

論 文

Vocabulary Levels Dictation Test の信頼性の再分析

¹ 今井 由美子 ² 三根 浩¹ 同志社女子大学・表象文化学部・英語英文学科・准教授² 同志社女子大学・学芸学部・情報メディア学科・教授A reanalysis of the reliability
of the Vocabulary Levels Dictation Test¹ Yumiko Imai ² Hiroshi Mine¹ Department of English, Faculty of Culture and Representation,
Doshisha Women's College of Liberal Arts, Associate Professor² Department of Information and Media, Faculty of Liberal Arts,
Doshisha Women's College of Liberal Arts, Professor

キーワード：EFL、語彙、ディクテーション、信頼性、項目分析

1. はじめに

語学教育は、4つのスキルの指導により構成される。受容的スキルであるリーディングとリスニング、そして発表的スキルであるライティングとスピーキングの総合的な運用力が外国語教育に求められている。これらのスキルを育成する上で重要な学習目標は発音、語彙、文法である (Nation, 2001, p.1)。語学学習において指導者が学習者の語彙力を把握し、そのレベルに合わせた指導法を取り入れていくことが重要である。語彙学習や語彙サイズ測定のために活用されている学習用語彙リストには A teacher's word book of 30,000 words (Thorndike and Lorge, 1944)、General Service List (West, 1953)、北海道大学英語基本語彙表 (園田, 1996)、Academic Word List (Coxhead, 2000)、JACET8000 (大学英語教育学会基本語改訂委員会, 2003) などがある。

語彙テストには、単語の文字形式、すなわち単語の綴りを提示し意味を選択させたり知識をチェックさせたりする再認テスト形式で受容語彙を測定するテストと、単語や文章を書かせる再生テスト形式で発表語彙を測定するテストがある。代表的な受容的語彙テストとして望月 (1998) による語彙サイズテスト、Nation (2001) や Schmitt,

Schmitt, and Clapham (2001) らによる語彙レベルテスト、Meara and Buxton (1987) のチェックリスト・テストなどが、また発表語彙テストとしては、Laufer and Nation (1995) の語彙頻度プロフィール法や Laufer and Nation (1999) の統制的発表語彙サイズテストなどがある。しかしながら一般的に、再生テストは再認テストよりも難しいことが知られている。発表的側面を測定する再生テストは、日本語を母語とする EFL 学習者にはむずかしすぎると考えられており、発表的側面を有する語彙テストが使用された研究例は少ない。

2. 先行研究

2.1 発表的語彙力測定の試み

Waring (1997) は、ESL 学習者の受容的語彙知識と発表的語彙知識の比率に焦点を当てた実験を計画した。日本語を母語とする中級レベルの EFL 学習者に対し VLT と統制的発表語彙サイズテストを実施し、その結果、受容語彙と発表語彙の比率は44.8% (表1、p.60) で、「受容的語彙知識は発表的語彙知識よりも大きいことを明確に示すデータを見いだした」と報告した (p.63)。

望月 (2003) は、発表語彙サイズを測定するテストの日

本における適用例を紹介しているが、その中で、「統制的発表語彙テストは、受験者がある程度の読解力、文法力を持っていることを前提」にすることと、「語彙頻度プロフィールの問題点としては、作文の長さが400語以上」であり「比較的長めの作文でない」と処理できない」ため初級学習者に向いていないことを指摘している (pp. 191-192)。

古荘・今井・三根 (2009) は、Fountain and Nation (2000) がヒアリングの要素を含めた語彙知識測定のために開発した語彙レベルディクテーションテスト (Vocabulary Levels Dictation Test; 以下 VLDT) により発表語彙力の検証を試みた。VLT および VLDT で測定される語彙力と、CELT や TOEIC などの英語標準テストで測定される語学力、および語彙調べ課題の効果について検証し、「語彙力と語学力は関連していること」、そして「VLDT は語彙力の指標として信頼性があること」を報告した (pp. 146-147)。

今井・三根 (2010) は、VLDT について日本語を母語とする EFL 学習者への適用可能性と信頼性を検証した。その結果、VLDT が十分に高い信頼性を示した一方で、その調査協力者の平均点が非常に低いことから、語彙力よりも音韻ループの容量に制約を受けていた可能性が高いこと、つまり長いフレーズを聞き取る際に前半部分の単語が音韻ループの容量を超えることでセンテンスの再構築ができないということを指摘した (p. 66)。

今井・三根 (2011) は、VLDT の妥当性について検証した。その結果、VLDT の総得点が語彙テストとして十分な信頼性を有すると報告した。また、VLDT により測定される発表的語彙サイズは VLT の受容的語彙サイズの約50%であることと、英語力の高い学習者は、VLT で測定された受容的語彙サイズから期待されるよりも高い発表的語彙サイズを持つ可能性を指摘した。一方で、Test-C の難易度が比較的に高く結果の妥当性に疑問があることを示唆した (p. 92)。

2.2 VLDT

VLDT は Fountain and Nation (2000) により開発されたテストである。特定の言語項目に個別に焦点を当てるのではなく、リスニング (ヒアリング) という「受容的」スキルと、聞き取った単語のつづりを書くという「発表的」スキルの両面を測定するためのディクテーションテストにより構成されている。このテストの解答者は、一連の音の流れから単語を一つずつ文節化し、順次に記憶し、解答用紙の上にセンテンスを再構成することが求められる。

VLDT は A、B、C、D の4つの並行テストが作成されており、各テストは5つのパラグラフからなる文章で構成されている。導入パラグラフは基本語彙1~500語、第1パラグラフは基本語彙501~1,000語、第2パラグラフは1,001~2,000語、第3パラグラフは2,001~3,000語、第4パラグラフは4,000~6,000語にグレード分けされた語彙をターゲット語としている。Fountain and Nation (2000) は、これらの4並行テストの相互相関は0.95以上、語彙レベルテストとの相関は0.78、Grammar test との相関は0.70以上であり、高い信頼性と妥当性を有することを報告している。

今井・三根 (2010) は、VLDT のオリジナル版での「Test-A の平均得点20.88は Test-B の平均点29.19に比べて極端に低かった」こと、とくに第2パラグラフにおける差が大きく、1フレーズに含まれる単語数が多いことから、Test-A は日本語を母語とする EFL 学習者には難しすぎることを指摘した (p. 59)。ディクテーションでは「6語目以降急激に誤りが増える」と竹蓋 (1984, p. 235) が指摘しているように、単語数の多さが Test-A の第2パラグラフの得点の低さに影響していると考え、Test-A のテキスト中、6語以上の語数からなるフレーズを、発音の切れ目かつ文法的に切り分け可能な箇所分割した。切れ目は録音された音声の波形に基づいて明確に波形が途切れている箇所だけを分割しポーズを挿入し、これを Test-A の修正版とした。

語彙力の測定には VLT の2,000語、3,000語、Academic 語、5,000語レベルを使用し、解答時間は30分とした。VLT と VLDT は共に語彙レベル別にテストが構成されているが、それらの一番の違いは、VLT が語彙の視覚情報による再認テストという形で受容的スキルを測定するのに対し、VLDT では音韻情報をヒアリングする受容的課題と語彙の綴りを再生する発表的課題の2種類のスキルを測定することである。

2.3 目的

本研究においては、Test-B と Test-D はオリジナル版を使用した。Test-A は、今井・三根 (2010) による修正版を使用した。Test-C は、今井・三根 (2011) により難易度が比較的に高く結果の妥当性に疑問があることが示唆されたためポーズを追加することで難易度の調整を試みた。本研究の目的は、日本人 EFL 学習者を対象として、VLDT の4つの並行テストのうち Test-C の難易度を修正することにより、並行テスト間のテスト得点の同質性を検

証すること、および、VLDT の語彙テストとしての信頼性について再検証を試みることである。

3. 方法

3.1 調査協力者

調査協力者は2010年度英語専攻の2年生で週1回90分のリスニング科目を受講した女子大学生である。リスニングの授業の中で、春学期始めに CELT (Listening Part Form B ; 30分)、VLT (Version 1 ; 30分)、VLDT (Test-A ; 15分)、春学期末に CELT (Listening Part Form B ; 30分)、VLT (Version 2 ; 30分)、VLDT (Test-B ; 15分)、および秋学期始めに CELT (Listening Part Form A ; 30分)、VLT (Version 1 ; 30分)、VLDT (Test-C ; 15分)、秋学期末に CELT (Listening Part Form A ; 30分) と VLT (Version 2 ; 30分)、VLDT (Test-D ; 15分) を実施し、4回のテストを全て受験した99名を調査の対象とした。調査協力者には、春学期の始めに口頭で、秋学期の最後に書面で調査への参加意思を確認した。

3.2 手続き

VLDT の課題音声は日本人に英語を教えた経験豊富なアメリカ人女性に録音を依頼した。さらに、Fountain and Nation (2000, p. 30) の指示通りの切れ目の位置に解答時間のためのポーズを入れ約12分の長さに編集し、録音テープに記録した。ただし、Test-C (Appendix A) は Test-A (再修正版) と同様に、音声的・文法的に切り分け可能な箇所ではポーズを追加したものを使用した。追加した箇所は、第1パラグラフで1カ所(6語文)、第2パラグラフで2カ所(6語文、6語文)、第3パラグラフで2カ所(8語文、9語文)、第4パラグラフで1カ所(7語文)である。これを Test-C の修正版とした。

課題音声は、調査協力者に1回だけ提示した。解答用紙は自作し (Appendix B)、日本語による簡単な指示文の後

に、それぞれのフレーズごとに聞き取った音声を書き取るための解答欄を用意し、全ての書き取りが終了した後、見直しのための時間を3分間与えた。採点方法は、Fountain and Nation (2000, pp.32-33) に準じた。採点の対象となるターゲット語は各パラグラフに20語ずつ含まれ、導入パラグラフを除き、第1パラグラフから第4パラグラフまで各20点ずつ、合計80点満点で採点された。採点基準は、スペリングが正確に綴られている場合だけを正解とし、名詞の複数形の -s と -es、規則動詞過去形の -d と -ed は、語幹の綴りが正しい場合だけ許容した。

4. 結果

Table 1 は CELT、VLT、VLDT の平均点の推移をまとめた。CELT 平均点は回を重ねるごとに上昇し、1年を通して調査協力者の英語力が向上したことを示した。VLDT 総得点は1回目が一番低く、次いで3回目が2回目より低くなり4回目が一番高くなるという点で、VLT-2000、VLT-3000、VLT-Academic とほぼ同様の得点変化を示した。

Figure 1 は VLT2000と VLT3000および VLDT 総合点の1年を通じた成績の伸びについて、各テストの1回目平均点を100%としてまとめた。基本語彙である VLT の2,000語レベルの伸び代は約5%となった。VLT の3,000語レベルでは約15%、VLDT 総得点の平均点では約10%上昇した。VLT の2,000語レベルでは30点満点に近い平均点に達したため伸び代は少ないが、3,000語レベルの平均点は全体の6~7割程度であったため、1年の授業を通して約15%の伸びを示した。VLDT 総得点の平均点においても3,000語レベルと同様の伸びが認められた。Test-C (修正版) で難易度の調整をしたことにより今井・三根 (2011) で指摘された「Test-C は、4つの平行テストの中では極端に平均点が低く、基準関連妥当性を欠いている」(p.92) という問題点は解消された。

Table 1 CELT、VLT、VLDT の平均点の推移

	1	2	3	4	F(3/291)
CELT	27.72	29.06	31.31	33.44	63.44
VLT2000	25.00	26.73	25.92	26.95	24.62
VLT3000	19.17	21.90	20.16	22.40	34.18
VLT-Academic	19.35	20.01	19.92	20.86	6.13
VLT5000	16.33	15.60	16.97	16.12	4.05
VLDT-T	26.71	29.67	27.47	30.12	11.72

Table 2にはVLDTのパラグラフ別と総得点の α 係数を示した。その結果、パラグラフごとの α 値は0.51~0.74であり十分な信頼性を示す数値に至らなかった。一方、VLDTの総得点は0.85~0.89という高い α 値を示し十分な信頼性を示した。VLDTは「パラグラフ別の得点より総得点の方が語彙力の指標として適切である」という今井・三根(2011, p. 92)の結論を支持した。

VLDTの語彙レベルでの問題点を明らかにするために、JACET8000の語彙ランクに基づいて各ターゲット語を分類することでテストの構成概念について再検討を行った。[以下、()内はJACET8000の語彙レベルを示す。]

JACET8000を指標とすることで明らかになった点は、レベル4を超えると聞き取りが困難になることである。

Test-CとTest-Dにおいては、解答率0%がexceed (4)、2%がapproximately (4)、distribution (4)、rural (4)、providing (5)、striving (5)、3%がinevitable (4)、compel (5)、acquaintance (6)であった。また、レベルが低い単語でもスペリングが難しいものoccurred (1) 2%、occurs (1) 9%、debts (2) 9%、obvious (2) 1%、threaten (2) 2%と、複数形のものcosts (1) 0%、cities (1) 9%、experts (2) 3%、institutions (2) 3%などは解答率が極端に低かった。一方で、高いレベルの単語でも、reduction (4) 62%、quest (6) 52%、slender (6) 34%などは解答率が高かった。

Table 3は、各テスト形式におけるJACET8000のレベル1と2の語彙数を段落ごとに示したものである。本来、ターゲット語が4,000語レベル以上であるはずの第4段落

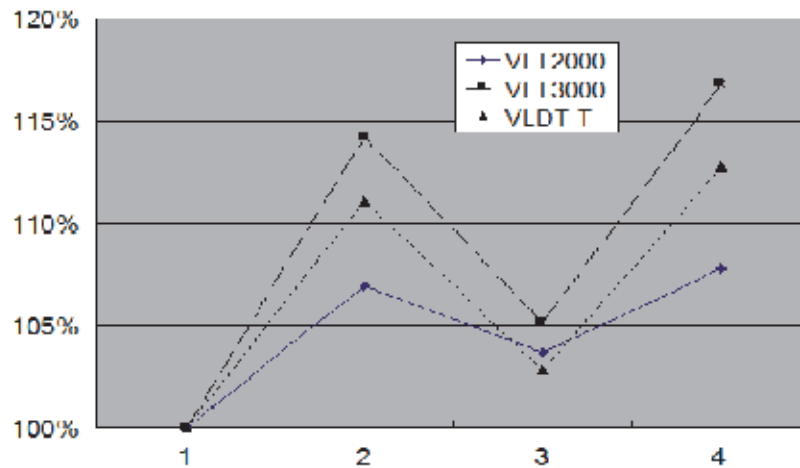


Figure 1. VLT と VLDT の成績の伸び。

Table 2 VLDT の α 係数

	Test-A	Test-B	Test-C	Test-D
Para-1	0.57	0.73	0.74	0.63
Para-2	0.58	0.73	0.66	0.58
Para-3	0.67	0.62	0.63	0.68
Para-4	0.51	0.58	0.61	0.59
VLDT-T	0.85	0.89	0.89	0.87

Table 3 段落ごとのレベル1-2語彙の数

	Test-A	Test-B	Test-C	Test-D
Para 1	20	20	20	18
Para 2	15	16	18	15
Para 3	10	7	8	10
Para 4	7	6	8	6
Total	52	49	54	49

においても、数多く含まれているレベル1の語彙に対しては正答可能であった。このことが、各段落の信頼性係数(α)を下げている原因の一つと推測された。また、第1段落において、レベル1～2の語彙の規則変化複数形はTest-AとTest-Cで6個、Test-Bで2個、Test-Dで0個であり、Test-AとTest-Cで平均点が低かった原因の一つになった。テスト形式ごとの平均点に影響を及ぼした別の要因として4音節以上の語彙の存在が挙げられる。4音節以上の語彙が最も多かったのはTest-Cの19個であった。それに対し最も4音節以上の語彙が少なかったのはTest-Dの10個であった。

5. 考 察

日本人EFL学習者にとって、発表的語彙力テストの側面を有するVLDTは適用が困難であると考えられてきたが、今回の研究結果は、語彙力の指標としてVLDTのパラグラフ別の得点ではなく全項目の総得点を採用する方が適切であることを再確認した。VLDTのTest-A修正版と同様にTest-Cの難易度を調整することで、VLDTが発表的語彙テストとして十分な信頼性と妥当性を確保した結果が得られた。VLTとVLDTを用いることで日本人EFL学習者の受容的および発表的語彙力の双方からの測定が可能であることを示した。

JACET8000を指標にした場合、日本人EFL学習者は4,000語レベルを超える単語はほとんど解答できないことが明らかになった。高いレベルの語彙は一般的に綴りが長く発音の規則性が低い。出現頻度の低い高レベル語を含んだリスニング教材の場合、日本人EFL学習者の動機づけや興味を維持することは困難であることが示唆される。一方で、スペリングが難しい単語では、高頻度語であっても子音連続の不規則性が見られる単語(occur)、黙字を含む単語(debt)、子音が連続するとともに母音の発音が不規則な単語(threaten)などの綴り字が困難であった。また、レベル6以上の単語でもquestやslenderのように、近年の流行の影響でインプットする機会が多くなった単語は正確に綴れる学習者が増えてきた。

日本人EFL学習者において、ディクテーション形式の語彙テストでは、1)高レベルの語彙のヒアリング、2)発音と綴り字が不規則に対応する単語のスペリング、3)複数形の認識、4)長い音節数の語彙のヒアリング、などに困難が見いだされた。とくに、複数形の困難性は、日本人EFL学習者がディクテーションにおいて文脈的・

文法的処理を行わず、単語を逐語的に処理していることを示唆している。

6. 課題と展望

今回の研究では、Test-A修正版と同様にTest-Cの難易度を調整することで、VLDTのTest-A～Dの4つのテストは日本人EFL学習者の発表的語彙テストとしての信頼性があることが確認された。同時にVLDTは受容的スキルであるヒアリングに依存する側面が高いため純粋な発表語彙力を測定できているかどうか疑念がある。Laufer and Nation (1999)は、VLTをベースに、文脈から正しい語彙を推測し、数文字の頭字手がかりを基にスペリングを完成させる発表的スキルの測定を目的とした統制的発表語彙サイズテストを開発している。今後はより発表的スキルへの依存度が高い、頭字手がかりのみによる発表的語彙テストのEFL学習者への実用性の検証を試みたい。

注：この研究は、外国語教育メディア学会(LET)第52回全国研究大会(甲南大学、2012年8月)で報告した研究発表を加筆修正したものである。

引用文献

- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- 大学英語教育学会基本語改訂委員会(編)(2003).『大学英語教育学会基本語リスト(JACET8000)』大学英語教育学会.
- Fountain, R. L., & Nation, I. S. P. (2000). A vocabulary-based graded dictation test. *RELC Journal: Guidelines*, 1, 76-80.
- 古荘智子・今井由美子・三根浩(2009).「大学生における語彙力と英語標準テストの関連性(3) — ディクテーションテストと語彙力 —」.『第49回外国語教育メディア学会全国研究大会』146-147.
- 今井由美子・三根浩(2009).「大学生における語彙力と英語標準テストの関連性(2) — 3000語レベルとアカデミック語レベルの比較 —」. *Asphodel*, 44, 189-205.
- 今井由美子・三根浩(2010).「語彙レベルディクテーションテストの日本人学習者への適用」. JACET 関西紀要, 12, 56-67.
- 今井由美子・三根浩(2011).「ディクテーション形式によ

- る語彙レベルテストの妥当性の検討」. *Language Education & Technology*, 48, 83-94.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Laufer, B. & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4, 142-151.
- 望月正道 (1998). 「日本人学習者のための英語語彙サイズテスト」. 『語学教育研究所紀要』 12, 27-53.
- 望月正道 (2003). 「語彙はどうテストするのか」. 望月正道・相澤一美・投野由紀夫 (編), 『英語語彙の指導マニュアル』 大修館書店, 181-209.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the vocabulary levels test. *Language Testing*, 18 (1), 55-88.
- 園田勝英 (1996). 「大学生用英語語彙表のための基礎的研究」. 『言語文化部研究報告叢書』 7, 北海道大学言語文化部.
- Thorndike, E. L. & Lorge, I. (1944). *A teacher's word book of 30,000 words*. Teachers College, Columbia University.
- Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata; The Occasional Papers of Notre Dame Seishin University*, 53-68.
- West, M. (1953). *A general service list of English words*. Longman, Green & Co.

Appendix A VLDT Test-C の冒頭部分

Test C

This is a dictation test. You will hear a passage read once. Write what you hear. Remember, the passage will only be read once. It will not be repeated. Are you ready to begin writing?

INTRODUCTORY PARAGRAPH

During the last twenty years/ many people left the country/ to work in the towns./

PARAGRAPH 1

Various reasons are advanced/ to explain this movement./ The use of machines on farms/ has reduced the demand for workers./ while the growth of industry/ has increased the jobs in the cities./ This partly accounts for the change./

(Source : Fountain and Nation, 2000, p. 43より).

Appendix B VLDT Test-C の日本人 EFL 学習者用解答用紙の冒頭部分

ディクテーションテスト (C)

Class: _____ ID: _____ 氏名(漢字): _____ Booth#: _____

以下の指示を聞いて、Introductory Paragraph から最後まで、全文を書き取ってください。英文は区切って、1 度だけ読まれます。区切りのあとに書き取りのための時間がとってあります。最後にスペル確認のための時間を用意しますので、今は消しゴムや修正ペンを使わないようにして、書き間違いなどは後で修正してください。解答用紙の () の番号は区切りの位置を示しています。() 中の番号は読まれません、区切りごとに出来るだけ速く書き取ってください。

This is a dictation test. You will hear a passage read once. Write what you hear. Remember, the passage will only be read once. It will not be repeated. Are you ready to begin writing?

Introductory Paragraph (1).....

(2)..... (3).....

Paragraph 1 (1)..... (2).....

(3).....

(4).....

(5).....

(6).....

(7).....

