

# Cloze Test の有効性

内 藤 徹

(2008年1月21日受理)

## Effectiveness of Cloze Test

NAITO, Toru

キーワード (key words)

総合性 (comprehensiveness)、信頼性 (reliability)、実用性 (practicality)

### 1. はじめに

テスト形態には、能力テスト (Proficiency Test) と学力テスト (Achievement Test) がある。入学試験や受験生の英語の能力をバランスよく測定しようとする実用英語検定試験 (英検)、TOEFL、TOEIC などは典型的な能力テストと言えよう。これは、学習者が受けてきた色々な学習条件や興味などは関係なく、テストを行うものが要求する英語の能力を基準に問題が作成される。これに対して、学習者が一定期間、ある学習条件 (すなわち、教師、教授法、教材など) のもとで学んだ知識が到達目標に対して、どの程度定着したかを測定するのは学力テストである。これは、通常教師が作成し、教えた内容の中から出題するものである。

さて、テストは、信頼性と妥当性でいかに優れていても、実用性に乏しくてはその価値は半減すると言わざるを得ない。そのテストが、能力テストであろうと学力テストであろうと、それを行うものが、許容される一定時間の中で、テストを実施し、測定し、評価ができなければならないと言える。そして、この実施から評価までの過程は、公平で客観的であることが必要である。

このようなことを考えた場合、Cloze Test (以下CT) は、通常ある文章の中からN番目に当た

る単語を消去し、被験者にそれを埋めさせるという方法を取り、極めて客観的で機械的に行いうるテストである。従って、Lange (1981) も述べているように、CTは他の形式のテストと比較すると、実用性の面で利点が多いと言える。

CTは、総合的な言語能力を測定するのに非常に良い方法であるという報告が多くなされている。しかし、その一方でそれ程良い方法ではないとも言われている。この小論の中では、CTの有効性について検証していきたい。

### 2. 背景および先行文献研究

CTは、Taylor (1953) によって開発され、最初は読解力の測定に使用されたテストである。これは、ゲシュタルト心理学の "closure" という語の概念、つまり不完全な円や模様を見れば、完全な形に修復したくなる人間の心理を応用して、文章中の空所を再生、補充するものである。佐藤 (1988)、内藤 (1990) も述べているように、このような再生・補充作業が成功するためには、文法の知識や語彙力を伴った言語能力や様々な思考力 (類推、分析、統合、帰納、演繹など) に支えられた予測能力 (expectancies) を必要とすると思われる。そして、これらの能力が言語の持つ余剰性 (redundancy) に助けられ、総合的に動員された時に再生が可能になると考えられる。

CTの方法について、Oller (1979) は fixed-ratio method と variable-ratio methodをあげている。前者はある一節のN番目ごとの単語を削除するもので、後者は特定の意図を基盤に空所を設けるものである。後者のように、特定の品詞や文章を理解する上で重要と思われる単語を予め消去する方法は、文章の特定の部分のみに焦点を当てることになり、当該文章全体の難易度を決定する際に客観性を維持する妨げになると思われる。また、ある文章の中から特定の基準に基づいて単語群を選択することは、それだけで主観性を介入させることにつながるだけでなく、その時の主観

的判断は、別の文章には応用できないことから、客観性を欠くことになる。従って、私の今までのデータからも fixed-ratio method が適当と考えられ、今回のCTはこの方法を採用している。

また、採点方法は Exact-word method (以下E.W.)、Acceptable-word method (以下A.W.) が考えられる。Lange and Clausen (1981) は Nth Cloze Format について E.W.による Exact Scoring および A.W.による Acceptable Scoring により採点を行った場合の信頼度係数を示している。いずれの採点法でも .83 と信頼性は高いと言えるようである。

#### Reliability Coefficients for Nth Format

	Spearman-Brown	Guttman Split Half
Nth Format		
Exact Scoring	.83	.83
Acceptable Scoring	.83	.83

E.W.というのは正語法とも訳されているように、省略した語以外は正答と認めない採点方法である。そして、A.W.は適語法と言われるように、内容から判断して、適当と思われる語を正答とする採点方法である。例えば、

How long does it take you to ( ) to the town? ...

この文において、E.W.は get であるが、

A.W.では go も正答となる。上述のように、E.W.も A.W.もあまり変わらないが、Miller and Coleman (1967), Klare et al (1972), Oller (1972), Naccarato and Gilmore (1976) によると、A.W.の方がより正確に英語の総合能力を測定できるということである。従って、この小論の中ではA.W.を採用した。

さて、A.W.の1つの方法である Weighting Degree of Appropriateness は次のようなものである。

Joe is a freshman and he (1) IS having all of the problems that most (2) FRESHMEN have. As a matter of fact, his (3) PROBLEMS started before he even left home. (4) HE had to do a lot of (5) THINGS that he didn't like to do (6) JUST because he was going to go (7) AWAY to college...

例えば、Oller (1979) は、この(3)の項目には次の5段階の点数を与えている。

- a. problem : exact word 4 points
- b. difficulties : second best 3 points
- c. bewildering : 2 points

- d. methods : 1 point
- e. before : 0 point

しかし、採点が複雑になるため、実用性の面からこの Weighting Degree of Appropriateness の方法は除外した。

### 3. 実証的研究

#### (1) 仮説

Cloze Test は英語の学力を測定するのに有効な方法である。

#### (2) 被験者・実施月・被験者数等

高等学校 1 回目 (7 月) 1 年普通科 2 クラス (73名)  
 2 回目 (11月) 1 年普通科 2 クラス (74名)  
 大 学 (7 月) 1 回生 (20名)  
 (7 月) 2 回生 (16名)

#### (3) 分析・統計処理方法 [Hatch (1982), 内藤 (1996)]

相関係数: Pearson product moment correlation

有意水準は 0.1% (\*\*\*) , 1 % ( \*\* ) , 5 % ( \* ) の 3 つ

参考までに相関係数は次のとおりである:

±. 00 ~           ほとんど相関なし  
 ±. 20 ~           低い相関あり  
 ±. 40 ~           かなり相関あり  
 ±. 70 ~           高い相関あり  
 ±. 90 ~ ±1.00 極めて高い相関あり

テストの信頼度係数:  $rt = reliability: KR-21$  (Kuder-Richardson 21 formula)

#### (4) テスト

##### \* 高校生

E G = English Grammar (英文法): 1 学期と 2 学期の 2 回分

E I = English I (英語 I): 1 学期と 2 学期の 2 回分

M E = Mock Exam. (模擬試験): 1 学期と 2 学期の 2 回分

C T = Cloze Test (7 番目毎の語を抜いたクローズ) 2 回分

##### \* 大学生

R = Reading (総合的なリーディング)

C T = Cloze Test (7 番目毎の語を抜いたクローズ)

T E = Term Exam (Multimedia English I (1 回生)、Multimedia English II (2 回生) という総合的な英語): 前期分

LSD = Listening Spot Dictation (適当な箇所の語を抜いた英文に、英語を聞いて書き入れるテスト)

CTの作成方法: 最初の文はそのまま残し、第 2 番目の文から 7 番目毎の語を削除・空所 (blanks) にし、全部で空所は 25 個の問題文章を 2 つ用いた。すなわち、合計で空所は 50 個である。各文章の最後の文には空所が入らないだけの長さ

の文章を用いた。なお、空所にすべき箇所に理解を困難とする固有名詞や数詞がきた場合には、その語を跳ばして次の語とした。また、各問題文章の下に空所に入れる語の語群をアトランダムに置いた。(資料参照)

## (5) 結果

この表中における相関係数は、\*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$  である。

Table 1 (高等学校1年生：7月)

1回目 (n=73)

	EG	E I	ME	CT
EG		.69	.58	.43
E I			.65	.61
ME				.55
CT				

EG = English Grammar (1学期)

E I = English I (1学期)

ME = Mock Exam. (1学期)

CT = Cloze Test (1回目)

CT - EG = .43\*\*\*

CT - E I = .61\*\*\*

CT - ME = .55\*\*\*

ME - EG = .58\*\*\*

ME - E I = .65\*\*\*

E I - EG = .69\*\*\*

rt : EG = .90 E I = .91 ME = .88 CT = .96

Table 2 (高等学校1年生：11月)

2回目 (n=74)

	EG	E I	ME	CT
EG		.70	.76	.67
E I			.59	.53
ME				.78
CT				

EG = English Grammar (2学期)

E I = English I (2学期)

ME = Mock Exam. (2学期)

CT = Cloze Test (2回目)

CT - EG = .67\*\*\*

CT - E I = .53\*\*\*

CT - ME = .78\*\*\*

ME - EG = .76\*\*\*

ME - E I = .59\*\*\*

E I - EG = .70\*\*\*

rt : EG = .93 E I = .92 ME = .89 CT = .95

Table 3 (大学1回生：7月)

(n=20)

	R	CT	TE	LSD
R		.55	.24	.45
C			.48	.51
TE				.43
LSD				

R = Reading

(Fukui-ken English Reading Test)

CT = Cloze Test

TE = Term Exam I (1<sup>st</sup> Term)

LSD = Listening Spot Dictation

CT - LSD = .51\*

CT - TE = .48\*

CT - R = .55\*

R - LSD = .45\*

R - TE = .24

TE - LSD = .43\*

rt : R = .97 CT = .98 TE = .93 LSD = .95

Table 4 (大学2回生：7月)

(n=16)

	R	CT	TE	LSD
R		.51	.36	.30
CT			.83	.40
TE				.70
LSD				

R = Reading

(Fukui-ken English Reading Test)

CT = Cloze Test

TE = Term Exam II (1<sup>st</sup> Term)

LSP = Listening Spot Dictation

CT - LSD = .40  
 CT - TE = .83\*\*\*  
 CT - R = .51\*  
 R - LSD = .30  
 R - TE = .36  
 TE - LSD = .70\*\*  
 rt : R = .83   CT = .97   TE = .96   LSD = .95

## (6) 考察

Table 1 : CTはEG、EI、MEとかなり相関がある (.43, .61, .55 で0.1%水準)。EIもCT、EG、MEと相関がかなりある (.61, .69, .65 で0.1%水準)。さて、上記のように全て相関係数は.43以上であるので、CTを含めてEI、ME、EGは互いにかなり相関があり、いずれの方法でも英語の学力を測定するのには信頼性があると言えよう。なお、テストの信頼度係数は全て .88 以上で信頼性が高い。

Table 2 : CTはEG、EI、ME とかなりの相関があり (.67, .53, .78 で0.1%水準)、特に ME とは高い相関が見られる。また、全て相関係数は.53以上であるので、CTを含めてEI、ME、EGは互いにかなり相関があり、いずれの方法でも英語の学力を測定するのには信頼性があり有効と言えよう。なお、テストの信頼度係数は全て .89 以上で信頼性が高い。そして、この2回目のCTは、教科担当者から「入れ終わった語は斜線で消しながら問題を進めた方がよい」とのアドバイスをを行ったこともあり、1回目よりもCTの相関が高くなっている。

Table 3 : CT は LSD、TE、R と相関がかなりあるので (.51, .48, .55 で5%水準)、CTは英語の学力を測定するのには信頼性があると言えよう。しかし、R は TE と弱い相関 (.24 で有意性なし) しかみられなかった。なお、各テストの信頼度係数は全て .93 以上で信頼性は高い。CTについては、「入れ終わった語は斜線で消しながら問題を進めた方がよい」とのアドバイスをを行っている。

Table 4 : CT は LSD と .40 の相関で、R とはかなりの相関 (.51で5%水準) があり、TE とはさらに高い相関 (.83で0.1%水準) がある。R は CT と比べると他との相関は低くなる (LSD と .30, TEと .36で共に有意性なし)。ま

た、LSD は全般的に他との相関は低い、TE とは高い相関 (.70で1%水準) が見られた。これは、TEの中に spot dictation が含まれていたからと考えられる。CTについては、1回生と同様、「入れ終わった語は斜線で消しながら問題を進めた方がよい」とのアドバイスを行っている。

全体をまとめると、CTは1つの listening (Table 4 の LSD との相関 .40) を除く他のテストとかなり有意な相関があり、十分に英語の学力測定の信頼性があると考えられる。従って、仮説「Cloze Test は英語の学力を測定するのに有効な方法である。」というのは文字言語においては支持されていると言える。

## 4. おわりに

今回、CTの有効性について再度調べてみた。CTは、特に実用性に優れ、英語の学力を測定するのに充分可能なテストであると言えよう。内藤 (2001) によれば、総合的学力を測定している「要約」とも相関が高かった (.40\*~ .52\*\*\*) とのことである。

さて、内藤 (1990) も述べているように CT の長所をあげれば次のようにまとめられよう。

- (1) 客観性がある。(特にfixed-ratio method の場合)
- (2) 項目数を多く含むことができる。
- (3) N番目を空所にすることで作成が容易である。
- (4) 採点が正確で早い。
- (5) 空所の場所を変えることにより、同じテストを複数回使用することも可能である。

また、優れた testing device であるとともに、優れた teaching device にもなりうると思われる。今後、CTを用いた英語の文字言語だけでなく音声言語上の教育実践 (ex. listening cloze) についてもデータを収集し、その効果に

についても研究をすすめていきたいと考えている。  
最後に、C-test という方法もあるが、CTと比べて実用性（作成上）に若干問題もあり作為的でもあると考えられるので、今回はより自然と思われる CT に焦点を当てて考察した。

〔CT、C-test については資料参照〕

#### 引用文献

- Hatch, Evelyn and Hossein Farhady, *Research Design and Statistics for Applied Linguistics*, Newbury House Publishers, Inc., 192-214. (1982)
- Lange, Dale L. and G. Clausen, "An Examination of Two Methods of Generating and Scoring Cloze Tests with Students of German on Three Levels", *MLJ*, 65, 3, Autumn, 254-261. (1981)
- 内藤 徹, 「CLOZE TEST の可能性－READING CLOZE から AURAL CLOZE へ－」 *STEP BULLETIN. VOL. 2* 日本英語検定協会出版, 29-41. (1990)
- 内藤 徹, 『新しい 英語教育ハンドブック』リーベル出版, 19-25. (1997)
- 内藤 徹, 「要約、クローズ、多肢選択と段落構造が学習者におよぼす影響」『中部地区英語教育学会紀要』31, 43-48. (2001)
- Oller, John W. Jr., *Language Tests at School : A Pragmatic Approach*. London : Longman Group Ltd., 345, 330-332, 370-373. (1979)
- 佐藤史郎, 『クローズテストと英語教育』 南雲堂, 10-11. (1988)

#### 参考文献

- Brown, David, "Conversational Cloze Tests and Conversational Ability", *ELT Journal* 37, 158-161. (1983)
- Cohen, Andrew D., *Testing Language Ability in the Classroom*, Newbury House Publishers, Inc. (1980)
- Hughes, Arthur, "Conversational Cloze as a Measure of Oral Ability," *ELT Journal* 35, 161-167. (1981)
- Madsen, S. Harold, *Techniques in Testing*, Oxford University Press. (1983)
- Richards, C. Jack and Schmidt, W. Richard, *Language and Communication*, Longman Group Limited. (1984)
- Richards, Jack, John Platt and Heidi Weber, *Longman Dictionary of Applied Linguistics*, Longman Group Limited. (1985)
- Rivers, M. Wilga, *Teaching Foreign-Language Skills*, The University of Chicago Press. (1981)
- Templeton, H., "A New Technique for Measuring Listening Comprehension", *ELT Journal* 31, 4, July, 292-299. (1977)
- Underhill, Nic, *Testing Spoken Language : A handbook of oral testing techniques*, Cambridge University Press. (1987)
- Widdowson, H. G., *Teaching Language as Communication*, Oxford University Press. (1985)
- Widdowson, H. G., *Explorations in Applied Linguistics*, Oxford University Press. (1979)

資料：

Cloze Test (前半 25 blanks 後半 25 blanks の合計 50 blanks)

前半

1. 次の英文を読み、前後関係から推測して括弧を埋めなさい。語群は下にあります。

Two Americans were traveling in Spain. They did not speak a word ( ) Spanish. One day the train in ( ) they were traveling stopped for several ( ) for repairs in a small town. ( ) pass the time, the two men ( ) out of the train and took ( ) walk through the town. At last, ( ) went into a small restaurant to ( ) something to eat. But the only ( ) on the menu that they could ( ) was coffee. So, they ordered a ( ) of coffee. The young man who ( ) on them brought them the coffee ( ) once; he was very much interested ( ) the two Americans. But he brought ( ) black coffee. Americans, of course, seldom ( ) black coffee. They prefer milk with ( ) coffee. But the two men did ( ) know the word for milk. They ( ) different gestures with their hands, but ( ) boy did not understand what they ( ). At last, one of the men ( ) a picture of a cow on ( ) piece of paper. Then, again with ( ), he explained that milk always comes ( ) a cow. The boy studied the picture a long time. Then, suddenly he ran out of the door. He returned with two tickets for a bull-fight, which is the most popular sport in Spain.

語 群： made, which, to, from, waited, wanted, they, a, a, gestures, hours, of, their, understand, drink, not, at, cup, the, got, in, get, drew, word, them

Marks

No. \_\_\_\_\_ Name \_\_\_\_\_ / 25

参考

C-test (2<sup>nd</sup> word deletion)

下の英文の空所に意味が通じるように1文字または数文字を入れなさい。下線を含んだ語は、1語を前半と後半に分け、後半の部分に下線を引いてあります。下線部と残りの部分で意味をなす語となります。

There are usually five men in the crew of a fire engine. One o\_\_\_\_ them dri\_\_\_\_ the eng\_\_\_\_. ………. 以下略  
[最初と最後の文、また数詞・固有名詞を削除の対象からはずすのは Cloze Test と同じである。]

また、削除される語が奇数の場合、その語を n とすると (n-1)/2 次は (n+1)/2 という削除法もある。例をあげると、  
stout . . . phone . . . mouth . . . overt . . . のようになる。

