また、過剰学習説を支持する考えとして、学習者が学習中にどの項目が再生されやすいかを選択しており、再生の容易な項目ほど選択されやすく、かつ当該の項目に付加的な練習つまり過剰学習を行って検索を容易にするという。しかしながら、この項目選択の考えでは、例えば、直後テストでは反復学習群の成績がよいが、遅延テストでは反復テスト群の成績がよくなるというクロスオーバーの交互作用結果を説明することはできないであろう（例えば、Roediger & Karpicke, 2006b）。

3-2 検索仮説

検索仮説とは、検索時の様々な取り組みによる処理の困難さや努力がテスト効果を生み出すとするものである。検索時に学習内容を想起しようとして努力を要したり、検索の困難の水準がそれほど容易ではなく、かつ適切な望ましさの困難さによって検索が可能であるとき、テスト効果が認められるとする（Gardiner, Craik, & Bleasdale, 1973; Jacoby, 1978）。例えば、先述の Jacoby (1978) の研究例では、反応項を生成する条件群のテスト効果が顕著であった。また、Gardiner et al. (1973) は実験参加者に一般知識問題をたずね、回答に要する時間を測定した。実験のセッションの終わりに、実験参加者に回答についての最終自由再生テストを実施した。その結果、質問に対して回答に時間がかかるほど、最終自由再生テストでは多くの回答を再生した。

上記の結果は、反応項を生成したり、回答するのに苦勞して時間をかけた場合、検索時に様々な努力を要して反応項や回答を復元することとなる。また、反応項を生成したり、回答に時間をかけることは、記憶痕跡の精緻化を高め、検索のルートを増やすことに結びつくのかもしれない。更に、項目の生成ないし回答の再生を必要とするテストが、再認や単語同定を含むテストよりもテスト効果が大きいことが示されている（Roediger & Karpicke, 2006a）。このような結果を説明するために、再生テストが再認テストよりも検索努力ないしは深い処理の水準を必要とするからであるといってもよい。検索仮説はこれらの内容を含むテスト効果の説明仮説といえる。

3-3 転移適切性処理仮説

転移適切性処理仮説は、エピソード記憶の研究においてよく知られている仮説である（Morris, Bransford, & Franks, 1977; Tulving, 1983）。転移

適切性処理仮説の基本的な概念は、学習課題の後の記憶成績を規定するのは、学習時の処理の深さよりも学習とテストの関係性であるとするものである。たとえ学習の符号化処理が浅くとも（例えば、音韻的処理）、検索時に浅い水準（音韻的処理）に関連するテストが要求されるとき、深い水準（意味的処理）に関連するテストが要求されるよりもよい成績を示す。換言すると、転移適切性処理仮説は、記憶テストの成績は先行学習中に従事した符号化処理と検索処理がマッチする程度に恩恵を受けることに言及する仮説である。転移適切性処理仮説に類似の概念が符号化と検索過程の関係性を強調する符号化特定性原理（encoding specificity principle、多鹿, 1989; Tulving, 1983）である。

では、転移適切性処理仮説によってテスト効果をどのように説明するのか。学習時に記銘材料を符号化処理しその内容をテストすることを1つの作業サイクルに捉えると、反復テスト群では、この作業サイクルでテストされることと後の遅延最終テストで実施されるテストとは、処理も意味内容も一致する場合が一般的である。これに対して、学習時に学習試行のみを反復することを1つの作業サイクルとして項目処理する反復学習群では、通常は1回程度のテストを実施されたとしても、遅延最終テストとの一致は前者の一致に比べて強くないだろう。それ故、反復テスト群では、学習時の作業サイクルで何度もテストを実施することによって、同一タイプの最終テストへ適切に転移がなされ、結果的にテスト効果を生み出すといえる。

なお、転移適切性処理仮説は、テスト効果のみを説明するために提案された仮説ではない。検索仮説と併用するなどして、得られた結果に適用すべきものといえる。

4 結論

1998年に告示された学習指導要領は、学校週5日制のもと、「ゆとり」の中で「特色ある教育」の展開を目指し、児童・生徒に「生きる力」を育成することを基本とした。しかしながら、その基本的なねらいを達成するために、各教科の授業時数が削減された。その方針の実施に呼応するかのよう、学力（低下）論争が生じてきた。既述のように、2007年4月実施の全国学力テストはこのような背景から実施されたものである。また、2007年に公表さ

れた次の新しい学習指導要領では、ゆとり教育による学力低下の反省に立ち、各教科の授業時数が現行の1割程度増加されるようである。

本論文は、学力低下の原因の1つとされる授業時数の削減、換言すれば学習時間の削減が、本当に学力の低下を招いた主因の1つであるのかどうかに関係する問題に深くかかわるテーマを扱ったものである。学力の定義は多様であり、一義的に学力の定義を規定することはできない。しかしながら、学力の主要な側面を学習した結果と捉えるとき、学力テストは、学習成果の定着の確認のために実施されるテスト(学習内容の保持や理解のテスト)として、前述したように、学習内容の単なる事実の記憶や事実の記憶以上の知識を測定するものと考えられる。

本論文は、認知心理学の記憶研究の成果から、学習内容の長期の保持に関して、学習とテストに同程度の時間をかけた場合、学力テストの一部を構成する学習した内容を確認するためのテストを繰り返す方が、学習内容の長期の定着にはよいことを文献展望によって示した。

これらの文献展望では、削減された授業時間を以前の授業時間までもどすことが、学力テストの成績の向上に寄与するかどうかを吟味したものではない。同じ授業時間が設定されている場合、授業を繰り返して一度の学力テストで授業内容を確認するよりも、ある授業を終えれば確認のテストを繰り返す方が(ある意味では形成的テストと呼ぶことができるが、テストの工夫が必要である)、学習内容の長期の定着にはよいようであることを示したに過ぎない。また、学習内容も単語レベルの材料が中心で(例外もあるが)、授業で提示される教材の内容ではない。しかしながら、記憶研究で得られた結果は、授業時間の利用の仕方に関して、ある種の方向性を示したといえるであろう。今後は、学習とテストの更なる吟味が必要である。

5 引用文献

- Allen, G.A., Mahler, W.A., & Estes, W.K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8, 463-470.
- Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49, 415-445.
- Anderson, M.C., Bjork, R.A., & Bjork, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063-1087.
- Dempster, F.N. (1996). Distributing and managing the conditions of encoding and practice. In E.L. Bjork & R.A. Bjork (Eds.), *Memory* (pp.317-344). San Diego, CA: Academic Press.
- Gardiner, J.M., Craik, F.I.M., & Bleasdale, F.A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, 1, 213-216.
- 橋本重治 (1956). テスト効果について 教育心理学研究, 4, 96-101.
- 橋本重治 (1959). テストが再学習の成果に及ぼす影響 教育心理学研究, 7, 90-95.
- 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. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17, 649-667.
- Karpicke, J.D., & Roediger, H.L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57, 151-162.
- 小柳恭治・石川信一・大久保幸朗・石井栄助 (1960). 日本語三音節名詞の熟知価 心理学研究, 30, 357-365.
- McDaniel, M.A., Roediger, H.L., III, & McDermott, K.B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14, 200-206
- 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—教育実践上の問題点と解決のための視点— 教育心理学研究, 54, 265-279.
- Remmers, H.H., & Remmers, E.M. (1926). The negative suggestion effect on true-false examination questions. *Journal of Educational Psychology*, 17, 52-56.
- Roediger, H.L., III. (1978). Recall as a self-limiting process. *Memory & Cognition*, 6, 54-63.
- Roediger, H.L., III, & 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., III, & Karpicke, J.D. (2006b). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.
- Roediger, H.L., III, & Marsh, E.J. (2005). The positive and negative consequence of multiple-choice

- testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1155-1159.
- Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, 12, 43-50.
- Slamecka, N.J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, 76, 504-513.
- Spitzer, H.F. (1939). Studies in retention. *Journal of Educational Psychology*, 30, 641-656.
- Tajika, H. (1986). Influences of the encoding instruction on retrieval processes in recall and recognition memory. *Psychologia*, 29, 240-246.
- 多鹿秀継 (1989). 記憶の検索過程に関する研究 風間書房
- Thompson, C.P., Wenger, S.K., & Bartling, C.A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 210-221.
- Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, 69, 344-354.
- Tulving, E. (1967). The effects of presentation and recall of material in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175-184.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press. (太田信夫 (訳))
- (1985). タルヴィングの記憶理論 教育出版
- 山森光陽・荘島宏二郎 (編) (2006). 学力ーいま、そしてこれからー ミネルヴァ書房