

# 高校生の自由英作文はどのように評価されているのか —分析的評価尺度と総合的評価尺度の比較を通しての検討—

## **How are High School Students' Free Compositions Evaluated by teachers and Teacher Candidates?: A Comparative Analysis between Analytic and Holistic Rating Scales**

山西博之（やまにし ひろゆき）  
広島大学

The aim of this paper was to clarify the characteristics of the evaluation of high school students' free compositions by high school teachers and university students who are teacher-candidates. This was done by analyzing a comparison of the results of analytic and holistic rating scales. In addition, information about their consciousness regarding the evaluation and instruction of free compositions was obtained with the help of a questionnaire. The pedagogical implications were explored with the help of the analyses and the questionnaire.

In recent years, free compositions in English (for example, explaining a particular situation and arguing for or against a statement) have been increasingly required in university entrance examinations in Japan. Such kinds of activities are also required in situations of actual communication. Moreover, the renewed *Course of Study* (Ministry of Education, 1999) has been conducted in high schools since the academic year of 2003. The overall objective of the *Course of Study* is to develop students' practical communication abilities such as expressing their own ideas through written English. To satisfy these requirements, the chances of instructing students on free compositions in high school classrooms will be increased. In such a situation, it is important to obtain information on how teachers and teacher-candidates instruct and evaluate students' free compositions.

The survey consisted of several analyses of the teachers' and teacher-candidates' evaluation data, and a questionnaire concerning their consciousness of the evaluation and instruction of free compositions. The raters who evaluated high school students' free compositions were 10 high school teachers from a national high school and two public high schools and 6 university students who belonged to a teacher-training course in a national university and had experience with practical teaching.

First, the raters evaluated 40 compositions written by 20 high school students who belonged to the first and the second grade of a national high school in the *Chugoku* region in Japan (all 20 students wrote two kinds of compositions). The evaluations were carried out using holistic rating scales and two kinds of analytic rating scales. In this survey, two subscales given by Ishida and Mori (1985), that is, the *Objective Holistic Rating Scale* (How good do you think the composition is?) and the *Subjective Holistic Rating Scale* (How do you personally like the composition?) were adopted as holistic rating scales. At the same time, two kinds of rating scales, that is, Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey's (1981) ESL Composition Profile and National Institute for Educational Policy Research's (2002) *kantenbetsu-hyoka* (an analytic rating), were adopted as analytic rating scales. The following two points were discussed on the basis of a comparison of the result of the evaluation of the holistic and analytic rating scales using descriptive statistics, Pearson's product-moment correlation coefficients, and Cronbach's alpha coefficients: 1) To clarify the characteristics of the evaluations performed by the teachers and teacher-candidates, and 2) To examine the reliability and the validity of the *kantenbetsu-hyoka* scale. From the results, it appeared that the teachers' evaluations were consistent within and beyond the rating scales, while those performed by the university students who were teacher-candidates were slightly inconsistent within and beyond the rating scales. At the same time, it also appeared that the validity of the *kantenbetsu-hyoka* scale was not higher than that of the other scales.

Second, an open-ended questionnaire, which included two parameters, was given to the same raters. The first question was what they regarded as being of greatest importance while instructing students on free compositions, and the second was what they regarded as being of greatest importance when they evaluated them. The following point was discussed on the basis of the questionnaire: 3) To obtain information on the consciousness of teachers and teacher-candidates concerning the evaluation of and instruction on free compositions. From the results of the questionnaire, it appeared that teachers were more sensitive to the conditions of high school students than the university students who were teacher-candidates.

These results led to the pedagogical implications as follows: 1) The validity of the *kantenbetsu-hyoka* scale was not higher than the other scales; therefore, it would be better to use it in combination with other scales, 2) It would be better for teachers to carefully consider each parameter of the analytic rating scales, in order to make good use of the merits of the scale, and 3) It would be better

for university students who are teacher-candidates to have more opportunities of evaluating high school students' compositions and of instructing them on writing compositions, before they become teachers.

本研究では、複数の評価尺度を組み合わせることで、教員と大学生による高校生の自由英作文評価の実態が調査された。総合的評価尺度として「客観的総合評価」と「主観的総合評価」が、分析的評価尺度としてESL Composition Profileと「観点別評価」が使用された。調査では、10名の教員と6名の大学生が、20名の高校生によって書かれた40編の自由英作文を評価した。評価結果を記述統計量、信頼性係数、相関係数から分析することで、教員の評価は大学生の評価よりも一貫性が高いことが示された。また「観点別評価」の信頼性と妥当性も議論され、妥当性がやや低かったことが示された。同時に、自由記述の質問紙による指導と評価の意識調査も行われ、教員は大学生よりも指導現場に根ざした意識を有していることが確認された。これらの結果に対する考察が行われ、そこから教員と大学生の双方に対する教育的示唆が示された。

近年の大学入試では英語の試験において、「～についてあなたの考えを述べなさい」や「～について説明しなさい」といった自由英作文 (free composition) 課題が多く出題されている。このような自由英作文課題は「実践的コミュニケーション能力」の養成を標榜する高等学校での新学習指導要領の実施 (2003年度より) 以降の大学入試では、今後多く出題されることが予想される。このような状況において、実際に指導の一環として高校生の自由英作文を評価する立場にある高等学校の教員と教員を目指す大学生によって、自由英作文がどのように評価されているのかという情報を得ることは重要であると言える。そこで本研究では、教員と大学生によって実際にどのように評価が行われているのかを、複数の評価尺度を組み合わせた評価を行うことで調査した。そして、彼らがどのような観点を重視し、自由英作文の指導と評価を行っているかを自由記述の質問紙によって調査した。そこから、教員と大学生の持つ自由英作文の指導や評価の特徴や意識を捉えることが行われた。本研究では、このような調査から教育的示唆を得ることが目指された。

## 評価方法

### 分析的評価と総合的評価の特徴

自由英作文の評価は難しい問題であるとされている (Cumming, 1997)。評価方法には大きく分類して、分析的評価 (analytic rating) と総合的評価 (holistic rating) の2種類がある。両者の違いについてはCarr (2000) やCumming (1997) に詳しいが、ここでは簡単にまとめている。

分析的評価の利点は、測定する技能を異なった項目として定めることで、その項目ごとの詳細な評価を行うことができる点である。その結果として、項目ごとの診断的 (diagnostic) なフィードバックを行うことが可能になる。欠点としては、どのような下位技能を設定すべきかの決定が困難なことや、項目が増えすぎてしまうと丁寧な評価をした場合、労力が要求されることなどが挙げられる。また、第2言語や外国語での作文においては、全体的な「言語能力」 (language ability) の中から「作文技能」 (writing expertise) や「第2言語能力」 (second-language proficiency) を区別するべきであるか否かの決定も難し

い点である (Cumming, 1989)。分析的評価の代表的なものに、Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981) のESL Composition Profileがあるが、それは「作文技能」のみを対象とする評価尺度である。

その一方、総合的評価の利点は、項目数が少なく簡便に評価できるため労力が少なく済む点である。また、全体的な印象に基づくため様々な要因を汲み取った評価を行うことができる点も利点であると言える。欠点としては、利点の裏返しで、様々な要因が混在しているため評定者ごとに評価基準が異なる可能性が高いことが挙げられる。また、複数の作文間の得点に差があった場合、どうしてその差が生じたのかを説明することが困難である点も欠点であると言える。

### 分析的評価尺度と総合的評価尺度の組み合わせ

ここから、上で見るように分析的評価と総合的評価が異なる特徴を持つことは明らかであり、両者の適切な使い分けが重要である。しかし、どちらの方法を用いるにしても、両者を比較することでそれぞれの特徴を捉えてから使用する必要がある (Hamp-Lyons, 1995)。そのために、分析的評価の影響を評定者に与えることを避けるために先に総合的評価を行い、それに続いて分析的評価を行い、両者の比較をするという研究がある。このような研究の目的とするところは、評定者が作文を評価する時にどの分析的評価尺度項目を重視しているのかを、分析的評価尺度の影響を受けていない状態の総合的評価との関連から調査することである。

日本語での児童の作文の評価研究で、石田・森 (1985) は、総合的評価尺度と分析的評価尺度の下位項目との比較を行った。石田・森では、小学校教員10名と大学生10名が、児童の作文に対して「客観的総合評価」と「主観的総合評価」という2つの尺度を用いて総合的評価を行った。そして、それぞれの結果を目的変数とし、学習指導要領などから決定した4つの分析的評価尺度項目 (説得力、表現力、構成力、独創性) を説明変数とした重回帰分析を行い、標準偏回帰係数からの分析・考察を行った。石田・森のいう「客観的総合評価」とは「客観的にみてどのくらいまい作文といえるか」に関するもので、「主観的総合評価」は「主観的にみてどのくらいその作文が好きか」というものである。その結果、教員は日常的な指導経験の中から得た教員特有の教育観や児童観を反映した評価を行い、また、大学生は一般的、常識的な評価を行っているということが示された。

同様の研究として、梶井 (2001) では、21名の小学校教員が、国語科の学習指導要領などから作成した18の分析的評価尺度を用いて、児童の作文を評価した。そして、その結果と石田・森 (1985) の2つの総合的評価尺度を用いた評価結果がどれほど関連しているかを重回帰分析や相関係数の分析などから比較することで、分析的評価尺度項目の信頼性と妥当性を検討した。その結果、学習指導要領に基づいた分析的評価尺度の信頼性と妥当性は各項目とも低く、わずかな項目のみが「使用できる項目」であることが示され、学習指導要領に基づいた分析的評価の困難さを浮き彫りにした。なお、梶井 (2001) は「客観的総合評価」を「うまさ」、「主観的総合評価」を「好み」と呼んだ。同様に本研究でも今後「うまさ」と「好み」という表記を用いる。

そこで本研究では、自由英作文の総合的評価尺度として「うまさ」と「好み」を使用し、分析的評価尺度との比較を行うことで、評定者が高校生の書き

た自由英作文をどのように評価しているのかという実態を探る。併せて、本研究では自由英作文の評価結果からのみでは見ることの難しい、評定者が指導と評価に関してどのようなことを重視しているかという意識を、自由記述形式の質問紙調査の実施によって見ることにする。

## 調査

### 目的

高校生が書いた自由英作文を素材として、総合的評価尺度と分析的評価尺度（ESL Composition Profileと「観点別評価」）の関係を比較することで、以下の3点を目的とする。

- 1) 教員による評価の特徴と大学生による評価の特徴を明らかにする。
- 2) 「観点別評価」の信頼性と妥当性を明らかにする。  
また、同時に行った自由記述形式の指導と評価に関する質問紙から、
- 3) 教員と大学生の自由英作文の指導と評価において何が重視されているかを明らかにする。

### 調査方法

#### 評定者

評定者は中国地方の高等学校の現職の英語科教員10名（1校の国立高等学校教員8名と2校の公立高等学校教員各1名；教員歴8-32年、平均教員歴17.3年；男性7名、女性3名）、そして英語教員になる意志が強く実際の高校での指導を教育実習で経験済みの大学生6名（国立大学の学部4年生4名と修士課程の大学院生2名；男性1名、女性5名）であった。なお、大学生6名中、非常勤講師などでの定期的な高校生への指導の経験を有している者はいなかった。

#### 対象者

評価の対象者は、中国地方の国立高等学校の高校生20名（高校1年生10名、男女5名ずつ；高校2年生10名、男女5名ずつ）であった。彼らは週5時間の英語の授業を受けており、そのうち2時間の「ライティング」の授業時間中に自由英作文を書くための指導を受け、書いた自由英作文に対し教員からのフィードバックを得ていた。

#### 素材

評定者が評価する素材は、高校生20名が書いた2種類の自由英作文をそれぞれ1編ずつ、計2編の自由英作文をA4の用紙1枚に見開きで印刷したものであり、評価は1-10点で行われた。高校生に課された課題は大学入試の自由英作文課題を模したもので、そのうち1つは自らの経験や感想を表現するもので（課題1）、もう1つは絵を見てその内容を説明するもの（課題2）であった。

#### 評価の手順

作文の評価の際には、渡部・曹（1992）が指摘するように、字の丁寧さが評価に影響を与えることが考えられるため、生徒が書いた自由英作文はあらかじめ

めすべてワープロでタイプし直した上で評定者に手渡した。具体的には1人の生徒あたりにA4用紙の見開きを1枚ずつ用い、左側のページに課題1を、右側に課題2をそれぞれ印刷した。そして、自由英作文各1編の下側に各評価尺度用のスケールを掲載した。この20名分の自由英作文（計40編）と評価尺度の用紙の1つづりを1セットとした。

本研究では、各評価尺度（総合的評価、ESL Composition Profile、観点別評価）が他の尺度に与える影響を低減するために、上記の自由英作文と評価尺度のセットを3つの評価尺度ごとに、計3セット作成した。その際、特に分析的評価尺度の項目を評定者が知ることによって総合的評価に影響が及ぶことを避けるために、「総合的評価」→「ESL Composition Profile」→「観点別評価」のセット順に評価を行うように指示した。

そして、「総合的評価」セットの最終ページにおいて、教員に対しては普段の自由英作文の指導と評価においてどのようなことを重視しているのかについて、大学生に対しては実際に教員になった時どのようなことを重視して自由英作文の指導と評価を行いたいのかということについて、それぞれ自由記述形式で尋ねた。自由記述の質問紙を「総合的評価」セットの最終ページで行った理由は、2種類の分析的評価尺度項目の影響を受けない状態でのデータの収集を目指したためである。

### 評価尺度

本研究では総合的評価尺度との比較のための分析的評価尺度として、Jacobson et al. (1981) のESL Composition Profileと文部科学省による学習指導要領で「学力観」として示されている「観点別評価」を用いることとする。ESL Composition Profileを用いる理由は、過去の多くの研究で用いられている尺度であるためで、「作文技能」(Cumming, 1989) に特化した尺度であるためである。また、「観点別評価」を用いる理由は、学校現場で実際に用いられる尺度であるためで、導入されて間もない「観点別評価」そのものの信頼性と妥当性を調査する必要があるためである。

「総合的評価」セットに含まれるものは、客観的総合評価尺度の「うまさ」（「客観的に見て、どのくらいこの作文がうまいと思いますか？」）と主観的総合評価尺度の「好み」（「主観的に見て、どのくらいこの作文が好きですか？」）の2項目である。また、他の評価尺度との比較のために、従来の研究において通常は総合的評価として取り扱われている印象的な評価を見るための「総合的に見て、この作文に何点を与えますか？」という1項目（以後、「印象的評価」と表記）も「総合的評価」のセットに加えた。

「ESL Composition Profile」セットに含まれるものは、Jacobs et al. (1981) の挙げた5項目（content, organization, vocabulary, language use, mechanics）である。評価はJacobs et al. が示した4つの基準（poor, fair, good, very good）に基づいた。Jacobs et al. では、それぞれの項目に異なる重みづけがなされ、全項目の合計で34-100点の間で点数がつけられるが、本研究では重みづけを廃し各項目とも一律に10点満点にした。10点満点にした場合の4つの基準の当てはめ方として、Sasaki and Hirose (1999) がESL Composition Profileを模して作成した日本語説明文用の評価尺度である「国語説明文評価表」と同様に、poor = 1-2点、fair = 3-5点、good = 6-8点、very good = 9-10点とした。

「観点別評価」セットに含まれるものは、国立教育政策研究所が中学校用の評価規準（国立教育政策研究所，2002）を基に現在検討中の高等学校用の観点のうち、外国語科英語科の必修科目「英語I」の中の技能である「書くこと」で用いられる表1に示された6項目である。<sup>注</sup> 松浦（2002a, 2002c, 2002d）によると、本来「観点別評価」とは、測定しようとする技能を表1で挙げた「規準」（何を評価するのかという評価の際の切り口＝「観点」）によってあらかじめ設定し、その到達の「基準」（どの程度目標に到達したのかに関する水準）によって評価するという「目標基準準拠評価」であるため、本研究で行うような事後的な評価のみに用いられるためのものではなく、普段の授業中の取り組みなどと併せて評価する際の基準である。しかし、実際の高等学校での評価現場では、実力テストや校内模擬試験などのテスト実施後に「観点別評価」を行うこともあり得ると考えられる。そのような場合において、「観点別評価」の「観点」が書かれた結果のみから自由英作文の評価を行うといった用途に使用できるかどうか、つまり「観点別評価」そのものの信頼性と妥当性を調査する必要がある。そのため、本研究では「観点別評価」に関してもESL Composition Profileと同様に総合的評価との比較を行うものとした。

表1: 本研究での「観点別評価」の内訳(「書くこと」の評価規準)

関心・意欲・態度	表現の能力	知識・理解
【言語活動への取組】 「書くこと」の言語活動に積極的、主体的に取り組んでいる。	【正確な表現の能力】 日常的な話題について、得た情報や自分の考えを英語で正しく書くことができる。	【言語についての知識】 言語や言語の運用についての基本的な知識を身に付けている。
【コミュニケーションの継続】 さまざまな工夫をする中で、コミュニケーションを主体的に続けている。	【適切な表現の能力】 日常的な話題について、場面や目的に応じて、英語で適切に書くことができる。	【文化についての理解】 日常的な話題について、英語の学習において取り扱われた文化について理解している。

### 分析の手順

2種類の課題についての得点は合計され、すべての評価尺度項目において生徒1人あたりの得点は20点満点とした。評定者の与えた得点は項目ごとに合計し平均値を求めた。課題を合計した理由は、複数の課題を評価することでより一般性を高めるためで、評定者の得点を平均した理由は個々人の評定者の評価特徴ではなく、集団としての教員と大学生の評価特徴を見るためである。20名の生徒それぞれの得点を合計しなかった理由は、教員と大学生の対応のある比較を可能にするためである。また、本研究の目的は上述の通り、集団としての教員と大学生の評価特徴を見ることであるため、対象となる生徒の学年、性別は分析の対象とはしなかった。

## 結果

## 全体的傾向

はじめに、教員による評価と大学生による評価の特徴の全体的な傾向として、表2に示されたように、すべての評価尺度のすべての項目において教員による評価の平均値（20点満点）は大学生による評価の平均値を下回った。この平均値の差に対して対応のある平均値の差の検定（*t*検定）を行った結果（ $\alpha = .05$ ）、すべての項目でその差は統計的に有意なものであった（ $t(19) = 5.41 - 22.58, p < .001$ ）。

表2: 項目ごとの記述統計量（20点満点）（ $N = 20$ ）

	最小値 ( <i>Min.</i> )	最大値 ( <i>Max.</i> )	平均値 ( <i>Mean</i> )	標準偏差 ( <i>SD</i> )	<i>t</i> 値
うまさ（客観的総合評価）	6.90 / 9.00	13.60 / 16.17	9.35 / 12.48	1.46 / 1.62	14.14
好み（主観的総合評価）	6.70 / 9.33	13.10 / 15.67	9.09 / 12.13	1.43 / 1.51	13.92
印象的評価	7.30 / 9.33	13.40 / 16.33	9.73 / 12.48	1.41 / 1.55	11.57
content	7.50 / 11.83	12.60 / 16.50	9.72 / 13.58	1.21 / 1.39	18.56
organization	7.30 / 11.00	13.60 / 16.33	9.71 / 13.20	1.32 / 1.40	15.05
vocabulary	8.10 / 11.33	13.00 / 16.00	9.58 / 13.41	1.13 / 1.15	21.84
language use	7.70 / 10.33	12.80 / 14.83	9.30 / 12.36	1.05 / 1.29	12.16
mechanics	8.60 / 10.83	13.50 / 15.17	10.12 / 13.62	1.04 / 1.09	13.17
言語活動への取組	10.30 / 10.17	15.00 / 17.33	12.48 / 13.58	1.14 / 1.85	5.41
コミュニケーションの継続	9.60 / 9.50	14.30 / 17.00	11.82 / 13.32	1.15 / 1.92	6.45
正確な表現の能力	9.20 / 11.67	13.20 / 15.00	10.59 / 13.13	0.97 / 1.15	12.88
適切な表現の能力	8.80 / 11.00	13.20 / 16.00	10.50 / 12.81	1.00 / 1.36	15.13
言語についての知識	9.10 / 11.17	13.00 / 16.33	10.44 / 12.84	0.92 / 1.42	14.43
文化についての理解	8.22 / 11.50	12.82 / 15.50	10.02 / 12.92	1.02 / 0.99	22.58

注1) 教員による評価／大学生による評価

注2) *p*値は全項目  $p < .001$

## 評価尺度の信頼性

次に、評価尺度の信頼性を見た。Jacobs et al. (1981) は、信頼性の高い自由英作文用の分析的評価尺度とは、尺度内の信頼性係数（ $\alpha$  係数）が高く、項目間の相関係数がそれぞれ高い、つまり内部一貫性（internal consistency）が高いものであるとしている。なぜなら、分析的評価尺度の各項目は、自由英作文の中心として測定したい能力を少しずつ異なる側面から評価するためのものであるから（Jacobs et al., 1981）、ここで大きく評価の異なる項目が含まれていると測定したい能力に対する信頼性のある評価結果を得ることはできない。今回の調査で用いた分析的評価尺度のうちESL Composition Profileは「作文技能」

を、観点別評価は「学力観」を測定するためのものであると言える。そこで本研究では、この2種類の分析的評価尺度と総合的評価尺度の内部一貫性を尺度内の信頼性係数と項目間の相関係数という2つの観点から見た。併せて、得られた結果を用いて教員と大学生の比較を行った。

#### 各評価尺度内の信頼性係数

評価尺度の内部一貫性を調べるための第1の観点として、尺度内の信頼性係数（クロンバックの $\alpha$ 係数）を求めた。その結果、表3のように、どの評価尺度においても高い内部一貫性があったことが示され、尺度間の差異は大きくなかった（教員 $\alpha = .96-.98$ ；大学生 $\alpha = .90-.98$ ）。また、教員と大学生の評価間の差異も大きなものではなかった。

表3: 各評価尺度内の信頼性係数（ $\alpha$ 係数）（ $N = 20$ ）

	総合的評価 (2項目)	ESL Composition Profile (5項目)	観点別評価 (6項目)
教員	.96	.97	.98
大学生	.98	.90	.93

#### 各評価尺度の項目間の相関係数

第2の観点として、各評価尺度の項目間におけるピアソンの積率相関係数を求めた。結果は表4に示された通りである。なお、 $N = 20$ であるため $r > .44$ の場合、相関係数は5%水準で有意である。

まず、「総合的評価」セットでは、教員・大学生ともに「うまさ」と「好み」の間に非常に高い相関があることが確認された（教員 $r = .97$ 、大学生 $r = .96$ ）。

次に、「ESL Composition Profile」セットの項目間の相関係数を求めた結果、教員の評価は各項目間でかなり高い相関（ $r = .74-.93$ 、平均.86）を持ち項目間の差が小さい（最大差.19）ことが、一方大学生の評価は中程度の相関（ $r = .39-.84$ 、平均.58）を持ち項目間の差が大きい（最大差.45）ことが示された。

最後に、「観点別評価」セットを見ても、教員と大学生の評価を比較した結果において、ESL Composition Profileと同様の項目間の相関係数の関係が示された（教員の評価は $r = .74-.97$ 、平均.89、最大差.23；大学生の評価は $r = .46-.97$ 、平均.69、最大差.51）。

#### 評価尺度の妥当性

続いて、各評価尺度の妥当性を「総合的評価」のセット中で測定した「印象的評価」との相関係数を見ることで求めた。作文を一見して得られた「印象的評価」は、上述したように作文の様々な要因を汲み取った全体的な評価を行うのに適した尺度であると言える。そのため、「印象的評価」との相関係数が高い評価尺度は、各項目の得点を合計することによって、作文の全体的な評価を

表4: 各評価尺度セット内項目の相関行列 ( $N=20$ )

「総合的評価」セット						
	うまさ	好み	印象的評価			
うまさ (客観的総合評価)	-	<u>.96</u>	<u>.95</u>			
好み (主観的総合評価)	.97	-	<u>.91</u>			
印象的評価	.98	.97	-			
「ESL Composition Profile」セット						
	content	organization	vocabulary	language use	mechanics	
content	-	<u>.84</u>	<u>.76</u>	<u>.55</u>	<u>.39</u>	
organization	.90	-	<u>.80</u>	<u>.72</u>	<u>.39</u>	
vocabulary	.86	.90	-	<u>.69</u>	<u>.70</u>	
language use	.74	.86	.88	-	<u>.67</u>	
mechanics	.78	.88	.91	.93	-	
「観点別評価」セット						
	取組	継続	正確	適切	知識	理解
言語活動への取組	-	<u>.97</u>	<u>.46</u>	<u>.70</u>	<u>.60</u>	<u>.79</u>
コミュニケーションの継続	.97	-	<u>.48</u>	<u>.71</u>	<u>.61</u>	<u>.80</u>
正確な表現の能力	.83	.82	-	<u>.84</u>	<u>.90</u>	<u>.71</u>
適切な表現の能力	.84	.84	.95	-	<u>.84</u>	<u>.87</u>
言語についての知識	.78	.74	.95	.97	-	<u>.85</u>
文化についての理解	.90	.87	.93	.96	.94	-

注1) 各相関行列の左下三角は教員、下線の右上三角は大学生による評定

注2) 相関係数は  $r > .44$  の場合、5%水準で有意

行うことのできる尺度である、つまり「印象的評価」との併存的妥当性の高い尺度であると言える。そこで本研究では、各評価尺度のすべての項目の得点の平均値と「印象的評価」の相関係数を求めることで両者の関連を分析した。併せて、得られた結果を用いて教員と大学生の比較を行った。

結果は表5に示された通りである。まず、「うまさ」と「好み」の合計である「総合的評価」は「印象的評価」と非常に高い相関があった（教員  $r = .98$ 、大学生  $r = .94$ ）。次にESL Composition Profileは、教員・大学生ともにかかなり高い相関があることが示された（教員  $r = .88$ 、大学生  $r = .82$ ）。最後に「観点別評価」は、教員の評価はかなり高い相関を示したのに対し、大学生の評価では中程度の相関を示すにとどまった（教員  $r = .81$ 、大学生  $r = .67$ ）。

表5: 「印象的評価」と各評価尺度の項目合計との相関係数 ( $N = 20$ )

	総合的評価 (2項目)	ESL Composition Profile (5項目)	観点別評価 (6項目)
教員	.98	.88	.81
大学生	.94	.82	.67

注1) 相関係数は  $r > .44$  の場合、5%水準で有意

### 「うまさ」と「好み」との比較による検討

その次に、各分析的評価尺度の下位項目と「うまさ」「好み」との相関係数による比較を行うことで、教員と大学生の評価特徴の違いを見た。

今回の調査のデータでは、表4に示されたように「うまさ」と「好み」の相関が非常に高く、また「うまさ」と「好み」それぞれと印象的評価の相関も非

表6: 「うまさ」「好み」と分析的評価尺度項目との相関係数 ( $N = 20$ )

	ESL Composition Profile			
	教員		大学生	
	うまさ	好み	うまさ	好み
content	.84	.86	.82	.85
organization	.86	.85	.81	.81
vocabulary	.94	.92	.67	.70
language use	.88	.84	<u>.65</u>	<u>.46</u>
mechanics	.89	.85	<u>.46</u>	<u>.59</u>
ESL Composition Profile合計	.93	.92	.79	.83
	観点別評価			
	教員		大学生	
	うまさ	好み	うまさ	好み
言語活動への取り組み	.72	.75	<u>.54</u>	<u>.69</u>
コミュニケーションの継続	.68	.74	<u>.53</u>	<u>.83</u>
正確な表現の能力	.84	.84	<u>.67</u>	<u>.80</u>
適切な表現の能力	.89	.90	.74	.80
言語についての知識	.90	.89	.78	.82
文化についての理解	.83	.86	<u>.75</u>	<u>.88</u>
観点別評価合計	.85	.87	.72	.79

注1) 下線は「うまさ」「好み」間で .10 以上の差がある項目

注2) 相関係数は  $r > .44$  の場合、5%水準で有意

常に高い。そのため、分析的評価尺度の下位項目と「うまさ」との相関と、その下位項目と「好み」との相関には大きな差は生じないはずである。ここで「うまさ」との相関係数と「好み」との相関係数が大きく異なる分析的評価尺度の下位項目があった場合、その項目は「うまさ」か「好み」に片寄った評価がなされた項目であると考えられる。そこで、そのような項目があるかどうかを見るために、教員と大学生による分析的評価尺度の各下位項目と「うまさ」「好み」との相関係数を求めたものが表6である。

ここでは、分析的評価尺度の各下位項目と「うまさ」と「好み」との間の相関係数に差異があるかどうかを個別に見た。具体的には、1つの目安として「うまさ」と「好み」の間で.10以上の相関係数の差がある項目を見た。その結果、教員による評価では.10以上の差がある項目は無かったが、大学生の場合にはESL Composition Profileのlanguage useとmechanicsの2項目で、また「観点別評価」の「言語活動への取り組み」「コミュニケーションの継続」「正確な表現の能力」「文化についての理解」の4項目で.10以上の差があった。特に「観点別評価」における差は大きく、それらはすべて「好み」との相関が高いことが分かった。

### 質問紙調査で得られたデータ

最後に、自由記述形式の質問紙から得られた、評定者が指導の際と評価の際にどのようなことを重視するかというデータは、アイデア・ユニットの単位で分析された。具体的には、1文で1つの内容を表している回答はそのまま1つのアイデア・ユニットとし、複数の内容を含む文は内容ごとに複数のアイデア・ユニットに分割した。そして、それぞれのアイデア・ユニットを整理し、KJ法を用いて表7に示された6つのカテゴリーに分類した。

その結果、教員・大学生にかかわらず、指導と評価ともに ESL Composition Profile で取り扱われる技能 (content, organization, vocabulary, language use, mechanics) と類似した回答をした評定者が多かったことが確認された。しかし、中にはそれらの技能つまり「作文技能」(Cumming, 1989) 以外の項目を重視した評定者もいた。主に教員による自由記述から、どのような課題を設定して指導するかに関する記述や、生徒の自由英作文がその課題状況に合致しているかどうかを評価するという記述など、日常的な指導や評価現場に根ざした記述が得られた。また、教員の自由記述からは、指導や評価の際に生徒に対する配慮を行っていることを示唆させる記述が得られた。

### 考察

ここでは、以上の結果に対する考察を行っていく。

まず、評価尺度の信頼性のうち、信頼性係数には教員・大学生とも、どの尺度でも大きな差はなかった。しかしながら、項目間の相関係数に関しては、ESL Composition Profileにおいて教員の評価では項目間の差が小さく、大学生では差の大きな項目同士があることが確認された。この結果を、1つの基準として Jacobs et al. (1981) が提示した599人のESLの生徒に対する評価における項目間の相関係数 ( $r = .57-.88$ , 平均.73) と比較すると、教員による評価は項目間の幅が小さく、逆に大学生の評価は差が大きなものであったと言える。基準が存在しないものの「観点別評価」においても同様の構造であったと言える。

表7: 自由記述で得られた指導と評価における重視項目

カテゴリー	指導において重視することの 具体例		評価において重視することの 具体例	
	教員 (N=10)	大学生 (N=6)	教員 (N=10)	大学生 (N=6)
文章の組立てに関するもの	<ul style="list-style-type: none"> <li>・構成</li> <li>・ディスコースマーカ―を効果的に用いること</li> <li>・論理の展開の仕方</li> <li>・パラグラフの構成</li> <li>・ディスコースの工夫</li> <li>・構成 (文と文の関連)</li> <li>・易しい構文であわすこと</li> </ul>	<ul style="list-style-type: none"> <li>・意味の流れが論理的であること</li> <li>・話の流れがきちんとしていること</li> <li>・流れを意識して書くこと</li> <li>・文章全体の構成</li> <li>・相手に伝わるための書き方</li> <li>・文学性を出すような書き方</li> </ul>	<ul style="list-style-type: none"> <li>・構成 (文と文の関連)</li> <li>・文と文とのつながりが明確かどうか</li> <li>・文構成がはっきりしていること</li> <li>・表現に工夫がみられること</li> </ul>	<ul style="list-style-type: none"> <li>・文章としてのまとまり</li> <li>・一文一文につながりを持たせる意味が伝わっているかどうか</li> <li>・理解されやすい構文の使用</li> </ul>
伝達内容に関するもの	<ul style="list-style-type: none"> <li>・何を言いたいのが相手にわかるように書くこと</li> <li>・伝えたいメッセージは何かを明確にさせること</li> <li>・表現したい内容を整理してから書くこと</li> <li>・論旨がはっきりと伝わること</li> <li>・内容のおもしろさ</li> </ul>	<ul style="list-style-type: none"> <li>・伝えたい内容を正確に書くこと</li> <li>・伝えたい内容を面白く書くこと</li> <li>・伝えたいことが適切に伝えられること</li> <li>・伝えたいことが正確に伝えられること</li> </ul>	<ul style="list-style-type: none"> <li>・内容のおもしろさ</li> <li>・内容がユニークであるか</li> <li>・内容</li> <li>・表現したい内容が明確に伝わるかどうか</li> <li>・話の内容が持つアピール度</li> <li>・内容を十分に伝える表現力</li> </ul>	<ul style="list-style-type: none"> <li>・伝えたいことができちゃんと伝わるかどうか</li> <li>・自分が言いたいことを諦めずに書いているか</li> </ul>
文法に関するもの	<ul style="list-style-type: none"> <li>・文法の正しい使用</li> <li>・文法的な正確さ</li> </ul>	<ul style="list-style-type: none"> <li>・正しく文法を使うこと</li> </ul>	<ul style="list-style-type: none"> <li>・文法的な正確さ</li> <li>・基本的文法事項が正しいか</li> </ul>	<ul style="list-style-type: none"> <li>・文法の使い方</li> </ul>
語彙に関するもの	<ul style="list-style-type: none"> <li>・易しい語彙であわすこと</li> <li>・語彙・イディオムの適切な使用</li> </ul>	<ul style="list-style-type: none"> <li>・正しく語彙を使うこと</li> </ul>	<ul style="list-style-type: none"> <li>・語彙が適切に使われているか</li> </ul>	<ul style="list-style-type: none"> <li>・語彙の使い方</li> <li>・理解されやすい語の使用</li> <li>・知らない単語を言い換えて書いているか</li> </ul>
課題に関するもの	<ul style="list-style-type: none"> <li>・書きやすいトピックの選定</li> <li>・書きたくないトピックの選定</li> </ul>	<ul style="list-style-type: none"> <li>・誰に対して書くかという場面の設定</li> <li>・何を目的に書くのかという場面の設定</li> </ul>	<ul style="list-style-type: none"> <li>・問いで求められていることに答えているかどうか</li> <li>・与えられた課題に添った内容であるかどうか</li> <li>・課題の趣旨にあっている内容か</li> </ul>	
生徒への配慮に関するもの	<ul style="list-style-type: none"> <li>・テーマに縛られない自由な発想</li> <li>・ミスを恐れずに伝えたいことを表現しようとする勇氣</li> <li>・数多く書いてみる</li> </ul>		<ul style="list-style-type: none"> <li>・一貫性 (公平性があるようにする)</li> <li>・アカウントビリティ</li> </ul>	

次に、各評価尺度と「印象的評価」との併存的妥当性であるが、教員・大学生とも「うまさ」と「好み」の合計点と「印象的評価」との相関が高く、この2項目による総合的評価で「印象的評価」と同様の全体的な評価を行うことは容易であったと言える。また、ESL Composition Profileの各項目の合計点で作文を評価することでも、「印象的評価」の結果とかなり近い全体的な評価を行うことができたと言える。しかしながら、大学生が「観点別評価」尺度で自由英作文評価を行うことでは、「印象的評価」との相関係数がやや低くなることから併存的妥当性は高いとは言えず、結果を全体的な評価に結びつけることはやや困難であった可能性が示された。

また、分析的評価尺度の個々の項目と「うまさ」と「好み」の相関係数を見た結果、大学生の評価のうち特に「観点別評価」の下位項目の多くにおいて、「うまさ」よりも「好み」寄りで評価が行われた可能性が強いことが示され、大学生による分析的評価は「うまさ」と「好み」の差が大きなものであった可能性が示された。逆に、教員による評価はESL Composition Profileと「観点別評価」の双方で、「うまさ」と「好み」の差が小さなものであったことが示された。

最後に、自由記述のデータは、教員は大学生と同様に生徒の「作文技能」を把握した上で、それに加えて必要と考えられる指導や評価に関する意識を有していると解釈することが可能であろう。

以上の考察から、目的に関連したさらなる考察を行っていく。

目的1)の教員と大学生の評価の特徴に関して、教員はどのような評価尺度やその下位項目を用いても評価の幅が少ないのに対して、大学生は分析的評価尺度の下位項目で評価に幅があり、さらに「うまさ」よりも「好み」を重視した評価を行っていたことが示された。このことは石田・森(1985)が指摘したように、教員は日常的な作文に対する指導と評価の中で独自の評価観を確立し、評価尺度間・評価尺度内の双方で一貫性の高い評価を行ったと解釈することができよう。逆に大学生は分析的評価尺度の下位項目1つ1つに応じた評価を行ったものの、結果としては「観点別評価」と「印象的評価」の併存的妥当性や、「うまさ」「好み」の比較からも明らかのように、特に評価尺度間の一貫性が高くなかったと解釈することができよう。このことは多分に教員という集団と大学生という集団とでは、自由英作文の指導や評価の経験の絶対量が異なることに起因すると考えられる。しかしながら、今回の分析は両者を集団として捉え個々人を平均したデータであるため、より詳細な評価の実態を解明するためには、今回は行わなかったような個々の評定者のデータの分析結果も併せて議論する必要があるだろう。そこで今回は分析は行わないものの、参考資料としてAppendixに今回の評定者全員の相関行列を集団ごとに示すものとする。

目的2)の「観点別評価」の信頼性と妥当性は、他の評価尺度と比較して、信頼性においては大きな違いはなかったが、妥当性（「印象的評価」との併存的妥当性）においては、大学生による評価でやや低いものとなった。そのため、今回の調査結果からは「観点別評価」の信頼性と妥当性は、必ずしも高いものであったとは言えないのではないだろうか。しかし、今回は敢えて「目標基準準拠評価」であるという特徴を無視した事後的な評価を行ったため、「観点別評価」の本来の目標とは異なった使用場面での信頼性と妥当性の調査であった。そのため、今回の調査結果のみから「観点別評価」の信頼性と妥当性に関して結論づけることは早計であると言え、「目標基準準拠評価」尺度として使用した際の結果との比較を行うことは今後の課題の1つであろう。

目的3) の自由記述による指導と評価の際の意識の違いも、目的1) の考察と関連して、両集団の自由英作文の指導と評価の経験の絶対量の差に起因するものと考えられる。

### 教育的示唆

結論として、本研究の調査で得られた結果に対する考察からの教育的示唆について述べる。

まず、「観点別評価」を本研究のような事後的な評価に用いると、特に評定者の評価経験が十分ではない場合に、評価結果の「印象的評価」との併存的妥当性が低くなる可能性がある。そのため、「観点別評価」を自由英作文の事後的な評価に使用する際には、他の評価尺度と組み合わせるなどの配慮が必要である。

次に、自由英作文評価における教員と大学生の違いは、指導と評価の経験の絶対量に起因する部分が大きいと解釈されることから、教員と大学生の双方に対する教育的示唆を述べる。まず、評価尺度間の一貫性の高い評価を行ったことが示された教員は、しかしながら評価尺度内の一貫性も高い傾向にあり、分析的評価尺度を用いる場合にはその特色を生かすために各項目の特徴により意識的にすることが望まれるのではないだろうか。次に、評価尺度間と評価尺度内の一貫性のあまり高くなかった大学生は、例えば大学での教員養成課程において実際に高校生の書いた自由英作文を評価するという機会を得ることで、擬似的な経験を積むことが重要なのではないだろうか。そのような経験は、大学生が評価に対する実感を得るきっかけとなり評価の一貫性を高めることに貢献すると考えられる。本研究での調査に参加した大学生の評定者の全員が教育実習を経験済みであることを考慮に入れると、教員になるまでに擬似的な指導・評価の経験を積むことは重要であると指摘できよう。

評定者の人数や、評価対象になった高校生が1校の国立高校の生徒であるといった制約から、本研究で得られた結果を一般化することには慎重にならざるを得ないが、1つの事例として自由英作文評価の実態を示すことができたものと考えられる。本研究が高校生の自由英作文の評価や指導の一助になれば幸いである。

### 注

高等学校の外国語科英語科の必修科目「英語I」では、「聞くこと」「話すこと」「読むこと」「書くこと」という「内容のまとまり」ごとに「関心・意欲・態度」「表現の能力」「理解の能力」「知識・理解」の「観点」を定め、その「評価規準」として表1の【 】で囲んだ下位区分を設けている（松浦、2002b）。表1に示されているのは、目安として定められた評価規準例であり、実際の授業時には目標に応じてより具体的な評価規準を作成し使用する。本研究では表1に示された各観点を評価尺度項目として、その信頼性と妥当性を見ることを目的とした。ただし「書くこと」においては問われない「理解の能力」は省略し、計6項目とした。（追記：高等学校用の評価基準は、2004年6月に国立教育政策研究所のウェブサイトの下記ULTにおいて公開された。  
<http://www.nier.go.jp/kaihatsu/kou-sankousiryoku/html/tobira.htm>）

## 謝辞

本論文の掲載までには多くの方々のご尽力を賜りました。特に調査に評定者として参加して下さいました高等学校の先生方ならびに大学生の皆さまには、この場を借りまして心よりの感謝を申し上げます。

山西博之 (YAMANISHI, Hiroyuki) は現在、広島大学大学院博士課程後期の大学院生である。また、広島大学附属中・高等学校で英語科の非常勤講師を務めている。主な関心領域は、高校生の英語ライティング・プロセス研究である。

## 参考文献

- Carr, N. T. (2000). A comparison of the effects of analytic and holistic rating scale types in the context of composition tests. *Issues in Applied Linguistics*, 11, 207-241.
- Cumming, A. (1989). Writing expertise and second-language proficiency. *Language Learning*, 39, 81-141.
- Cumming, A. (1997). The testing of writing in a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 51-63). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Hamp-Lyons, L. (1995). Rating nonnative writing: The trouble with holistic scoring. *TESOL Quarterly*, 29, 759-765.
- 石田潤・森敏昭 (1985) 「小学生の文章表現の発達的变化」 『広島大学教育学部紀要』 33, 125-131.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- 梶井芳明 (2001) 「児童の作文はどのように評価されるのか? 評価項目の妥当性・信頼性の検討と教員の評価観の解明」 『教育心理学研究』 49, 480-490.
- 国立教育政策研究所 (2002) 「評価規準の作成、評価方法の工夫改善のための参考資料—評価規準、評価方法等の研究開発 (報告)—」 Retrieved October 16, 2003, from <http://www.nier.go.jp/kaihatsu/houkoku/saiyu.htm>
- 松浦伸和 (2002a) 「絶対評価の導入に向けて—第1回その意義と評価規準—」 『STEP英語情報』 1・2月号, 32-37.
- 松浦伸和 (2002b) 「絶対評価の導入に向けて—第2回「内容のまとめ」ごとの評価規準—」 『STEP英語情報』 3・4月号, 34-39.
- 松浦伸和 (2002c) 「絶対評価の導入に向けて—第3回評価計画の立案—」 『STEP英語情報』 5・6月号, 32-37.
- 松浦伸和 (2002d) 「絶対評価の導入に向けて—第4回評価の総括—」 『STEP英語情報』 7・8月号, 24-29.
- 文部省 (1999) 『高等学校学習指導要領解説 外国語編 英語編』 開隆堂出版.
- Sasaki, M., & Hirose, K. (1999). Development of an analytic rating scale for Japanese L1 writing. *Language Testing*, 16, 457-478.
- 渡部洋・曹亦薇 (1992) 「小論文評価における字の美しさの影響について」 『東京大学教育学部紀要』 32, 253-256.

(Received December 15, 2003; revised July 29, 2004)

## Appendices

Appendix A: 評価尺度セットごとの教員の相関行列 ( $N = 20$ )

	教員1	教員2	教員3	教員4	教員5	教員6	教員7	教員8	教員9	教員10
教員1	-	<u>.61</u>	<u>.39</u>	<u>.45</u>	<u>.54</u>	<u>.48</u>	<u>.70</u>	<u>.25</u>	<u>.30</u>	<u>.65</u>
教員2	.67/.67	-	<u>.38</u>	<u>.64</u>	<u>.54</u>	<u>.40</u>	<u>.63</u>	<u>.66</u>	<u>.19</u>	<u>.59</u>
教員3	.36/.24	.67/.58	-	<u>.52</u>	<u>.46</u>	<u>-.02</u>	<u>.51</u>	<u>.46</u>	<u>.19</u>	<u>.41</u>
教員4	.65/.41	.36/.46	.43/.50	-	<u>.43</u>	<u>.18</u>	<u>.44</u>	<u>.56</u>	<u>.06</u>	<u>.33</u>
教員5	.64/.51	.44/.45	.63/.39	.77/.57	-	<u>.66</u>	<u>.66</u>	<u>.52</u>	<u>.45</u>	<u>.53</u>
教員6	.53/.33	.21/.57	.43/.48	.55/.54	.56/.34	-	<u>.49</u>	<u>.13</u>	<u>.45</u>	<u>.53</u>
教員7	.46/.31	.33/.29	.29/.06	.60/.35	.68/.40	.26/-.10	-	<u>.46</u>	<u>.06</u>	<u>.53</u>
教員8	.52/.42	.45/.47	.16/.51	.57/.64	.50/.59	.08/.40	.45/.50	-	<u>.15</u>	<u>.50</u>
教員9	.46/.51	.63/.65	.55/.68	.17/.40	.27/.39	.47/.54	.09/-.01	.20/.47	-	<u>.43</u>
教員10	.58/.43	.35/.61	.40/.56	.53/.47	.60/.70	.55/.64	.42/.24	.50/.56	.33/.47	-

注1) 左下三角はESL Composition Profileセット／観点別評価セット、下線の右上三角は総合的評価セット

注2) 相関係数は  $r > .44$  の場合、5%水準で有意

Appendix B: 評価尺度セットごとの大学生の相関行列 ( $N = 20$ )

	大学生1	大学生2	大学生3	大学生4	大学生5	大学生6
大学生1	-	<u>.37</u>	<u>.60</u>	<u>.46</u>	<u>.07</u>	<u>.21</u>
大学生2	.56/.68	-	<u>.13</u>	<u>.24</u>	<u>.40</u>	<u>.15</u>
大学生3	.66/.61	.72/.60	-	<u>.58</u>	<u>.11</u>	<u>.22</u>
大学生4	.50/.72	.29/.51	.36/.61	-	<u>.10</u>	<u>.23</u>
大学生5	.24/.59	.41/.68	.42/.59	.20/.57	-	<u>-.14</u>
大学生6	.05/.25	.34/.18	.40/.46	.37/.22	.21/.02	-

注1) 左下三角はESL Composition Profileセット／観点別評価セット、下線の右上三角は総合的評価セット

注2) 相関係数は  $r > .44$  の場合、5%水準で有意

