

# テストの測定精度を飛躍的に高める新しい測定技術 - 任天堂 DS 用英単語学習ソフトによる実証研究 -

寺澤孝文<sup>1</sup>

高木伸也<sup>2</sup>

寺前謙治<sup>3</sup>

原 健二<sup>4</sup>

<sup>1</sup>岡山大学 教育学部

<sup>2</sup>(株)D3 パブリッシャー

<sup>3</sup>ヴァンテアンシステムズ(株)

<sup>4</sup>(株)リクルート テクノロジーマネジメント開発室

**Key words : 動的テスト法、連続測定、学習用ソフト、マイクロステップ技術、形成的評価**

従来のテスト法は、少なくとも2つの点で、学力の推定精度に限界を有している。本論では、その限界を越える新たな学習効果の測定技術を紹介する。またその方法を、ゲーム端末で展開する英単語学習に適用し、学習効果を測定した実験結果を紹介する。

## テスト項目の抽出誤差の問題

従来のテスト法では、たくさんある学習内容から一部を取り出して、短時間でテストをし、その成績から全体の到達度を推定している。それゆえ、易しい問題がたまたま取り出されれば、成績は高くなるなど、抽出誤差が大きいという問題を有している。また、一部の学習内容から学力を推定している限り、個々の学習内容ごとに学習者の到達度を推定することは難しい。学習者にとっては、本来、具体的にどの問題を勉強する必要があるのかという指針が重要な情報になるが、そのような指針は現在の方法では提供できない。

## 学習とテストのインターバルの問題

テストの成績は、学習からどのくらい経過しているのかによって大きく影響を受けるが、これまでのテスト法では、その点は全く考慮されていない。前日に勉強した内容や、1ヵ月、1年前などに学習した内容が、学習の時期などは全く考慮されずにテストが構成されている。当然、インターバルの効果による誤差が現在のテストの成績には混入することになる。

## 学習とテストの一体化

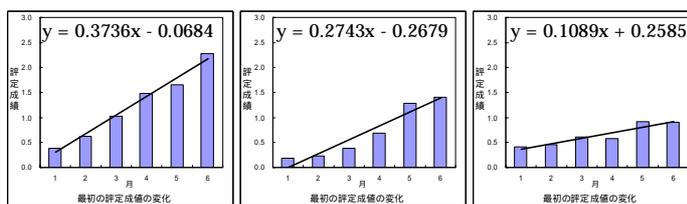
寺澤・太田・吉田は(2007)は、従来のように単発的になされるテストで学習者の到達レベルを推定する方法に代え、連続して実施するテストの成績で、到達レベルを推定する方法を開発した。さらに、テストのみならず、テストを構成する問題項目一つ一つについて、それ以前に、いつ、どのような学習がなされ、どのくらいのインターバルをあけてテストがなされ

るのか、すなわち学習スケジュールを考慮したテストの構成法を確立した。言い換えれば、テストと学習という区別を排し、学習内容の一つ一つに関して、どのようなイベントがいつ生起し、どのような反応データが収集されるのかを、あらかじめ詳細に、また年単位でスケジューリング原理。また、そのスケジュールにしたがってイベント(学習とテスト)を引き起こし、全ての反応データを収集する方法論を確立し、その方法論を具体化した大規模なデータベースシステムを開発した。

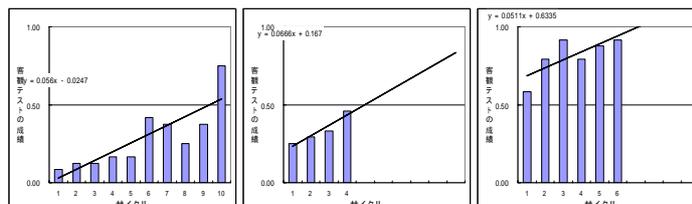
## 到達度の時系列変化を個別に描き出すことが可能に

これまで、上記データベースシステムを、高校生の英単語学習や小学生の漢字の読みのドリル学習へ適用し、長期にわたる学習実験を多数実施してきた(寺澤・太田・吉田[2007]、寺澤[2007]参照)。その結果、英単語学習の到達度の変化(図1)や、小学生の漢字の読み学習の到達度の変化(図2)に、大きな個人差が存在することが明らかになってきた。

学習には積み重ねが大切であると言われてきたが、実際に英単語や漢字のような多数の学習内容について、学習の積み重ねの効果を客観



学習者 A (約8ヵ月) 学習者 B (約1年) 学習者 C (約2年)  
図1 高校生の英単語学習の到達度の変化



児童 A 児童 B 児童 C  
図2 小学生の漢字読み学習の到達度の変化

的に描き出している研究は、世界的にもみられない。さらに、図 1.2 に示したように、その到達度の変化を個別に描き出す研究は、皆無といえる。その意味で、図のような結果が数多く得られている事実は、上記の新しい方法論の有効性を示すものといえる。

### **マイクロステップ計測技術**

上述した新しい測定法は、マイクロステップ計測法と呼んでいるが、そのコアは2つある。その一つは、学習とテストのインターバルや、学習を行うタイミングなど、時間軸上に想定される要因(時間次元の要因)の影響を排除するための新しい実験計画法(種まき法、インターバル相殺法)である。

例えば、説明のため、一つのパズルを紹介する。1000個の英単語をそれぞれ5回ずつ学習し、それから1カ月のインターバルをあけてテストをする場合、従来の方法によれば、ある日に延べ5000個の英単語を学習して、1ヵ月後に1000個についてテストをすることになるが、それでは学習に何時間も要してしまう。したがって、学習を何日かに散らばせる必要がある。ところが、学習を散らばせた場合、テストまでのインターバルがそれぞれ変わってきてしまう。一日の学習時間は短く、なおかつ全ての学習項目について、テストまでのインターバルが等しくなるような方法を考える必要がある。このようなパズルをクリアして、インターバルの影響を排除して学習到達度を正確に測定していく、従来にない新たな測定法(実験計画法)が複数導入されている。

もう一つは、上記の実験計画法に基づき、何万というイベントを、コンピュータにより生起させ、イベントに対応して反応データを収集するために必要な新しいスケジューリング原理とその技術である。一般にスケジュールというなじみがあるが、現在一般に使われているスケジュールの表現方法は、非常に冗長であり、コンピュータの処理には全くなじまないものである。この点は、寺澤・太田・吉田(2007)をご参照いただきたいが、何万というイベントがいつどのように生起するのかを定義し、それに従ってイベントを生起させ、さらに反応データを解析していくための方法論は、これまで一切科学的な研究の対象とされてこなかった。スケジュールの表記方法から、具体的にスケジュールに従ってイベントを生起させ、学習者ごとに反応データを収集・記録する技術は、これまで全く研究されてこなかったといえる。

### **OCR教材と情報端末を活用した学習支援**

寺澤(2007)は、マイクロステップ計測法を、WEBを経由したe-learningに関連付け、公立の小学校において、オーダーメイド的OCRの紙教材と、携帯情報端末(PDA)を利用した漢字読みドリルの学習支援を実現した。すなわち、学習到達度の変化を個別に測定した結果、これまでと同様、非常に高い精度で学習到達度を個別に測定できることが明らかになった。さらに、クラスの児童の成績の変化を一つのエクセルのファイルとしてまとめ、教師を通じてフィードバックする支援も実現した。現在、岡山県内の1つの中学校をモデル校とし、全校規模でこの支援を開始している。学校のドリル学習の支援で、特に注目すべき結果は、到達度データの個別フィードバックにより、児童の自己効力感(学習意欲に重要な影響を及ぼすと考えられている尺度)が上昇することが明らかになってきたことである(寺澤[2007]参照)。

しかし、ここで問題も出ている。インターネットとPDAを利用した学習は、ネットワークの管理・維持が難しく、また、学習のためのインターフェイスもまだ十分とはいえない。また、なにより、インターネット接続ができない家庭の子どもに、同様の支援を提供することはできない。一方、一人ひとりに紙教材を印刷しそのデータをOCRソフトを利用し収集する仕組みは、家庭学習を全員に提供できる反面、教材の印刷とデータ収集にコストがかかることが問題となっている。

### **携帯ゲーム端末を活用した学習支援**

将来的には、自宅にインターネット接続できる環境がない子どもであっても学習ができ、また、容易に操作が可能で、データの収集が容易になる仕組みを、大規模に導入することが理想と考えた。そこで、現在、一般の家庭に広く普及し、画面操作のインターフェイスも非常に充実し、子どもの乱暴な操作にも耐えられる、携帯ゲーム端末に着目した。最近のゲーム端末は、文字認識技術などにより手書きの入力も可能であり、学習の効果を視覚的にわかりやすく提供できる機能を十分持っている。学校でPDAを利用した実験を実施した際は、PDAの充電やインターネットの接続環境の維持が大きな問題となったが、ゲーム端末の場合、個人が学習する状況に限れば、それらの問題はほとんど解決される。本研究は、そこから携帯ゲーム端末をマイクロステップ計測に活用し、個別学習支援を実現することを目指した。

## 2 種類の難易度の区別

多少話は前後するが、テスト項目の難易度は、学習者の学習量に応じて変化していく特徴を持つ。それゆえ、テスト項目の難易度は、本来、学習者の学習履歴に対応して想定できる概念である。しかし、現状では、特定の受験者を想定したテストの成績で難易度を特定する方法が一般に採られている。言うまでもなく、より理想的には、問題の難易度は学習履歴に対応した成績の時系列的な「変化パターン」で表すべきものといえる。例えば、問題の正答率が極端に低い問題であっても、1、2度その問題の解き方を理解することで、それ以降の正答率がいつになっても100%であるような問題もあれば、英単語の意味のように、短時間で覚えて簡単に正解できても、1週間後には思い出せないような学習内容もある。難易度といっても、特定の受験者が問題に正解できる可能性をとらえた「テストのための難易度」と、実質的に解答できるようになるまでにどのくらいの学習を要すのかという、「習得のための難易度」の2種類が想定できる。

従来のテスト法は、テストのための難易度を推定することはできたが、習得のための難易度を推定することはできなかった。後者は、連続的になされるテストの成績を、学習履歴に対応付けて初めて描き出されるものであり、単発的になされるテストでは、描き出すことができないものといえる。本研究では、マイクロステップ計測技術により、習得のための難易度を推定するための基礎データを得ることを目指す。

また、テストのための難易度と、習得のための難易度は対応するのであろうか？本研究では、マイクロステップ計測技術を導入した携帯ゲーム端末用の英単語学習ソフトを開発し、長期学習実験を実施することで、この疑問に対する知見を得ることを目的とする。

以上、本研究の目的は、携帯ゲーム用の英単語学習ソフトを用い、従来同様、学習到達度を個別に描き出せるか否かを検討すること、一般的な英単語の難易度と、英単語の意味を習得するまでの到達度の時系列変化パターンを対応付け、テストのための難易度と、習得のための難易度の関係を検討することにある。

### 方法

**実験参加者:** 岡山大学の学部学生および大学院生17名、民間企業の社員5名。

**学習コンテンツ:** 一般の大学受験に必要とされる英単語の参考書9冊に収録されている英単

語のうち3冊以上で重複して掲載されている2252語の英単語をピックアップした。その英単語について、全国13の国立大学(岡山、埼玉、千葉、鳴教、高知、新潟、山形、信州、東京学芸、大阪教育、茨城、奈良教育、鹿児島)の大学1、2年生計846名を対象に、英単語の熟知度と日本語訳の学習経験について調査を行った(寺澤・太田・吉田[2007]参照)。すなわち、英単語の熟知度評定では「大学入学前ほどの程度目にしたことがあったか」を5段階(目にしなかった~いつも目にした)で評定することが要求された。日本語訳の評定では「高校の頃の勉強ででてきたように思う日本語訳」にチェックするよう教示された。熟知度は、過去の学習経験量を反映することがこれまでの研究で明らかになっている(寺澤, 1995, 1996)。また学習量と難易度は対応すると考えられる。そこで、英単語の難易度の指標として、国立大学に入学した学生が評定した熟知度を用いることにした。また、複数の日本語訳がある英単語の場合、大学生によって、より多くチェックされた日本語訳が、学習時に表示される日本語訳の最初になるよう、コンテンツを作成した。さらに、以下に説明するスケジューリングにあわせるために、複数の参考書で重複した数が少ない英単語で、かつ熟知度が低い単語を除いた、2170語の英単語とその日本語訳のリストが作成された。2170語の英単語を7つの難易度のカテゴリーに分け、それぞれの難易度について、以下で示す学習スケジュールを適用した<sup>1</sup>。

**学習スケジュール:** 実験では、複数の学習スケジュールが設けられ、並列的に学習が進められた。本研究でターゲットとしたスケジュールは、10日を1サイクルとし、1日の学習で1回と4回学習が求められるスケジュール条件であった(寺澤・太田・吉田[2007]のスケジュール表記法によると、イベントユニットとインターバルユニットが10日、条件ユニットが1日、呈示条件が1回と4回繰り返し条件、タイミング条件の表記は[E010\_I010\_J001\_D000])。このスケジュールに、計280個(1回と4回条件それぞれ140個)の英単語が割り振られた。

<sup>1</sup> 最終的に抽出された2170語の英単語の難易度のカテゴリーごとに、TOEICの得点レベルの対応が明記されている3冊の参考書との対応を調べたところ、難易度のカテゴリーレベルとTOEICの得点レベルが対応することが明らかになっている(未発表)。

**手続き**：実験参加者は、自分の英単語力よりも難易度が高いレベルの難易度条件に割り振られた。その結果、上記、7つの難易度のレベルのうち、難易度が高いレベルから5つのレベル（C, B, A, S, SS：SSが最も難易度が高い）が学習対象となった。学習者の英単語力にもともと大きな差があったため、割り振りはランダムではなく、個人が学習を希望するレベルを極力尊重し、それ以外はランダムに割り振りを行った。各自が所有している、任天堂DSもしくはDS Liteを使い、所定のスケジュールで、できる限り1日に1日分の英単語の学習を、40日以上継続することが求められた。都合で学習ができなかった場合は、まとめて複数日分の学習ができるようになっていたが、できる限り1日の間に、2日分を越えて学習をしないように注意が与えられた。単語学習は、単語カード的な学習であり、表示される英単語と日本語の意味を見て、その英単語に対する自分の学習の到達度を「よい」「もう少し」「だめ」「全くだめ」の4段階で自己評定する学習方法であった。自己評定反応が指標とされた。

### 結果および考察

実験参加者のうち1名を除く全員が40日分の学習を完了した。また、多くの学習者が50日分までを完了していたため、結果の分析は50日分を完了した学習者のデータに関して行われた。各難易度に割り振られた参加者の所属（大学/企業）が均質になるよう（それ以外はランダムに）分析対象者が選ばれた。結果として、5つの難易度レベル（C, B, A, S, SS）に3名の大学生と1名の社会人が配置され、計20名のデータに基づき分析が行われた。

学習者の個別の到達度の自己評定値の変化を概観したところ、全ての学習者の到達度は、右上がりに上昇していた。図3に難易度SSのレベルの学習を行った大学生と社会人の到達度の変化を示した。どの学習者も到達度のデータは、図に示した例と同様の傾向を示している。

得られた結果を見る限り、これまで高校生や小学生のデータに見受けられた大きな個人差は、顕著とはいえなかった。実験参加者の属性が均質になったこと、共通の携帯端末によって学習条件が等質化されたことなどが原因として考えられる。

5つの難易度ごとに、学習者の到達度の平均を算出し、サイクルに対応させてプロットした図を図4に示した。難易度とサイクルを要因とする、5×5の2要因混合計画による分散分析

を行ったところ、難易度の要因、サイクルの要因それぞれに主効果が有意であり（ $F(4,15)=3.78, p<.05$ ;  $F(4,60)=112.39, p<.01$ ）交互作用は有意ではなかった（ $F(4,16)=0.50$ ）。

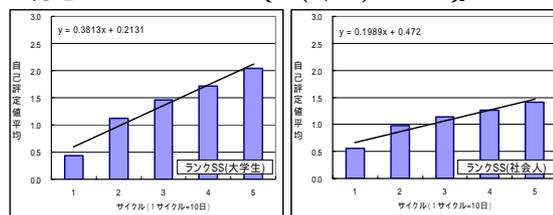


図3 学習到達度の変化（個人データ）

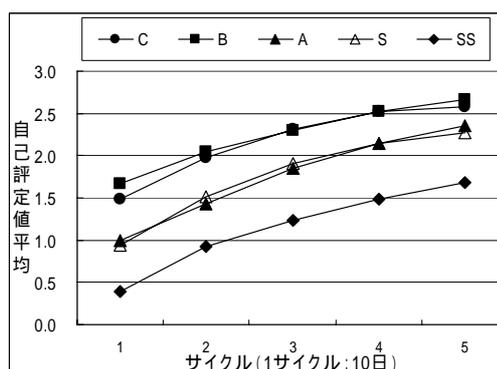


図4 難易度ごとに平均化された到達度変化

難易度の要因の主効果は、学習者の学習レベルの影響を受けることは明らかであるため、この主効果が、英単語の一般的な難易度の影響を示しているとは言えない。ここで重要なポイントは、難易度とサイクルの要因の交互作用に有意差が認められなかった点である。図4を見る限り、どの難易度のグループも、サイクルに対してほぼ同程度に到達度が上昇しているといえる。これは、予備調査で得られるような一般的な難易度が高いとあって、英単語の習得の進み方が遅いとはいえないことを意味している。すなわち、一般の難易度が高い単語と、低い単語で、「習得の難易度」に違いが認められないことを意味している。となく一般的な難易度が高い単語は、習得に時間を要するような気がするが、実際は同様の学習を行えば（同様の時間をかければ）同程度の到達度の上昇が見込めることを示唆している。

**引用文献** 寺澤孝文 1996 漢字の熟知度 福沢周亮(編) 言葉の心理と教育 教育出版, pp.51-56. / 寺澤孝文 1995 中学生の漢字の熟知度と読書量の関係 日本教育心理学会第37回総会発表論文集,195. / 寺澤孝文・太田信夫・吉田哲也 2007 マイクロステップ計測法による英単語学習の個人差の測定 風間書房 / 寺澤孝文 2007 「経験の変数化」を念頭においた実験計画法に基づく客観的絶対評価の実現 平成14~17年度科学研究費補助金基盤研究(A)(1)研究成果報告書.