

採点基準の違いが読解のリコールに及ぼす影響 ——高校生（下位者）の場合——

平野 紗枝*

(平成10年4月30日受理)

要旨

読解力を測定するテストの一つにリコール (recall) テストがある。平野 (1998) では、学力上位校の高校生にリコールテストを課した。学力が異なると、採点基準の違いが読解のリコールに及ぼす影響が異なるかどうかを調べるために、本研究では、学力の低い高校生を対象に、リコール・プロトコル (recall protocol) の異なった採点基準による採点（厳密な採点基準とややゆるやかな採点基準）が、(1)テクスト全体、(2)テクスト中の重要な情報（主題文全体）、(3)各パラグラフ全体、(4)各パラグラフの主題、の再生量にどのような影響を及ぼすかを検証した。その結果、(1)2種類の採点基準の相関は極めて高かった。すなわちテクスト全体のリコールでは、2種類の採点基準の相関係数は .983 で、重要な情報量（主題文全体）のリコールでは、.942 であった。また(2)各パラグラフ間の再生率の違いが、採点基準の違いによって異なった。しかし各パラグラフの主題間の再生率の違いは異なった採点基準の影響を受けなかった。

KEY WORDS

recall	リコール	recall protocols	リコール・プロトコル
assessment	評価	strict criterion	厳密な採点基準
loose criterion	ゆるやかな採点基準		

1. はじめに

読解力を測定するテストの一つにリコールテスト（再生テスト）がある。これは学習者にテキストを読ませた後、テキストの内容を想起、再生させるテストである。

第二言語学習者のリコールに関する実験研究では、リコールに影響を及ぼす要因として、テキスト構造、リコールする時の言語、英文を読む前に行なうリコールの指示の影響、英語学力のレベル、採点基準の違い等が Lee (1986) や平野 (1996) で指摘されている。

リコール・プロトコルを分析する際、再生量の評価に及ぼす採点基準の影響は、学習者の英語学習経験年数や年齢の異なる中学生、高校生、大学生、大学院生によって違いがみられるのかは、読解力の評価の観点から大変興味深いが、平野 (1996) の研究は日本人の大学院生を、平野 (1998) では上位校に在籍する高校生を対象にしており、学力の低い日本人 EFL 高校生を対象にしてリコール量の測定に異なった採点基準がどのような影響を及ぼすかを調べた研究は見当たらない。

* 言語系教育講座

従って、学力が異なると、採点基準の違いが読解のリコールに及ぼす影響が異なるのかどうかを調べるために、本研究では、学力の低い高校生を対象に、リコール・プロトコル (recall protocol) の異なる採点基準による採点（厳密な採点基準とややゆるやかな採点基準）が、(1) テキスト全体、(2) テキスト中の重要な情報（主題全体）、(3) 各パラグラフ全体、(4) 各パラグラフの主題、の再生量にどのような影響を及ぼすかを検証する。

本研究の目的は、具体的に次の3点を明らかにすることである。英語学力の低いEFL高校生において、

- (1) ①文章全体のリコールの量、つまりトータル・リコール (total recall) と、②テキスト中の重要な情報、つまり主題 (main idea) 全体のリコールにおいて、厳密な採点基準による採点法とややゆるやかな採点基準による採点法の相関がどの程度であるか。
- (2) 厳密な採点基準とややゆるやかな採点基準を使用した場合、採点基準の違いによって、文章全体のリコール及び主題全体のリコールの得点に差がみられるか。
- (3) 厳密な採点基準とややゆるやかな採点基準を使用した場合、採点基準の違いによって、各パラグラフ間及びパラグラフの主題間のリコール率の差に違いが生ずるか。

2. 方 法

2.1 被験者

長野県内の県立高校普通科2年生1クラス36人。長野県内では下位校の一つに在籍する高校生である。英語圈滞在経験者は0人であった。被験者全員リコール・タスクはこれまで一度も受けたことがなく、またリコール・テストの英文と同じ（様な）内容の英文を以前読んだことがないと報告した。平均年令約16歳（15歳～16歳）。

2.2 実施時期

1995年7月

2.3 材料

大学院生を対象にした平野（1996）、学力の高い高校生の場合の平野（1998）と同様、リコール・テストは *Reading Power* (Mikulecky and Jeffries, 1986:100) の英文 (Appendix 参照) を使用した。リコール・タスクは、語彙や統語が容易な243語の英文（4つのパラグラフから構成）を用意した。リコール・テストの際、はじめに英文をよく黙読して、英文に書いてあることを理解し、その後英文を見ないで、英文に書いてあったことをできるだけ多く、詳しく、そして正確に思い出して、解答用紙に日本語で書くように指示した。英文の黙読には、20分与え、日本語では15分で書くようにした。

2.4 採点法

2.4.1 文章全体

Carrel (1985) のアイデア・ユニット (idea unit) の分析法に従って、筆者が説明文 (expository text) を66個のアイデア・ユニット (Appendix) に分けて採点した。10年以上の教師経験のある中学校英語教員1名にアイデア・ユニットの数の設定と被験者6人（全体の約17%）のリコ

ル・プロトコルの採点を依頼した。その結果 inter-rater reliability は、idea unit の設定では 1.0, リコール・プロトコルの採点では、strict criterion で, .85 ($p < .05$), loose criterion では, .97 ($p < .01$) であった。

2.4.2 採点基準

詳細な採点基準は平野（1996）参照。厳密な採点基準では、もとの文章の内容と完全にあつていれば 1 点とし、誤りは 0 点として中間点は与えなかった。ただし、もとの文章にないような別の文や節が付加されている場合はその情報は採点しないで除外した。どちらの採点基準も、言い直し (paraphrase) は許容して採点した。Carrel (1983) ではゆるやかな採点基準では、もとの意味の大部分か一部が正確にリコールされるかパラフレースされた際は正解とみなしたが、本研究では一つのアイデア・ユニットの意味の大部分（約 3 分の 2 の情報）が正しく再生されていると考えられる場合、正解として 1 点を与え、それ以外は 0 点とした。

2.4.3 主題

英文は 4 つのパラグラフから構成されているので、パラグラフの主題は 4 つあり、合計 10 個のアイデア・ユニットからなる。第 1, 第 2 パラグラフの主題はそれぞれ 3 つのアイデア・ユニット、第 3, 第 4 パラグラフは各 2 つのアイデア・ユニットからなる。主題文の満点は計 10 点。各パラグラフの主題は以下のとおりである (Mikulecky and Jeffries 1986:264)。

- (1) Main Idea 1=Many people moved/out of American cities / after World War II.
(3 idea units) (3 点)
- (2) Main Idea 2=American businesses (also) left the city / during the 1950s / and 60s. (3 idea units) (3 点)
- (3) Main Idea 3=Cities began / to have money problems. (2 idea units) (2 点)
- (4) Main Idea 4=Many people believed/ that American cities were dying. (2 idea units) (2 点)

3. 結 果

3.1 トータル・リコールと主題のリコールの平均値・標準偏差・分散分析

厳密な採点法とややゆるやかな採点法の 2 種類の異なった採点法による文章全体のリコール、すなわちトータル・リコール (total recall, 66 点満点) と 4 つのパラグラフの主題 (計 10 個のアイデア・ユニットから構成、10 点満点) のリコールの得点の平均値、標準偏差、分散分析の結果が表 1 で示されている。

表 1 からわかるように分散分析の結果、トータル・リコールでもすべてのパラグラフの主題文の再生でも得点は、当然ゆるやかな採点法のほうが、厳密採点法より有意に高かった (それぞれ $F(1, 35) = 34.15, p < .01$; $F(1, 35) = 11.67, p < .01$)。

3.2 相関係数

表 2 が示すように厳密な採点基準とややゆるやかな採点基準の相関は、文章全体のトータル・リコールでは、.98 ($p < .01$) であり、主題を構成するアイデア・ユニットの総数では、.94 ($p < .01$) であり、いずれも極めて高い相関係数を示した。

Table 1 Means and SDs of all ideas and main idea units recalled

	Strict Criterion			Loose Criterion		Difference
	n	M	SD	M	SD	F
Total Recall ^{a)}	36	5.42	4.83	6.44	5.31	34.15**
All Main Ideas ^{b)}	36	1.44	1.21	1.69	1.29	11.67**

**p<.01

^{a)} Maximum possible score=66^{b)} Maximum possible score=10

Table 2 Correlations of recall

	Loose Criterion	
	Total Recall	Total Main Ideas
Strict Criterion		
Total Recall	.98**	—
Total Main Ideas	—	.94**

**p<.01

Table 3 Mean percentage of the recall of each paragraph

Paragraph	Strict Criterion	Loose Criterion
	Mean (SD)	Mean (SD)
Paragraph 1	24.20% (19.29)	27.98% (22.30)
Paragraph 2	8.99% (14.10)	10.95% (15.88)
Paragraph 3	1.92% (4.92)	3.42% (7.57)
Paragraph 4	1.14% (2.91)	1.26% (2.96)

3.3 パラグラフごとのリコールの平均率・標準偏差・分散分析

各パラグラフの平均再生率と標準偏差は表3の示すとおりである。採点基準の違い（厳密な採点法とややゆるやかな採点法）が各パラグラフ全体の再生率間の差に影響を及ぼすかどうか調べるために、2（採点基準）×2（パラグラフ）の分散分析を行った。その結果、表4が示すように、採点基準（A）とパララグラフ（B）の交互作用が有意であった（F(3, 105)=7.65, p<.01）。そこで、各水準ごとに単純主効果を分析した結果、第4パラグラフ以外のいずれのパラグラフにおいても、ややゆるやかな採点基準によるほうが、厳密な採点基準よりリコール率が有意に高かった（パラグラフ1 F(1, 35)=20.68, p<.01; パラグラフ2 F(1, 35)=14.00, p<.01; パラグラフ3 F(1, 35)=8.45, p<.01; パラグラフ4 F(1, 35)=1.00, ns）。またLSD法による多重比較の結果、厳密な採点基準の場合、パラグラフ1>パラグラフ2>パラグラフ3=パラグラフ4の順の再生率となり第3と第4パラグラフの間以外の全ての対間に有意差がみられた(MSe=123.69, p<.05)。同様にゆるやかな採点基準の場合、パラグラフ1>パラグラフ2>パラグラフ3=パラグラフ4の順であった(MSe=169.75, p<.05)

Table 4 Results of ANOVA for paragraphs recalled

Source	SS	df	MS	F
Sub	20053.78	35	572.97	
Criterion(A)	243.29	1	243.29	38.16**
S X A	223.14	35	6.38	
Paragraph(B)	28087.90	3	9362.63	32.50**
S X B	30251.23	105	288.11	
A X B	122.30	3	40.77	7.65**
S X A X B	559.80	105	5.33	
Total	79541.42	287		

**P<.01

Table 5 Mean percentage of the recall of each paragraph main idea

	Strict Criterion	Loose Criterion
	Mean (SD)	Mean (SD)
Main Idea 1	36.11% (29.79)	40.74% (32.50)
Main Idea 2	11.11% (23.57)	12.04% (26.24)
Main Idea 3	1.39% (8.22)	4.17% (13.82)
Main Idea 4	0 % (0.00)	1.39% (8.22)

Table 6 Results of ANOVA for paragraph main ideas recalled

Source	SS	df	MS	F
Sub	31277.62	35	893.65	
Criterion(A)	425.40	1	425.40	11.06**
S X A	1345.61	35	38.45	
Main Idea(B)	65080.30	3	21693.43	25.68**
S X B	88707.12	105	844.83	
A X B	149.55	3	49.85	1.01 ns
S X A X B	5163.43	105	49.18	
Total	192149.03	287		

**p<.01

3.4 パラグラフの主題ごとのリコールの平均率・標準偏差・分散分析

表5には、各パラグラフの主題の平均再生率と標準偏差が示されている。採点基準の違いが、各パラグラフの主題の復元率間の差に影響をおよぼすかどうかを検討するため、2(採点基準)×2(主題)の分散分析を行った。その結果が表6に示されている。すなわち、採点基準と主題の主効果は有意であった(それぞれ $F(1, 35) = 11.06, p < .01$; $F(3, 105) = 25.68, p < .01$)。しかし採点基準と主題の交互作用は有意でなかった。ゆるやかな採点基準の再生率の

ほうが、厳密な採点基準の再生率より、有意に高かった。LSD法による多重比較によれば ($MSe=844.83$, $p<.05$) , リコール率は主題1が他の3つの主題より有意に高く、また、主題2は主題4よりも有意に高かった。しかし主題2と主題3、及び主題3と主題4の差は有意でなかった。

4. 考察

4.1 2つの採点基準の相関

4.1.1 トータル・リコール

学力の高い高校生ではトータル・リコールで厳密採点とややゆるやかな採点基準による得点の相関が.96であったが(平野 1998), 本研究での学力の低い高校生でも同様に非常に高い相関係数($r=.98$)が検出された。高校生においては、学力の高低に関係なく、非常に高い相関であることがわかる。日本人EFL大学院生を対象にした平野 (1996) の場合も、 $r=.95$ を報告している。このことから、被験者が大学院生でも、高校生(学力上位・下位者両方)でもトータル・リコールではどちらの採点基準を使っても、得点の相対的順位にはほとんど変化がないといえるので、文章全体のリコールの成績を用いて被験者の読解力の順位を出すときは、どちらの基準を採用するかは、神経質にならなくてもよいようである(平野 1998)。

4.1.2 主題全体(主題の総数)のリコール

テクスト内の重要な情報量(主題文を構成するアイデア・ユニットの総数)においても英語高校生の学力の高低にかかわらず、厳密採点法とややゆるやかな採点法の相関は高いようである。すなわち学力の高い高校生の場合、 $r=.81$ でかなり高い相関係数が検出された(平野 1998)し、本研究の下位レベルの高校生でも.94と高かった。このことから、被験者が高校生の場合、学力上位者でも下位者でも、テクスト内の主題文の得点の順位に焦点をあてる際は、採点基準の厳しさの程度をどうするかは、あまり心配しなくてもよいのではないだろうか。

4.2 採点基準の違いの影響

4.2.1 テクスト全体及び主題全体のリコール

採点基準をややゆくすれば許容される情報が増えるので、平野 (1996) の大学院生、また平野 (1998) における学力上位校の高校生と同様に、ややゆるやかな採点基準では、テクスト全体のリコールも、すべての主題のリコールも、両方の得点が厳密採点法より有意に高くなつた。ただしどの程度得点が上昇するかは、学力によって違いがあるようである。学力の高い高校生では、厳密採点によるテクスト全体のリコール得点の平均点が35点から、ややゆるやかな採点法では41.66点に上昇したが(平野 1998), 学力が低くなると、厳密な採点基準で採点された平均点は5.42点で、採点基準がややゆるやかになると平均点が6.44点で、1点しか上昇しなかつた。また主題の総数においても同様であった。すなわち学力の高い高校生では、厳密採点法で平均点が6.02点、ややゆるやかな採点法では7.43点であったが(平野 1998), 学力の低い高校生では採点基準がゆるやかになっても有意差はあったが1.44点から1.69点と極めてわずかな上昇であった。

4.2.2 パラグラフ間及びパラグラフの主題間のリコール率の比較

各パラグラフのリコールの比較では、採点基準×パラグラフの交互作用が有意であった。つ

まり採点基準が異なると各パラグラフ間のリコール率の差のパターンが異なることが明らかになった。これは大学院生(平野 1996), および、学力の高い高校生(平野 1998)における結果を支持するものであった。しかし、大学院生や学力の高い高校生の結果(いずれのパラグラフでも2つの採点基準による得点間に有意差がある)と比較して違いがみられたのは、学力が低い高校生になると、パラグラフのなかでも第4パラグラフだけ、リコール率が厳密な採点法とややゆるやかな採点法の間に有意差がみられなかった。第4パラグラフのトータル・リコールの平均点が両者の採点法で、約1%の正解率であり、学力が低くてほとんど正しくリコールされなかつたせいであろう。

これまでの結果と本研究の結果をまとめると、次の点が明らかになった。大学院生でも学力下位・上位の高校生でも、①各パラグラフ間のリコール率の比較を行う時は、採点基準を厳密にする場合とやや甘くする場合とでは、再生率の差のパターンに違いがあるので、どちらの採点基準を使用するか配慮する必要があるだろう。②テクストのトタル・リコールでは、その点はあまり考慮しなくてもよいであろう。

読解力を測定する上で、パラグラフの位置も関係しているかどうかについて、本研究でもパラグラフの位置によって再生率が有意に異なることがわかった。平野(1996, 1998)と同様に、いずれの採点基準を使っても、第1パラグラフのリコール率のほうが他のパラグラフより有意に高く、最後の第4パラグラフのリコール率が他より有意に低かった。このことから第1パラグラフがより理解されたということが考えられよう。また各パラグラフの主題も第1パラグラフの主題が他の主題よりも再生率が有意に高かった(平野 1996, 1998)。

しかし、興味深いことに、低い学力の高校生の場合、採点基準×主題の交互作用は有意でなかった。つまり、採点基準の厳しさが違っても、パラグラフの主題間のリコール率の差のパターンは同じかった。一方大学院生や学力の高い高校生では採点基準の違いが各主題間のリコール率の差に影響を与えた(平野 1996, 1998)。学力の低い高校生の場合、各主題のリコール率が全般的に低く、主題の把握は学力を反映するのではないだろうか。

5. おわりに

本研究では、英語学力の低い高校生の場合、読解力を測定するリコールテストの採点基準の厳しさの程度が、リコール・プロトコルの成績にどのような影響を及ぼすかどうかを検討した。その結果、次の主な点が明らかになった。

- (1) (a)文章全体のリコールの得点では、厳密な採点法とややゆるやかな採点法の相関は.98で非常に高いことから、文章全体のリコール、テクスト内の重要な情報の採点ではどちらの採点基準を使用するかは、それほど神経質にならなくてよいであろう。又(b)テクスト内の重要な情報(主題の総数)のリコールの採点でも(a)と同様のことが言えるであろう。
- (2) (a)採点基準の違いによって、パラグラフ間のリコール率の差が異なったことから、テクストの中でどのパラグラフがよりよくリコールされるか比較する際は、どちらの採点基準を使用するかは考慮する必要があるだろう。しかし、(b)パラグラフの主題間のリコール率の差は、採点基準の厳しさの違いによる影響を受けなかった。
- (3) どちらの採点基準を使用しても第1パラグラフのリコール率が一番高かった。なお(1)(a), (2)(a), 及び(3)の結果は平野(1996), 平野(1998)を支持するものであった。

本研究では、英語学力の低い高校生を対象にした場合、英語学力のかなり高い高校生と類似の結果がみられるかどうかを考察したが、被験者が少ないと、英語学力を標準テストで測定していないことで一般化が難しい。今後さらに検討していく必要がある。

参考文献

- Barnett, M. 1989. *More than meets the eye*. Englewood Cliffs, New Jersey: Prentice Hall Regents.
- Bernhardt, E. 1983. Syntactic and lexical/semantic skill in foreign language reading comprehension: the immediate recall protocol. *Die Unterrichtspraxis*, 16, 27-33.
- Bernhardt, E. 1991. *Reading development in a second language*. Norwood: Ablex Publishing.
- Bransford, J. 1979. *Human cognition: Learning, understanding, and remembering*. Belmont, CA: Wadsworth.
- Brown, A., and Smiley, S. 1977. Rating the importance of structural units of prose passages: a problem of metacognitive development. *Child Development*, 48, 1-8.
- Carrel, P. 1983. Three components of background knowledge in reading comprehension. *Language Learning*, 33, 183-207.
- Carrel, P. 1984. The effects of rhetorical organization on ESL readers. *TESOL Quarterly*, 18, 441-469.
- Carrel, P. 1985. Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19, 727-752.
- 平野絹枝. 1996. 「採点基準の違いが読解のリコールに及ぼす影響——大学院生の場合——」『上越教育大学研究紀要』第15巻, 455-467.
- 平野絹枝. 1998. 「採点基準の違いが読解のリコールに及ぼす影響——高校生（上位者）の場合——」『上越教育大学研究紀要』第17巻, 713-724.
- Johnson, R. 1970. Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 9, 12-20.
- Lee, J. 1986. On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 83-93.
- Lee, J. and Ballman, T. 1987. FL learners' ability to recall and rate the important ideas of an expository text. In B. VanPattern et al. (Eds.), *Foreign language learning* (pp. 108-118). Rowley, MA: Newbury House.
- Mikulecky, B. and Jeffries, L. 1986. *Reading power*. Reading, Mass.: Addison-Wesley Publishing Company.
- Riley, G. L. and Lee, J. F. A comparison of recall and summary protocols as measures of second language reading comprehension. 1996. *Language Testing*, 13, 173-189.
- Swaffar, J. et al. 1991. *Reading for meaning: An integrated approach to language learning*. Englewood: Prentice Hall.
- Takahashi, T. 1994. The role of lexical and syntactic knowledge in understanding a

- well-organized prose passage. *ARELE*, 5, 121-130.
- Wolf, D. 1993. A comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal*, 77, 473-489.

Appendix: Idea Unit Analysis

(Paragraph 1)

1. American cities changed
2. in many ways
3. after World War II.
4. Many people moved
5. out of the city.
6. They moved
7. to the suburbs,
8. the areas around a city.
9. Most of the people...were rich.
10. who moved [9 and 10 = Most of the people who moved were rich.]
11. Poor people usually did not have enough money
12. to move.

13. They stayed
14. in the cities.

(Paragraph 2)

15. During the 1950s
16. and 60s
17. there was another important change
18. in American cities.
19. Businesses began..., too.
20. to leave the city
21. They left
22. because the people...were poorer.
23. in the city
24. Poor people do not have much money
25. to buy things.
26. So, many shops
27. and restaurants moved
28. to the suburbs.
29. People...had more money
30. in the suburbs
31. to spend. [29, 30 and 31 = People in the suburbs had more money to spend.]

(Paragraph 3)

32. Cities began
33. to have many serious problems.
34. The rich people
35. and the businesses did not pay city taxes anymore.
36. The poor people could not pay much money
37. in taxes.
38. So cities had less money
39. for schools
40. and housing.
41. Sometimes they could not pay their police officers
42. or firefighters.
43. And they could not take good care of their streets
44. and parks.
(Paragraph 4)
45. But money was only part
46. of the problem.
47. Many people believed
48. that American cities were dying.
49. They had good reason
50. to believe this.
51. City streets were sadly empty.
52. Many neighborhoods...were dirty
53. and parks
54. and dangerous. [52, 53, 54 =Many neighborhoods and parks were dirty and dangerous.]
55. In some places
56. buildings were even falling down.
57. And nobody seemed
58. to care.
59. This was the real problem.
60. Most of the people
61. and businesses...were
62. with money
63. in the suburbs. [60, 61, 62, and 63 = Most of the people and businesses with money were in the suburbs.]
64. They did not care
65. what happened
66. to the cities.

The Effect of Scoring Criteria on Recall Protocols: Focusing on Japanese EFL High School Students with Low Proficiency

Kinue HIRANO*

ABSTRACT

This study investigates the effects of two different scoring criteria on the written recall of an expository text. Using both a strict and a loose criterion of recall, the written recalls of 36 Japanese EFL high school students (focusing on students with low proficiency) were scored. The results revealed that the correlation between the strict and the loose criterion was extremely high for the total number of idea units recalled correctly ($r = .98$, $p < .01$). This finding for Japanese high school students with low proficiency supported the very high correlation found with Japanese graduate students in Hirano (1996), and Japanese high school students with high proficiency (Hirano 1998). For the total number of main ideas recalled correctly, the correlation was also high ($r = .94$, $p < .01$). In addition, the criterion of scoring recall protocols applied proved to be a significant factor which affected the comparisons of the recall of each paragraph. This finding also reaffirmed the results of Hirano (1996), and Hirano (1998). That is, significant differences were found between the strict and the loose criterion in terms of the percentage of correct recall across paragraphs. For the recall of each paragraph main idea, however, there were no significant differences between the strict and the loose criterion in terms of the percentage of correct recall across paragraph main ideas.

* Division of Languages: Department of Foreign Languages