

# Cloze Test の難易度に影響を与える 要因について\*

長 加奈子

## Abstract

This study investigates the factors that determine the difficulties of a cloze test for Japanese learners of English. A total of 196 Japanese university students who were studying English in Japan were chosen for this study. The passage used in this study was consisted of 220 words. Every 7th word from the beginning was deleted and this passage was made into a 31-item cloze test. The participants were asked to fill in the blanks with appropriate words, and their responses were scored based on the exact-answer scoring method. The results of this study indicate high reliability estimates (*Cronbach*  $\alpha = .78$ , *KR-21* = .78) and high criterion-related validity. Correlation coefficient between the cloze test and the TOEIC® scores is .82. Each item is analyzed based on Item Response Theory, and the results show that six items are too difficult to be included in the test and one item is excluded from the analysis because no participant answered it correctly. The article concludes with discussion of the implications of these findings.

---

\* 本稿は、平成18年度～20年度科学研究費補助金（若手研究B）：課題番号18720156「学習者の日本語力・学習動機・個性・英語力が英語学習活動に及ぼす影響」による研究の一部である。

## 1. はじめに

英語を教えるもの者にとって、今、目の前にしている学生の英語力を授業開始時に手軽に測定できる方法があればありがたいであろう。もちろん、さまざまな業者テストも1つの方法として考えられるであろうが、残念ながら時間と費用がかかる。例えば、学期の途中に、教員が行っている授業の教育効果を測定するために実施するというのも容易ではない。しかし、教育効果を客観的に測定する何らかの手段は必要であろう。

比較的容易に学習者の英語力を測定する方法の1つとして、英語の cloze test があげられる。ゲシュタルト心理学の“closure”という概念からヒントを得て作成された cloze test は、もともと英語のネイティブスピーカーを対象として、読解力を測定する指標として開発され、学校現場に導入された (e.g., Ruddell, 1964, Bormuth, 1967, Gallant, 1965; Crawford, 1970; Taylor, 1953)。その後、英語学習者に対しては、学習者の英語総合力を測定するテストとして様々な研究がなされている (e.g., Alderson, 1979; Mullen, 1979; Brown, 1980, 1983, 1984, 1988, 1993; Hinofotis, 1980)。これは cloze test が空欄に単語を埋めさせるという単純な作業にもかかわらず、正解を得るために、文法の知識、文章の把握力など様々な力を必要とするためである。また日本語を母語とする英語学習者を対象とした研究も、様々な角度からなされている。中でも cloze test が学習者のどのような英語力を測定しているのかという点に注目が集まっている。例えば中川 (2001, 2002) では cloze test のスコアと英語のリスニング力、リーディング力および文法運用能力の関係に焦点を当て分析している。その結果、cloze test とリスニング力の間には .64 の相関が、リーディング力との間には .69 の相関が、文法力とは .65 の相関が存在すると報告している。一方、中郷 (2007) では多肢選択式タイプの cloze test と TOEIC® のスコアとの関係を分析している。その結果、cloze test のスコアと TOEIC® の合計スコア、リスニング・スコア、リーディング・スコアの間には .08 から .33 とほとんど相関関係が存在しないか、存在しても弱い相関関係しか存在していなかった。中川 (2001, 2002) と中郷 (2007) では一見矛盾した結果が示されているように感じるかもしれないが、ここでは cloze test に用いられ

ている素材文の難易度の差や出題形式の違いが影響していると考えられる。中郷（2007）の cloze test はもともと大学入試問題として作成されているため、比較的難易度の高い単語（例えば “expose” や “instinctive” など）が出題されている。また、問題形式が4つの選択肢から適当なものを1つ選ぶ形式を取っており、空欄に適当な単語を書かせる一般的な cloze test とは異なる。一方、中川（2001, 2002）では比較的易しい素材文が用いられており、文章の内容を表すイラストと一緒に提示されている。また、空欄に単語を書かせる記述タイプの出題形式が取られている。

そこで本研究は、まず比較的容易な素材文を用いて cloze test を作成し、cloze test の難易度を決定する様々な要因について探る。特に日本の大学で英語を学ぶ学生を対象として調査を行い、日本語を母語とする英語学習者の cloze test におけるパフォーマンスに影響を与えている要因を分析する。本研究の research question は次の5つである。

- (1) 本調査で作成した cloze test の信頼性および妥当性は高いか
- (2) 空欄の語彙レベルは cloze test の難易度に影響を与えているか
- (3) 空欄の文法項目は cloze test の難易度に影響を与えているか
- (4) 項目困難度の高い（低い）項目に何らかの傾向が存在するか
- (5) 項目弁別力の高い（低い）項目に何らかの傾向が存在するか

## 2. 調査方法

### 2.1 被験者

本調査の被験者は日本の私立大学で英語を学ぶ学生196名である。学生たちの所属は英語系学科または社会系学部である。

### 2.2 素材

日本語を母語とする英語学習者を対象とした英語学習教材を作成している英語のネイティブスピーカーに、素材文の作成を依頼した。作成に当たっては、英語のネイティブスピーカーではなく英語学習者向け、特に日本人英語学習者

に多くみられる lower-intermediate レベルの学習者を対象に素材文を作成するよう依頼した。素材文には内容を示すタイトルがついており、タイトル部分を除いた総語数は220語である。素材文の語彙レベルを表1に示す。本研究では、素材文の語彙難易度を検討するデータベースとして、大学英語教育学会基本語改訂委員会（2003）によるJACET 8000を利用した。JACET8000は、British National Corpusをベースに日本人英語学習者の学習環境を考慮して作成された学習者用語彙データベースである。JACET8000は中学校の英語教科書に類出する基本的な単語であるLevel 1から始まり、高等学校初級レベルのLevel 2、高等学校教科書レベルのLevel 3と続き、日本人英語学習者の一般的な単語学習の最終到達目標であるLevel 8まで区分されている。なお、JACET 8000に基づくレベルを算出する際に、短縮形はもとの短縮されていない形式に戻しているため、総語数が225となっている。なおこのJACET8000に存在しないものは、『Level 8 超』と区分した。

表1：JACET8000に基づくレベルとその出現数

| 単語レベル     | index |       | token |       |
|-----------|-------|-------|-------|-------|
|           | 数     | 割合    | 数     | 割合    |
| Level 1   | 96    | 78.69 | 191   | 84.89 |
| Level 2   | 12    | 9.84  | 14    | 6.22  |
| Level 3   | 2     | 1.64  | 2     | 0.89  |
| Level 4   | 1     | 0.82  | 1     | 0.44  |
| Level 5   | 2     | 1.64  | 2     | 0.89  |
| Level 6   | 4     | 3.28  | 4     | 1.78  |
| Level 7   | 0     | 0.00  | 0     | 0.00  |
| Level 8   | 2     | 1.64  | 2     | 0.89  |
| Level 8 超 | 3     | 2.46  | 9     | 4.00  |
| 固有名詞      | 0     | 0.00  | 0     | 0.00  |
| 総数        | 122   |       | 225   |       |

表1で使用したindex数とは、単語の種類数をさす。例えば、be動詞のis, were, been等はすべてbeの変化形ととらえられ、indexを数える場合は1つと数えられる。一方、tokenの方は、純粋な単語の出現数である。例えば、appleという単語がテキスト中に3回出現した場合、token数は3となる。今回作成し

た素材は、index数で全体の78.69%、token数で全体の84.89%がLevel 1、つまり中学校の教科書に頻出する基本的な単語群であった。出現単語の難易度から考えると、易しい英文であると言える。今回、Level 8を越える単語が、index数で3、token数で9見られたが、これは本素材のトピックにもなった“hostel(s)”に加え、“dormitories”、“inexpensive”という単語であった。

### 2.3 Cloze Test

2.2で示したレベルの素材文を使用し、cloze testを作成した。空欄の空け方については諸説あるが、本研究では一般的なルールである、機械的に7語目を空欄とする方法を取った。<sup>1</sup> また、文章の最初の文と最後の文の扱いについても諸説あるが、本研究調査で用いた素材文の語彙レベルが易しいため、最初と最後の文も空欄の対象とした。なお空欄はインターネット上の無料のソフトウェアを用いて7単語おきに、自動的に作成した。<sup>2</sup> 空欄となった単語のレベルを表2に示す。

表2：空欄部分のレベル一覧

| 単語レベル     | 数  | 割合    |
|-----------|----|-------|
| Level 1   | 25 | 80.65 |
| Level 2   | 2  | 6.45  |
| Level 3   | 0  | 0.00  |
| Level 4   | 1  | 3.23  |
| Level 5   | 0  | 0.00  |
| Level 6   | 0  | 0.00  |
| Level 7   | 0  | 0.00  |
| Level 8   | 0  | 0.00  |
| Level 8 超 | 3  | 9.68  |
| 固有名詞      | 0  | 0.00  |
| 総数        | 31 |       |

<sup>1</sup> cloze testには、機械的にn番目の単語を削除し空欄とする「通常型クローズ (random cloze)」と問題作成者がテストの目的により意図的に削除する項目を選ぶ「意図的削除クローズ (rational cloze)」とがある。

<sup>2</sup> 名古屋大学大学院杉浦正利教授が開発されたソフトウェアであるCloze Test Creator 2 (<http://oscar.glad.nagoya-u.ac.jp/program/pearl/cloze2.html>) を利用した。

表2から明らかなように、空欄となった単語は80.65%がLevel 1と中学校の教科書に頻出する基礎的な単語と判断される。またLevel 8を越える単語は“hostel”のみである。

本テストの素材文についていたタイトルは、cloze testと一緒に学生側に提示した。これは、読解にあたって学生の頭の中に認知フレームを導入し、文章全体の概念化を容易にするためである。学生は、空欄に適当な単語を1つ、書き入れるよう指示されている。本調査で用いたcloze testをAppendix 1に示す。<sup>3</sup>

## 2.4 結果

学習者の解答は、正語法 (exact-answer scoring method) で採点された。Cloze testを採点する採点方法としては、今回採用した正語法のように答えを1つだけ定めて、それ以外はすべて不正解とする方法と、模範解答ではないが文法的観点および意味的観点から正解としてみなしてよいものも正解とする適語法 (acceptable-answer scoring method) とがある。さらに適語法の一つで文脈により適した語とそうでない語に点差を設ける適語加点法もあるが、採点方法がかなり複雑になり実用的ではない。Brown (1980) は、cloze testの採点方法とし

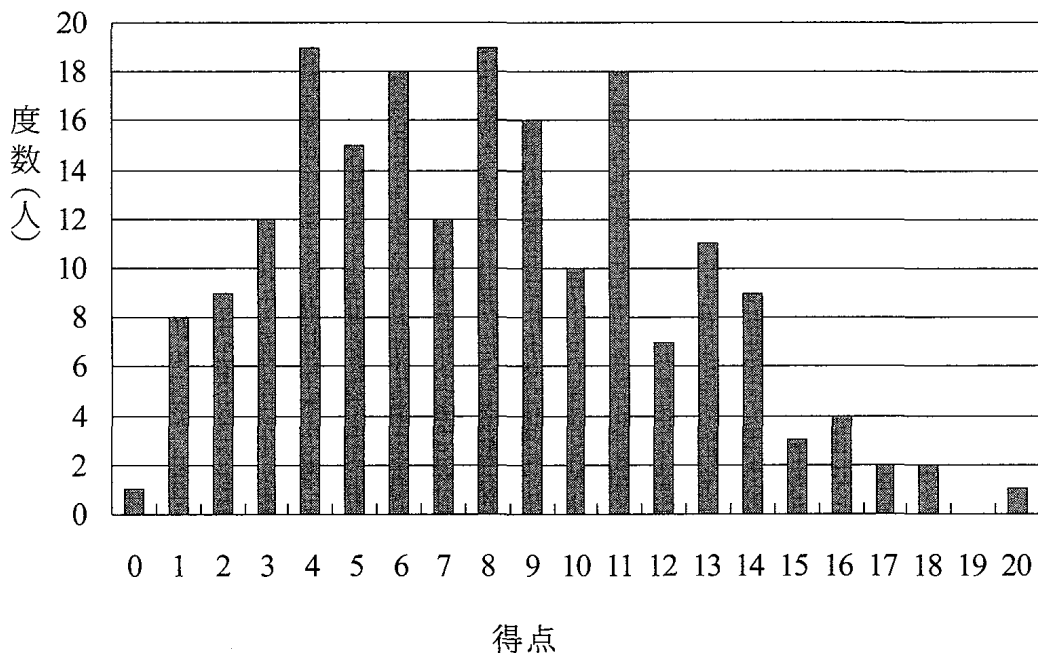


図1：受験者の総合得点の分布

<sup>3</sup> 著作権の関係上、cloze testの掲載は全文ではなく、一部にとどめる。

て、正語法、適語法、適語加点法、そしてあらかじめ受験者に選択肢を提供する選択肢提供型採点法の4つを上げ、測定の標準誤差、妥当性、信頼性、項目難易度、項目弁別力、テストの開発のしやすさ、採点のしやすさの7つの観点から比較し、それぞれの項目に対して、1点 (best) から4点 (worst) の間で点数を付けている。<sup>4</sup> その結果、正語法の合計得点が13点、適語法の合計得点が12点となり、2つの採点方法の間に有意な差はなかった。また門田 (2002) においても、正語法と適語法に統計的に有意な差が見出されていない。そのため、本調査では正語法を採用した。

学生のスコアは31点を満点として算出された。設問に対する重みづけは行っていない。学生のスコア分布を図1に示す。最小値は0、最大値は20、平均値は8.00、標準偏差は4.26であった。さらに本調査で用いた cloze test の信頼度は  $Cronbach\ \alpha = .78$  および  $KR-21 = .78$  で、比較的信頼度が高いテストであると言える。

次に妥当性の問題であるが、本研究では criterion-related validity として TOEIC® のスコアとの相関を算出する。被験者196名のうち、6ヶ月以内に TOEIC® の受験経験があるものは142名で、そのスコアの平均は489.25点 (最小170、最大860、標準偏差153.74) であった。TOEIC® のスコアを持つ学生の TOEIC® のスコアとの相関を算出したところ、 $r = .82$  という結果になり、本研究で用いた cloze test の妥当性が高いことがわかった。次に cloze test のそれぞれの項目に対し、項目応答理論の枠組みを用いて分析を行った (Hambleton, Swaminathan, & Rogers, 1991; Liden & Hambleton, 1997; McNamara, 1996; 大友, 1996; 高橋, 2002)。項目応答理論とは、古典的テスト理論に替わり、近年用いられているテスト理論である。以前の古典的テスト理論では、テスト項目の難易度はそのテストを受験する集団に依存していた。テストから受験者の学力を測定する値として、素点、偏差値、順位等があるが、これらはテストの難易度や受験者の集団によって左右されてしまう値であり、絶対的な値であるとは言いがたい。例えば、100問のテストを3つの集団に実施した場合、集団の能力が異なれば、3つのきわめて異なる平均点が出てくるだろう。また、

<sup>4</sup> Brown (1980) で比較する要素となっている項目難易度および項目弁別度は古典的テスト理論の枠組みで算出されたものである。

ある集団で順位が1位だとした場合、その集団においては能力が高いという相対的な結論には至ることができるが、果たしてどの程度の能力なのかという絶対的な結論に到達することはできない。このような古典的テスト理論で生じる問題点を解決するのが、項目応答理論である。項目応答理論は、テストの問題の特性と受験者の能力との間にある種の確率的な関係を仮定すれば、受験者の能力を絶対的に測定することが可能であるという立場に立つ理論である。Educational Testing Service (ETS) が開発しているテストはTOEFL<sup>®</sup>やTOEIC<sup>®</sup>も含め、すべてこの項目応答理論に基づいて分析され作成されている。そのため、TOEIC<sup>®</sup>で算出されている「スコア」というものは、いわゆる素点ではなく、項目ごとに与えられた項目特性を総合したものをスコア化しているものである。

大友（1996）では、項目応答理論の利点として次の3つをあげている。1つは、test-free person measurementと呼ばれるもので、テストを受験する学習者の能力推定値は、その学習者が受験した特定のテスト項目とは切り離して独立して求めることができるという点である。第2の利点は、項目特性が受験者の集団とは切り離して求めることができるという点、sample-free item calibrationである。項目特性とは、項目困難度パラメーターや項目弁別力パラメーター、当て推量パラメーターのことである。項目困難度パラメーターは、設問項目ごとの難易度をさしている。また項目弁別力パラメーターとは、ある設問項目が能力の高い受験者とそうでない受験者を弁別することができる度合いのことを指す。3つ目の当て推量パラメーターは、特に多肢選択式（multiple-choice）のテストに関わるもので、受験者があてずっぽうで答えたときの正解の得やすさの度合いを示す。項目応答理論において、これらの項目特性は、項目応答理論で用いられているモデルに一致していれば、受験者集団に関わらず1つの普遍的な値として示すことが可能である。3つ目の利点がmultiple reliability estimationで、測定の精度というものを、受験者集団に対して算出された標準誤差や信頼性係数からしか検討できない古典的理論とは異なり、項目応答理論では、受験者一人ひとりについて検討することが可能であるという点である。

本研究では、各項目の項目特性を分析するため項目応答理論を用いる。今回は受験者数がそれほど多くないこと、また全テスト項目が記述式であるため第3のパラメーターである「当て推量パラメーター (guessing factor)」を考慮する必要



性が高くないことから、項目困難度と項目弁別能力とを算出する2パラメーターモデルを用いて分析する。分析用のコンピュータプログラムとしてXcalibre 1.10 (Assessment System Corporation, 1997) を用いた。被験者の能力推定法には最尤推定法 (Maximum Likelihood Estimation) を用いた。

本研究で用いた cloze test のそれぞれの項目に対して空欄に入れるべき単語、文法項目、項目弁別力、項目困難度を表3に示す。項目数は全部で31項目である。

表3：設問項目の項目特性

| 番号        | 解答             | 文法項目            | 語彙レベル    | 項目弁別度       | 項目困難度       |
|-----------|----------------|-----------------|----------|-------------|-------------|
| 1         | like           | 形容詞             | 1        | 0.78        | 1.87        |
| <b>2</b>  | <b>really</b>  | <b>副詞</b>       | <b>1</b> | <b>0.82</b> | <b>3.00</b> |
| 3         | an             | 冠詞              | 1        | 0.67        | 1.01        |
| 4         | hostels        | 名詞 (複数)         | 8+       | 0.60        | 2.32        |
| 5         | but            | 接続詞             | 1        | 0.67        | 2.87        |
| <b>6</b>  | <b>need</b>    | <b>動詞</b>       | <b>1</b> | <b>0.79</b> | <b>3.00</b> |
| 7         | they           | 代名詞 (主格)        | 1        | 0.57        | 0.17        |
| 8         | places         | 名詞 (複数)         | 1        | 0.79        | 2.49        |
| 9         | or             | 接続詞             | 1        | 0.80        | -1.30       |
| 10        | as             | 比較              | 1        | 0.55        | 0.62        |
| <b>11</b> | <b>provide</b> | <b>動詞</b>       | <b>1</b> | <b>0.74</b> | <b>3.00</b> |
| 12        | are            | be 動詞           | 1        | 0.76        | -0.72       |
| 13        | young          | 形容詞             | 1        | 0.87        | 1.66        |
| 14        | are            | be 動詞           | 1        | 0.97        | -0.02       |
| 15        | you            | 代名詞 (主格)        | 1        | 0.55        | 1.80        |
| 16        | hostel         | 名詞              | 8+       | 0.67        | 1.40        |
| 17        | a              | 冠詞              | 1        | 0.96        | 0.02        |
| 18        | there          | there 構文        | 1        | 0.96        | 0.97        |
| 19        | bed            | 名詞              | 1        | 0.62        | 0.47        |
| <b>20</b> | <b>sheet</b>   | <b>名詞</b>       | <b>2</b> | <b>0.75</b> | <b>3.00</b> |
| <b>21</b> | <b>top</b>     | <b>形容詞</b>      | <b>1</b> | <b>0.79</b> | <b>3.00</b> |
| 22        | rent           | 動詞              | 4        | 0.73        | 2.73        |
| 23        | morning        | 名詞              | 1        | 0.53        | 0.71        |
| 24        | of             | 前置詞             | 1        | 0.60        | 2.69        |
| 25        | to             | 不定詞             | 1        | 0.55        | 1.44        |
| 26        | where          | 関係副詞            | 1        | 0.84        | 0.44        |
| 27        | and            | 接続詞             | 1        | 0.52        | -0.79       |
| <b>28</b> | <b>offers</b>  | <b>動詞 (三単現)</b> | <b>1</b> |             |             |
| <b>29</b> | <b>tours</b>   | <b>名詞 (複数)</b>  | <b>2</b> | <b>0.86</b> | <b>3.00</b> |
| 30        | hostels        | 名詞 (複数)         | 8+       | 0.68        | 1.71        |
| 31        | to             | 不定詞             | 1        | 0.56        | 0.17        |

今回の分析で用いた Xcalibre では、項目弁別度が 0.30 を下回る項目および項目困難度が 2.95 を上回るまたは -2.95 を下回る項目は改良が必要な項目としてマークされる。項目弁別度が 0.30 を下回るということは、その項目の弁別能力が低いため、本来区別すべき能力が高い受験者と能力が低い受験者を適切に区別できていないということである。また、項目困難度が高すぎる（または低すぎる）項目も同様に、正答者が少なすぎて（または多すぎて）その項目をテストの中に入れてもあまり意味がないということを示している。テストというものは、できるだけ少ない項目を用いてできるだけ信頼性の高いテストを構成することが一番望ましい。項目数が多くなればなるほど、受験者の負担が大きくなり、様々なノイズが受験者のスコアに影響を及ぼす可能性が大きくなるためである。しかしその一方、項目数が多くなればなるほど、テスト自体の信頼性が高くなる。その両者のバランスをいかにとるかという点がテストを作成する者には求められるが、項目応答理論で改良が必要と求められる項目は、テストにおいて、それらの項目で得られる情報量が極端に少ないことを示している。つまり難しすぎる（易しすぎる）ため、受験者の選別に不適切であったり、弁別力が小さすぎてテスト項目の役割を十分に果たしていなかったりするということである。今回の cloze test では、項目困難度が高すぎて改良が必要とされる項目が 6 項目（項目 2, 6, 11, 20, 21, 29）、また、正解者がいないため分析の対象とならなかった項目が 1 項目（項目 28）存在した。これらの項目は、いわゆる「難しすぎる」項目であるため、もう少し難易度の低い項目に改良する必要がある。

さらに項目困難度が高い項目 10 項目と低い項目 10 項目を、それぞれ表 4 および表 5 に示す。困難度が高い項目として、動詞、名詞を中心とする内容語が目立つ一方、困難度が低い項目には冠詞、前置詞、接続詞、be 動詞のような機能語を中心とする意味内容の希薄なものが目立った。

表4：項目困難度が高い項目

| 番号 | 解答      | 文法項目   | 語彙レベル | 項目困難度 |
|----|---------|--------|-------|-------|
| 2  | really  | 副詞     | 1     | 3.00  |
| 6  | need    | 動詞     | 1     | 3.00  |
| 11 | provide | 動詞     | 1     | 3.00  |
| 20 | sheet   | 名詞     | 2     | 3.00  |
| 21 | top     | 形容詞    | 1     | 3.00  |
| 29 | tours   | 名詞(複数) | 2     | 3.00  |
| 5  | but     | 接続詞    | 1     | 2.87  |
| 22 | rent    | 動詞     | 4     | 2.73  |
| 24 | of      | 前置詞    | 1     | 2.69  |
| 8  | places  | 名詞(複数) | 1     | 2.49  |

表5：項目困難度が低い項目

| 番号 | 解答    | 文法項目    | 語彙レベル | 項目困難度 |
|----|-------|---------|-------|-------|
| 10 | as    | 比較      | 1     | 0.62  |
| 19 | bed   | 名詞      | 1     | 0.47  |
| 26 | where | 関係副詞    | 1     | 0.44  |
| 7  | they  | 代名詞(主格) | 1     | 0.17  |
| 31 | to    | 不定詞     | 1     | 0.17  |
| 17 | a     | 冠詞      | 1     | 0.02  |
| 14 | are   | be動詞    | 1     | -0.02 |
| 12 | are   | be動詞    | 1     | -0.72 |
| 27 | and   | 接続詞     | 1     | -0.79 |
| 9  | or    | 接続詞     | 1     | -1.30 |

次に項目弁別度が高い項目10項目と低い項目10項目をそれぞれ表6および表7に示す。

表6：項目弁別度が高い項目

| 番号 | 解答     | 文法項目     | 語彙レベル | 項目弁別度 |
|----|--------|----------|-------|-------|
| 14 | are    | be 動詞    | 1     | 0.97  |
| 18 | there  | there 構文 | 1     | 0.96  |
| 17 | a      | 冠詞       | 1     | 0.96  |
| 13 | young  | 形容詞      | 1     | 0.87  |
| 29 | tours  | 名詞（複数）   | 2     | 0.86  |
| 26 | where  | 関係副詞     | 1     | 0.84  |
| 2  | really | 副詞       | 1     | 0.82  |
| 9  | or     | 接続詞      | 1     | 0.80  |
| 6  | need   | 動詞       | 1     | 0.79  |
| 21 | top    | 形容詞      | 1     | 0.79  |

表7：項目弁別度が低い項目

| 番号 | 解答      | 文法項目    | 語彙レベル | 項目弁別度 |
|----|---------|---------|-------|-------|
| 19 | bed     | 名詞      | 1     | 0.62  |
| 24 | of      | 前置詞     | 1     | 0.60  |
| 4  | hostels | 名詞（複数）  | 8+    | 0.60  |
| 7  | they    | 代名詞（主格） | 1     | 0.57  |
| 31 | to      | 不定詞     | 1     | 0.56  |
| 15 | you     | 代名詞（主格） | 1     | 0.55  |
| 25 | to      | 不定詞     | 1     | 0.55  |
| 10 | as      | 比較      | 1     | 0.55  |
| 23 | morning | 名詞      | 1     | 0.53  |
| 27 | and     | 接続詞     | 1     | 0.52  |

項目弁別度については、語彙のレベル、文法項目、内容語と機能語の差などの観点から検証したが、今回の研究では特に何らかの傾向は見出されなかった。

### 3. 考察

今回の調査では、5つのresearch questionを立てた。まず、cloze testの信頼性と妥当性であるが、信頼度係数はCronbach  $\alpha = .78$ およびKR-21 = .78であり、信頼度が高いテストであると言っていいだろう。さらに妥当性係数として、

criterion-related validity を算出した。本研究で用いた criterion は TOEIC® のスコアである。本研究の cloze test のスコアと TOEIC® のスコアとの間の相関関係は  $r = .82$  と統計的に有意で高い相関関係がみられた。これは、TOEIC® が測定している英語力の 66% を本研究で用いた cloze test も測定しているということである。TOEIC® はリスニングとリーディングのセクションに分かれていることを考えると、妥当な結果であると考えられる。以上のことから、本研究で用いた cloze test は信頼性、妥当性ともに高いテストであると結論づけられる。

さらに空欄の語彙レベルと cloze test の難易度との関係についてであるが、語彙レベルと項目困難度の間に見られる相関は  $r = .17$  であり、また項目弁別度との間に見られる相関は  $r = -.15$  で、どちらも統計的に有意な相関関係は見られなかった。今回用いられた素材文の単語の 84.89% がレベル 1 の語彙であることを考えると、当然の結果であるだろう。語彙のレベルが cloze test の難易度に影響を与えているのであれば、今回の cloze test における被験者たちのスコアは高いはずである。しかし実際は、最高点が 20 点、平均点が 8 点と低い結果であった。満点が 31 点であることを考えると、被験者にとって難しいテストであったと言えるであろう。このことから、語彙のレベル以外の要因が難易度を決定する要因として存在していると結論づけられる。では、空欄の文法項目はどうであろうか。文法項目を独立変数として項目困難度および項目弁別度を従属変数として、回帰分析を行ったところ、項目困難度との間に統計的に有意な関係が見出された ( $p < .05$ )。今回の調査から、cloze test に影響を与えている大きな要因として文法項目が存在すると考えてよいであろう。

さらに、項目困難度の高い項目または低い項目に見られる傾向として、語彙の機能があげられる。表 4 および表 5 で見たように、項目困難度の高い項目は、動詞、名詞を中心とする内容語が目立つ一方、困難度が低い項目は冠詞、前置詞、接続詞、be 動詞のような機能語を中心とする意味内容の希薄なものが目立っている。今回の cloze test では 7 項目について改良が必要だという結果になったが、そのいずれもが内容語であった。これは、採点方法の影響が存在することをうかがわせる結果である。例えば、前置詞や冠詞が入る空欄については、「別解」が存在することはまずないが、名詞や動詞が入る空欄については、意味内容および文法的な関係から複数の回答が存在しうる。今回の調査で用い

た cloze test の項目 6 を見るとより明らかであろう。正語法で採点する場合、この空欄に当てはまるのは動詞の need だけである。しかし文の流れから考えて want でもなんら問題がない項目である。今回作成したような通常型 cloze test では、機械的に n 番目に空欄を空けていくため、問題項目自体を改良することは難しい。そこで考える必要があるのが採点方法である。表 4 で見たように、改良が必要な項目は内容語ばかりである。つまり、採点法を正語法から適語法に変えることで項目困難度を下げることが可能になるのではないかと考えられる。

次に、項目弁別力の高い項目または低い項目について、何らかの傾向が見出されるか検証したが、今回の cloze test においては、特に傾向は見出されなかった。これは語彙レベルがほぼ Level 1 であるため、語彙レベルと項目弁別度との関連性が検証できなかつたためである。今後、語彙レベルを上げて検証することも考えられるが、今回の正答率の低さを考えると、語彙レベルを高くすることはあまり意味がないだろう。

#### 4. まとめ

今回の研究では、cloze test の難易度に問題項目の文法事項が影響を与えており、語彙レベルは影響を与えていないことがわかった。また、内容語なのか機能語なのかという語彙の情報量も難易度に影響を与えていた。さらに、項目応答理論に基づいて分析した結果、改良が必要だと判別された項目困難度が高すぎる項目は、内容語に偏っており、これは採点方法を検討する必要があることを示唆している。今回の研究では、先行研究の知見をもとに、正語法による採点を行ったが、今後、適語法による採点結果と比較検討し、先行研究で述べられているように、採点方法による違いが本当に存在しないかどうかを検証する必要があるだろう。

## 参考文献

- Alderson, J. C. (1979). Scoring procedures for use on cloze tests. In C. A. Yorio, K. Perkins, & J. Schachter (Eds.), *On TESOL '79* (pp. 193-205). Washington, DC: TESOL.
- Assessment System Corporation (1997). *Xcalibre marginal maximum-likelihood estimation program (Version 1.10)*.
- Bormuth, J. R. (1967). Comparable cloze and multiple-choice comprehension tests scores. *Journal of Reading*, 10, 291-299.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal*, 64, 311-317.
- Brown, J. D. (1983). A closer look at cloze: Validity and reliability. In J. W. Oller, Jr. (Ed.) *Issues in Language Testing Research* (pp. 237-250). Rowley, MA: Newbury House.
- Brown, J. D. (1984). A cloze is a cloze is a cloze? In J. Handscombe, R. A. Orem, & B. P. Taylor (Eds.), *On TESOL '83* (pp. 109-119). Washington, DC: TESOL.
- Brown, J. D. (1988). Tailored cloze: improved with classical item analysis techniques. *Language Testing*, 5, 19-31.
- Brown, J. D. (1993). What are the characteristics of natural cloze tests? *Language Testing*, 10, 93-116.
- Crawford, A. (1970). *The cloze procedure as a measure of reading comprehension of elementary level Mexican American and Anglo-American children*. Unpublished doctoral dissertation, University of California, Los Angeles.
- 大学英語教育学会基本語改訂委員会 (2003) 『大学英語教育学会基本語リスト JACET List of 8000 Basic Words』 大学英語教育学会.
- Gallant, R. (1965). Use of cloze tests as a measure of readability in the primary grades. In J. A. Figurel (Ed.), *Reading and inquiry* (pp. 286-287). Newark, Delaware: International Reading Associates.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publication.
- Hinofotis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller Jr. & K. Perkins (Eds.), *Research in language testing* (pp. 121-128). Rowley, MA: Newbury House.
- 門田修平 (2002) 「日本人大学生を対象とした英語クローズテストの項目分析：ディスコース処理能力との関係を中心に」 『言語と文化』 第5号, 63-78.
- Liden, W. J. van der & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
- Mullen, K. (1979). More on cloze tests as tests of proficiency in English as a second language. In E. J. Briere and F. B. Hinofotis (Eds.), *Concepts in language testing: Some recent studies* (pp. 21-32). Washington, DC: TESOL.

- 中川武 (2001) 「クローズテストと予測文法能力」『つくば国際大学研究紀要』vol. 7, 79-96.
- 中川武 (2002) 「クローズテストと文法運用能力」『つくば国際大学研究紀要』vol. 8, 37-51.
- 中郷慶 (2007) 「TOEIC®スコアで測定される英語運用能力と発音テストおよびクローズテストとの相関関係について」『愛知淑徳大学論集 文化創造学部・文化創造研究科篇』第7巻, 77-93.
- 大友賢二 (1996) 『項目応答理論入門』東京：大修館.
- Ruddell, R. B. (1964). A study of the cloze comprehension technique in relation to structurally controlled reading material. *Improvement of Reading Through Classroom Practice*, 9, 298-303.
- 高橋正視 (2002) 『項目反応理論—新しい絶対評価』東京：イデア出版局.
- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 30, 414-438.



## Appendix 1: 本研究で用いた cloze test の一部

内容を考えながら、空欄に当てはまる単語を1語書き入れなさい。

*Youth Hostels*

The word *hostel* looks a lot (1) \_\_\_\_\_ the word *hotel*, and a hostel (2) \_\_\_\_\_ is something like a hotel. It's (3) \_\_\_\_\_ inexpensive place to stay while traveling. (4) \_\_\_\_\_ are usually designed like school dormitories, (5) \_\_\_\_\_ they are for travelers who only (6) \_\_\_\_\_ to stay for a short time. (7) \_\_\_\_\_ are likely to be located in (8) \_\_\_\_\_ where people often travel....