主観的評価法の信頼性について：サマリーにおける採点法

若本　夏美　同志社女子大学短期大学部
枝澤　康代　同志社女子大学短期大学部
福地美奈子　金光第一高等学校
野口ジュディー　武庫川女子大学
竹内　理　関西大学
梅田　巖　京都産業大学

1．はじめに

近年、大学英語入試において日本語訳や要旨を書かせる、またエッセイを課すなど主観的採点を要する出題をすることある。また平常の授業において英語の理解力を確認するためにサマリーを書かせ評価することも少なくない。海外でも様々な報告があるが (Kirkland and Saunders, 1991)、その評価にどれだけの信頼性があるのであろうか。本研究では、実際に大学生に英文のパッセージを読ませ、その後日本語のサマリーを書かせたものを以下に述べる2つの方法を利用して複数のraterにより評価し、その結果がどの程度一致しているかについて検証し、同時にrater自身が感じた困難点についても検討する。

2．研究の目的

1) 全体的な印象により評価する場合（Holistic Evaluation 、以下HEと略記）の信頼性を検討する。
2) 採点基準を決めて評価する場合（Criteria Evaluation、以下CEと略記）の信頼性を検討する。
3) HEとCEを比較検討し、サマリーの評価方法として有効な方法を検討する。
4) サマリー評価の困難性に関し、rater がどの点で食い違いを見せているか検討する。

3．方法

3.1　Ratersと調査対象

Raters：近畿圏で教える大学教員11名
調査対象：近畿圏14大学よりランダムに抽出した学生50名の書いた日本語による
　　サマリー

3.2　使用した英文パッセージ

1996年11月30日にJapan Times に掲載された " Destruction of Mt. Fuji continuing apace" というタイトルのエッセー（約900語）

3.3　手順

大学生にエッセイを読ませた後（25分）、250字以内でそのサマリーを日本語で書かせた（20分）。その後、以下2つの方法でその評価をおこなった。

1) HE：全体的印象により採点する（5点満点）
2) CE：評価基準（以下4項目）を決めて採点する（10点満点）
   A. 富士山の環境破壊に関する事実関係についての言及（4点）
   B. 観光と自然環境の保護の関係についての言及（2点）
   C. 結論部分についての言及（2点）
   D. 首尾一貫性（2点）

その他の条件：字数オーバーは関知しない。ただし、余分なことを書いている場合、Dから2点まで減点。またA～Dのポイントを与えた箇所をマーカーで印をつける。

3.4　採点をする上での条件

採点をする前にrater 11名の間で以下の条件を確認した。

1) HEとCEの混同防止：HEを採点した後、少なくとも3日間あけ、CEによる採点をおこなう。
2) 基準の一貫性：HEとCEそれぞれ連続して一気に採点する。途中に長時間の休憩（5分以上）をはさまない。
3) その他：1枚目から5枚目までは評価せずに読み、採点基準をraterの中で確定する。採点は6枚目からはじめ、最後まで採点した後、最初の1枚目から5枚目を採点する。一端採点したものは見直さない。

4．結果

4.1　HEにおけるrater間の関係

Table 1　HEにおけるrater間の相関（Peason correlation coefficients）

	A	B	C	D	E	F	G	H	I	J	K
A	1.00	0.51	0.39	0.54	0.52	0.52	0.28	0.57	0.40	0.21	0.52
B		1.00	0.61	0.65	0.56	0.66	0.53	0.64	0.57	0.37	0.59
C			1.00	0.67	0.58	0.68	0.57	0.59	0.57	0.31	0.71
D				1.00	0.70	0.73	0.62	0.73	0.54	0.18	0.75
E					1.00	0.71	0.54	0.70	0.56	0.39	0.66
F						1.00	0.55	0.65	0.59	0.29	0.71
G							1.00	0.67	0.62	0.32	0.67
H								1.00	0.70	0.44	0.62
I									1.00	0.49	0.60
J										1.00	0.24
K											1.00

A-Kは11名のraterを無作為に並べたもの（以下のTableにおいても同様）

4.2　CEにおけるrater間の関係

Table 2　CEにおけるrater間の相関（Peason correlation coefficients）

	A	B	C	D	E	F	G	H	I	J	K
A	1.00	0.84	0.66	0.69	0.76	0.73	0.72	0.69	0.72	0.74	0.68
B		1.00	0.69	0.73	0.77	0.74	0.77	0.75	0.69	0.76	0.73
C			1.00	0.80	0.72	0.73	0.80	0.74	0.78	0.79	0.71
D				1.00	0.73	0.67	0.81	0.79	0.77	0.78	0.75
E					1.00	0.73	0.78	0.75	0.74	0.70	0.78
F						1.00	0.80	0.76	0.72	0.73	0.65
G							1.00	0.80	0.85	0.76	0.79
H								1.00	0.78	0.75	0.72
I									1.00	0.79	0.73
J										1.00	0.74
K											1.00

5.　考察

5.1　HEとCEのrater間の比較

HE、CEという2つの評価方法においてrater間の評点はどの程度一致しているのであろうか。度数分布をTable3に示した。CEの場合の*r*値は、HEに対し相対的に高く、rangeも .60 ～ .90の間であるのに対し、HEのrangeは .10 ～ .80と大きい。

Table 3 HEとCEの相関係数の度数分布の比較

相関係数（*r*）	HEの度数	HEの累積相対度数	CEの度数	CEの累積相対度数
0-.10	0	0%	0	0%
.10-.20	1	1.90%	0	0%
.20-.30	3	7.40%	0	0%
.30-.40	5	16.7%	0	0%
.40-.50	3	22.2%	0	0%
.50.-.60	19	57.4%	0	0%
.60-.70	14	83.3%	8	14.8%
.70-.80	9	100%	39	87.0%
.80-.90	0	100%	7	100%
.90-1.00	0	100%	0	100%

では、rater内のHEとCEの一致度はどうであろうか。Table4に示したとおり、最も相関の高いraterではr = .936であるのに対し、r = .462と食い違いの大きなraterまで、かなりの個人差があることがわかる。

5.2　評価方法と要する時間の関連

次に採点に要する時間と評価方法の関係について検討しよう（Table5）。まず、評価方法と同様raterの個人差が大きく見られるが（例えば、HEにおいては一枚あたりに要する時間は、最短で36秒、最長で2分；CEにおいては、最短で1分20秒、最長で4分24秒）、全体としてはHEの方が短い[1)]。

Table 4　CEとHEのrater内における相関

Rater	Peason *r*
A	0.568
B	0.848
C	0.739
D	0.936 (Max.)
E	0.729
F	0.843 (Med.)
G	0.726
H	0.919
I	0.741
J	0.462 (Min.)
K	0.867

Max.: 最大値, Med.: 中央値
Min: 最小値

Table 5　CEとHEそれぞれの採点に要する時間

Rater	HE (total time)	min/sheet	CE (total time)	min/sheet
A	90	1.8	150	3
B	100	2	130	2.6
C	45	0.9	70	1.4
D	60	1.2	165	3.3
E	60	1.2	180	3.6
F	36	0.72	75	1.5
G	95	1.9	220	4.4
H	60	1.2	120	2.4
I	31	0.62	67	1.34
J	83	1.66	76	1.52
K	80	1.6	140	2.8
Mean	67.27	1.35	126.64	2.53

6.　結論

2つの評価方法のrater間のばらつきを見る限り、基準を決めてサマリーの評価をする方が信頼性が高くなることを今回の結果は示している。入試において英文和訳や要旨説明を採点する際には、HE はより危険性をはらんでいるといえるだろう。しかし、一方でCEでは要する時間も2倍近くになることも事実である。今回は相関係数を中心に検討したが、実際にどのようなポイントで採点がずれているのかという細部にいたる考察の詳細については発表時に報告をする。また、raterに対するトレーニングにより採点の信頼性が上がるのか、それともあまり変わらないのか、などは今後の課題である。

References

JACETハンドブック作成特別委員会　『大学設置基準改正に伴う外国語（英語）教育改善のための手引き（2）』（1996): 44-50, 大学英語教育学会.
Kirkland, M. R. & M. A. P. Saunders. "Maximizing student performance in summary writing: managing cognitive load." *TESOL Quarterly* 25 (1991): 105-121.
Perkins, K. "Using objective methods of attained writing proficiency to discriminate among holistic evaluations." *TESOL Quarterly*14 (1980): 61-67.

注：1）CEによる採点時には、基準ごとに色の異なるマーカで印を付けながら進めが、その分時間が若干多くかかった可能性も否定できない。

附記：本研究はLLA 関西支部外国語学習ストラテジー研究グループに所属する会員による共同研究成果の一部である。発表者以外の会員は以下の通りである。門田修平（関西学院大学）、三島篤志（帝塚山学院大学）、中西義子（大阪国際女子短期大学）、高島淳子（大阪国際女子短期大学：非常勤）、田中朋子（相愛女子短期大学）、吉村満知子（京都教育大学：非常勤）。