

Cloze Test の可能性

— Reading Cloze から Listening Cloze へ —

キーワード: クローズテスト, 読解クローズ(テスト), 聴解クローズ(テスト), 音声語上の障害

内 藤 徹

1. はじめに

テストは、信頼性と妥当性でいかに優れていても、実用性に乏しくはその価値は半減すると言わざるを得ないであろう。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と書く)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いうテストである。従って、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、総合的に言語能力を測定するのに良い方法であるという報告が多く出されているが、これ迄のものは reading を中心とするものであった。この小論では listening にCTを利用出来るかどうか、またその有効性についても考えていきたいと思う。

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。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。(注1)

CTの方法について、Oller(1979)はfixed-ratio methodとvariable-ratio methodをあげている。(注2)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。しかし、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。これらのことから、fixed-ratio methodが良いと考え、この小論の中のテスト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. と書く)、Acceptable-word method(以下A. W. と書く)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

(注 3)

さて、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 も正答である。この小論の中では A. W. を採用した。これは上述のように、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事実 (注 4) にもよるものである。

また、listening cloze を与える方法は、Oller(1979) や Templeton(1977) がある。そして、Oller(1979) は、次のように示している。(7th word)

Each year since the 1960s hospitals in the United States have had to accommodate about one millions additional/(patients). As a result, hospitals across the/(country) have searched for new ways to/(be) more efficient in order to provide/(the) best possible patient care.

(注 5)

Templeton(1977) は、ポーズを文単位で置き、解答ポーズの長さは 4 秒としている。点線はブラック、//はポーズの箇所である。(15th word)

By way of introduction, let's have some information on just how far British industry commerce have been computerised. // How many machines are in use here compared with other of the world? // Well, the statistics of national computer populations are notoriously difficult to // Differences in the ways different countries make up their figures lead to further problems making comparisons.//

(注 6)

また、Templeton(1977) は、英国の大学に入学して来る外国人留学生に C T を行い、次の様な結果を報告している。

Validity coefficients against sections of an English examination

	Essay	Grammar	Reading	Vocab.	Aural
cloze	.62	.71	.67	.54	.80

(注 7)

日本人の高校生の場合はどうであろうか。phonological barriers の問題などデータをもとに考察してみたい。

それでは、listening cloze の長所と短所について考えてみよう。まず、長所は、(1) 客観性がある。(2) 項目を多く含むことができる。(3) N 番目ということで作成が簡単である。(4) 採点が正確で早くできる。(5) 空所の場所を変えることにより、同じテストを何回も使用できる。短所は、録音に際して、beeps を入れるのに技術面で時間を要することがある。しかし、慣れれば困難なことではないし、AET と協力してやればより容易に行うことができる。

CT は、空所の箇所で expectancies を必要とする点で reading cloze も listening cloze も共通点を持つが、後者の場合は耳で聞くだけであるので、自分で文章を聞き返したりすることができず、類推の clues も掴みにくいし、short term memory や concentration が強く要求されると思われる。従って、reading cloze より易しいレベルの教材を用いたり、復習用として用いるのが適当であると考えられる。

3. 実践研究

3. 1 仮説

(1) Cloze test は英語の学力を測定するのに有効な方法である。

(2) Listening cloze も phonological barriers を低くすればする程 reading cloze と同様、有効な測定方法になりうる。

3. 2 クラスおよび被験者数

平成元年度

2年1組 普通科理科系進学クラス 45名

2年4組 普通科文科系進学クラス 45名

3. 3 使用テスト

平成元年度 ((1)(2)はそれぞれ第1回 [7月実施]、第2回 [1月実施]を表す)

(1)8.RC=第1回 8th word Reading Cloze (8番目の語を削除) [7月実施]

(2)8.RC=第2回 8th word Reading Cloze (8番目の語を削除) [1月実施]

F.RC=Function word Reading Cloze (機能語を削除) [7月実施]

C.RC=Content word Reading Cloze (内容語を削除) [7月実施]

(1)L.C.=第1回 Listening Cloze [7月実施] (削除した10番目の語の所でチャイムを入れ4秒間のポーズを置く。そして、同じ material を3回繰り返してきく。)

(2)L.C.=第2回 Listening Cloze [1月実施] (削除した10番目の語の所でチャイムを入れ4秒間のポーズを置く。そして、同じ material を3回繰り返してきく。)

(1)H.T.=第1回福井県英語放送テスト D [7月実施]

(2)H.T.=第2回福井県英語放送テスト D [1月実施]

(1)模試=第1回進研模擬試験 [7月実施]

(2)模試=第2回進研模擬試験 [1月実施]

3. 4 結果 (ここにおける相関係数の検定は ** $P < 0.001$ * $P < 0.005$ である。)

Table 1 2年1組 (各テストの相関)

	(1)8.RC	F.RC	C.RC	(1)L.C.	(1)H.T.	(1)模試
(1)8.RC	1.00	.48**	.63**	.47*	.44*	.66**
F.RC		1.00	.64**	.33	.50**	.45*
C.RC			1.00	.33	.41	.58**
(1)L.C.				1.00	.65**	.24
(1)H.T.					1.00	.39
(1)模試						1.00

Table 2 2年4組 (各テストの相関)

	(1)8.RC	F.RC	C.RC	(1)L.C.	(1)H.T.	(1)模試
(1)8.RC	1.00	.55**	.69**	.49**	.45*	.71**
F.RC		1.00	.58**	.47*	.51**	.52**
C.R			1.00	.51**	.50**	.64**
(1)L.C.				1.00	.64**	.31
(1)H.T.					1.00	.46*
(1)模試						1.00

Table 3 2年1組 (各テストの相関)

	(2)8.RC	(2)L.C.	(2)H.T.	(2)模試
(2)8.RC	1.00	.51**	.52**	.69**
(2)L.C.		1.00	.59**	.39
(2)H.T.			1.00	.42*
(2)模試				1.00

Table 4 2年4組 (各テストの相関)

	(2)8.RC	(2)L.C.	(2)H.T.	(2)模試
(2)8.RC	1.00	.71**	.55**	.70**
(2)L.C.		1.00	.65**	.61**
(2)H.T.			1.00	.51**
(2)模試				1.00

Table 5 (両クラスの点数と標準偏差)

	1組		4組	
	点数	標準偏差	点数	標準偏差
(1)8.RC	36.3	11.9	43.7	9.9
(2)8.RC	45.7	11.1	50.1	10.2
F.RC	30.2	9.7	37.2	7.9
C.RC	31.4	11.0	36.9	9.4
(1)L.C.	20.5	12.6	36.1	13.4
(2)L.C.	29.3	11.9	48.9	14.3

(1)H.T.	49.3	13.3	60.8	13.0
(2)H.T.	50.5	13.9	68.2	12.8
(1)模試	27.5	10.9	36.8	9.6
(2)模試	33.9	11.5	42.0	10.2

Table 6 ((1)L.C. と (2)L.C. の1組と4組のT検定)

	1組	4組
(1)L.C. 点数	20.5	36.1
標準偏差	12.6	13.4
(2)L.C. 点数	29.3	48.9
標準偏差	11.9	14.3
	T=3.368	T=4.333
	df=88	df=88
	** p<0.005	*** p<0.001

Table 7 (KR-21 による各テストの信頼度係数)

	1組	4組
(1)8.RC	.85	.76
(2)8.RC	.81	.77
F.RC	.78	.63
C.RC	.83	.74
(1)L.C.	.91	.88
(2)L.C.	.86	.89
(1)H.T.	.87	.87
(2)H.T.	.88	.88
(1)模試	.84	.76
(2)模試	.84	.77

3.5 考察

第1回目のテストの中において、reading cloze ((1)8.RC、F.RC、C.RC)の中では nth cloze format を用いた 8.RC が最も点数が高い。(43.7、36.3) [Table 5 参照] これは、学習者にとって function words や content words を blanks にした CT よりも nth CT の方がやり易いと思われる。また、これらの reading cloze の中で模試と最も相関が高いのも 8.RC である。(.66、.71) それで、模試が学習者の学力を測定するのに適当なテストであるとすれば、8.RC もその妥当性は高いと言えよう。従って、第2回目のテストでは 8.RC のみを用いた。(1)L.C. については、同じ音声言語を用いた H.T. 以外とはあまり相関はない。(1組においては (1)8.RC とは .47、F.RC とは .33、C.RC とは .33、模試とは .24、H.T. とは .65；4組においては (1)8.RC とは .49、F.RC とは .47、C.RC とは .51、模試とは .31、H.T. とは .64) [Tables 1, 2 参照]

第2回目のテストでは、2年1組において、(2)L.C. は他のテストと相関が若干上がっている。(8.RC とは .47 から .51、模試とは .24 から .39) [Tables 1, 3 参照] しかし、LLやAETなどを積極的に利用して音声言語に慣れさせてきた2年4組においては(2)L.C. は他のテストとの

相関がかなり高くなっている。(8.RC とは .49 から .71、模試とは .31 から .61) [Tables 2, 4 参照]そして、点数においても (2)L.C. は1組と比べるとかなり高くなっている。(1組においては 20.5 から 29.3 点差 8.8: 4組においては 36.1 から 48.9 点差 12.8 : T検定では、1組 ** $P < 0.005$ 、4組 *** $P < 0.001$ で4組の方がより有意に差がある。) [Tables 5, 6 参照]

ただし、両クラスとも同じ音声言語テストである H.T. とはあまり相関は変わっていない。[1組においては (1)L.C. と (1)H.T. は .65、(2)L.C. と (2)H.T. は .59: 4組においては(1)L.C. と (1)H.T. は .64、(2)L.C. と (2)H.T. は .65]

また、KR-21による各テストの信頼度係数はすべて高く、信頼性があるテストである。

[Table 7 参照] また、L.C. と H.T. との相関係数(併存的妥当性係数)も高い方であり、音声言語としての妥当性があるといえる。

以上のことから、CT、特に nth CT、は学習者の学力を測定するのに有効な方法であると言える。また、L.C. も LLやAETを積極的に利用して音声言語に多く触れさせるなど phonological barriers を低くすれば reading cloze とほぼ同様有効な測定方法になりうる。従って、仮説(1)(2)はほぼ証明されたことになる。

4. おわりに

CTは、言語の総合的な能力を測定するのに適当な方法であると言われているが、この実践の中でそれはほぼ証明された。そして、reading CTと同様 listening CTも listening の方から学習者の学力を測定するのに適当な方法であることも確認されたと思う。さらに、測定のためだけでなく、実際教室で学習者に listening CTを行ってみて、集中力という観点から listening の訓練としても非常に効果のある方法であるとも思っている。今後、その効果の方の研究も行っていきたい。

(福井県立 鯖江高等学校)

注

- 1 佐藤史郎(1988), 「クローズテストと英語教育」 南雲堂, 10-11
- 2 Oller, John W. Jr.(1979), Language Tests at School: A Pragmatic Approach. London : Longman Group Ltd., 345
- 3 Lange, Dale L. and G. Clausing (1981), "An Examination of Two Methods of Generating and Scoring Cloze Tests with Students of German on Three Levels", MLJ, 65, 3, Autumn, 254-261
- 4 吉田一衛(1984), 「英語のリスニング」 大修館書店, 183
- 5 Oller, John W. Jr.(1979), 330
- 6 Templeton, H.(1977), "A New Technique for Measuring Listening Comprehension", ELTJ, 31, 4, July, 292-299
- 7 Templeton, H.(1977), 292-299