

論文

テストへの反応形式がテスト効果に及ぼす影響:

手書きとタイピングの比較¹

The response format for the testing effect:

Handwriting vs. Keyboard typing

高橋 功¹⁾・岩木 信喜²⁾

Isao Takahashi, Nobuyoshi Iwaki

キーワード: 記憶, テスト効果, 対連合課題, 手がかり再生, 手書き, タイピング

Keywords: memory, testing effect, paired-associate learning, cued recall, handwriting, keyboard typing

要旨: テストへの反応形式が手書き (HW: handwriting) の場合とタイピング (KB: keyboard typing) の場合とでの, テスト効果への影響を比較した。大学生 36 名を対象とし, カタカナの対連合課題を学習課題とした。原学習後の学習を参加者内で操作し, HW による initial test (HW 条件), KB による initial test (KB 条件), 再学習 (RO 条件: read only) を行わせた。1 週間後の final test は参加者間で 2 群に分け, HW による final test (HW' 群) もしくは KB による final test (KB' 群) を実施した。final test の正答率は, KB' 群では両条件ともに RO 条件より正答率が高く同程度であった (i.e., 同程度のテスト効果が両条件で確認された)。一方, HW' 群では HW 条件 (i.e., テストへの反応形式が initial-final 間で一致) の正答率のみ RO 条件より有意に高く, KB 条件 (i.e., 不一致) の正答率は RO 条件より高かったものの統計上の有意差が曖昧であった。全体として条件差が見られなかったことは, テスト効果が主に内的な想起に起因し, 外的反応がもたらす影響は付加的なものに過ぎないという解釈を支持すると考えられた。しかし, initial test と final test が不一致の場合に KB 群のみ効果が曖昧になったことは, 転移場面で外的反応の影響が露呈するという点かもしれず, 外的反応の想起への深い関りも否定しきれないと考えられた。再現性を確認するために, final test も参加者内で操作した追試が必要と考えられた。理論的な検討のためには, 難易度の高い学習課題, covert 反応も含めた比較が必要と考えられた。

¹⁾ 山陽学園大学総合人間学部生活心理学科

²⁾ 聖学院大学人文学部児童学科

問題と目的

テスト効果

一般に、教育現場で実施されるテストは、測定を主目的としており、学習活動そのものと見なされないことが多い。例えば、大学の教職課程のテキストで、テストに焦点が当てられるのは評価の文脈においてである (e.g., 遠藤, 2014)。「指導と評価の一体化」という言葉も、主として評価活動と指導改善や学習者自身の自己評価との一体化に言及したものであり、テストはあくまでも診断やフィードバックのための道具のひとつとして位置づけられている (e.g., 中央教育審議会, 2019; 梶田, 2010)。大学教育においても、「定期試験は授業に含まれない」と解釈されていることが多い (仲井, 2016)。しかしながら、テストには学習活動そのものとしての有用性もあり、その一つに学習内容の保持の促進がある (多鹿, 2008)。

ここでいうテストによる学習内容の保持の促進とは、継続的なテストの実施が学習者の動機づけを高めて結果的に保持が促進されるといった「間接的な効果」によるものではなく、テストを受ける行為自体が記憶の定着を促す「直接的な効果」によるものを指す。このことは認知心理学の領域では古くから知られており (e.g., Spitzer, 1939; Tulving, 1967)、特に近年、よく統制された実験に基づく証拠が提出されている (Roediger & Karpicke, 2006a; Karpicke & Roediger, 2008; Butler, 2010; Karpicke & Blunt, 2011)。この現象は、テスト時の想起過程こそが効果の源泉であるという理解から“想起に基づく学習 (retrieval-based learning)”と呼ばれもするが (Karpicke et al., 2014)、一般的には“テスト効果 (testing effect)”と呼ばれることが多く (Roediger & Karpicke, 2006b)、本稿でも後者を採用する。

テスト効果を示す近年の実験の基本的な手続きは、学習者に学習内容を呈示した後、もう一度それを読ませる場合 (restudy 条件) と、テストを受けさせる場合 (initial test 条件) とで、遅延テスト (final test) の成績を比較するものである。例えば、Roediger & Karpicke (2006a) の実験 1 は、学習の 2 条件 (restudy vs. test) × 遅延テストまでの 3 期間 (5 分, 2 日, 1 週間) の混合計画 (参加者内×参加者間) でデザインされ、大学生 120 名が参加した。最初の 7 分間は、いずれの条件においても参加者に短文 (TOEFL の参考書から引用した 260 語前後の文章) を読ませた。次の 7 分間は、restudy 条件では再び同じものを読ませ、test 条件では読んだ内容について思い出せることをできるだけ多く書かせた。遅延テストでは、10 分間、読んだ内容について思い出せることをできるだけ多く書かせた。短文の内容を 30 単位に分けて遅延テストを採点したところ、5 分後群の正答率は restudy 条件の方が test 条件より高かったが (81% vs. 75%)、2 日後群は逆の結果であり (54% vs. 68%)、更に 1 週間後群ではその差が顕著であった (42% vs. 56%)。学習内容、test の回数や方法、遅延期間の長さなど設定は様々ではあるが、他の研究でも同様の結果、つまり遅延期間の短い場合を除き、final test の成績は test 条件で高くなることがほぼ一貫して報告されている。

この現象に関する有力な説明として、Roediger & Karpicke (2006b) は大きく 2 つの仮説を挙げている。一つは、“検索仮説 (retrieval hypothesis)”であり、テスト学習時の検索の努力によって、記憶痕跡が精緻化されたり検索経路が増えたりするので、後の検索はたとえ流暢でなかったとしても強度が高まるというものである。もう一つは、“転移適切性処理仮説 (transfer-appropriate processing)”であり、転移において重要なのは記録や想起時の処理の深さよりも転移場面との適合性にあるとしたうえで、学習時に行われる認知処理と

記憶の転移場面 (final test) で求められる認知処理について考えたとき, *restudy* 条件ではかなり異なるが, *initial test* 条件では類似しているので, 転移が起こりやすくなるというものである。ただし, これらの 2 つの仮説は相互排他的なものではなく, テスト効果はそれら両方のメカニズムの合成によってもたらされるものであると考えられている。

なお, 近年の多くの実験は, *restudy* 条件を統制群として設定したうえで, 学習時間も条件間で統制しており, *test* 条件でテスト後に正答をフィードバックしない場合でもテスト効果が確認されているので, 単に *test* 条件の方が学習内容に接触する時間や学習に費やす時間が長い, すなわち有利であるという説明は否定的に捉えられている。また, Roediger & Karpicke (2006b) によれば, *test* 条件の学習者が学習内容の一部のみを選択的に学習した結果であるという説明も, そうした項目選択効果を除去した分析によって否定されている。

テストへの反応形式とテスト効果

テスト効果に関する問題の一つに, テストへの「反応形式 (response format) の問題」がある。これは, 選択肢問題か自由再生かといった「テスト形式 (test format) の問題」ではなく, 回答を口に出した方がよいのか, 書いた方がよいのかといった問題である。学習活動としてのテストの教室実践への導入を考えたとき, これは興味深い問題のひとつである。

実のところ, この問題については, 種々の反応形式の比較以前の問題として, そもそもそういう外在化 (*overt*) がよいのか, それとも頭の中で思い出す, つまり内的な処理だけ (*covert*) でよいのかという「*overt vs. covert* の問題」が議論されている。そしてその実証研究において, *covert* 反応でも十分なテスト効果が見られたという実験結果が既に報告されている (e.g., Putnam & Roediger, 2013; Sundqvist et al., 2017)。このため, *overt* 反応はテスト効果を少し補強するだけの付加的なものに過ぎないと見なす見解が示されている。

反応を外在化させる *overt* 反応と, 「思い出した」という報告を求めただけの *covert* 反応の比較は, 実験手続き上難しいところがあり, 特に *covert* 反応で学習者がどの程度明確に思い出したか知り得ない点が難しい。つまり, *covert* 反応でテスト効果が見られなかった場合, *initial test* 時の想起の不備に帰して説明できてしまう。*covert* 反応にテスト効果が見られず, *overt* 反応にのみ見られたことを示す報告もあるが (e.g., Whitten & Bjork, 1977; Jönsson et al., 2013), それらはこの点で *overt* 反応の優位性に疑問を残すものであった。

このため, Sundqvist et al. (2017) は, *initial test* において, 想起できたことを宣言してから *overt* 反応を求める (あるいは求めない) という巧みな手続きを用いた実験を行った (詳細は本稿で後述)。その結果, *final test* の成績は, *covert* 反応でも統制群 (*restudy*) より十分高く, *overt* 反応は更に高かったものの効果量が小さかった。Sundqvist et al. (2017) は, この結果と *overt* 反応の優位性を示す先行研究でも効果量の小さいこと (彼らによれば平均 $d=0.07$) を根拠に, テスト効果は主として想起過程によってもたらされるものであり, *overt* 反応は明瞭化によって小しだけそこに恩恵を与えるに過ぎないと結論づけている。

しかしながら, 記銘段階における産出効果 (MacLeod & Bodner, 2017) の恩恵, 空書行動 (佐々木・渡辺, 1983) のような外顯的行動の認知処理への恩恵の存在を考えると, *overt* 反応が, 想起後に想起内容を明瞭化したり単に視覚情報や音韻情報を加算したりするだけではなく, 想起過程そのものにも深く関わり検索経路の豊富化に重要な役割を果たしていると期待するのはさほど不合理ではない。*covert* 反応で十分なテスト効果が得られること

は確かだとしても、overt 反応の優位性については検討の余地があるのではないだろうか。

そしてその余地のひとつに、手書き (handwriting <以下, “HW” と表記>) による initial test の検討が挙げられる。というのも、overt vs. covert の問題に焦点があったためか、テストへの反応形式に焦点を当てたテスト効果の実験では、HW による initial test がそれほど多く採用されていないのである。すなわち、その多くはタイピング (keyboard typing <以下, “KB” と表記>) か口頭報告であった。Sundqvist et al. (2017) が実験 1 と 2 で HW による initial test を採用しているものの、この実験には手続き上の疑問があった (後述)。

また、転移適切性処理仮説の観点で考えれば、final test の反応形式も HW と KB を用いて比較する必要があるが、先行研究の多くは 1 つの形式のみで行われており、その多くも KB であった。Putnam & Roediger (2013) が final test の形式を操作しているものの、用いられたのは KB と口頭報告であった。この点でも HW について検討しておく余地がある。

更にいえば、HW は歴史的伝統のある方法であり、教育場面でも重要な学習方法として推奨されてきたものである。実際、近年でも、講義を聞いてまとめるような場面では HW が KB よりも有益であることを示す実験結果が報告されている (Mueller & Oppenheimer, 2014)。また、特に日本では、子どもの頃から書いて覚えることが推奨されている (Naka & Naoi, 1995)。したがって、テスト効果においても、HW について詳細に検討しておくことは、仮にその効果が KB と同程度であったとしても、その確認に一定の意義があるだろう。

手書き (HW) とタイピング (KB)

HW と KB は、それに伴う運動と認知上の処理を考えると、かなり異なる行為である。その相違について、Mangen et al. (2015) は、「手の受動感覚」、「視覚的注意と感覚運動の統合の程度」、そして「運動制御時に払われる注意」の相違を挙げている。すなわち、KB は両手とすべての指を使うが、HW は特定の手指のみを使用する。また、書いている最中にペン先に注意が向けられる HW は、視覚的注意と感覚運動が時間的・空間的に統合されているが、キーボードとモニタが離れている KB はその統合が失われがちである。そして、HW は運動制御の識別性が高いが、KB のそれは低い (i.e., HW では文字固有の手の動きが求められるが、KB で求められるのは文字に関わらず同じ動きである)。また、通常、大学生以上であれば、HW よりも KB の方が速い (Mueller & Oppenheimer, 2014) ことも、運動と認知上の処理の違いといえる。すなわち、KB は HW よりも行為に費やされる時間が短い。

各行為への慣れの程度も関わる問題ではあるが、現代の一般的な成人を想定すると、総じて HW は KB より多くの身体的・心的労力を要する行為であるといえる。そうした特徴づけには議論の余地もあるが、HW の書字文化がより便利なものとしての KB に代わってきたことは事実であろう。文化庁 (2015) によれば、世代を問わず 9 割前後の者が手書きの習慣を「大切にすべき」とした一方で、実際に日常生活で手書きをする機会については「よくある」が 38.2%、「時々ある」が 34.4%であり、変化を読み取ることができる。その変化は、デジタル化された情報の利便性によるところが大きいであろうが、42.0%の者が情報機器の普及の影響として「手で字を書くことが面倒くさく感じるようになった」としており (文化庁, 2012)、労力の軽減も読み取ることができる。経験的にも、より「楽な」KB 中心の書字文化の中で、多くの人々が HW ではあり得ないような不注意な誤字・脱字に悩まされている。

では、そうした労力は記憶過程にどう関わるのだろうか。これは一概には言えない。それ

は記憶痕跡や検索経路を豊富にしているかもしれないし、身体や認知のコストとなって、本来の符号化や検索を毀損しているかもしれない。しかし少なくとも記銘段階においては、HW の優位性を示す実験結果が報告されている (Smoker et al., 2009; Mangen et al., 2015)。

Mangen et al. (2015) は、36 名の大学生を対象に、簡単な動詞や名詞で構成された 84 (各条件 28) の単語リストを学習課題とした実験を行った。これらの単語はすべてヘッドフォンからの音声によって継時呈示された。音声を聞いた後のメモの取り方が参加者内で操作され、HW 条件、KB 条件、そしてタッチキーボード条件が設けられた。妨害課題後の事後テストは、自由再生もしくは再認課題の形式で行われ、いずれも実験者との口頭のやりとりで行われた。その結果、再認課題の成績は条件間に差がなかったが、自由再生の成績は HW 条件が他条件より高かった (KB との差の効果量 $r = .37$)。この結果について、Mangen et al. (2015) は、視覚的注意がモニタとキーボードを行き来してしまう KB では、運動と視覚情報の連合が弱くなり、それゆえ単語の心的表象も弱まったのではないかと考察している。

記銘時にこうした違いが生じるのであれば、想起時においても同様のメカニズムにより、KB による反応では記憶の心的表象が弱まっていると考えられる。KB による overt 反応の成績が covert 反応の場合とさほど変わらなかった先行研究の結果もこの点から説明できる。

本研究の目的

そこで本研究は、テストへの反応形式が HW の場合と KB の場合とでの、テスト効果への影響を比較検討した。もしも initial test 時の overt 反応が想起後の付加的なものに過ぎないのであれば、final test の成績は HW でも KB でも同程度と予測される。すなわち、視覚情報の加算についていえば、どちらの反応でも紙面上あるいはモニタ上で得られるのだから、キーボードとモニタの間で目線が行き来する KB に多少の不利があるものの、時間的な差は僅かであろうから、HW が優位だったとしても効果量は極めて小さいと考えられる。しかし、initial test 時の overt 反応が想起過程にも深く関わるものであれば、より大きな労力を要する HW の成績は、検索経路をより豊富化するため、その優位性が大きな効果量となって現れるだろう。また、initial test と final test の反応形式が一致する場合に成績が高くなると考えられるが、それは心的表象の強い HW でより顕著になると予測される。

ここで、同じく HW と KB が initial test で採用されたものの、HW の優位性が見られなかった Sundqvist et al. (2017) の実験の問題点について述べておく必要がある。Sundqvist et al. (2017) では、実施された 4 実験のうち、最初の実験 1 と実験 2 で HW と KB が採用された。ここでは実験 1 の詳細を述べたうえで、その問題点と本実験での改善点を述べる。

この実験は、学習の 4 条件 (reading vs. covert vs. KB vs. HW) × 遅延テストまでの 2 期間 (5 分, 1 週間) の参加者内計画でデザインされた。大学生 32 名が参加し、cue と target からなる 48 対の対連合課題 (連想価 .02) が学習課題として用いられた。最初、48 対の課題が 6s 間隔で継時呈示され、3 回繰り返された (原学習)。その後の学習は 1 課題につき 12s であり、条件ごとの指示がランダムに与えられた。すなわち reading 条件では cue と target が再呈示された。他の 3 条件では、cue のみが呈示され、参加者は最初の 5s 以内に target を想起することが求められた。参加者は想起できたら ENTER キーを押下した。時間内に押下がなければ次の課題が呈示された。押下されると、残りの時間内 (e.g., 3s で想起の報告となる押下があれば残りの 9s) に条件に応じた反応が求められた。すなわち、covert 条

件では何もせずに残りの時間を待った。KB 条件では想起した **target** をタイピングした。HW 条件では手元の用紙に **target** を記入した。**final test** は、5 分後と 1 週間後の 2 回に分けて行われ、順次呈示される **cue** に対する **target** を 15s 以内にタイピング (i.e., KB) した。

なお、実験 2 もほぼ同じ手続きであり、相違は学習回数を 3 回に増やした点、最初の原学習の呈示時間を 5s にした点であった。結果は両実験とも概ね同じであり、1 週間後の **final test** の成績は、reading 条件よりも他の 3 条件の方が高く (i.e., テスト効果)、reading 以外の 3 条件間に有意な違いはなかった (i.e., covert 反応だけでもテスト効果が得られる)。

この実験は巧みなものではあるが、HW の優位性の有無を検討する上では、手続きに問題があると考えられる。第一に、先に想起の完了を宣言させてから反応を求めた点である。この手続きは、covert 反応を操作する実験としては大変洗練されたものである。というのも、参加者は overt 反応が求められるかもしれないという前提で想起できたことを報告しなければならないので、不明瞭な想起を報告しにくくなるからである。しかし一方、このような手続きは overt 反応の再現という点では歪である。すなわち、人が記憶にあるものを表出する際、通常そうしたメタ認知的な行為を切り分けてリソースを割かない。この歪さが、overt 反応、特により労力を要する HW が検索経路にもたらす恩恵を毀損した可能性がある。

第二に、初期テストの同型性が十分ではない点である。というのも、基本的な実験の進行は PC を用いて行われていたため、当然のことながら **cue** は常にモニタ上に呈示されていたのである。参加者はそれを見て HW で紙に書いたり KB で入力したりするわけであるが、そうすると HW 条件の場合、手元のシートとモニタの両方に視覚的注意を行き来させ、キーボードとペンを持ち替える必要が生じる。加えて、与えられた 12s 以内にモニタに注意を戻さなければならず、認知的負担となる。こうした状況は、HW に期待される統合的な処理を大きく損なった可能性がある。実際、Sundqvist et al. (2017) も、実験 3 以降で HW を採用しなかった理由にこの忙しさを挙げている。また、モニタ上で作業が完結する KB 条件の方が、反応で生成された **target** の視覚情報の加算時間が必然的に長くなり有利になる。

第三に、**final test** が KB でのみ行われている点である。転移適切性処理仮説の観点で考えれば、これは明らかに KB に有利で HW に不利である。

第四に、2 要因 (4×2) の参加者内計画でありながら、用いられた対連合記憶課題が 48 対と少なかった点である。つまり、1 つの条件に割り当てられた単語が 6 対しかなく、必然的に成績のばらつきが大きくなる。実際、1 週間後の **final test** の平均正答率と *SD* を見ると、実験 1 の場合、HW 条件が .50 (0.31)、KB 条件が .47 (0.32)、実験 2 の場合、HW 条件が .55 (0.30)、KB 条件が .58 (0.27) であり (括弧内が *SD*)、*SD* が比較的大きい。

そこで本実験においては、これらの問題点を踏まえた手続きを採用した。第一の問題点については、本実験では HW と KB の比較のみを行い、covert 反応を求めないため、想起と反応を分ける手続きは必要とせず、問題は生じない。第二の問題点については、HW 条件においては、**cue** をモニタ上に呈示するのではなく、問題冊子を用いて用紙に呈示、すなわち完全なペーパーテスト形式とした。第三の問題点については、**final test** にも HW と KB を被験者間要因として設定した。第四の問題点については、課題を 1 条件につき 12 対とした。

なお、本実験においては、**initial test** の再生率を向上させるため、**initial test** 時は、**target** の頭文字 1~2 文字をヒントとして呈示した。これは、**initial test** 時の十分な再生に失敗し、床効果が生じることでテスト効果が検出されにくくなることを危惧したためであった。

方法

実験参加者

18-22 歳 ($M = 20.4$, $SD = 1.27$) の私立 S 大学の学生 36 名 (男性 13 名, 女性 23 名) が実験に参加した。各参加者には謝礼として 1000 円分のプリペイドカードを進呈した。実験は、2018 年 7~11 月の期間に、大学構内の教室において、1~4 名の小集団単位で実施した。

倫理的配慮

いずれの実験参加者も書面により参加に同意した。同意書には、実験内容、所用時間、プライバシーの保護、結果の公表、同意の撤回方法に関する説明が記されており、参加者自身が署名した。研究内容は山陽学園大学研究倫理審査委員会の承認を受けた (平 30 大 009)。

実験計画

学習後の初期テスト (initial test) の反応形式 3 条件 (HW, KB, RO; 参加者内) × 1 週間後の遅延テスト (final test) の反応形式 2 群 (HW', KB'; 参加者間) の混合計画であった。なお、initial test の RO 条件とは、test ではなく学習課題を再読するだけ (read only) の統制条件 (i.e., restudy) であった。参加者間変数は各群 18 名であった。

学習課題

弱い連想関係にある手がかり語 (cue) とターゲット語 (target) からなる対連合課題 36 対を課題とした (Appendix 1)。単語は、水野 (2011) の連想価表のうち、連想価が .02~.04 ($M = .029$, $SD = .006$) のものから選出した。cue は全てカタカナ 3 文字とした。target は、initial test 時にヒントとして頭文字を最大 2 文字まで呈示するため、カタカナ 4-6 文字とした。選出においては、カタカナ表記によって意味が伝わりにくくなるかどうかを実験協力者 3 名で協議し、問題がないとされた単語に限定した。また、長音と母音の混同による間違い (e.g., “ボーリング” と “ボウリング”), あるいはキーボードの打ち間違いを避けるため、長音の「ー」が登場するものも target から除外した。36 対の中に単語の重複はなかった。課題は、初期テストの各条件に 12 対ずつ参加者単位で無作為に割り当てた。

実験手続き

実験は、学習セッションと一週間後の final test セッションの 2 回に分けて実施した。HW 条件の initial test および HW' 群の final test 以外は全てデスクトップ PC と外付けキーボードを使用し、モニタ上に学習課題等を呈示した。実験教示は口頭で述べるとともに画面上にも呈示した。KB においては、日本語入力モードのローマ字変換を使用させた。

学習セッション 最初に一週間後の final test の実施を予告した。手続きは条件ごとにブロック化して 1 ブロックずつ実施した。各ブロックは、原学習フェイズ、妨害課題、initial test フェイズで構成された。3 条件の実施順序は参加者間でカウンターバランスをとった。

原学習フェイズでは、最初に「左の単語を見て、右の単語を思い出せるように覚えて下さい」と教示したうえで、画面中央部の左に cue, 右に target を呈示する形式で 12 対の学習課題を継時呈示した (Figure 1.)。呈示時間は 1 対につき 7s, 呈示間隔は 1s とした。

妨害課題は、順次呈示される 4 枚の顔写真から笑顔のものを早く発見して高得点を狙うゲームのような課題であり、2 分間繰り返させた。4 枚の顔写真のうち 3 枚は怒りの表情、1 枚が笑顔の表情であった。参加者には、できるだけ早く笑顔の写真を見つけ、該当する番号をキーボードで入力することを求めた。入力後に正解・不正解のフィードバックを与え、次の課題を呈示した。顔写真は、Lyons et al. (1998) が作成した database を利用した。

続く initial test フェイズは、条件により手続きが異なったが、低正答率による床効果を回避するため、いずれの条件も手続きを 2 セットずつ繰り返した。

HW 条件では、表紙を含む 13 枚綴りの A4 冊子を配布した (Figure 1)。冊子の各頁は、学習課題呈示時の PC モニタの画面を模してデザインされており、中央部の左方に cue のみを示し、右の target 部は空欄とした。空欄の上には target の頭文字 2 文字 (1 セット目) もしくは 1 文字 (2 セット目) をヒントとして記載した。参加者には各頁 10s の制限時間内に target を思い出して記入することを求めた。制限時間は事前に作成した合成音声の音源を用いて統制し、記入の有無に関わらず、10s のタイミングで次の頁を捲る指示を出した。

KB 条件では、画面中央部の左に cue のみを呈示し、右の target 部は空欄を 2 つ呈示した (Figure 1)。上の空欄は入力用のものであり、入力中の内容が呈示された。ENTER 押下により入力を確定させると、上の空欄の内容が消去され、下の空欄に確定内容が表示された。HW 条件と同様に、空欄の上には target の頭文字 2 文字 (1 セット目) もしくは 1 文字 (2 セット目) をヒントとして呈示した。参加者には 10s の制限時間内に target を思い出してタイピングすることを求めた。制限時間は PC のプログラム上で制御し、入力の有無に関わらず、10s 経過後に自動的に次の課題を呈示した。

RO 条件では、原学習フェイズ時と同様に画面上に学習課題を継時呈示した。ただし、呈示時間は他条件の initial test の時間と同じ 10s とした。

なお、いずれの条件においても、課題の呈示順序は、テストごとに無作為化し直した。**final test セッション** final test は学習セッションから一週間後に実施した。final test は initial test 時の条件に関わらず 36 の cue すべてをプールし無作為な順序で呈示した。回答制限時間は各課題 12s とした。target の頭文字のヒントは与えなかった。それら以外は、HW' 群は HW 条件、KB' 群は KB 条件の各 initial test と同一の手続きであった。

最後に 1 分間のタイピングテストを 2 回実施した。これは画面上に順次呈示されるカタカナ 3 文字の単語をできるだけ早くその通りに回答欄に打ち込んでいくものであった。

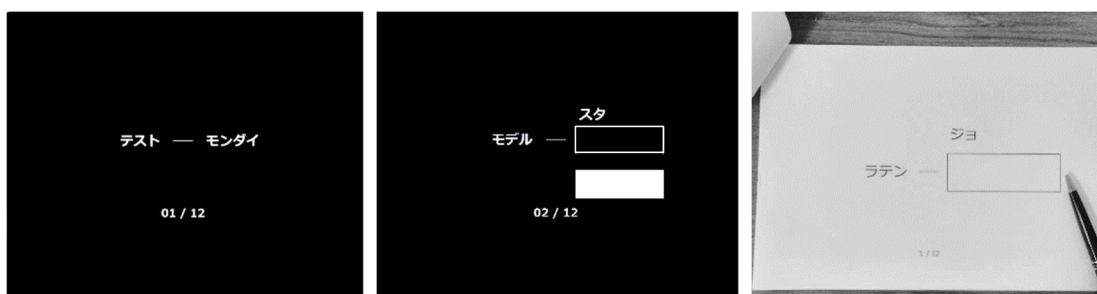


Figure 1. 原学習フェイズにおける学習課題呈示画面 (図左), KB 条件の initial test (1 セット目) の想起課題呈示画面 (図中), HW 条件の initial test (1 セット目) で用いた冊子 (図右),

結果

test の採点方法

initial test および final test の採点は、HW のものも KB のものも全て実験協力者 3 名の協議を通して行った。一般に、本研究のように平易なカタカナ語を用いた実験でそうした手続きを経ることは少ないと思われるが、これは KB による回答の過小評価を避けるためであった。というのも、KB の回答の中には正しい単語が入力されているにも関わらず、余計な文字が入ってしまっているものが一部に見受けられたからである。多くは ENTER キーの隣のキーの文字 (“j”) や余分なスペースが含まれているケースであった。こうした回答は、target の単語が含まれている限りすべて正答とした。また、HW において、ひらがなで回答したものが一部に見受けられたが、カタカナの正答と一致している限り正答とした。

なお、KB' 群の 1 名は、2 回目の KB 条件の initial test において 12 問中 11 問の回答を確定させておらず、記録がとられていなかった。このため、この参加者のこの部分の測定値には、同群、同回、同条件の initial test の平均正答数 (11.06) を充当して分析を行った。

タイピングスコア

タイピング能力の群間での等質性を確認するために、各群のタイピングスコアを算出した。算出にあたっては、参加者ごとに正しく入力された単語数をカウントし、1 回目と 2 回目で数の多かった方をタイピングスコアとして採用した。スコアの群間差は見られず ($t(30) = 1.23, n.s.$)、タイピング能力は群間で等質であったことが確認された (Table 1)。

なお、タイピングテストは 1 分間にカタカナ 3 文字の単語をできるだけ多く入力するものであったので、文字換算すると全体平均で 1 分間あたり 44.7 文字となる。測定方法が異なるので安易な比較はできないが、文部科学省 (2017) によれば、高校生の 1 分間あたりの文字入力数は平均 24.7 字であり、これと比較する限り十分高い。また、日本商工会議所が行うビジネスキーボード検定試験では、10 分間で 400~599 字の入力が合格の基準となっており、この基準から考えても、参加者のタイピングスキルは概ね十分であったと考えられる。

Table 1
各群のタイピングスコア

	<i>n</i>	<i>Min</i>	<i>Max</i>	<i>M</i> ^a
HW' 群	18	9	23	14.0 (0.89)
KB' 群	18	9	29	15.9 (1.25)
全体	36	9	29	14.9 (0.77)

^a 括弧内は SEM。

initial test の正答率

initial test の成績の群間・条件間での等質性を確認するために、各群、各回、各条件の initial test の平均正答率を算出した (Table 2)。3 要因の混合分散分析を行ったところ、群と回の交互作用が有意であった ($F(1, 34) = 12.61, p < .01, \eta^2 < .01$)。このため下位検定を行ったところ、HW' 群の回の単純主効果が見られ ($F(1, 34) = 22.90, p < .01, \text{partial } \eta^2 < .01$)、

1回目よりも2回目の initial test の正答率が高かった。

回による差は効果量が極めて小さく、無視できるものと考えられる。いずれにせよ、少なくとも2回目の initial test に群間、条件間の差が見られなかったことから、群間、条件間は概ね等質であったと考えられる。また、全体的に正答率が高く、とりわけ2回目の initial test では9割を超えており、テスト効果の検証の前提としては十分であったと考えられる。

Table 2
initial test の平均正答率^a

	n	1回目		2回目	
		HW条件	KB条件	HW条件	KB条件
HW'群	18	.90 (.025)	.86 (.034)	.92 (.021)	.93 (.019)
KB'群	18 ^b	.93 (.030)	.92 (.027)	.93 (.028)	.92 (.033)
全体	36	.92 (.019)	.89 (.022)	.92 (.017)	.93 (.019)

^a 括弧内は SEM。^b KB'群の2回目のKB条件における1件の欠損値には、同群、同回、同条件の平均値を充当した。

final test の正答率

テストへの反応形式のテスト効果への影響を検討するために、各群、各条件の final test の最終テストの平均正答率を算出した (Table 3, Figure 2.)。2要因の混合分散分析を行ったところ、条件の主効果が見られた ($F(2, 68) = 20.43, p < .01, \eta^2 = .18$)。Bonferroni 法により調整された有意水準 ($p < .0167$) を用いて多重比較を行ったところ、HW条件とKB条件はRO条件よりも有意に大きく、HW条件とKB条件の間に有意な差は見られなかった。

分散分析による交互作用は確認されなかったものの、HW'群のKB条件の平均正答率が低く、条件の主効果の効果量が小さかった。そこで事後分析として、群ごとに条件間で平均正答率の一対比較を行った (Table 4)。その結果、KB'群ではHW条件 (49%) とKB条件 (51%) がともにRO条件 (28%) よりも有意に大きかったが、HW'群ではHW条件 (52%) のみRO条件 (28%) よりも有意に大きく (i.e., KB条件のテスト効果が見られなかった)、HW条件 (52%) のほうがKB条件 (40%) よりも有意に大きかった。HW'群におけるHW条件とKB条件 (52% vs. 40%) の差の効果量は $dD = 0.54$ であり中程度の大きさであった。

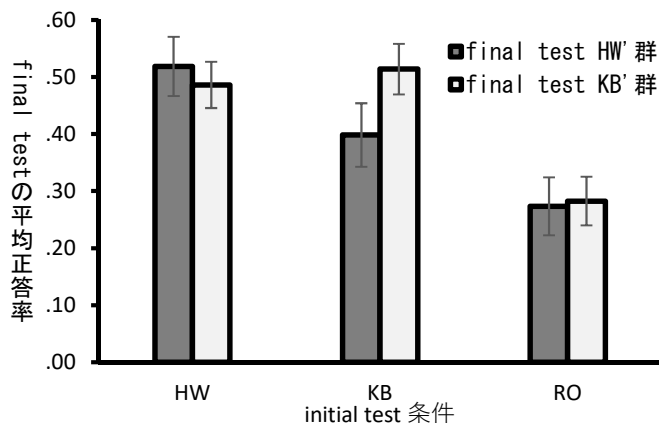


Figure 2. final test の平均正答率

Table 3
final test の平均正答率^a

	<i>n</i>	HW 条件	KB 条件	RO 条件
HW' 群	18	.52 (.052)	.40 (.056)	.27 (.051)
KB' 群	18	.49 (.041)	.51 (.044)	.28 (.043)
全体	36	.50 (.033)	.46 (.036)	.28 (.033)

^a 括弧内は SEM。Table 4
群ごとの final test の平均正答率の一对比較

final test	<i>n</i>	initial test	<i>M</i>	<i>t</i>	<i>p</i>	<i>dD</i>
HW' 群	18	HW vs. KB	52% vs. 40%	2.30	.034*	0.54
		HW vs. RO	40% vs. 27%	3.77	.002**	0.89
		KB vs. RO	52% vs. 27%	2.07	.054	0.49
KB' 群	18	HW vs. KB	49% vs. 51%	0.57	.579	0.13
		HW vs. RO	49% vs. 28%	4.96	.000**	1.17
		KB vs. RO	51% vs. 28%	5.46	.000**	1.29

タイピングスコアと正答率の相関

タイピング能力による影響について検討するため、条件ごと、群ごとにタイピングスコアと各テストの正答率について、Pearson の積率相関係数をそれぞれ算出した。その結果、initial test においても final test においても有意な相関は確認されなかった。唯一、KB' 群の HW 条件においてのみ、やや正の相関 ($r = .42$, $p = .08$) の傾向がみられた。

少なくとも本実験では、タイピング能力による影響は皆無であったと考えられる。

考察

テストへの反応形式が HW の場合と KB の場合とでの、テスト効果への影響を実験比較した。もしも initial test 時の overt 反応が想起後の付加的なものに過ぎないのであれば、HW 条件と KB 条件の final test の成績は同程度になると予測した。しかし、テストへの overt 反応が想起過程にも深く関わるのであれば、より大きな身体的・認知的労力を要する HW 条件の成績が KB 条件を上回ると予測した。それは更に final test と initial test の反応形式が一致する場合、すなわち HW 群の HW 条件で顕著になると予測した。

結果を見ると、全体としては final test の成績に条件間の差がなかった。この結果は、HW 条件に求められる大きな労力が検索経路を豊富化するという仮説を棄却し、Sundqvist et al. (2017) の主張を支持するものであった。initial test は想起こそが重要であり、反応形式はさほどの問題ではないのかもしれない。ただし、天井効果の可能性も否めないで、課題の難易度をあげる、final test までの遅延期間を長くする事態での検証に検討の余地がある。

一方、final test と initial test の反応形式の一致・不一致については、興味深い結果が見られた。すなわち、final test の反応形式が KB' の群では、initial test の反応形式に関わらず明確なテスト効果が確認されたが、final test の反応形式が HW' の群においては、initial test が KB のときのテスト効果が曖昧なものとなった。HW' 群の HW 条件の成績は KB'

群の KB 条件を超えるようなものではなかったため、この結果は、HW' 群の HW 条件に高い効果が見られたというよりは、HW' 群の KB 条件の効果が低かったと見るべきであろう。

これはどのように説明できるだろうか。実験で用いられた学習課題は難解な漢字等ではなく 6 文字までのカタカナであるから、initial test がタイピングであったために書き方自体が分からなくなったということは考えにくい。むしろ、この結果は、KB を用いた検索特有の心的表象の弱さが転移場面において露呈したことを示唆するものであろう。その理由としては、KB における運動の識別性の低さが挙げられる。逆にいえば、HW における運動の識別性がテスト効果に寄与しているともいえる。すなわち、HW によって検索を練習した場合、目の前に紙とペンがなくても運動に関わる記憶痕跡を使うことができるが、KB によって検索を練習した場合はその痕跡を使うことができない。しかし目の前にモニタとキーボードがあればその他の記憶痕跡を使うことができる。この説明は、空書行動のタイピング版のような「空打行動」が一般的に行われにくいこととも合致する。ただし、本実験の KB' 群と HW' 群は参加者間変数であったため、参加者内での再現を確認する必要があるだろう。

ところで、英語圏以外の文化特有の事情となるが、指摘しておかなければならない KB の重要な特徴がある。すなわち、日本では「ローマ字入力」による変換を用いる場合、入力時に押すキーは、出力される日本語の文字そのものを表すキーではなく、アルファベットのキーであることである。更に、このため内的には先行して出力させたい日本語をローマ字に変換しなければならない。これは HW にはない KB 固有の認知処理である。もちろん、「かな入力」という入力方法もある。しかし、詳細が公表されていない通俗的な web 調査ではあるが、9 割以上の者が「ローマ字入力」を用いていることを示した記事がある（遠藤, 2015）。また、学習指導要領（文部科学省, 2016）においても「ローマ字入力」の指導が想定されている。そして実際、本稿の実験はすべて「ローマ字入力」を用いて行われた。

KB のこの特徴がテスト効果にどう影響したかは一概には言えない。KB で求められる運動の識別性の低さを考えれば、キーの種類は検索経路の豊富化にさほど寄与しないだろうから、キーと出力させる文字の不一致については影響を想定しにくい。しかし、先行する認知処理上のローマ字変換は KB にのみ求められる心的努力であり、それは検索経路の豊富化に寄与しているかもしれないし、認知的負担となり検索を毀損しているのかもしれない。

もし豊富化への寄与があったとすれば、本実験において KB 条件と HW 条件で同等のテスト効果が見られたことは、それぞれ固有の検索経路の豊富化が行われたためであると説明することも可能となる。ただし、この考え方では、先行研究で示された covert 反応と KB の成績の同等性を説明できない。一方、ローマ字変換にそのような寄与はなく、認知的負担でしかないと考えれば、それにより KB の成績は毀損されているということになる。であれば、KB による overt 反応には検索経路の豊富化への寄与と認知的負担による毀損が混在し相殺が起きているという可能性も考え得る。こうした考え方の妥当性を論じるためには、HW 条件でもローマ字変換を求めるか、数値や英単語など日本語を用いない学習材料を用いた検討、そして何よりも covert 反応も含めた比較実験を実施しなければならない。

利益相反

本論文に関し、開示すべき利益相反関連事項はない。

引用文献

- 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.
- 文化庁 (2012). 平成 23 年度 国語に関する世論調査 [平成 24 年 2 月調査] Retrieved from https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/kokugo_yoronchosa/pdf/92701201_08.pdf (2022 年 12 月 1 日)
- 文化庁 (2015). 平成 26 年度 国語に関する世論調査 [平成 27 年 1 月調査] Retrieved from https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/kokugo_yoronchosa/pdf/92701201_05.pdf (2022 年 12 月 1 日)
- 中央教育審議会 (2019). 児童生徒の学習評価の在り方について (報告) Retrieved from https://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/afieldfile/2019/04/17/1415602_1_1_1.pdf (2022 年 12 月 1 日)
- 遠藤 諭 (2015). 日本人は“大人”になるとローマ字入力になるらしい by 遠藤諭
Retrieved from <https://weekly.ascii.jp/elem/000/002/631/2631699> (2022 年 12 月 1 日)
- 遠藤 司 (編) (2014). 新・教育課程シリーズ 教育心理学 一藝社
- Jönsson, F. U., Kubik, V., Sundqvist, M. L., Todorov, I., and Jonsson, B. (2013). How crucial is the response format for the testing effect? *Psychological Research*, *78*, 623–633.
- Karpicke, J.D., & Blunt, J.R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772–775.
- Karpicke, J.D., Lehman, M., & Aue, W.R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, *61*, 237–284.
- 梶田 叡一 (2010). 教育評価 (第 2 版補訂 2 版) 有斐閣双書
- Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. In Proceedings Third IEEE international conference on automatic face and gesture recognition, pp. 200–205.
- MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*, 390–395.
- Mangen, A., Anda, L.G., Oxenburgh, G.H., & Brønneck, K. 2015 Handwriting versus keyboard writing Effect on word recall. *Journal of Writing Research*, *7*, 227–247.
- 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
- 文部科学省 (2016). 小学校学習指導要領 (平成 29 年告示) Retrieved from <https://erid.nier.go.jp/files/COFS/h29e/index.htm> (2022 年 12 月 1 日)
- 文部科学省 (2017). 情報活用能力調査 (高等学校) 調査結果 Retrieved from https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/_icsFiles/afieldfile/2017/01/18/1381046_02_1.pdf (2022 年 12 月 1 日)

- Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological science*, *25*, 1159-1168.
- 仲井 邦佳 (2016). 大学の単位制度と学年歴－「1 単位=45 時間」と「1 科目=1350 分説 (15 週論)」－ 立命館産業社会論集, *51*, 1-11.
- Naka, M., & Naoi, H. (1995). The effect of repeated writing on memory. *Memory & cognition*, *23*, 201-212.
- 日本商工会議所 ビジネスキーボード試験方法と技能の評価 Retrieved from <https://www.kentei.ne.jp/bus> (2022 年 12 月 1 日)
- Putnam, A.L., & Roediger, H.L., III (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition*, *41*, 36-48.
- Roediger, H.L., III, & Karpicke, J.D. (2006a). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249-255.
- Roediger, H.L., III, & Karpicke, J.D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181-210.
- Smoker, T. J., Murphy, C. E., & Rockwell, A. K. (2009). Comparing Memory for Handwriting versus Typing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*, 1744-1747.
- Spitzer, H.F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641-656.
- Sundqvist M.L., Mäntylä, T., & Jönsson, F.U. (2017) Assessing Boundary Conditions of the Testing Effect On the Relative Efficacy of Covert vs. Overt Retrieval. *Frontiers in Psychology*, 1-15.
- 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. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175-184.
- Whitten II, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 465-478.

Appendix 1

学習課題として用いた対連合課題と連想価

cue	target	連想価	cue	target	連想価
カジノ	スロット	.0266	ノルマ	エイギョウ	.0332
カルテ	シンダン	.0233	パイプ	ゲスイドウ	.0233
クイズ	シツモン	.0299	バトン	リクジョウ	.0299
グラム	タイジュウ	.0299	バナナ	フィリピン	.0233
グルメ	レストラン	.0233	ピアノ	エンソウ	.0365
コイン	ウラオモテ	.0266	ブラシ	カミノケ	.0399
コンビ	アイカタ	.0332	ベンチ	カップル	.0266
シャツ	ネクタイ	.0233	ボトル	イレモノ	.0233
シルク	チュウゴク	.0266	マイク	オンキョウ	.0233
スパイ	テイサツ	.0299	ミイラ	ガイコツ	.0266
タオル	センタク	.0332	ミシン	カテイカ	.0399
テスト	モンダイ	.0399	メダカ	スイソウ	.0266
テント	サンカク	.0266	メロン	キミドリ	.0233
ドラマ	レンアイ	.0365	モデル	スタイル	.0365
ドレス	シンデレラ	.0399	ラテン	ジョウネツ	.0266
ナイフ	サバイバル	.0233	ランプ	アラジン	.0332
ネオン	ハンカガイ	.0299	ワイン	コウキュウ	.0233
ノイズ	オンガク	.0299	ワルツ	クラシック	.0233

¹ 本研究は、日本学術振興会より科学研究費助成事業（基盤研究（C）課題番号 16K04331）の助成を受けた。内容の一部は、日本教育心理学会第 61 回総会における高橋 功・岩木 信喜のポスター発表、山陽学園大学における伊勢 和史、弘中 宏輝の各卒業論文で公表された。