

日本人英語学習者における5タイプのスペリングテスト比較： 音と文字の対応規則に注目した予備的研究

高波 幸代

1. はじめに

英語の綴りを習得することは英語母語話者にとって困難が伴うといわれている。英語の書記体系は *alphabetic* に分類され、音と文字の対応関係が密接な言語である (Taylor, 1981)。英語の綴りには音と文字の関係が規則的なもの (*regular spellings*) と不規則なもの (*irregular spellings*) が含まれ、特に不規則な綴りには難しさを感じる傾向がある。とはいえ、発音と綴りの関係を規則的に理解できていると、英単語を正確に読めるようになるため、フォニックスなどの綴りと発音の規則を教える教授法が開発されてきた。

日常の様々な場面において、読み書き能力 (*Literacy*) は重要であるが、英語母語話者にとって、「正しく発音できること」と「正しく綴りを書けること」は教育を受けたことの表れとなるため、幼少期から訓練されるべき重要項目となっている。そのため、学習初期の幼児や小学生を対象とした綴りの学習方法や指導方法などの研究も行われ、幼児が読み書きを習得する過程が注目されている (e.g., Bruck & Treiman, 1990; Drake & Ehri, 1984; Ehri, 1997; Ehri, Wilce, 1987; Harley, 2001; Griffith, 1991; Stanovich, Cunningham & Cramer, 1984; Stuart & Masterson, 1992; Yopp & Yopp, 2000)。

日本では、綴りの指導法の一つとしてフォニックスの有用性が提唱されているが (e.g., 相澤・望月, 2010)、英語母語話者を対象とした様々な研究と

比較すると、現段階では、英語の綴りに関する研究が広く行われているとは言い難い。綴りの習得が母語話者にとって難しいのであれば、異なる書記体系を持つ日本人英語学習者にとっては相当難しいはずであるが、森（2007）も指摘しているように、スペリングに関する注目度は低いと言わざるを得ないのが現状である。しかし、スペリング能力はライティングの下位技能でもあるように、英語を書く上では無視できない要素であることも忘れてはならない。

学習者が既習の綴りを習得できているかを調べるための有効な方法として、スペリングテストを挙げることができる。スペリングテスト形式の比較研究は、L1において1930年代から行われている（Allan & Ager, 1965; Allred, 1984; Brody, 1944; Carpenter & Carpenter, 1978; Croft, 1982; Freyberg, 1970; Moore, 1937; Northby, 1936; Westwood, 1999）。最も一般的なスペリングテストとして古くから用いられている形式は、単語の書き取りテストである（i.e., 教師が単語を1語ずつ発音しそれを学習者が書き取る方法）。この他にも、1文を書き取る形式（sentence dictation）や正しい単語の綴りを選択肢から選ぶ形式（Multiple-choice）などもこれまで用いられてきた。スペリングテストの形式比較を行った研究は、学習者のスペリング能力を評価する方法として、どの形式が効果的か、またそれぞれのテスト形式はどのような特徴を持っているかを明らかにすることを目的としていた。しかし、研究の対象となったテスト形式は様々であり、その結果、得られた結論も多様であった。加えて、テストに用いた単語の抽出方法も、研究により大きく異なっており、一貫性が見られない。

以上の点から、スペリングテストの比較研究に関する一貫した結果は未だ得られていないことは明らかである。また、研究対象もL1に限られており、L2またはEFLを対象としたテスト形式の比較は行われていない。

本研究は、日本人英語学習者を対象として、異なるスペリングテストを用いたテスト形式の比較研究を行う。それぞれのテストがどのように綴りの能力を測定するかを検証し、学習者の綴りの習得度に焦点を当てた診断的評価

を行う。発音された音から文字に書き起こす **Sound-Letter Correspondences** (音と文字の対応規則)¹ の理解は、綴りの能力の基盤となっていることが考えられるため、この対応規則の習得度がスペリングテストの結果と、どのように関係しているのかについて考察していく。

2. スペリングテストに関する先行研究

スペリングテスト形式の比較研究は、前述のとおり 1930 年代から行われてきた。しかし、これまでの先行研究で用いられたテスト形式の種類は様々であり、研究結果も異なっている。本節ではこれまでの先行研究を概観していく。

Northby (1936) は、「学習者がある単語を習得している場合、幾つかの異なる形式のスペリングテストを与えても全て同様の結果を示すはずである (i.e., 正答なら一貫して正答、誤答なら一貫して誤答)」という仮説に基づき、診断的評価を目的とした調査を行った。Northby が調査で用いたテスト形式は、文章内の単語補充、ディクテーション、単語の書き取り、正しい綴りの選択、口頭で綴りを述べる、というように様々な角度からスペリング能力を測定するものであった。表 1 は Northby の研究で用いられたテスト形式をまとめたものである。Oral Form 以外のテストは written form (書き言葉形式) のスペリングテストである。実験の結果、5 つのテスト形式には難易度に違いが見られた (Story Form > Timed Dictation > List Form > Oral Form > Multiple-choice)。難しさという点では、Timed Dictation と Story Form が適切なテスト形式であると Northby は述べている。ただし、綴りの間違いの一貫性を見るためには、他の要因 (文脈・文章) からあまり影響を受けない List Form の形式が適切であると述べている。Northby の仮説に反して、受験者のテスト得点には一貫性が見られなかったが、これはテストの性質が production test (i.e., Story Form, Timed Dictation, List Form) と recognition test (i.e., Multiple-choice) に分かれていたためと思われる。つまり、受験者のテスト得点はテスト形式の影響を受けていたことを意味する。

表 1. Northby(1936)で用いられたテスト形式

テスト形式	テスト内容
Story Form	短い文章の中で目標語 20 語が穴埋めの形式。
Timed Dictation	一文ディクテーション。用意された文は 12 文 ² 。
List Form	一般的に行われる綴りテストの方法で、調査者が単語を読み上げ、それを学習者が書き取る形式。
Multiple-choice	5 つの選択肢の中から正しい綴りを選ぶ。正しい綴りは一つ。
Oral Form	調査者により指定された単語の綴りを、学習者が口頭で解答する形式。

Moore (1937) は Northby (1936) の研究にならい、Oral Form を除いた 4 つの written form spelling tests に限定して調査を行った。結果は Northby と同様で、Story Form が最も難しく、次いで Timed Dictation, List Form と続き、Multiple-choice が最も簡単であった。

教師が単語を 1 語ずつ読み上げ、その綴りを学習者が書き取る、という List Form の形式は、先行研究 (Moore, 1937; Northby, 1936) で適切な方法として挙げられており、最も広く用いられているテスト形式である。しかし Brody (1944) は、単語 1 語の綴りのみ答えさせるだけでは不十分であると指摘し、形式の異なる数種類のスペリングテストを学年の異なる学習者に実施した。表 2 は、Northby (1936) や Moore (1937) 以降の年代に行われた、スペリングテストの形式比較に関する研究をまとめたものである。ほとんどの研究において、dictation が用いられてはいるものの、使用されたマテリアルは多様であり比較の仕方もそれぞれの研究により異なるため、一貫した結果はやはり得られていない。

表 1. Northby (1936) で用いられたテスト形式

研究	テスト形式	結果と結論
Brody (1944) [L1: grade IV through IX (n = 1231)]	(1) list test, (2) sentence test (= Timed dictation), (3) dictated paragraph, (4) proof-reading test ※ dictated paragraph と proof-reading test は easy, average, difficult の 3 段階で構成。	・ dictated paragraph は sentence test や list test と比較して難しかったが、文章の影響を強く受けている。 ・綴りの能力を測定する場合、1 つのテスト形式が他のテスト形式の代わりに機能を果たすことはできない。
Allan and Ager (1965) [L1: twelfth grade volunteers (n = 100)]	(1) a check-list approach (Recognition), (2) a dictation-in-context approach (Oral), (3) a write-in correction approach, (4) Multiple-choice (Wellesley Spelling Scale) 上記の 4 タイプのスペリングテストの他に、認知機能に関わるテスト (personality, spatial, perceptual, intellectual), 習熟度テストとの関連を比較。	・スペリング能力と認知機能との関わりについて調査したが、結果から、スペリング能力は独立したものであった。 ・スペリングテスト間では差がなく general spelling ability としてまとめられる。 ・スペリング能力が独立した能力であるならば、特定の指導が必要。

研究	テスト形式	結果と結論
Freyberg (1970) [L1: intermediate (n = 506)]	(1) writing test, (2) dictated word test, (3) Multiple-choice test	・ライティングにおける綴りの正確さと, dictated word test の相関が高かった。 ・2つのテスト間 (dictated word test と Multiple-choice) の相関は高く, 診断的に用いるべき。
Carpenter and Carpenter (1978) [L1: fourth-grade (n = 45)]	スペリングテストと習熟度テストの比較。 (1) Larsen-Hammill Test of Written Spelling (TWS) (dictation format (predictable and unpredictable words)のテスト) (2) California Achievement Test (CAT) (Reading, Mathematics, Language の構成で, Language には Spelling も含まれる。)	・TWS の得点が CAT の得点を予測できるかを焦点とした。 ・TWS は一般的な規則に基づく predictable な綴り, 規則に基づかない unpredictable な綴り, を書く能力を測定しているもので, CAT の得点を予測するものとは言い切れない。

研究	テスト形式	結果と結論
Croft (1982) [L1: standard III and IV groups, (n = 40)]	綴りの間違いを指摘する recognition, 正しい綴りを書 く production を区別。 (1) a proof reading/correction test, (2) a Multiple-choice/cloze format, (3) a traditional dictated word test, を比較。 (1), (3)は production, (2)は recognition。	productionの方が recognitionより難しい。 ライティングにおける綴 りを評価する際には, productionを含むテスト が妥当。
Allred (1984) [L1: one through six grade (n = 3024)]	(1)standardized spelling test (= proofread), (2)written spelling test 2つのテスト結果が, 学年・ 性別・テストを受験した順 番によって影響を受けるか どうか, を検証。	・(1)と(2)のテスト間には 有意な相関が見られた。 proofreadのタイプも適切 なスペリングテストと言 える。教師が, 学習者の 相対的, 一般的なスペリ ング能力について調べた い場合は, proofreadのタ イプで良い。 ・実際のスペリング能力 について調べたい場合 は, 綴りを書かせること が重要である。

研究	テスト形式	結果と結論
Westwood (1999) [L1: year 2 to 5 primary school (n = 93)]	(1) a dictated spelling list, (2) a proofreading exercise, (3) the correction of spelling errors, (4) a Multiple-choice spelling test, (5) unaided writing of a story	<ul style="list-style-type: none"> ・それぞれのテストの相関は全て高く、妥当なテスト形式である。 ・5つのテスト全てが、<i>general spelling ability</i> を測定している。 ・これらのテストを一緒に用いて測定することで、より良い診断的評価を行うことができ、学習者のスペリングの能力の発達状況を探ることができる。

これまでの先行研究の結果から、学習者の綴りの能力は1つのテストを用いるだけでは適切に測定できないことが理解できる。綴りの能力とライティング能力との関係を調査した研究もあったが (e.g., Freyberg, 1970; Westwood, 1999), ライティングには文法の知識などが必要であり、綴りの正しさや語彙選択の適切などの観点も含まれる。また、ライティングの場合、綴りを間違えない単語を学習者が選んでいる可能性も否定できないため、スペリングテストの結果とライティングで用いる綴りの質は異なるものであることを認めなければならない (Croft, 1982)。即ち、ライティングテストで学習者のスペリング能力を測定することには限界がある、ということが明らかである。

綴りを見て正しいか正しくないかを判断する *proofreading* や *Multiple-choice* のような *recognition test* は難易度が低いテスト形式として扱われる場合が多い。しかし、Brody (1944) で用いられた *recognition test* のように文章を用いたテストの場合は、与える文章の難しさに学習者が影響を受けてしまうこと

も考えられるため、文章の難易度によっては、**recognition test** も難しくなることがある。

どの研究においても、一貫して妥当だと判断されたものには **List Form** (または **list test**) や **Timed Dictation** (または **dictated word test**) などの **production test** が挙げられる。とはいえ、多くの学習者が **Multiple-choice** のような **recognition test** で高い得点を示すということは、綴りを正しく書くことができなくても、正しい綴りを選択することができるという証拠となる。つまり、**recognition test** に正答する能力と **production test** に正答する能力が一貫していないことが明らかになっている。スペリング能力が、**recognition** から **production** に向けて段階的に習得される可能性も否定できない。

スペリングテストの形式は様々であるが、各テストが測定している能力は異なっており、幾つかのテスト形式を組み合わせる用いることが好ましいと考えられる (e.g., Brody, 1944; Calhoun, Greenberg, & Hunter, 2010; Freyberg, 1970; Westwood, 1999)。

これまでの先行研究で用いられたテスト形式のうち、それぞれの特徴が最も分かりやすいものは Northby (1936) の 5 つのテスト形式であった。物語、文、単語、選択肢、など様々な観点から偏りなくテストが作成されており、その全てが授業内で使いやすい形式であった。

本研究は、音と文字の対応規則の習得状況に注目し、スペリングテストの形式比較を行うことを目的とする。音と文字の対応規則の習得状況に関しては、森 (2007) の「中学校段階で習得しておくべきであると思われる規則的な音とつづりの連関」にある 90 語のリストを用い、書き取りテスト³を行う。また、スペリングテストの形式比較に関しては、Northby (1936) で用いられたスペリングテストの形式 (i.e., **Story Form**, **Timed Dictation**, **List Form**, **Multiple-choice**) に単語の意味を選ばせる **Matching Task** を加え、合わせて 5 つのテストを検証する。

3. 方法

3.1 調査質問

これまでの先行研究を踏まえ、調査質問 (リサーチクエスチョン: RQ) を以下のように設定した。

RQ1: Story Form, Timed Dictation, List Form, Multiple-choice, Matching Task の 5 つのテスト形式の得点には差があるか?

RQ2: 5 つのテストと Sound-Letter Correspondence の理解度との間には、どのような関係が見られるか?

RQ3: 受験者の各スペリングテストの得点には一貫性が見られるか?

3.2 参加者

高校 1 年生の 2 クラス計 78 名を対象として調査を行った。全 3 回のセッションに参加できなかった 5 名の生徒は分析から除外した。その結果、最終的に計 73 名 (男子 38 名・女子 35 名) を分析対象とした。

本研究では、音と文字の対応規則に関する習得度に焦点を当てたため、平均学力が同程度の 2 クラスで調査をすることにした。データ収集を行った 2 クラスはいずれも特進クラスに属さず、同高校の英語教員らの授業観察などから同程度とみなされる。

3.3 マテリアル

3.3.1 音と文字の対応規則に関する知識

音と文字の対応規則に関しては、森 (2007, p. 102) の「中学校段階で習得しておくべきであると思われる規則的な音とつづりの連関」リストから全 90 語を用いて測定した。このリストは子音 25 項目に関わる 45 語と母音 24 項目に関わる 45 語から構成されている。スペリングに関する包括的な習得状況を測定するためには、50 語から 100 語の語数が必要であると指摘されているため (Masterson & Apel, 2000), リスト内の全ての単語について書き取りテ

ストを行った。しかしながら、語数が90語と多いことから学習者に与える負荷を軽減し、より正確なデータを得るために、子音と母音それぞれ45語ずつを1週目と2週目に分けて実施した。調査者は各単語を2回ずつ読み上げ、学習者にその発音に該当する綴りを書かせた。ここでは一般的な綴りの書き取りテストの方法を採用した。

3.3.2 スペリングテスト

表1のNorthby (1936) のマテリアルから Oral Form を削除したものに (i.e., Story Form, Timed Dictation, List Form, Multiple-choice), Matching-Task を加え、計5つのテストを実施した。

Matching Task を加えた理由は以下の通りである。英語を母語とする学習者の場合、文字から正しく発音できる (または発音を理解できる) ことは、その単語の意味を理解していることを表すが、第二言語あるいは外国語として英語を学ぶ学習者にとっては、これは必ずしも当てはまらない。そのため、本研究では英語の綴り (形) と意味を結ばせる Matching Task を加え、単語の意味を理解できているかどうかを測定することとした。この形式は、単語の意味を書かせる場合よりも簡単になってしまうことが懸念されたが、意味を書くという負荷から無解答となる学習者が多くなることを考慮し、綴りと意味を結びつける比較的単純な形式を用いた。

3.3.3 目標語

学習者の習熟度を考慮し、英検4級からマテリアルとして適切な英文を選び、その中から学習者が間違えやすいと思われる単語10語を選び出した。英文の一部に同じ単語が数回使われている箇所には文章を一部加工 (i.e., 必要な場合は2文を1文にまとめるなど) して、同じ単語が出現しないよう配慮した。目標語の語彙レベルと音節数、また発音記号は表3に示す⁴。

表3. 目標語のレベルと音節数, 発音記号

目標語	JACET 8000 (1-8)	SVL (1-12)	音節数	発音記号
because	1	1	2	/bikɔ':z/
before	1	1	2	/bifɔ':r/
birthday	2	1	2	/bɛ':rθðeɪ/
found	2	3	1	/fáund/
guitar	3	1	2	/gita':r/
pair	2	1	1	/péər/
party	1	1	2	/pa':rti/
piece	1	1	1	/pí:s/
shoes	2	shoe(1)	1	/ʃu:z/
wanted	1	want(1)	2	/wɑ'ntid/

注. 目標語はアルファベット順に記載した。

3.4 手順

本研究の手順は図1の通りである。「音と文字の対応規則 (Sound-Letter Correspondences) の理解」に関するテスト (以下, SLC test) は子音と母音のテストを45語ずつ, 2週かけて実施した。調査のうち, 1クラス目はセッション1にて子音のテストを先に実施し, セッション2にて母音のテストを実施した。2クラス目はセッション1にて母音を, セッション2にて子音, というようにカウンターバランスをとった。セッション3では, 3.3.2で挙げた, Story Form, Timed Dictation, List Form, Multiple-choice, Matching Task, の5つのテストを実施した。このテストの実施順序は, 先行研究 (Northby, 1936) に基づき, 練習効果を最小限に抑えることを意図して配列されている。また, 本研究で著者が加えた Matching Task は, 単語の意味と日本語の意味を結びつけるという課題の性質上, 単語の音 (発音) の情報よりも単語の意味の情報から綴りを想起してしまう可能性が否定できないため, 最後に配置した。

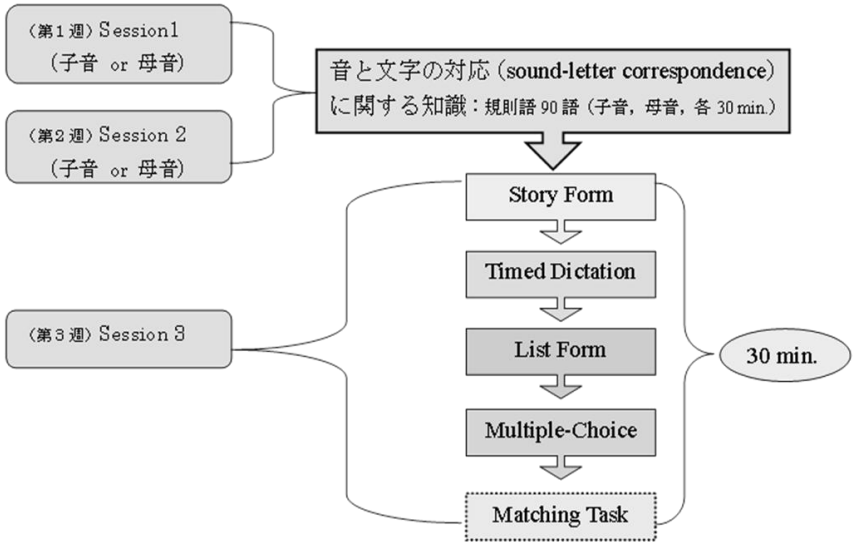


図1 調査手順

3.5 採点方法

各テストの特性を考慮して、表4に示す基準で採点を行った。基本的に採点方法は正答に「1点」、不正答は「0点」という配点である。ただし例外として、List Form に関しては、文章 (i.e., Story Form) や文単位 (i.e., Timed Dictation) での手がかりが得られないため、同じ発音の綴りに関しても得点を与えることとした。例えば、*pair*, *pear*, *pare* の3つは同じ発音 (/pɛər/) であり、*piece* と *peace* も同じ発音 (/pi:s/) である。これらの単語に関しては、発音の同じ単語を書いた場合には List Form のときに限って、全て正答として扱った。

表 4. 5つのスペリングテストにおける採点基準

テスト形式	正	誤	発音が同じ別の単語
Story Form	1	0	0
Timed Dictation	1	0	0
List Form	1	0	1 ex. <i>pair, pear, pare</i> と <i>piece, peace</i>
Multiple-choice	1	0	
Matching Task	1	0	

3.6 分析

分析には SPSS 16.0 を用いた。主な分析の手順は以下の通りである。

- (1) Story Form, Timed Dictation, List Form, Multiple-choice, Matching Task の5つのスペリングテスト得点に差があるかを、分散分析を用いてその差を比較する（→ RQ1）。
- (2) SLC test (i.e., 音と文字の対応規則の理解) と各テスト得点の関係性を相関分析で調べる（→ RQ2）。
- (3) 各テスト得点（1点～10点）を順位化し、スピアマンの順位相関係数を用いて、学習者の得点に一貫性が見られるかを検証する（→ RQ3）。

4. 結果

4.1 音と文字の対応規則（Sound-Letter Correspondence）

テストを実施した単語90語のうち、0分散が含まれるなどした3語を削除し残りの87語に関して信頼性分析を行った。その結果、 $\alpha = .891$ と高い信頼性が得られた。

4.2 テストの記述統計, 信頼性, 弁別力

本研究は Story Form, Timed Dictation, List Form, Multiple-choice, Matching Task のテストを用いて学習者のスペリング能力を測定し, 診断的評価を行うことを目的としている。各スペリングテスト形式の信頼性, 弁別力を算出し, 表5にまとめた。

表5. 記述統計, 信頼性, 弁別力 ($N=73$)

テスト形式	k	M	SD	Min	Max	Cronback's alpha	DIS
Story Form	10	5.80	2.09	1	10	.67	.64
Timed Dictation	10	6.11	2.07	1	10	.63	.60
List Form	10	7.47	1.72	2	10	.54	.51
Multiple-choice	10	8.91	1.23	5	10	.46	.42
Matching Task	10	9.93	0.30	8	10	.55	---

注. DIS (弁別力) は point-bi-serial を用いた。 k = 項目数。 Min = 最小値。 Max = 最大値。

4.3 因子分析

先行研究で用いられていた4つのテスト (i.e., Story Form, Timed Dictation, List Form, Multiple-choice) が第1因子として決定された。綴りと意味を結びつける Matching-Task が第2因子となった (表6参照)。

表6. 因子分析の結果(主成分法, バリマックス回転後の因子負荷構造)

テスト形式	第1因子	第2因子	共通性
Story Form	.880	.348	.775
Timed Dictation	.844	.309	.809
List Form	.840	.265	.896
Multiple-choice	.770	-.207	.636
Matching Task	.082	.558	.318
因子寄与	2.792	.641	3.433
累積寄与率	55.847	68.664	

4.4 テスト得点の比較

参加者全てが5つのテストを受験したため、被験者内計画の分散分析を行った。球形性が満たされなかったため、Greenhouse-Geisserの修正された値を用いた。結果は有意であった ($F(2.51, 180.73) = 188.60, p < .001$)。その後、5つのテストのどの水準間で差が見られるかを検証するため、多重比較を行った結果、Story FormとTimed Dictationの間には有意な差が見られなかったが ($p = .437$)、Story Form, Timed Dictationと全てのテスト間で、またList Form, Multiple-choice, Matching Taskのそれぞれのテスト間すべてにおいて、有意な差が認められた ($p = .000$)。結果をまとめると、Story Form \approx Timed Dictation < List Form < Multiple-choice < Matching Task, のようになった。

4.5 テスト得点の相関分析

4.5.1 各テストの関係性

テスト間にどのような関係性が見られるかを調査するため、ピアソンの積率相関係数を用いて算出した。相関分析の結果、音と文字の対応規則と5つのテスト形式の間には、低い相関から中程度の相関が見られた。また、Story FormとTimed Dictation, List Form, Multiple-choiceの4つのテストの間には中程度から非常に高い相関が見られた。しかし、Matching Taskに関しては、いずれのテストとも相関が低かった（表7参照）。

表7. 音と文字の対応規則と各スペリングテストの相関

Variables	1	2	3	4	5	6
1 S-L Correspondences	—	.400**	.502**	.518**	.287*	.268*
2 Story Form		—	.801**	.821**	.592**	.218
3 Timed Dictation			—	.851**	.576**	.233*
4 List Form				—	.615**	.275*
5 Multiple-choice					—	-.052
6 Matching Task						—

注. * $p < .05$., ** $p < .01$, 両側検定. S-L CorrespondenceはSound-Letter Correspondenceを略したものである。

4.5.2 参加者の得点に関する一貫度

次に、参加者のテスト得点に一貫性が見られるかどうかを調べるため、スピアマンの順位相関係数を用いて検証した。順位相関の結果から、学習者の得点には中程度から高い相関が見られたことが分かった。Story Form, Timed Dictation, List Form においては、いずれも高い相関が見られたが、Multiple-choice との相関は中程度にとどまった。また、Matching Task との相関はどれも低く、有意な相関とはならなかった。これは学習者の Matching Task に関する得点順位が他のテストとは傾向が異なることを意味している（表 8 参照）。

表 8. スピアマンの順位相関係数

Variables	1	2	3	4	5
1 Story Form	—	.785**	.813**	.560**	.137
2 Timed Dictation		—	.845**	.568**	.136
3 List Form			—	.595**	.115
4 Multiple-choice				—	-.124
5 Matching Task					—

注. ** $p < .01$, 両側検定.

5. 考察

まず、これまでの分析結果を以下にまとめる。

- (1) 音と文字の対応規則 (Sound-Letter Correspondence) のテスト結果は信頼性が高かった ($\alpha = .891$)。
- (2) 因子分析を用いた結果、本研究で使用した 5 つのテストは、Story Form, Timed Dictation, List Form, Multiple-choice で第 1 因子、Matching Task で第 2 因子となった。
- (3) 本研究で使用した 5 つのテスト得点を、分散分析により比較した結果、

Story Form ≒ Timed Dictation < List Form < Multiple-choice < Matching Task, という結果になった。Matching Task には天井効果が見られた。

- (4) ピアソンの積率相関係数を用いてテスト間の関係性を分析した結果、Matching Task と他のテスト (i.e., Story Form, Timed Dictation, List Form, Multiple-choice) との相関は低かった。しかし、SLC テスト (i.e., 音と文字の対応規則) と5種類のスペリングテスト (i.e., Story Form, Timed Dictation, List Form, Multiple-choice, Matching Task) との相関は、全てにおいて低い相関から中程度の相関が見られた。
- (5) 参加者のテスト得点における一貫性を、スピーアマンの順位相関係数を用いて算出したところ Matching-Task 以外のテスト間では中程度から高い相関が見られた。

(1) に関して、「音と文字の対応規則」テストの信頼性が高かったことは、森 (2007) で挙げられた「中学校段階で習得しておくべきであると思われる規則的な音とつづりの連関」のリストが機能したということを示唆している。つまり、中学校での習得段階を確認する単語リストとして非常に有効だといえる。

(2) に関して、因子が2つに分かれたことは、Northby (1936) で用いられたテスト形式が全体として、1つのスペリングテストとして考えられることを示している。本研究で新たに加えた Matching Task のみ別の因子となったことは、他の4つのテストと性質が異なっていたことを意味している。しかし、単語の意味を選ばせるのではなく学習者自身に書かせる課題を与えた場合、全体を合わせて1つの因子として決定されるかは、今後検証する必要がある。

(3) に関して、先行研究の結果と同様、recognition test にあたる Multiple-choice の得点が高く、production test にあたるその他のテスト得点が低かった。今回加えた Matching Task では、ほぼ全員が正答であり ($M = 9.93$, $SD = 0.30$)、天井効果が見られたが、用いた単語の意味を理解できていたこ

とが示された。つまり、本研究で用いた **Matching Task** が単語の意味を知っていたかどうかを確認する課題として十分に機能したことを意味する。

Multiple-choice の得点が高かったことは、適切な意味が選べるにも関わらず、正しく綴りが書けない、という点で先行研究と類似した結果が得られた (Brody, 1944; Moore, 1937; Northby, 1936)。

(4) の結果について、**SLC test** (音と文字の対応規則: **Sound-Letter Correspondence**) と 5 つのスペリングテスト全てとの間に、弱から中程度の相関がみられた。これは、中学校段階までの音と文字の対応規則が理解できていた受験者は、本研究で用いたスペリングテストでも高い得点を取ることができていた、ということである。つまり、音と文字の対応規則を理解できているかを調べることは、スペリング能力の調査を行う上では重要であることが分かった。先行研究で **production test** と分類されていた **Story Form**, **Timed Dictation**, **List Form** の 3 つのテスト間において、中程度から高い相関がみられたことは予測通りであった。しかし、これら 3 つのテストとは性質の異なる **recognition test** として分類されていた **Multiple-choice** との間にも中程度の相関が見られていたことは、大変興味深い結果であった。**Multiple-choice** は性質の異なるテスト形式ではあるが、それぞれのテスト間に関連性が示されたことから、**recognition test** も綴りの理解度を測定する上で必要な課題であると言える。

(5) の順位相関の結果、**Story Form**, **Timed Dictation**, **List Form**, **Multiple-choice** の 4 つのテスト間に中程度から高い相関が見られたことは、学習者の解答が **Matching Task** 以外では一貫していたことを示す。これまでの先行研究では、**production test** と **recognition test** とが本質的に異なるものであり、**production test** の方が綴りの能力を測定する方法として妥当であると判断されてきたが、本研究の調査結果から、高校 1 年生の日本人英語学習者においては、綴りの能力は **production test** においても **recognition test** においても一貫していることが示唆された。スペリングテストに用いる目標語の難易度を調整するなどし、今後はさらに上の学年 (や年齢) を対象とした調査でも同

じ結果が得られるかを、慎重に検討する必要がある。

6. 結論

本研究は、高校1年生の日本人英語学習者を対象にスペリングテストの形式比較を行った。Northby (1936) で用いられた Story Form, Timed Dictation, List Form, Multiple-choice, また今回加えた Matching Task のテスト得点を比較した結果, production test として分類される Story Form, Timed Dictation, List Form の得点が低く, recognition test として分類される Multiple-choice の得点は高くなった。Story Form と Timed Dictation の得点が低いという結果は, 先行研究 (Moore, 1937; Northby, 1936) と同様であった。Matching Task は単語の綴りと正しい意味とを結ばせるテスト形式であったが, このテスト得点には天井効果が見られた。本調査で用いた単語は英検4級の読解問題から出題し全て既習語であった。綴りを見て意味は選べるものの, 綴りを書くというテスト形式において得点が低くなったことは, 学習者の全体的な傾向として, 綴りを書く能力が綴りを認識する能力と比較して劣っていることが明らかとなった (→ RQ1)。

SLC test (=音と文字の対応規則[Sound-Letter Correspondence]の理解)と各スペリングテストとの関係性を分析した結果, SLC test と各スペリングテストには相関が認められた。このことから, 学習者の中学校段階までの音と文字の対応規則の習得度と綴りの能力には関係があることが分かった。スペリングテスト間の相関については, production test 同士の相関が高かったが, これに加え production test と recognition test との間にも相関が見られた。これは, recognition test も綴りの能力を測定する上で必要であることを示している。本調査で新たに加えた Matching Task は, 他のテスト形式とは傾向が異なるため, 弱い相関または無相関であったが, 今後は単語の意味を書かせるなどの, 別の課題を用いて比較することも必要と思われる (→ RQ2)。

音と文字の対応規則に関しては, 異なる学年を調査対象とする場合は単語の難易度を調整する必要があるため, 今後の研究の際にはその学年に応じた

単語リストを用いることが必要となるであろう。スペリングテストで用いる目標語も同様に、学年を考慮して慎重に選ぶ必要がある。今回調査で用いた単語の音節数は1音節から2音節であったが、今後の調査においては音節数を統一し、単語の発音ごとにより細かい調査を行うことも重要となってくる。

各テストの解答が学習者ごとに一貫しているかを分析した結果、**Matching Task** 以外のテストで中程度から高い相関がみとめられた。このことは、**Story Form**, **Timed Dictation**, **List Form**, **Multiple-choice** の4つのテストに関しては「学習者の解答は、異なる形式においても一貫した結果を示す」という Northby (1936) の仮説を支持するものとなった(→RQ3)。**Matching Task** との結果が一致しなかった理由としては、**Matching Task** の正答率に天井効果が見られ、全体としてほぼ全ての学習者が単語の意味を理解していたことが挙げられる。しかし、**Matching Task** は学習者が単語の意味を理解できていることを確認する上で、必要なテストであった。また結果から注目すべきことは、単語の意味を分かっている(a)意味を理解していること、(b)正しい綴りを選ぶこと (**recognition**)、(c)正しい綴りを書けること (**production**)、とが一致していない、という点が挙げられる。スペリングテストは「正しい綴りが書けること」が前提であり、主に **production** 能力が重視される傾向にある。しかし、学習者の習熟度を考慮し、必要に応じて **recognition test** や意味理解を確認する課題を与えることで、学習者がどの段階でどの能力に関してつまづいているかを明らかにすることができる。今後は異なる年齢の学習者を対象とし、スペリング能力と他の技能(英語熟達度、リーディング能力、ライティング能力)との比較をすることも重要である。

注

1. 森(2007)は先行研究と独自の調査の結果から「中学校段階で習得しておくべきであると思われる規則的な音とつづりの連関」というリストを作成した。これには90語の単語が含まれており、いずれも音素一書記素

対応に一貫性がみられる単語であった。これらの規則が習得できていれば中学校段階までの理解できていることになる。Sound-Letter Correspondences は発音された音を文字に書き取る場合、つまり音から文字（音→文字）の方向性を指す場合に用いられる言葉である。一方で、印刷された文字から音を発音する場合、つまり文字から音（文字→音）の方向性を指す場合は、Letter-Sound Correspondence と表し、書記素—音素対応（Grapheme-Phoneme Correspondence: GPC 規則）と同様の意味で用いられる。これら2つは処理の方向性を区別して用いるよう指摘されている（Valtin & Naegele, 2001）。本論文では、発音された音から綴りを書き取る、という処理を扱っているため、Sound-Letter Correspondence に統一した。

2. 文によっては目標語が2語含まれるものもあったため、Timed Dictation の文は全体で12文となっている。
3. 書き取りテストとは、調査者が単語を1つずつ読み上げ、それを学習者が書き取る、というテスト形式である。
4. 語彙レベルと音節数は JACET 8000 と ALC 社の Standard Vocabulary Levels (以下 SVL) のリストから参照し、発音記号に関してはオンラインで使用可能な無料の発音記号変換ソフト (<http://www.freeenglish.jp/pronunciation.html>) とジーニアス英和大辞典を参照して記載した。

参考文献

- Allan, D., & Ager, J. (1965). A factor analytic study of the ability to spell. *Educational and Psychological Measurement*, 25, 153-161.
- Allred, R. A. (1984). Comparison of proofreading-type standardized spelling tests and written spelling test scores. *The Journal of Educational Research*, 77, 298-303.
- Brody, D. S. (1944). A comparative study of different forms of spelling tests. *The*

- Journal of Educational Psychology*, 35, 129-144.
- Bruck, M. & R. Treiman. (1990). Phonological awareness and spelling in normal Children and dyslexics: The Case of Initial Consonant Clusters. *Journal of Educational Child Psychology*, 50, 156-178.
- Drake, . A., & Ehri, L. C. (1984). Spelling acquisition: Effects of pronouncing words on memory for their spellings. *Cognition and Instruction*, 1, 297-320.
- Calhoun, M. B., Greenberg, D., & Hunter, C. V. (2010). A comparison of standardized spelling assessments: Do they measure similar orthographic qualities? *Learning Disability Quarterly*, 33, 159-170.
- Carpenter, D., & Carpenter, S. (1978). The concurrent validity of the Larsen-Hammill Test of written spelling in relation to the California achievement test. *Educational and Psychological Measurement*, 38, 1201-1205.
- Croft, A. C. (1982). Do spelling tests measure the ability to spell? *Educational and Psychological Measurement*, 42, 715-723.
- Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 237-269). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Ehri, L. C., & Wilce, L. S. (1987). Cipher versus cue reading: An experiment in decoding acquisition. *Journal of Educational Psychology*, 79, 3-13.
- Freyberg, P. S. (1970). The concurrent validity of two types of spelling test. *British Journal of Educational Psychology*, 40, 68-71.
- Griffith, P. L. (1991). Phonemic awareness helps first graders invent spellings and third graders remember correct spellings. *Journal of Reading Behavior*, 23, 215-233.
- Harley, T. A. (2001). *The Psychology of Language: From data to theory*. New York, NY: Psychology Press Ltd.

- Masterson, J. J., & Apel, K. (2000). Spelling assessment: Charting a path to optimal intervention. *Topics in Language Disorders, 20*, 50-65.
- Moore, J. E. (1937). A comparison of four types of spelling tests for diagnostic purposes. *The Journal of Experimental Education, 6*, 24-28.
- Northby, A. S. (1936). A comparison of five types of spelling tests for diagnostic purposes. *The Journal of Educational Research, 29*, 339-346.
- Stanovich, K. E., Cunningham, A. E., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38*, 175-190.
- Stuart, M. & J. Masterson. (1992). Patterns of Reading and Spelling in 10-Year-Old Children Related to Prereading Phonological Abilities. *Journal of Educational Child Psychology, 54*, 168-187.
- Taylor, D. S. (1981). English Spelling: a Help rather than a Hindrance. *ELT journal, 35*, 316-321.
- Valtin, R., & Naegele, I. M (2001). Correcting reading and spelling difficulties: A balanced model for remedial education. *The Reading Teacher, 55*, 36-45.
- Westwood, P. (1999). The correlation between results from different types of spelling test and children's spelling ability when writing. *Australian Journal of Learning Disabilities, 4*, 31-36.
- Yopp, H. K., & Yopp, R. H. (2000). Supporting phonemic awareness development in the classroom. *The Reading Teacher, 54*, 130-143.
- 相澤一美・望月正道（2010）．「英語語彙指導の実践アイディア集」．東京：大修館書店
- 森 千鶴（2007）．「L2 学習者のスペリング能力とリーディング能力の関係」．広島：溪水社