

事前テストと記憶定着

—対連合学習を用いたプレテスト効果の検討—

田中紗枝子・宮谷真人

Pretest and enhancement of subsequent learning:
Pretesting effects in paired-associate learning

Saeko Tanaka and Makoto Miyatani

テストに記憶を促進させる効果があることが古くから知られている。従来は、正しく解答できた場合にのみ効果があるとされていたが、近年になって誤情報にも正情報の学習を促進する役割があることが明らかになっており、これは“プレテスト効果”と呼ばれている。プレテスト効果が個人の特性によりどのような影響を受けるかについて検討された研究はまだないが、その他の記憶に関わる研究の結果から、ワーキングメモリなどの個人特性との関連が考えられる。本研究では、今後個人差要因を検討するのに適切な刺激語リスト、および実験の手続き等を決定するため、先行研究と同様の手続きにしたがい、日本語刺激を用いても同様のプレテスト効果が確認できるかどうかを検討した。その結果、有意なプレテスト効果が確認でき、またディストラクタ課題として行った暗算課題の成績と No-Pretest 群の再生成績の間に正の相関があった。このことから、本研究で用いた刺激と手続きがプレテスト効果を検討するのに適切なものであることが確認できた。また、プレテスト効果とワーキングメモリ容量の間に関連がある可能性が示唆された。

キーワード：記憶、プレテスト効果、対連合学習、ワーキングメモリ

問 題

学校現場において、テストは授業内容がどれだけ定着しているかという評価のために行われ、また記憶研究においても、再生テストや再認テストがパフォーマンス評価のための指標として利用されている。しかし、テストには学習到達度の評価という側面だけではなく、最終的な評価テストの前に事前テストを行うことによって、記憶すべき材料を複数回学習（再学習）するよりも記憶の定着を促進する働きもあることが知られており、この現象は“テスト効果”の名で古くから研究がなされている（遠藤，2007；多鹿，2008）。

では、事前テストにおいて正答した場合と誤答だった場合とでは、どちらがより記憶を促進させるのだろうか。従来は、事前テストにおける誤答は正答（評価テスト時に解答すべき刺激）と競合を

起こし、その結果記憶を抑制してしまうため、事前テストでの正答時にのみ記憶の促進効果があるといわれてきた (Metcalf & Kornell, 2007)。その理由として、errorless 学習の存在が挙げられる。Errorless 学習はその名の通り、エラーを最小化するように学習をいくつかの段階に細分化して行う方法である。この学習方法は、古くは行動心理学的アプローチにおいて動物を対象として行われてきたものであるが、現在でも記憶障害の患者に対するリハビリテーションの分野で大きな成果を上げている (Baddeley & Willson, 1994; 三村, 1998; レビューとして Clare & Jones, 2008)。

しかし近年、事前テスト時に誤った解答をする、つまり誤情報の想起が、正しい情報の学習を促進させる効果があることが分かってきた。Kornell, Hays, & Bjork (2009) は、意味的に弱い関連のある2つの単語 (手がかり単語と、評価テスト時に想起すべきターゲット単語、例: tide と beach) の対連合学習課題を用いて実験を行い、誤情報想起による“プレテスト効果”を報告している。この研究では、実験参加者は Pretest (単語の学習前に事前テストを行う; 誤情報あり) 群と No-Pretest (事前テストは行わず、単語学習のみを行う; 誤情報なし) 群のいずれかに分けられ、対連合学習の後、手がかり単語をもとにターゲット単語を再生する評価テストを行った。その際、Pretest 群は評価テストの対象となる単語対の学習の前に、手がかり単語から連想される単語 (手がかり単語と意味的に強い関連のある単語、例: wave) を1つ生成する事前テストを行ったが、評価テスト時は生成した単語ではなく、手がかり単語をもとにターゲット単語 (この場合、tide から beach) のみを報告するよう求められた。その結果、Pretest 群は No-Pretest 群よりも評価テストにおける再生成績が有意に高かった。Kornell et al. (2009) の研究は刺激として単語対を使用したものであるが、同様の結果は、散文を読みその内容についてのテストを行った研究 (Richland, Kornell, & Kao, 2009) や、大学の講義内容に関する事前テストを行った群と行わなかった群の評価テスト成績について比較した研究 (Butler & Roediger, 2007) でも得られている。

ではなぜ、このような事前テストでの誤情報の想起が記憶を促進するのだろうか。それを説明するための仮説として、これまでに error correction 仮説、search set 仮説、additional-cue 仮説の3つが提案されている。1つ目の error correction 仮説では、学習者が、生成した誤情報と正情報との間の不一致を感じ取ることにより、学習が強化されると想定されており、例えば高確信度問題に対する過剰修正効果などが、これにもとづいて生起していると考えられている (Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011)。2つ目の search set 仮説では、事前テストでの手がかりの呈示が、ターゲットとなる単語の候補が含まれる検索ネットワークを活性化し、学習者はそのネットワークから想起すべき情報を引き出しているため、記憶の促進効果がみられるとされている。また3つ目の additional-cue 仮説は、誤情報検索時の学習者の反応が、後の評価テスト時に“もう一つの手がかり”として想起されるため、プレテスト効果が生起すると考えている (Grimaldi & Karpicke, 2012)。Grimaldi & Karpicke (2012) は、Kornell et al. (2009) の研究の追試を様々な刺激を用いて行い、これら3つの仮説を検討した。その結果、プレテスト効果は、(1) 手がかり単語とターゲット単語の間に意味的な関連のある場合のみ生起し、(2) 事前テストにおいて誤情報を自由に生成した場合に最も効果が大きくなり、逆に生成すべき誤情報を実験者が潜在的に指定すると、むしろ干渉が起こるため再生成績は悪くなり、(3) 事前テストの直後に単語学習が行われなければ、記憶の促進効果は得られ

なかった。もし error correction 仮説が正しければ、手がかり単語とターゲット単語の間に関連がなくとも参加者は正情報と誤情報の不一致を感じるはずなので、プレテスト効果が生じるはずである。また additional-cue 仮説に基づけば、生成する単語を実験者が指定した場合でもその指定した単語が手がかりとなり、記憶の促進効果が得られるはずである。そのため、上記の結果はプレテスト効果が“手がかりの呈示がターゲット単語を含む検索ネットワークを活性化する”という search set 仮説を支持するものであると結論付けている。

では、どの人にとっても、つまり個人の特性によらずプレテスト効果は出現するのだろうか。誤情報の検索が正情報の記憶を促進するというプレテスト効果について、影響を及ぼす学習者の個人差について直接検討した研究は見当たらないが、プレテスト効果以外の記憶に関わる研究においては、ワーキングメモリ容量の個人差と記憶成績の間に関連があることが知られている。ワーキングメモリとは、情報の処理と保持を同時に行う認知システムであり、言語処理や推論、問題解決などの高次認知活動の成績と関係があることから、これらの認知機能を支える機構であると考えられている (芋阪, 2006; 上田, 2009)。例えば、虚記憶に関する研究では、ワーキングメモリ容量の大きい人ほどソースモニタリング能力も優れており、虚再生や虚再認が起きにくいことが、様々なパラダイムを用いて示されている (Leding, 2012)。また、ワーキングメモリ容量を測定するためのスパンテストを遂行する際に使用する方略の個人差が、スパンテスト成績に影響を与え、容量の低い人ほど、効果的な方略が利用できない事が知られている (遠藤・芋阪, 2012; 齊藤・三宅, 2000)。

以上のような知見を踏まえれば、プレテスト効果にもワーキングメモリの容量が関連していると考えられる。そこで本研究では、今後個人差の要因を検討するのに適切な刺激語リスト、および実験の手続き等を決定するため、日本語の刺激を用いて Kornell et al. (2009) と同様の実験を行った。Kornell et al. (2009) の研究では、Nelson, McEvoy, & Schreiber (1998) の連想強度 (ある単語から、別の特定の単語を連想する人の割合) を参考に、手がかり単語とターゲット単語を選択していた。本研究でも先行研究同様の連想強度について調査している水野 (2011) を参考に、刺激となる単語を選択し、実験を行った。また、Kornell et al. (2009) の研究は、事前テストと評価テスト時の単語解答をキーボードからの入力によって行っていたが、日本語では変換の作業が必要となるため、実験中呈示する画面はそのままに、口頭で解答させる方法に変更した。

方 法

実験参加者 大学生、および大学院生 24 名 (うち男性 9 名、女性 15 名) が実験に参加した。年齢は 19—25 歳 (平均 20.1 歳, $SD=1.5$) であった。このうち、11 名 (うち男性 4 名) を Pretest 群、13 名 (うち男性 5 名) を No-Pretest 群にランダムに振り分けた。参加者には実験の内容を口頭で説明した後、書面による同意を得た。

刺激 水野 (2011) の中から、手がかり単語として 3 モーラの単語を、ターゲット単語として 2—4 モーラの単語を使用した。単語の表記は漢字、ひらがな、カタカナが含まれていた。

Kornell et al. (2009) の先行研究と同様、連想強度が .045—0.054 であるような低連想強度の手がかり単語とターゲット単語を 60 組 (120 単語) 使用した。実験で使用した単語を、Table 1 に示した。

Table 1

本研究の対連合学習課題で使した単語

手がかり単語	ターゲット単語	モーラ数	連想強度	1位連想語	1位の連想強度	連想強度 5%以上の 単語数
赤字	借金	4	.054	黒字	.185	4
悪魔	地獄	3	.054	天使	.313	3
移民	難民	4	.054	アメリカ	.115	3
汚染	川	2	.054	環境	.281	4
親子	関係	4	.051	家族	.147	5
価格	商品	4	.054	高騰	.166	4
科学	進歩	3	.051	理科	.169	3
火災	炎	3	.051	火	.179	5
家庭	円満	4	.051	家族	.288	4
看護	ナース	3	.054	病院	.358	3
気温	湿度	3	.051	上昇	.118	5
機械	精密	3	.051	工場	.144	4
着物	帯	2	.051	日本	.188	5
苦勞	労働	4	.054	仕事	.163	4
検査	健康	4	.051	病院	.160	4
個人	情報	4	.054	ひとり	.102	4
古典	昔	3	.051	国語	.243	3
自然	森林	4	.051	森	.198	6
時代	戦国	4	.054	歴史	.144	3
世界	国	2	.051	地球	.185	3
石油	資源	3	.054	高騰	.166	4
手紙	メール	3	.051	紙	.083	5
電話	会話	3	.051	携帯	.319	3
うどん	きつね	3	.053	麵	.141	6
おとこ	筋肉	4	.050	女	.353	2
かかと	くつ	2	.050	足	.575	2
かばん	通学	4	.050	学校	.156	5
きずな	仲間	3	.053	友達	.303	3
くるま	免許	3	.053	タイヤ	.091	3
こころ	からだ	3	.050	心臓	.081	8
ごはん	おかず	3	.050	米	.181	5
さかな	うろこ	3	.053	海	.372	2
しるし	マーク	3	.050	ミスチル	.116	5
そうじ	掃除機	4	.053	ほうき	.384	2
たばこ	害	2	.050	けむり	.281	2
たんず	収納	4	.050	服	.256	6
つくえ	教室	4	.050	勉強	.284	5
とうふ	豆	2	.053	白	.306	3
におい	花	2	.053	香水	.194	3
はさみ	刃物	3	.053	紙	.263	5
はんこ	名前	3	.053	印鑑	.247	5
ひずみ	地震	3	.050	ゆがみ	.144	2
ふとん	夜	2	.050	睡眠	.203	5
まつげ	化粧	3	.050	目	.241	7
まぶた	瞳	3	.053	目	.550	2
アルミ	銀色	4	.050	缶	.206	5
カエル	雨	2	.053	みどり	.299	5
キムチ	赤	2	.053	韓国	.615	2
ゲスト	招待	4	.050	客	.226	4
コラム	本	2	.053	新聞	.266	3
ジャズ	サックス	4	.050	音楽	.571	3
シルク	高級	4	.053	箱	.259	4
スリル	映画	3	.053	恐怖	.163	5
テスト	期末	3	.053	試験	.209	4
ドラマ	俳優	4	.053	テレビ	.329	4
パーマ	髪の色	4	.050	髪	.199	5
パズル	ゲーム	3	.053	ピース	.143	2
ビール	大人	3	.050	酒	.216	5
マイク	音声	4	.053	カラオケ	.193	5
モグラ	地中	3	.053	土	.389	4

課題 本実験で用いた単語学習課題は、学習フェーズ、ディストラクタ課題、評価テストの3つで構成した。

学習フェーズでは、Pretest 群は学習する単語の事前テストと、手がかり単語とターゲット単語の対連合学習を、No-Pretest 群は手がかり単語とターゲット単語の対連合学習のみを行った。Pretest 群の学習フェーズはまず、パソコンのディスプレイ上に手がかり単語と枠が7秒間表示され、その手がかり単語からターゲット単語になるような単語（以下“生成単語”）を連想して、口頭で解答した（事前テスト）。7秒経つと枠が消え、枠が消えた500ミリ秒後にさらに画面が切り替わり、正しいターゲット単語が手がかり単語とともに5秒間表示された。参加者は5秒の間に、先ほどの生成単語ではなく、ターゲット単語を手がかり単語と対にして覚えるよう指示された（対連合学習）。その後、500ミリ秒のISIを挟んで、次の手がかり単語に対する事前テストと対連合学習を行った。

一方のNo-Pretest 群の学習フェーズでは、Pretest 群と異なり対連合学習のみを行った。内容は全く同じものであった。パソコンのディスプレイ上に手がかり単語とターゲット単語が5秒間表示され、それらを対にして覚えた。その後、500ミリ秒のISIを挟んで、次の対連合学習を行った。なお、単語対の呈示順序は参加者ごとにランダムであった。

ディストラクタ課題では暗算課題を5分間行った。パソコンのディスプレイ上に足し算、引き算、掛け算、割り算いずれかの問題と枠が表示され、参加者はその解答を、テンキーを用いて画面上の枠の中に入力した。問題は3位数同士の足し算、3位数引く2位数、もしくは3位数引く3位数の引き算、2位数掛ける1位数の掛け算、整数商で割り切れる3位数割る2位数の割り算、全300問からランダムに出題した。

評価テストでは、Pretest 群もNo-Pretest 群も同一の手がかり再生課題を行った。ディスプレイ上に手がかり単語と枠が7秒間表示され、参加者はその間に手がかり単語と対になっていたターゲット単語を思い出し、口頭で解答した。7秒後、500ミリ秒のISIを挟んで次の手がかり単語に対する再生を行った。なお、評価テストにおいても手がかり単語の呈示順序は参加者ごとにランダムであった。

手続き まず全ての参加者に、これから単語対の学習を行い、後でどれだけ覚えているかテストを行うという実験の目的を説明した。単語学習課題の前に、Pretest 群もNo-Pretest 群も本課題とは関係のない手がかり単語とターゲット単語10組を用いて、生成単語を解答したり、正しいターゲット単語を覚えたりする練習を行った。ここでの成績は分析では使用しなかった。その後、Pretest 群は事前テストと対連合学習からなる学習フェーズを、No-Pretest 群は対連合学習のみの学習フェーズを行った。

単語対60組の学習がすべて終わったのち、5分間のディストラクタ課題を行った。この課題では暗算を行うことが目的であったため、計算用紙などは用意しなかった。参加者はできるだけ速く正確に、5分間計算問題を解き続けるよう指示された。

ディストラクタ課題を行った後、評価テストを行った。学習フェーズで使用した60組の単語対すべての再生が終わったところで実験を終了した。実験の所要時間は、Pretest 群が35~40分程度、No-Pretest 群が25~30分程度であった。

結 果

Pretest 群において、事前テスト時にターゲット単語を連想した（生成単語とターゲット単語が同一であった）割合は 6.0%であった。先行研究に倣い、ターゲット単語を連想した手がかり単語については、その参加者の分析から除外した。

事前テストが評価テストでの再生成績に及ぼす効果について検討するため、Pretest 群と No-Pretest 群の間で平均値に差が見られるかどうか、対応のない t 検定を行った。その結果、Pretest 群の方が No-Pretest 群より有意に成績が高かった ($t(20.18)=2.64, p<.05$, Table 2)。

Table 2
Pretest 群と No-Pretest 群の再生成績と t 検定の結果

	平均	SD	t 値	p
Pretest 群	0.88	0.11	2.64	<.05
No-Pretest 群	0.72	0.18		

また、Pretest 群に比べて No-Pretest 群の再生成績の標準偏差が大きくなっていった。No-Pretest 群の成績は、Pretest 群と同程度の成績である参加者もいれば、低い参加者もいた (Pretest 群の成績範囲.96—1.60; No-Pretest 群の成績範囲.95—1.43)。そこで、No-Pretest 群における成績のばらつきが何に起因するのか、つまりプレテスト効果に影響を及ぼす個人内の要因について検討するため、ディストラクタ課題の成績（5 分間での解答数、正答数、正答率）と評価テストでの再生成績の関連について、学習フェーズ時の群ごとにピアソンの積率相関係数を算出した。その結果、No-Pretest 群のディストラクタ課題の正答率と再生成績の間に中程度の正の相関が見られた (Table 3, Figure 1)。そこで、学習フェーズ時の群ごとに正答率が中央値より高い者を高群、低い者を低群に分割し、評価テストにおける再生成績 (Table 4) について、学習フェーズ時の群 (Pretest 群, No-Pretest 群) × 正答率 (高群, 低群) の 2 要因分散分析を行った。なお、正答率がちょうど中央値にあたる参加者は分析に含めなかった。その結果、群の主効果のみ有意 ($F(1, 17)=6.28, p<.05$) であり、正答率の主効果と交互作用は有意ではなかった。

Table 3
評価テストの再生成績とディストラクタ課題の成績の相関

	解答数	正答数	正答率
Pretest 群	.05	.13	.38
No-Pretest 群	.19	.32	.49 †

†... $p <.10$

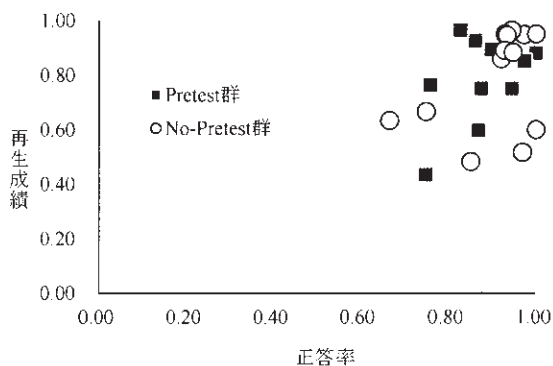


Figure 1. ディストラクタ課題の正答率と評価テストの再生成績の散布図

Table 4

学習フェーズ時の群とディストラクタ課題の正答率別の再生成績

	正答率高群		正答率低群	
	平均	SD	平均	SD
Pretest群	0.93	0.04	0.76	0.19
No-Pretest群	0.83	0.15	0.65	0.19

ディストラクタ課題について、最後に解いた問題の解答時間が、それまでの平均解答時間より3倍以上長くなっている参加者が1名 (No-Pretest群) いた。このことは、“ディストラクタ課題として、暗算を5分間解き続ける”という活動が十分なされておらず、分析対象とすることが不適切である可能性がある。そこで、すべての参加者を対象とした場合と同じ結果が得られるかどうか確認するため、この参加者1名を分析から除外したデータについて同様の分析を行った。

まずプレテスト効果について、Pretest群とNo-Pretest群の間で平均値に差が見られるかどうか、対応のない*t*検定を行った。その結果、Pretest群 (Table 2を参照)の方がNo-Pretest群 (平均: 0.84, SD: 0.17, $t(20.18)=2.27, p<.05$)より再生成績が高いという上述の分析と同様の結果が得られた。

また、ディストラクタ課題の成績と再生成績の関連について、ピアソンの積率相関係数を算出した。その結果、Table 3とは少し異なりNo-Pretest群のディストラクタ課題の正答数と再生成績の間に中程度の正の相関が見られた (解答数: .48, 正答数: .52†, 正答率: .48, †: $p<.10$)。そこで、学習フェーズ時の群ごとに正答数が中央値より多い者を高群、少ない者を低群に分割し、評価テストにおける再生成績について学習フェーズ時の群 (Pretest群, No-Pretest群)×正答数 (高群, 低群)の2要因分散分析を行った。その結果、群の主効果のみが有意 ($F(1, 18)=9.59, p<.01$)であり、正答数の主効果と交互作用は有意でなかった。ディストラクタ課題の解答時間から分析対象とすることが不適切である可能性がある参加者を除いた分析でも、全参加者を分析した場合と同様の結果が得られた。

考 察

本研究の目的は、Kornell et al. (2009) などの先行研究の結果が、日本語の刺激を用いても再現されるかどうかを確認し、今後プレテスト効果の個人差について検討する際に用いる刺激や手続きを決定することであった。Pretest 群の方が No-Pretest 群より評価テストにおける再生成績が高く、先行研究同様、事前テストを行ったことにより記憶の定着が促進されていた。本研究で用いた日本語の刺激は、プレテスト効果を検討するのに適切な刺激であったといえるだろう。

また、No-Pretest 群において、評価テストの再生成績と、ディストラクタ課題 (暗算) の成績の間に中程度の正の相関がみられ、ディストラクタ課題の成績がよい人ほど、再生成績もよいことがわかった。暗算の遂行は、ワーキングメモリ容量が深くかかわっているといわれている (齊藤・三宅, 2000) ことから、評価テストの成績と暗算課題の成績の関連性について、次のように推測できる。ディストラクタ課題 (暗算) の成績が高い、つまりワーキングメモリ容量の大きい参加者は、事前テストの有無にかかわらず評価テストの成績は高い。一方、ワーキングメモリ容量が小さい参加者は、事前テストを体験しないと容量の大きい参加者に比べて成績が低くなるが、事前テストという形で記憶する際の方略が示されることにより、容量の大きい参加者と同程度にまで成績が向上したと考えられる。

事実、No-Pretest 群で成績の良かった参加者数人に、単語対を記憶する際にどのような方略を使用していたかを尋ねたところ、手がかり単語とターゲット単語で1つの熟語を作る (“個人-情報” など)、手がかり単語とターゲット単語の意味の類似性を利用する (“はさみ-刃物” など)、手がかり語をカタカナ化する (“看護-ナース” など) など、呈示された単語対をグループ化し、評価テスト時の手がかりとしていたことが分かった。このことは遠藤・荻阪 (2012) で示されているとおり、ワーキングメモリ容量の大きい人ほど、記憶に効果的な方略を自ら選択し、使用することができることを表していると考えられる。

本研究では、ワーキングメモリ容量の個人差がプレテスト効果に及ぼす影響について直接的な検討を行っていない。プレテスト効果を学校現場など、実際の教育場面に応用する際には、個人差の問題を十分に考慮する必要がある。そのため、ワーキングメモリ容量などの個人差がプレテスト効果に及ぼす影響について、今後詳細に検討していく必要があるだろう。

引用文献

- Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: Amnesia and problem of error elimination. *Neuropsychologia*, *32*, 53-68.
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514-527.
- Clare, L., & Jones, R. S. P. (2008). Errorless learning in the rehabilitation of memory impairment: A critical review. *Neuropsychological Review*, *18*, 1-23.
- 遠藤香織・荻阪満里子 (2012). 日本語版リーディングスパンテストにおける方略利用の個人差 心理学研究, *82*, 554-559.

- 遠藤正雄 (2007). 再生課題によるテスト効果 近畿福祉大学紀要, **8**, 37-41.
- Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, **4**, 505-513.
- Kang, S. H. K., Pashler, H., Cepeda, N., J. Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, **103**, 48-59.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**, 989-998.
- Leding, J. K. (2012). Working memory predicts the rejection of false memories. *Memory & Cognition*, **20**, 217-223.
- Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, **14**, 225-229.
- 三村 将 (1998). 記憶障害のリハビリテーション —間違っただけがおぼえやすいか？ 努力した方がおぼえやすいか？— 失語症研究, **18**, 136-145.
- 水野りか (編) (2011). 連想語頻度表—3 モーラの漢字・ひらがな・カタカナ表記語— ナカニシヤ出版
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://w3.usf.edu/FreeAssociation/>.
- 芋阪満里子 (2006). ワーキングメモリ (1) 入門講座 臨床脳波, **48**, 691-696.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, **15**, 20-27.
- 齊藤 智・三宅 晶 (2000). リーディングスパン・テストをめぐる 6 つの仮説の比較検討 心理学評論, **43**, 387-410.
- 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 親和女子大学大学院研究紀要, **4**, 57-65.
- 土田幸男 (2009). ワーキングメモリ容量とは何か？ —個人差と認知パフォーマンスへの影響— 北海道大学大学院教育学研究院紀要, **109**, 81-92.