

# C テストの英語能力測定度

——語彙力、文法力、読解力および聴解力テストとの比較——

森 永 弘 司

## Abstract

The aim of this research paper is to examine the validity and reliability of C-test as a device for measuring English language proficiency. I compared C-test data with those of Vocabulary Levels Test (hereafter VLT), Vocabulary Size Test (VST), Productive Levels Test (PLT), Word Associates Test (WAT), The 7th version of Oba's Standard English Grammar Test of (EGT) and reading (RCT) and listening comprehension tests (LCT). I obtained the following correlation coefficient between C-test and the other seven tests mentioned above:  $r = .62$  with VLT,  $= .50$  with VST,  $= .65$  with PLT,  $= .50$  with WAT,  $= .62$  with EGT,  $= .61$  with RCT,  $= .71$  with LCT.

Although such standardized tests as TOEIC or TOEFL exhibit high validity and reliability in measuring examinees' English proficiency, it takes a great deal of time and money to sit for them. Although C-test does not match either TOEIC or TOEFL in validity or reliability, we can give C-test for only ten minutes and it is quite easy to prepare. I completely agree with Dörnyei & Katona's (1992) evaluation of C-test "not only as a reliable and valid measure of general language proficiency, but also as one of the most efficient language testing instruments in terms of the ratio between resources invested and measurement accuracy obtained."

## 1. 本稿の目的

近年学生の英語力に応じた効果的な指導をおこなうためにプレイスメント・テストを実施し、能力別クラス編成で授業をおこなう大学が増加してきている。また授業効果や学生の授業目標の達成度を検証するために学期末に

TOEIC や TOEFL のような標準テストで学生の英語力を測定する大学も増えてきている。しかしながらテストングを専門とする教員がいない大学で、信頼性・妥当性の高い英語力測定テストはもちろんのことプレイスメント・テストでさえも開発するにはかなりの困難が伴う。また TOEIC のような標準テストは信頼性・妥当性に関しては非常に高いといえるが、実施時間や費用の点で利便性に欠けるといえよう。信頼性・妥当性が高くかつ簡便性に優れたテストが理想といえるが、これらの条件を十分に満たしたテストは未だ開発されていない。

本稿の研究目的は総合的英語能力を判定するテストとして開発された C-test を 2 つの私立大学で実施した際のデータと、語彙力、文法力、読解力および聴解力を測定するために実施した諸テストのデータを回帰分析及び相関分析の視点から分析を試みることで、C-test が一般的な英語力を測定するうえでどの程度の有効性があるのかを検証することにある。

## 2. 先行研究

C-test の信頼性・妥当性の検証は、様々なテストとの比較でおこなわれてきた。Negishi (1987) はかつてアメリカの大学や大学院に留学するための語学テストとしてよく使用されていた Michigan English Language Assessment Battery (MELAB) と 4 つの異なったテキストから構成される 100 の空所の C-test を比較し、C-test が MELAB の文法部門との相関係数が .56、語彙部門との相関係数が .62、読解部門との相関係数が .80、テスト全体との相関係数が .76 であったことを報告している。Negishi はこの調査結果から、C-test が言語の基本的要素（文法、語彙）だけを測っているという Alderson (1979) の見解に反駁し<sup>1</sup>、C-test が L2 学習者の総合的言語能力を測っていると主張している。

Mochizuki (1994) は、STEP test の 2 級と A Comprehensive English Language Test for Learners of English (CELT) の Listening test、dictation

test と explanation、argumentation、description、narration の4つの異なる C-test を使用し、4つの C-test の中でどの C-test が一番信頼性が高いかの検証を試みた。その結果クロンバック・アルファ係数が最も高いのが narration を題材とする C-test で、信頼係数は  $\alpha = .93$  であった。またこの C-test と STEP test との相関係数は .44、CELT との相関係数は .003、dictation との相関係数は .30 であることが判明した。

Ikeguchi (1998) は1年以上の英語圏での生活経験のある大学生群と生活経験のない大学生群の2つの参加者グループを対象に、4つの異なったテキストから構成される100の空所の C-test と STEP test の得点との相関係数を調べる実験をおこなった。その結果1年以上の英語圏での生活経験のある大学生群の相関係数が .51、生活経験のない学生群の相関係数が .58 であったと述べている。また Ikeguchi は C-test が TOEFL の文法部門の得点との間に高い相関性が見られたことを報告している。

Dörnyei and Katona (1992) は、ハンガリーの大学の英語を専攻する1年生と高校生の2つの参加者グループを対象に C-test の信頼性を検証する実験をおこなった。大学生のグループに対しては大学独自に開発した語彙力テスト、文法力テスト、聴解力テストの3つの部門から構成される英語総合力テスト、TOEIC、会話力テスト、38の空所から構成される cloze test および91の空所から構成される C-test を実施し、高校生のグループに対しては TOEIC および C-test を実施した。その結果 C-test と語彙力テストとの相関係数が .38、文法力テストとの相関係数が .25、聴解力テストとの相関係数が .33、テスト全体との相関係数が .43、TOEIC のリスニング部門との相関係数が .51、リーディング部門との相関係数が .54、TOEIC 全体との相関係数が .62、会話力テストとの相関係数が .43、cloze test との相関係数が .33、上記の全てのテストとの相関係数が .57 であったことを報告している。また2人は C-test の方が cloze test よりも総合的な英語能力を測定するうえで、信頼性および妥当性の点でも高いと結論付けている。C-test と

C-test 以外の cloze test を比較した西田 (1989) も、テストの妥当性に関しては、C-test がそれ以外の cloze test より優れているとはいえないが、テストの信頼性に関しては cloze test よりも高いと報告している。Dörnyei & Katona は総合的な英語力を測定するうえで C-test を非常に高く評価しており、1992年の論文を次のような賞賛の言葉で結んでいる。“our conclusion about the C-test is that not only a reliable and valid measure of general language proficiency, but it is also one of the most efficient language testing instruments in terms of the ratio between resources invested and measurement accuracy obtained.”

筆者の今回の研究は、Dörnyei & Katona の響に倣い、多くのテストを援用することで、C-test の総合的英語力測定テストとしての信頼性の検証を試みたものである。

### 3. 実験内容

#### 3.1. 参加者

2つの私立大学の学生を対象に調査をおこなった。1つは共学の総合大学の1、2年生190名で、所属学部は、文学部（国文科、文化史学科）、社会学部、法学部、経済学部、商学部、文化情報学部、理工学部と多岐にわたる。もう一方は女子大学の1年生で現代社会学部と学芸学部に所属する81名の学生である。英語力を測定するうえで評価の高いTOEICやTOEFL等の標準テストは実施できなかったが、読解力と強い相関性が有ることが知られている受容語彙数の平均は共学の大学の参加者190名で4,643語、女子大の参加者81名で3,647語であった<sup>2</sup>。ちなみにEFLの学習者の場合、受容語彙数5,000語以上を読解力上級者、3,000語以上を中級者として解釈するのが一般的なので、共学の大学は全体としてみた場合、中級の上位レベル、女子大学の場合は中級の下位レベルのクラスとみなすことができる。なおテストの受験の際に、参加者に対してテストの目的（なぜそのデータを集めているのか、そ

れをどのように利用するのか) を十分説明して、データ使用に関して了承を得ている。

### 3.2. データ収集の手続き

2008年度と2009年度に担当していたリーディングクラスとリスニングクラスで5～6種類のテストを実施した。2008年度のリーディングのクラスでは、C-test、受容語彙数を推定するためのテストである Vocabulary Levels Test (以下 VLT と略す)、受容語彙数を測定するためのテストである Vocabulary Size Test (以下 VST と略す)、発表語彙数を測定するためのテストである Productive Levels Test (以下 PLT と略す)、語知識の深さを推定するためのテストである Word Associates Test (以下 WAT と略す) 及び読解力測定テストを実施した。同年のリスニングクラスでは C-test、4種類の語彙力測定テストと聴解力測定テストを実施した。2009年度のリーディングクラスでは、C-test、VLT、VPT、文法力テスト及び読解力テストを実施した。同年のリスニングクラスでは2つのクラスで以下のテストのデータの収集をおこなった。1つのクラスでは、C-test、VLT、PLT、WAT、文法力テスト及び聴解力テストを実施した。他方のクラスでは、C-test、VLT、PLT、文法力テスト及び聴解力テストを実施した。解答時間に関しては、C-test は10分、VLT は45分、VST は60分、PLT は50分、WAT は30分、文法力テストは30分とした。読解力測定テストに関しては90分で、聴解力測定テストは45分程度で実施した。

### 3.3. 使用したテストに関して

C-test は cloze タイプのテストで、言語の余剰性 (メッセージを伝える場合、その意味が理解されるのに必要な最低限度よりも多くの言語情報が含まれている度合い) を利用して作成されたテストである<sup>3</sup>。言語の余剰性が削減されると (例えば、印刷のある箇所が薄くて判読しにくいような場合)、

文意を推測するのが困難になってくる。cloze test はこの言語の余剰性を意図的に削減することで、総合的言語能力を測定しようと試みたテストである。C-test は1984年に Klein-Braley & Raatz によって考案されたもので、最初の文をそのまま残し、第2文の第2語から1語おきに語の後半部分（偶数文字数からなる単語は半分の文字数、奇数文字数からなる単語は前半部より1文字だけ多い文字数）を削除し、その部分を参加者に補わせる。Klein-Braley & Raatz が作成した C-test の1例を参考資料として掲載しておく。テスト作成の要点は以下の3項である。1) 異なる分野から通常5～6種類のテキストを使用して問題を作成する。2) 少なくとも合計100の削除箇所が必要だとされている。3) 実施に先立って、母語話者が満点を取れることを確認する。また採点は基本的に正語法（元のテキストにある語だけを正答とする）にもとづいておこなう。今回の研究調査では JACET 関西支部のライティング指導研究会によって開発された C-test を使用させていただいた<sup>4</sup>。この C-test は4つの異なったテキストを使用し、50項目より構成されている。このテストの信頼性を測定した Yoshimura (1999) は、クロンバック・アルファ係数が  $\alpha = .84$  であったことを報告している。採点は元のテキストにある語だけを正答とみなす正語法を採用し、1項目の正解で2点、計100点を満点とした。

VLT は読んで意味の理解できる受容語彙数 (receptive vocabulary) を測定するためのテストであり、2,000語、3,000語、5,000語、大学語彙（語彙数の算出には含めない）、10,000語の5つのレベルで構成されており、各レベル30問（大学語彙は36問）で総計156問から構成される。元々は受容語彙数の測定ではなく、受容語彙レベルの推定を意図して開発されたテストであるが、語彙数の測定に関して現在もこれを凌駕する信頼性の高いテストは開発されていないため、語彙力測定テストとして広く使用されている。設問は3つの単語の語義にあてはまる単語を、6つの単語の選択肢の中から選ばせる組み合わせ形式のものである。問題の1例を挙げておく。

1. copy
2. event — end or highest point
3. motor — this moves a car
4. pity — thing made to be like another
5. profit
6. tip

語彙数は屈折形や派生語を base form にひとまとめにして考える基本形 (word family) 方式に基づいて算出される。

VST は VLT の考案者の Nation & Beglar (2007) が開発した受容語彙数測定テストであるが、測定テストとしての評価は芳しくなく、語彙研究でも使用されることが少ない。このテストは1,000語レベルから1,000語単位で14,000語レベルまでの14レベルで構成されている。各レベルの設問数は10問で、14,000語レベルまで含めると140問になる。設問は文中の太字の単語とほぼ同じ意味の単語を4つの語(句)の中から選ばせる多肢選択形式のものである。問題の1例を挙げておく。

SEE : They **saw** it.

- a. cut
- b. waited for
- c. looked for
- d. started

PLT は発表語彙数 (productive vocabulary) を測定するためのテストで、Laufer & Nation (1999) によって考案された。VLT と同じく、2,000語、3,000語、5,000語、大学レベル (VLT と同様に語彙数算出に含めない)、10,000語レベルの5つのレベルで構成されている。各レベルは18問より成り、次の例のように完成させるべき英単語の最初の数文字が提示され、回答者は

文脈とこの数文字を手がかりにして単語を完成させる形式のものである。

1. I'm glad we had this oppo \_\_\_\_\_ to talk.
2. There are a doz \_\_\_\_\_ eggs in the basket.

WAT は語知識の深さを測る目的で Read (2000) によって開発されたテストである。問題は1つの形容詞と連合的關係にある語（同義語）を枠内の下の語群から1～3語と統語的關係にある語（コロケーションを形成する語）を右の語群から1～3語、合計で4語選ぶ形式のものである。左右の枠内の単語から最低1語選ぶことになっている。問題数は40ある。問題の1例を挙げておく。

strange

|                                    |                            |
|------------------------------------|----------------------------|
| beautiful quick surprising thirsty | change doctor noise school |
|------------------------------------|----------------------------|

この場合、連合的連想語は quick と surprising で、統語的連想語は change と noise で、この4語が正解ということになる。WAT に関しては、一般に語彙力上級者の場合、読解力の相関性は受容語彙数よりも語知識の深さの方が強いという研究が報告されており<sup>5</sup>、筆者の研究でも VLT が3,000語以上の学習者の場合、読解力と最も相関性が強いのが語知識の深さであることが判明した<sup>6</sup>。

文法力テストは大場（2008）が考案した60問から構成される標準英文法テスト第7f版を使用した。設問は会話文の下線の空所に3つの語（句）から1つを選ぶか、3つの語（句）のいずれにもあてはまらない場合は、none of the above の項を選ぶ多肢選択形式のものである。問題の1例を挙げておく。

A : The car \_\_\_\_\_ into the wall

B : How terrible !



- (a) crashed
- (b) crash
- (c) was crash
- (d) none of the above

大場の評価基準では、60問中高校生が40点以上、大学生が45点以上の得点で「合格」で、55点以上で「免許皆伝」となっている。

読解力測定テストの問題は、文中の10～12箇の空所に入る適語を4つの単語の中から選ばせる多肢選択式の空所補充問題、scanning で特定の情報を探しだす問題、および skimming でパラグラフの要旨に関して答えさせる記述問題から構成されている。1回目のテストでは2種類のテーマの異なる科学記事を扱ったテストを解答時間90分で実施した。2回目のテストでは初回とは異なる2種類のテストを同じく90分で実施した。この4つのテストのリーダビリティは Flesch Reading Ease で63～38、Flesch-Kincaid Grade Level で8th～15th なので、読解の難易度は中位から上位の問題といえるであろう。この読解力テストのクロンバック・アルファ係数は $\alpha = .86$ であった。

聴解力測定テストの問題はアメリカのCNNの科学番組 Science Desk のスクリプトの15ヶ所の空欄の単語を聞いて書き取らせる cloze 形式のテストである。Science Desk では10の異なったトピックが扱われている。番組を収録したDVDを5回視聴させ、15ヶ所の空所の単語を書き取らせた。1回のテストで2つの番組を視聴させ解答させるテストを2回実施した。この聴解力テストのクロンバック・アルファ係数は $\alpha = .82$ であった。

2回目の調査（2009年の後期の授業で実施）では、読解力テストを実施した116名と、聴解力テストを実施した2クラスの49名と20名の参加者を対象にデータを収集した。読解力クラスでは読解力テストの他に C-test および語彙力テストとして VLT、PLT を実施したが、VST と WAT は実施しなかった。読解力測定テストは初回の調査と同じ問題を使用した。聴解力測定

テストは調査対象の2クラスのうち20名のクラスでは初回と同じトピックを扱ったものを使用した、49名のクラスでは異なったトピックを扱ったテストを実施した。

## 4. 結 果

### 4.1. 1 回目の調査結果

1 回目の C-test の得点と読解力テストおよびその他の諸テストの得点との相関性を調べるために、ステップワイズ法（変数増減法）による重回帰分析をおこなった。今回の研究の目的は、受容語彙力を測定するうえで精度の高い VLT、VST、発表語彙力を測定するうえで精度の高い PLT、語知識の深さを測定するうえで精度の高い WAT、読解力と聴解力の測定を主眼として開発された読解力テストと聴解力テストが、どの程度の割合で C-test の得点を説明しえるかにある。従って今回の研究では C-test の得点を従属変数（目的変数）、その他の諸テストの得点を独立変数（説明変数）とする重回帰分析をおこなった。以下の重回帰分析も全てステップワイズ法にもとづく。これらの変数の記述統計量および相関係数は表 1、重回帰分析の結果は表 2 の通りである。参加者は57名であった。標準残差の絶対値が2を超える2つのデータは外れ値として調査から除いた。また以下全ての調査の相関係数が有意であることを上田（1997）による簡便法で検定をおこなって確認した<sup>7</sup>。

表 1. 1 回目の調査の記述統計量と C-test の得点と読解力テストおよびその他のテストの得点との相関係数

| 変 数         | 平均値   | 標準偏差  | 1   | 2   | 3   | 4   | 5   |
|-------------|-------|-------|-----|-----|-----|-----|-----|
| 従属変数：C-test | 57    | 15    | .63 | .61 | .79 | .66 | .64 |
| 独立変数：1. VLT | 4,986 | 1,053 |     | .74 | .73 | .69 | .62 |
| 2. VST      | 6,573 | 1,152 |     |     | .70 | .67 | .60 |
| 3. PLT      | 1,973 | 751   |     |     |     | .71 | .71 |
| 4. WAT      | 111   | 9     |     |     |     |     | .55 |
| 5. 読解力テスト   | 86    | 19    |     |     |     |     |     |

表 2. C-test の得点を従属変数とした重回帰分析の結果

| 変 数 | 偏回帰係数    | 標 準 誤 差  | 標準偏回帰係数 |
|-----|----------|----------|---------|
| PLT | .012931  | .00237   | .6440   |
| WAT | .325     | .187824  | .2042   |
| 切片  | -4.12334 | 17.86097 |         |

標準偏差の差がかなり大きいですが、これはそれぞれのテストの満点の差が大きいところに起因する。C-test と聴解力テストの満点は 100、VLT と PLT は 10,000、VST は 14,000、WAT は 160、文法力テストは 60、読解力テストは 120 である。

重回帰式の決定係数は  $R^2 = .64$  であった。決定係数  $R^2$  (寄与率) は独立変数 (説明変数) 全体が従属変数 (目的変数) をどの程度まで予測できるかを表す数値である。決定係数の明確な統一基準はないが、予測の精度に関しては  $r = .80$  以上が「非常に良い」、 $r = .50$  以上が「やや良い」、 $r = .40$  未満が「良くない」と考えることができるといわれている<sup>8</sup>。従ってこの調査の決定係数  $R^2 = .65$  は予測値として「良い」というレベルと解釈できるであろう。重回帰式の結果から、C-test の得点 =  $.012931 \times \text{PLT}$  による発表語彙数 +  $.325 \times \text{WAT}$  による語知識の深さ - 4.1234 という一次式が導き出せる。また標準偏回帰係数から、C-test の得点を説明する要因は PLT (.644) と WAT (.2090) で、発表語彙を測定する PLT と最も強い関連性があることがわかる。

次に C-test との単相関係数を見てみると、C-test との相関性が最も高いのが PLT との間であり、相関係数は  $r = .79$  であった。次いで WAT との間で、相関係数は  $r = .66$  であった。次が読解力テストとの間であり、相関係数は  $r = .64$  であった。相関係数に関しても決定係数、同様、統一的な基準はないが、 $r = .70 \sim 1.0$  (または  $r = -.70 \sim -1.0$ ) なら「かなり強い関連 (逆の関連)」がある、 $r = .40 \sim .70$  (または  $r = -.40 \sim -.70$ ) なら「中程度の関連 (逆の関連)」、 $r = .20 \sim .40$  または  $r = -.20 \sim -.40$  が「弱い関

連（逆の関連）」があるとされ、 $r=.20\sim.20$ が「ほとんど関連なし」と解釈されることが多い<sup>9</sup>。したがって PLT、WAT 及び読解力テストとの間の相関係数は  $r=.60$  超えているので、C-test はそれぞれのテストの代替テストとして使用しうるレベルにあると考えられる。

次に1回目のC-testの得点と聴解力テストおよびその他の諸テストの得点との相関性を調べるために、C-testの得点を従属変数、その他の諸テストの得点を独立変数とした重回帰分析をおこなった。これらの変数の記述統計量および相関係数は表3-1、重回帰分析の結果は表3-2の通りであった。参加者は49名であったが、2つの外れ値はデータから除外した。

表3-1. 1回目の調査の記述統計量とC-testの得点と聴解力テストおよびその他のテストの得点との相関係数

| 変 数         | 平均値   | 標準偏差 | 1   | 2   | 3   | 4   | 5   |
|-------------|-------|------|-----|-----|-----|-----|-----|
| 従属変数：C-test | 45    | 13   | .61 | .38 | .55 | .44 | .69 |
| 独立変数：1. VLT | 3,790 | 991  |     | .34 | .61 | .49 | .57 |
| 2. VST      | 4,829 | 996  |     |     | .44 | .25 | .33 |
| 3. PLT      | 1,311 | 597  |     |     |     | .55 | .65 |
| 4. WAT      | 95    | 15   |     |     |     |     | .57 |
| 5. 聴解力テスト   | 49    | 16   |     |     |     |     |     |

表3-2. C-testの得点を従属変数とした重回帰分析の結果

| 変 数    | 偏回帰係数    | 標準誤差     | 標準偏回帰係数 |
|--------|----------|----------|---------|
| VLT    | .004109  | .001597  | .3169   |
| 聴解力テスト | .410675  | .099775  | .5067   |
| 切片     | 9.177299 | 5.323501 |         |

重回帰式の決定係数は  $R^2=.54$  であった。この重回帰式の従属変数の予測値は「やや良い」と解釈できるレベルである。重回帰式の結果から、C-testの得点 $=.004109 \times \text{VLT}$ による受容語彙数 $+.410675 \times \text{聴解力測定テストの得点}$  $-4.12334$ という一次式が導き出せる。また標準偏回帰係数から、

C-test の得点を説明する要因は聴解力テスト (.5067) と VLT (.3169) で、聴解力測定テストと最も強い関連性があることがわかる。

単相関係数から見た場合、C-test との相関性が最も高いのが聴解力テストとの間であり、相関係数は  $r = .69$  であった。次いで VLT との間で、相関係数は  $r = .61$  であった。次が PLT との間であり、相関係数は  $r = .55$  であった。聴解力測定テスト及び VLT との間相関係数は  $r = .60$  超えているので、それぞれのテストの代替テストとして使用しうると考えられる。

また PLT に関しても使用に耐えうるレベルにあると解釈できる。

#### 4.2. 2 回目の調査結果

2 回目の C-test の得点と読解力テストおよびその他の諸テストの得点との相関性を調べるために、C-test の得点を従属変数、その他の諸テストの得点を独立変数とした重回帰分析をおこなった。これらの変数の記述統計量および相関係数は表 4-1、重回帰分析の結果は表 4-2 の通りであった。参加者は115名であったが、1つの外れ値はデータから除外した。

表 4-1. 2 回目の調査の記述統計量と C-test の得点と読解力テストおよびその他のテストの得点との相関係数

| 変 数         | 平均値   | 標準偏差 | 1   | 2   | 3   | 4   |
|-------------|-------|------|-----|-----|-----|-----|
| 従属変数：C-test | 53    | 13   | .59 | .63 | .60 | .57 |
| 独立変数：1. VLT | 4,231 | 981  |     | .68 | .66 | .63 |
| 2. PLT      | 1,590 | 590  |     |     | .72 | .65 |
| 3. 文法力テスト   | 39    | 7    |     |     |     | .66 |
| 4. 読解力テスト   | 78    | 20   |     |     |     |     |

重回帰式の決定係数は  $R^2 = .48$  であったので、従属変数の予測値としては「やや良い」と解釈できるレベルである。重回帰式の結果から、C-test の得点 =  $.002619 \times \text{VLT}$  による受容語彙数 +  $.006169 \times \text{PLT}$  による発表語彙数 +  $.307675 \times \text{文法力テストの得点}$  +  $.107177 \times \text{読解力テストの得点}$  +

表 4-2. C-test の得点を従属変数とした重回帰分析の結果

| 変 数    | 偏回帰係数    | 標 準 誤 差  | 標準偏回帰係数 |
|--------|----------|----------|---------|
| VLT    | .002619  | .001382  | .1937   |
| PLT    | .006192  | .002472  | .2757   |
| 文法力テスト | .307675  | .202744  | .1660   |
| 読解力テスト | .107177  | .064994  | .1640   |
| 切片     | 12.03925 | 5.522501 |         |

12.03925 という一次式が導き出せる。また標準偏回帰係数から、C-test の得点を説明する要因として一番大きいのが PLT (.2757)、次いで VLT (.1977)、最も小さいのが読解力テスト (.1640)、次に小さいのが文法力テスト (.1660) で、発表語彙を測定する PLT と最も関連性があることがわかる。

単相関係数から見た場合、C-test との相関性が最も高いのが PLT テストとの間であり、相関係数は  $r = .63$  であった。次いで C-test と相関が高いのは文法力テストで  $r = .60$ 、VLT との相関係数は  $r = .59$  であった。最も低いのが読解力テストとの間であり、相関係数は  $r = .57$  であった。最も相関係数の高い PLT と、最も相関係数の低い読解力テストとの差はわずか .06 であった。今回の調査では C-test が全てのテストの代替テストとしてかろうじて使用できるレベルにあると解釈できる。

2 回目の聴解力テストは 1 回目で使用したのと形式は同じであるが、扱ったトピックが異なるテストを用いて 2 つのクラスで実施した。49 名のクラスで実施した C-test の得点と、聴解力テストおよびその他の諸テストの得点との相関性を調べるために、C-test の得点を従属変数、その他の諸テストの得点を独立変数とした重回帰分析をおこなった。これらの変数の記述統計量および相関係数は表 5-1、重回帰分析の結果は表 5-2 の通りであった。3 つの外れ値はデータから除外した。

表 5-1. 2 回目の調査の記述統計量と C-test の得点と聴解力テストおよびその他のテストの得点との相関係数

| 変 数         | 平均値   | 標準偏差 | 1   | 2   | 3   | 4   | 5   |
|-------------|-------|------|-----|-----|-----|-----|-----|
| 従属変数：C-test | 44    | 10   | .63 | .60 | .40 | .61 | .66 |
| 独立変数：1. VLT | 3,556 | 918  |     | .64 | .40 | .58 | .59 |
| 2. PLT      | 1,187 | 497  |     |     | .58 | .66 | .66 |
| 3. WAT      | 103   | 12   |     |     |     | .44 | .46 |
| 4. 文法力テスト   | 36    | 7    |     |     |     |     | .53 |
| 5. 聴解力テスト   | 56    | 16   |     |     |     |     |     |

表 5-2. C-test の得点を従属変数とした重回帰分析の結果

| 変 数    | 偏回帰係数    | 標準誤差     | 標準偏回帰係数 |
|--------|----------|----------|---------|
| VLT    | .002674  | .001454  | .2530   |
| 文法力テスト | .383981  | .18554   | .2704   |
| 聴解力テスト | .218385  | .079486  | .3636   |
| 切片     | 8.382681 | 5.464688 |         |

重回帰式の決定係数は  $R^2 = .56$  で、この重回帰式の予測値は、「やや良い」と解釈できるレベルである。重回帰式の結果から、C-test の得点 =  $.002674 \times \text{VLT}$  による受容語彙数 +  $.383981 \times \text{文法力テストの得点}$  +  $.218385 \times \text{聴解力テストの得点}$  +  $8.382681$  という一次式が導き出せる。また標準偏回帰係数から、C-test の得点を説明する要因として一番大きいのが聴解力テスト (.3636)、次いで文法力テスト (.2704)、最も小さいのが VLT (.2530) で、聴解力テストと関連性が最も深いことがわかる。

単相関係数に関しては、WAT との間の  $r = .40$  を除けば、VLT、PLT、文法力テスト、聴解力テストとも全て  $r = .60$  を超える数値であった。今回の C-test は WAT を除くテストの代替として使用できるレベルにあると解釈できる。

他方の20名のクラスで実施した C-test の得点と聴解力テストおよびその他の諸テストの得点との相関性を調べるために、C-test の得点を従属変数、

その他の諸テストの得点を独立変数とした重回帰分析をおこなった。これらの変数の記述統計量および相関係数は表 6-1、重回帰分析の結果は表 6-2 の通りであった。外れ値は検出されなかった。

表 6-1. 2 回目の調査の記述統計量と C-test の得点と聴解力テストおよびその他のテストの得点との相関係数

| 変 数         | 平均値   | 標準偏差 | 1   | 2   | 3   | 4   |
|-------------|-------|------|-----|-----|-----|-----|
| 従属変数：C-test | 53    | 16   | .66 | .68 | .65 | .78 |
| 独立変数：1. VLT | 4,209 | 908  |     | .74 | .65 | .82 |
| 2. PLT      | 1,799 | 654  |     |     | .83 | .70 |
| 3. 文法力テスト   | 40    | 7    |     |     |     | .50 |
| 4. 聴解力テスト   | 66    | 18   |     |     |     |     |

表 6-2. C-test の得点を従属変数とした重回帰分析の結果

| 変 数    | 偏回帰係数    | 標準誤差     | 標準偏回帰係数 |
|--------|----------|----------|---------|
| 聴解力テスト | .540687  | .136964  | .6067   |
| 文法力テスト | .827274  | .366532  | .3469   |
| 切片     | -15.7487 | 12.89152 |         |

重回帰式の決定係数は  $R^2 = .70$  であった。これは「かなり良い」予測値であるといえる。重回帰式の結果から、C-test の得点 =  $.540687 \times$  聴解力テストの得点 +  $.827274$  文法力テストの得点 -  $15.7487$  という一次式が導き出せる。また標準偏回帰係数から、C-test の得点を説明する要因は聴解力テスト (.6067) と文法力テスト (.3469) で、聴解力テストと非常に強い関連性があることがわかる。

単相関係数に関しては、最も高いのが聴解力テストとの間であり、相関係数は  $r = .78$  であった。次いで PLT との間で、相関係数は  $r = .68$  であった。次が VLT との間で、相関係数は  $r = .65$  であった。最も低いのが文法力テストとの間であり、相関係数は  $r = .65$  であった。この調査では C-test が他の全てのテストの代替になりうる事が判明した。



## 5. まとめと考察

2回実施した C-test とその他の諸テストの標準回帰係数と単相関係数の結果を表 7-1 と表 7-2 にまとめてみる。

表 7-1. C-test とその他諸テストとの標準回帰係数 (括弧内は順位)

|      | VLT                    | VST | PLT         | WAT       | 文法力                                   | 読解力       | 聴解力                        |
|------|------------------------|-----|-------------|-----------|---------------------------------------|-----------|----------------------------|
| 1 回目 | .3169 (2)*             |     | .6440 (1)** | .2042 (2) |                                       |           | .5067 (1)**                |
| 2 回目 | .1937 (2)<br>.2530 (3) |     | .2757 (1)*  |           | .1660 (3)<br>.2704 (2)*<br>.3969 (2)* | .1640 (4) | .3636 (1)**<br>.6067 (1)** |
| 平均   | .2545                  | 0   | .4599       | .2042     | .2778                                 | .1640     | .4923                      |

\*  $p < .05$ ; \*\*  $p < .01$

表 7-2. C-test とその他諸テストとの単相関係数

|      | VLT               | VST        | PLT               | WAT        | 文法力               | 読解力 | 聴解力        |
|------|-------------------|------------|-------------------|------------|-------------------|-----|------------|
| 1 回目 | .63<br>.61        | .61<br>.38 | .79<br>.55        | .66<br>.44 |                   | .64 | .69        |
| 2 回目 | .59<br>.63<br>.66 |            | .63<br>.60<br>.68 | .40        | .60<br>.61<br>.65 | .57 | .66<br>.78 |
| 平均   | .62               | .50        | .65               | .50        | .62               | .61 | .71        |

表 7-1 から C-test が聴解力テストおよび PLT とかなり強い関連性があることがわかる。また単相関係数も聴解力テストとの間が .71、PLT との間が .65 という強い相関性が認められるので、C-test はこの 2 つのテストの代替テストとして使用できる可能性が一番高いといえる。また文法力テストと VLT の標準回帰係数も比較的高く、単相関係数も共に .62 あるので、この 2 つの代替テストとして C-test が使用できる可能性もあるとみなしうる。WAT と読解力テストの代替テストの可能性に関しては、当確線上にあると

考えられる。WAT は標準回帰係数の点では .2042 なので C-test を代替テストとしてかろうじて使用できるレベルにあるが、単相関係数は .50 なので代替テストとして使用するには若干苦しいレベルであると考えられる。逆に読解力テストは単相関係数の点では .61 なので代替テストとして使用できるレベルにはあるが、標準回帰係数の数値からみた場合、.1640 なので C-test を読解力テストの代替テストとして使用するにはいささか苦しいレベルにあるといえよう。

以上の結果から、C-test は聴解力テストおよび発表語彙数測定テストの代替テストとして使用できる可能性が十分あり、文法力および受容語彙数測定テストとして使用できるレベルにあると考えられる。また語知識の深さや読解力を測定する補助的なテストとしても使用できる可能性がある。C-test は他のテストと比較した場合、50問の設問数でもクロンバック・アルファ係数が  $\alpha = .84$  とテストの信頼性も高く (Yoshimura, 1999)、実施時間が 10~15分ですむという簡便性も持っている。しかしながら今回利用した諸テストと比較した際、C-test に関する欠点として指摘できるのは表面妥当性 (face validity) が低いことである。そのために受験者はこのテストが英語の総合的な能力を測定するテストであると考えにくいという短所がある<sup>10</sup>。C-test はこの表面妥当性の問題があるとはいえ、Dörnyei & Katona (1992) の研究結果と同様に、総合的な英語を測定するために使用できる利便性に優れたテストであると結論付けられる。

今後の研究課題としては、TOEIC や TOEFL のような信頼度の高い標準テストとの相関性およびスピーキング力とライティング力との相関性の検証があげられる。またテストングの知識を深めることで、Ishihara, K., Hiser, E. & Okada, T. (2003) において試みられたような精度の高い C-test の開発にもチャレンジしたい。

## 参考資料

There are usually five men in the crew of a fire engine. One o\_\_\_ them dri\_\_\_ the eng\_\_\_. The lea\_\_\_ sits bes\_\_\_ the dri\_\_\_. The ot\_\_\_ firemen s\_\_\_ inside t\_\_\_ cab o\_\_\_ the f\_\_\_ engine. T\_\_\_ leader kn\_\_\_ how t\_\_\_ fight diff\_\_\_ sorts o\_\_\_ fires. S\_\_\_, when t\_\_\_ firemen arr\_\_\_ at a fire, it is always the leader who decides how to fight a fire. He tells each fireman what to do. (Klein-Braley and Raatz, 1984)

## 注

1. Alderson 同様、C-test が言語の総合能力を測定しているのではなく、基本的要素しか測定していないとみなす研究として Little, D & Singleton, D. (1990). The C-test as an elicitation instrument in second language research. と Carroll, B. (1986). LT+25, and beyond? Comments. を挙げることができる。Little, D & Singleton, D. は C-test の伸びと語彙力の伸長の間に相関関係があることから、C-test が語彙力を測定するテストとしての要素が強いことを指摘している。Carroll は C-test が文法力測定テストとして色彩が強いと述べている。
2. 参考までに著名な研究者の数値を挙げると Pike (1979) は  $r = .84$  から  $.95$ 、Schooner, et al. (1998) は  $r = .86$ 、Qian (2002) は  $r = .68$  から  $.85$  という強い相関係数を報告している。この相関係数は2007年 JACET 主催の草津サマーセミナーに参加した際に、講師であった William Grabe 教授が配布された資料から引用させていただいた。
3. C-test の詳細に関しては、Klein & Raatz (1984) を参照されたい。Cloze test は William Taylor によってテキストの readability を測定する意図で提唱された。テキストの文頭から通例7語前後（5ないし10語）毎に機械的に削除して問題を作成するようになっている。内容、文脈、文法等の知識を活用して、削除された文字を推測するところから、総合的な言語能力を測定するうえでも有効なテストであるという主張もおこなわれている。C-test は基本的には、cloze test の弱点を補うべく開発されたテストであるといえる。
4. Writing Research Group, JACET Kansai Chapter. (Ed.) (1999). *Teaching writing in colleges and universities: Practical reports, 3.* の Appendix に収録されている C-test 1 を使用した。
5. 語知識の深さに関しては Meara (1996) が、VLT の受容語彙数で5,000語以上の学習者の場合、読解力との相関性が語彙数よりも語知識の深さのほうが強いことを報告している。Qian (1999; 2002) も受容語彙数3,000語以上の学習者を対

象に、VLT による受容語彙数と、語知識の深さを測定する WAT の得点と、TOEFL のリーディング部門の得点の相関係数を算出したところ、WAT とのほうが VLT より強い数値が検出されたと報告している。また中田 (2007) もオーストラリアのシドニー大学大学院で TESOL を専攻する現職英語教員33名を対象におこなった読解能力と受容語彙数および語知識の深さとの相関性を比較する調査で、受容語彙数よりも語知識の深さのほうが読解力との相関性が強いことを発表している。この3つの研究は、中級レベル以上 (3,000語以上) あるいは上級レベル (5,000語以上) の語彙数習得者の場合、語彙サイズよりも語知識の深さが読解力を決定付ける要因であることを示唆している。

6. 詳しくは森永 (2009a) を参照されたい。
7. 一般に相関があるかないかを検定するには統計量  $t = r(n-2)^{1/2} / (1-r^2)^{1/2}$  が自由度  $n-2$  の  $t$  分布に従うという公式を用いておこなわれるが、計算がいささか複雑でかつ  $t$  分布表が必要なので煩雑である。上田 (1997) はこの公式の代替となる簡便法を考案した。この簡便法は相関係数を  $r$  として、 $r^2 > 4 / (n+2)$  ( $n$ : データ数) が成立すれば相関があるとするものである。
8. 定係数の数値の解釈に関しては菅 (2001) に収録されているものを引用した。
9. 相関係数の数値の解釈に関しては田栗正章・藤越康祝・柳井晴夫・C. R. ラオ (2007) に収録されているものを用いた。
10. 20年くらい前までは C-test の研究が比較的多くおこなわれていたが、現在では C-test の発表を聞きく機会はめったになく、研究論文を目にする機会もほとんどない。この事に関してテストングを専門にされている立命館大学の清水裕子先生に質問した所、C-test を含む cloze test の表面妥当性が低いことが、C-test を含む cloze test の研究が衰微していった原因ではないかとのコメントをいただいた。白畑知彦・富田祐一・村野井仁・若林茂則 (2009) 『改訂版 英語教育用語辞典』の face validity (表面妥当性) の項には「例えば、口頭言語能力 (oral proficiency) を測る際に、筆記試験が口頭表現力を測るためにどれほど正確なものであっても、表面的には、別の能力を測っているように見えてしまう」ので、「テストを作成・実施する際には (表面妥当性に対する) 配慮が必要である」とある。

### 引用文献

- 上田太一郎. (1997). 「相関があるかを見つける簡便法」日本 OR 学会『オペレーションズ・リサーチ』97. 7. 493-496.
- 大場昌也. (2008). 『英語学習の DO'S, DON'T'S & MAYBE'S』金沢: 時鐘舎

- 白畑知彦・富田祐一・村野井仁・若林茂則. (2009). 『改訂版 英語教育用語辞典』  
東京：大修館書店
- 菅 民郎. (2001). 『Excel で学ぶ多変量解析入門』 東京：オーム社
- 田栗正章・藤越康祝・柳井晴夫・C. R. ラオ. (2007). 『やさしい統計入門』 東京：  
講談社
- 中田貴眞. (2007). 「英語学習上級者における語連想知識と学術的読解能力の関係  
——Quian & Schedl (2004) に関連して」 第13回日英・英語教育学会研究大会.
- 西田 実. (1989). 「Cloze test と C-test の比較」『広島大学総合科学部紀要. V.  
言語文化研究』 14. 196-209.
- 森永弘司. (2009a). 「語彙力増強の重要性——二つの大学の語彙力測定テストの  
データに基づく——」『主流』 70：85-105.
- Alderson, J. C. (1979). The cloze procedure and proficiency in English as a  
foreign language. *TESOL Quarterly*, 13, 219-227.
- Carroll, B. (1986). LT + 25, and beyond? Comments. *Language Testing*, 3,  
123-129.
- Dörnyei, Z. & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL  
learners. *Language Testing*, 9 (2), 187-206.
- Ikeguchi, C. (1998). Do different C-test discriminate proficiency levels of EL2  
learners? *JALT Testing & Evaluation SIG Newsletter*, 2 (2), 3-8.
- Ishihara, K., Hiser, E. & Okada, T. (2003). Modifying C-Test for Practical  
Purpose. *Doshisha Studies in Language and Culture*, 5-4, 539-568.
- Klein-Braley C. (1997). C-Tests in the context of reduced redundancy testing: An  
appraisal. *Language Testing*, 14, 47-84.
- Klein-Braley, C. & Raatz, U. (1984). A survey of research on the C-test.  
*Language Testing*, 1 (2), 134-146.
- Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive  
ability. *Language Testing*, 16, 36-55.
- Little, D & Singleton, D. (1990). The C-test as an elicitation instrument in second  
language research. Paper presented at *AILA* 1990. 16-22.
- Meara, P. (1996). The dimension of lexical competence. In. G. Brown, K.  
Malmkjer & J. Williams. (Eds.), *Performance and competence in second  
language acquisition*. (pp. 35-53). Cambridge: Cambridge University Press.
- Mochizuki, A. (1994). C-tests: Four kinds of texts, their reliability and validity.  
*JALT Journal*, 16 (1), 41-54.
- Nation, P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31

(1), 9-13.

- Negishi, M. (1987). The C-test: An integrative measure? *IRLT Bulletin*, 1, 3-26.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56, 282-307.
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning*, 52, 513-536.
- Writing Research Group, JACET Kansai Chapter. (Ed.) (1998). *Teaching writing in colleges and universities: Practical Reports*, 3.
- Yoshimura, T. (1999). Reliability of the C-Test. *Teaching Writing in Colleges and Universities: Practical Reports*, 3, 73-79.

#### 参考文献

- 佐藤史郎. (1988). 『英語能力の新しい測定法 クローズテストと英語教育』東京：南雲堂
- 望月明彦. (2008). 「クローズテストと英語教育」『新しい英語教育のために：理論と実践の接点を求めて』望月昭彦、久保田章、磐崎弘貞、卯城祐司編、東京：成美堂. 256-269.
- 望月正道. (1998). 「日本人英語学習者のための語彙サイズテスト」『財団法人語学教育研究所紀要』12：27-53.
- 森永弘司. (2009b). 「聴解力測定にはどのテストが最も有効なのか——望月語彙サイズ測定、Vocabulary Levels Test, Vocabulary Size Test, Productive Levels Test, Word Associate Test, C-Test との相関性の視点からの考察——」第49回外国語教育メディア学会全国大会発表要項集、312-313.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23, 290-325
- Ishihara, K., Okada, T. & Matsui, S. (1999). English vocabulary recognition and production: A preliminary survey report. *Doshisha Studies in Language and Culture*, 2-1, 143-175.
- Ishihara, K., Okada, T. & Matsui, S. (2000). Vocabulary Levels analysis: Survey results with university students. *Doshisha Studies in Language and Culture*, 3-1, 17-46.

- Morinaga, K. (2009). Reliability of C-Test as a Device Measuring Reading and Listening Comprehension Abilities. JACET 48<sup>th</sup> Annual Convention Proceedings. 199-200.
- Okada, T., Ishihara, K., Kinshi, K. & Pavloska, S. (2009). Practice Activities with Defining Vocabulary: Making English Dictionary Entries More Accessible. *Doshisha Studies in Language and Culture*, 12-1, 1-38.
- Oller, J. W., Jr. (1975). Cloze, discourse and approximations to English. In M. Burt & H. C. Dulay (Eds.), *New directions in second language learning, teaching, and bilingual education. Washington DC TESOL*, 345-355.
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the *Vocabulary Levels Test*. *Language Testing*. 18, 55-88.

## 謝 辞

今回使用させて頂いた C-test の開発者の 1 人である同志社大学名誉教授の岡田妙先生と cloze test の著名な研究者で筑波大学名誉教授（現在大東文化大学教授）の望月明彦先生には、本稿執筆にあたって有益なコメントを頂いたことに感謝申し上げます。

本稿は第49回外国語教育メディア学会全国大会（流通科学大学、2009年8月）におけるポスター発表および第48回大学英語教育学会（北海学園大学、2009年9月）における口頭発表の原稿に修正・加筆を施したものである。