

## ルーブリックの事前提示が ライティングパフォーマンスに与える影響

研究者：東京都／中央大学附属中学校・高等学校 教諭 岩本 祐樹

《研究助言者：小泉 利恵》

### 概要

本研究は、パフォーマンス課題などを評価する際に使用されるルーブリックが、ライティング力の向上に有効かどうかを調査した。ルーブリックは評価基準が記述されているので、それを参照した上でライティングに取りかかれれば、パフォーマンスの質が向上することが予想される。対象とした中学1年生(計60名)のうち1つのグループ(30名)には、ルーブリックを事前に提示した上で、ライティングを行ってもらい、フィードバックもルーブリックとスコアの両方を提示した。もう一方のグループ(30名)は、ルーブリックを提示されずにライティングをおこない、フィードバックとしてルーブリックに基づいて採点されたスコアのみを返却した。以上のライティングを3回実施し、その内容面と言語面のスコアを比較したが、グループ間で有意な差は見られなかった。以上の結果を踏まえると、ルーブリックの効果を高めるためには、ルーブリックの評価基準の明確化や、自己評価などのメタ認知活動や相互評価などを併用することが重要であるのかもしれない。

### 1 はじめに

現代社会における学校教育では、知識を習得するだけでなく、知識を使いこなしたり創造したりする力を育成すべきとの社会的要求が高まっている(石井, 2016)。知識を活用するといった場

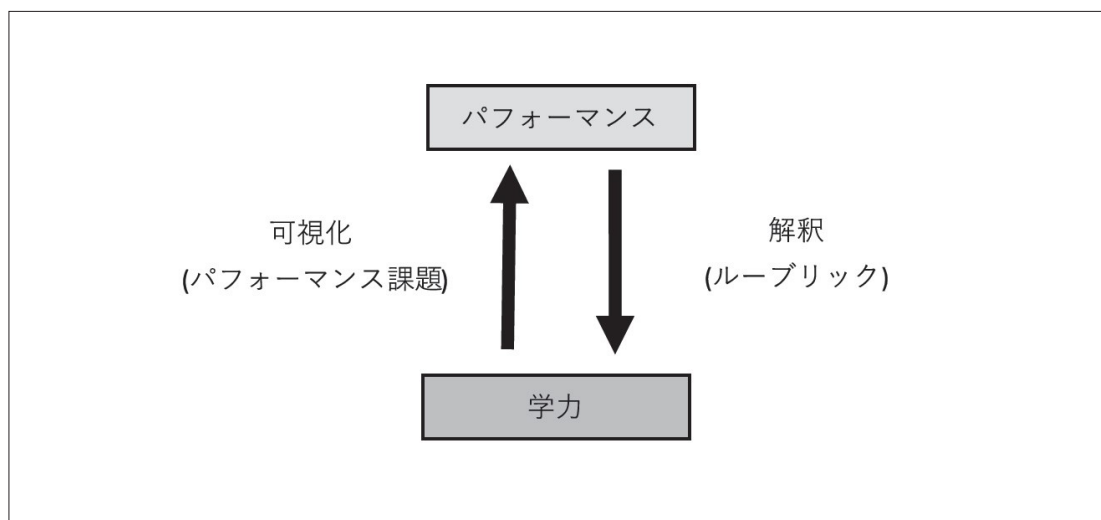
合、能力を階層的に捉えることが多い。例えば、石井(2015)は、教科内容に関する学びの深さは3つのレベルで捉えることができるとしている。1つ目は、知識・技能の習得状況を問う「知っている」レベル、2つ目は、概念の意味理解を問う「わかる」レベル、そして、3つ目は、現実世界の文脈における知識・技能の総合的な活用力を問う「使える」レベルである。

「使える」レベルの能力の重要性は、近年の知識と学びの「量」から「質」への転換を待たずとも、英語教育においては特に求められてきたことである。そして、このレベルの英語能力の育成を考えたときに不可欠であるのが、実際の英語使用場面を想定した指導や評価<sup>1</sup>である。これを可能にするのが、パフォーマンス評価である。Wiggins & McTighe(2005)は、パフォーマンス課題を、「効果的に行動するために知識を活用する課題、あるいは、ある人の知識と熟達化を明らかにするような複雑な感情作品を実現する課題」(Wiggins & McTighe, 2005, 邦訳 p. 405)と定義している。言語教育におけるパフォーマンス課題は、プレゼンテーション、ディベート、スピーチ、演技といった実演を求めるものや、レポート、論文、ポスターなどの作品も含まれるであろう。そして、このパフォーマンス課題を評価する際に用いられるのが、ルーブリックである。松下(2007)は、学力には「見えやすい学力」と「見えにくい学力」があるとしているが、「使えるレベル」の学力は、「見えにくい学力」と言える。松下は、この「見えにくい学力」を可視化するのがパフォーマンス課題であり、ルー

ブリックを使うことでこのパフォーマンス課題から学力を「解釈」することが、パフォーマンス評価だとしている(図1)。

ルーブリックを使用した評価の実践は各教科で行われている。しかしながら、沖(2014)は、初等

中等教育におけるルーブリックの活用状況について、「毎回の授業や授業内の学習活動で効果的に提示あるいはフィードバックされているとはいえない」と指摘しているように、その運用にはまだ問題点や実践的研究が不十分な点がある。



■ 図1: パフォーマンス評価の構図 (松下, 2007, p. 11)

また、言語教育におけるテストの観点からも、信頼性と妥当性のあるパフォーマンス評価の研究は行われている。教育現場では、ライティングやスピーキングのテストを実施する場合、教員がテーマを設定し、独自の評価尺度により判断を下す場合は多いが、藤森・小泉(2011)は、スピーキングテストの教員による全体評価と、生徒の発話を正確さ、流暢さ、複雑さの観点から客観的に分析した評価との関連を検証している。その結果、日本人教師と外国人英語講師の採点結果を総合したスコアは、正確さを厳しく客観的に評価した場合に有意な相関が見られた。

本実践研究では、日本語母語英語学習者のライティング課題に焦点を当て、評価におけるルーブリックの役割を検証する。後述するように、ルーブリックは、ペーパーテストでは測れない質的な評価をする道具であると同時に、教員が学習者に求める学習の到達レベルが記述されているため、形成的評価に使うことで教育効果をもたらす、指導と評価の一体化を具現化するツールとなりえる。ライティング課題は、語彙力、文法力、表現力、論理性、独創性などの総合的な知識が問われるパ

フォーマンス課題と言えるが、学習者がそのライティングをする際に、ルーブリックが提示され、求められていることを認識することができれば、直感的にはパフォーマンスの向上に有益であると推測できる。また、ルーブリックを用いたフィードバックにより、現在の到達レベルや足りない点を把握することができれば、次の課題にとりくむ際に有効に働くと考えられる。本実践では、ライティング課題の実施の事前にルーブリックを提示することが、パフォーマンスの向上に有効かどうかを中学1年生のライティング活動を通して、実践的に検証することを目的としている。

## 2 先行研究

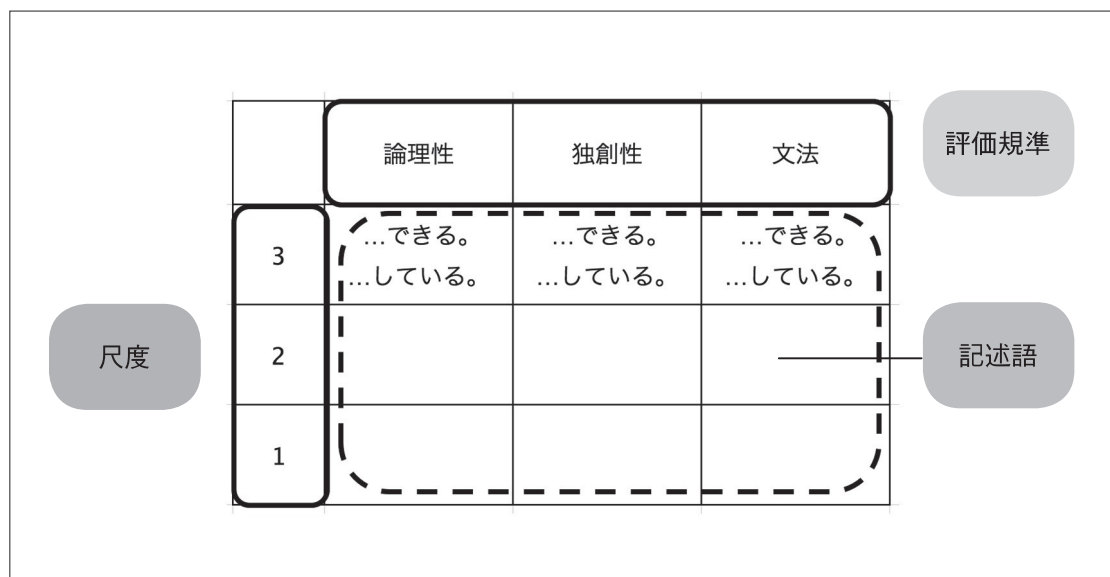
### 2.1 ルーブリックとは

ルーブリックは、主にパフォーマンス課題を評価する際に使われる道具のことであり、「成功の度合いを示す数値的な尺度(scale)と、それぞれ

の尺度に見られるパフォーマンスの特徴を示した記述語(descriptor)からなる評価指標」(田中, 2010, 48頁)と定義される。ルーブリックは、パフォーマンス課題のような質的評価をするための道具である。また、松下(2007)は、パフォーマンスから学力を解釈する際に、恣意的・独断的にならないようにするためにルーブリックを使う必要があるとしている。また、沖(2014)は、ルーブリック評価は、テスト法や質問紙法を除いて、思考・判断や関心・意欲・態度、技能・表現を評価するすべての評価方法に適しているといえることができ、と指摘している。

ルーブリックには、いくつかの種類がある。代表的なものが「分析的ルーブリック」と「全体的

ルーブリック」の2つである。前者は、ある課題を評価する際に、複数の評価規準<sup>2</sup>を設定し、それぞれについてレベル分けをして(尺度)、評価基準(判断基準)を記述しているものである。後者は、全ての規準を1つにして、レベルを設定し、評価基準を記述しているものである。図2は、分析的ルーブリックの概要を表している。また、ルーブリックの各レベルに対応する典型的なパフォーマンスの事例をアンカー作品と呼ぶ。アンカー作品をルーブリックに添付することで、学習者は課題が求めていることをより明確に理解することができる。このように、評価規準と評価基準を明示することにより、学習者は自身の到達度や目指すべきレベルとの差を知ることができる。



■図2: 分析的ルーブリックのイメージ

ルーブリックは質的なパフォーマンス課題を恣意的・独断的にならないように評価するために使用するのだが、当然ながら評価者の主観がまったくなくなるわけではない。同じルーブリックを用いて評価しても、評価する人によって結果が異なることはもちろんあり得る。ルーブリックによる評価が恣意的になるのを防ぐためには、主観のつき合わせや調整によって間主観性を担保することが求められる(松下, 2012)。そして、間主観性を担保するために用いられるのが、モデレーションである。モデレーションは、評価過程を統

一する方法と評価結果を統一する方法の2つがある。前者は、評価基準を作成する際に、複数の評価者で作品例を評価し、その結果を比較・検討し、共同でルーブリックの作成・修正を行い、評価規準と評価基準についての共通理解を図り、評価結果を統一することである。後者は、評価の結果を複数の評価者間で調整することである。

パフォーマンス評価の際に使用されるルーブリックには、様々な利点があることが指摘されている。Stevens & Levi(2005)は、ルーブリックの利点として、①タイミングの良いフィードバック

ク、②学生による詳細なフィードバックの利用、③批判的思考力のトレーニング、④他者とのコミュニケーションの活性化、⑤教員の教育技法の向上、⑥平等な学習環境作り、の6点を挙げている。この中で最も重要であると考えられる要素が、①と②のフィードバックの側面である。パフォーマンス課題を評価する場合、評価が曖昧になったり、コメントなどに時間がかかったりする場合が多くある。ルーブリックには、評価基準、到達の度合い、不足している部分がすでに記述されているため、教員は、多くの時間を割くことなく、タイミングの良いフィードバックをすることができる。必要に応じて、コメントを追加することも可能である。一方、生徒も比較的迅速なフィードバックを期待しており、学習を促進するためには、適切なタイミングでフィードバックをすることが必要となる。ルーブリックを使えば、時間がかかる質的な評価を多くの時間を割くことなくおこなうことができる。フィードバックにおいては、多くの情報が含まれていることと、どのような場合に最高の評価になるかが記載されていることが有効であるとされており、ルーブリックはこの両方を満たしていると言える(Stevens & Levi, 2005)。また、この他にも、後述するように形成的に使うことで多くの利点がある。

## 2.2 形成的評価としてのルーブリック

ここまでで述べてきたように、ルーブリックは、パフォーマンス課題のような質的なものを評価し、適切なフィードバックを可能にする道具であり、質を量に変換する装置という一面を持つ(松下, 2012)。しかしながら、安藤(2014)は、成績評価のための採点ルーブリックだけではなく、形成的評価の観点からルーブリックを位置付けなければ、ルーブリックの教育効果が十分に発揮されない、としている。つまり、必ずしも点数化する必要はなく、学習の最後ではなく、途上でルーブリックを学習者と共有し、期待する事柄を周知徹底することの必要性に言及している。Stiggins(2001)は、形成的に生徒中心の評価としてルーブリックを使うことの効果を指摘している。

Panadro & Jonsson(2013)は、小学校から大学までの21本のルーブリック研究を分析し、

ルーブリックの形成的な評価がパフォーマンスの向上に寄与する要素として、①評価の透明性、②フィードバックの省察、③自己効力感の向上、④不安の低減、の4点を挙げている。また、Panadro & Jonsson は、このパフォーマンスの向上に寄与する要素を効果的にする要因として、自己評価とメタ認知活動を挙げている。特に、後述の実証研究で紹介するように、ルーブリックを用いた自己評価と相互評価の有効性については多くの研究がある。山田・森・毛利・岩崎・田中(2015)は、ルーブリックを用いて、生徒を評価活動に参加させることで、生徒は求められる結果がどの程度達成される途上にあるか、またどの程度達成されたのかについて、より深い認識を持つことができるとしている。このように、評価を学びのプロセスの中に取り組み、評価活動に参加させることも、ルーブリックを用いると可能になる。

## 2.3 実証研究

前節で示したルーブリックの理論的な枠組みに対し、その効果を実証的に調査した研究がある。Arter, Spandel, Culham & Pollard(1994)は、「書き方の特性(writing traits)」<sup>3</sup>と呼ばれるルーブリックを書き方の指導に導入する効果を検証した。書き方の特性は、アイデア、構成、ボイス、言葉の選択、文章の流暢さ、句読点等の取り決めの6つに、オプションとして、プレゼンテーションを含んだルーブリックである。Arterらは、小学5年生を対象に、このルーブリックを用いて、自身の作品の自己評価に加え、サンプルや他の生徒の作品を評価したり、改善策を考えたり、また教員によるそれぞれの尺度についての指導などの授業を行ったりした。実験群(67名)と、ルーブリックを用いないで通常書き方の指導をした統制群(65名)の事前テストと事後テストの作文を比較した。その結果、書き方の特性のうち、実験群の方がアイデアの観点において有意に統制群より優れていた。以上の結果を踏まえて、Alterらは、自己評価と評価基準を指導することの重要性を指摘している。

ライティングの評価からは離れるが、鈴木(2011)は、数学の授業とテストの評価にルーブ

リック評価を用いることで、ループリックがどのように学習者のテストに対する認識に影響を与えるかを検証した。鈴木は、評価の目的や基準が実施者と受け手との間で合意がなされている場合、その評価の仕方(インフォームドアセスメント)は、学習者のテストに対する認識(テスト観)をポジティブなものにするとしている。鈴木が挙げているテスト観とは、テストは自分の学習改善に活用するものであるという「改善」、自発的に学習を行うためのものであるという「誘導」、他者と比較するためのものだという「比較」、勉強を強制させるためのものだという「強制」の4つの因子で構成されるとしている。そして、ループリックは高い水準のインフォームドアセスメントとなりうると指摘し、ループリックの提示が学習者のテスト観に与える影響について、中学2年生の5日間の数学の授業において検証した。その結果、「改善」のテスト観において、ループリックを提示されたグループは学習改善に活用するための目的や役割を有意に強く認識していた。また、5日間のプログラムの最後に実施された総合テストでは、5日間の授業で扱った問題では、ループリックを提示された群は、有意に高い結果を示したが、授業で扱ったテーマとは無関係な発展問題においては、有意な差は見られなかった。

寺嶋、林(2006)は、大学の演習型の授業において、ループリックを事前提示し、学習の途中や終了時に学習者自身がそれをもとに自己評価する授業を実施した。受講者へ行った質問紙の結果から、ループリックをもとに自己評価をする事で、学習目標を意識化し、それに従って学習を進め、その学習における成果や課題を把握できるといった、学習一般への効果もあるということが示唆されている。

一方で、ループリックの提示がライティングに有意な効果を示さなかった研究もある。例えば、Goodrich Andrade(2001)は、形成的評価としてループリックを提示する事が、学習者の良いライティングに関する知識に影響を与えるかどうか、また実際にライティングのスコアに影響を与えるかどうかを検証した。この実験において、中学2年生を対象にループリックを提示するグループと提示しないグループに3種類のライティング(説得型エッセイと旅行記と自分の挑戦、または歴史的物語)を課し、そのスコアを比較した。また

最後に、「優れているライティングと良いライティングの違いを教員がどのように判断していると思うか」という記述式の質問をして、良いライティングに対する知識の度合いを分析した。その結果、ループリックを提示されたグループの方が、ライティングの評価基準に関する深い知識・理解は見られたものの、実際のライティングのスコアの比較では、3つのうち1つのライティング(旅行記)でしか有意な差は見られず、良いライティングに関する理解が実際のライティングのスコアに反映されることは、より難しいことであるとしている。

また、Goodrich Andrade & Boulay(2003)は、中学1年生と2年生に対し、ライティング(歴史的物語と物語を読んだ感想)における自己評価がパフォーマンスに及ぼす影響を調査した。実験群と統制群の両方にループリックを提示し、ライティングを1ヶ月の間をあけて、2回実施させた。その中で、実験群は、ループリックに沿って自己評価をする授業を実施した上で推敲する時間を設けたが、統制群に対しては単に推敲の時間を設けるのみにした。しかしながら、最終的な両者のライティングのスコアに有意な差は見られなかった。Goodrich Andrade & Boulayは、自己評価の効果はあったかもしれないが、ループリックの提示の効果が大きく、自己評価の効果が顕著にならなかったのかもしれない、としている。

英語教育においてもループリックを活用する実践は多い。例えば、鶴田(2018)は、高校生のスピーキングの帯活動にループリックを用いて、自己評価させる活動を行なった。スピーキングを「発表」と「やりとり」の2つにわけ、各6回ずつ活動をし、回ごとにループリックから焦点を当てる項目を1つ選び、その項目に意識して活動を行うように指示し、計12回の活動の中で段階的な指導を実践した。その結果、回を追うごとに自己評価は高くなり、ループリックを自己評価と共に継続的かつ段階的に使用することで、自己肯定感を高めるスピーキング活動が可能になるとしている。

また、松井(2015)は、高校生のライティング活動において、ループリックの作成を生徒と共同で行い、また自己評価をさせる実践を行った。その実践から、教員が作成したループリックを使うよりも、生徒が能動的に作成に関わったループリックを用いた方が、ライティング能力の育成に有効



であると指摘している。また、生徒と作成したルーブリックで自己評価を行うことは、教員の評価とのずれが少なくなり、自己評価力が高まるとしている。

以上の先行研究を踏まえると以下のことが言える。

1. ルーブリックを事前提示したり、ルーブリックを用いて書き方の指導をすることで、学習者は目指すべきものを理解したり、良い作品についての理解を深めることができるので、パフォーマンスが向上するかもしれない。
2. 自分のパフォーマンス課題を自己評価したり、他者のパフォーマンス課題を評価したりすることで、モチベーションやテストに対する意識、また自己調整力、自己評価力などのメタ認知能力の向上に影響を与えるかもしれない。
3. 2の結果として、パフォーマンスが向上するかもしれない。

このように、ルーブリックの提示やルーブリックを使っのフィードバック、また、ルーブリックを用いての自己評価が、学習者のメタ認知能力の向上に寄与しそうであることは、学習者に対する質問紙などの結果から示唆されている。しかし、実際のパフォーマンスの向上に影響があるのか、またあるとしたらどのような面に効果があるのかは明確ではなく、ルーブリックの有無で、パフォーマンスのスコアを比較した研究は多くない。また、中学生のライティング活動における検証も少ない。そこで、本実践研究は、上記1の点に焦点を当て、日本語を母語とする英語学習者である中学1年生を対象にしたライティング課題を取り上げ、ルーブリックの事前提示やフィードバックがパフォーマンスの質の向上に有効であるかどうかを検証する。

## 3 実践方法

### 3.1 目的

本実践の目的は、ライティング活動をルーブ

リック評価する際に、ルーブリックを学習者に事前提示することが、ライティングのパフォーマンスにどのような影響を与えるかを調査することである。英語のライティングは、文法知識、語彙力、論理性や独創性などの総合的な力が問われるパフォーマンス課題である。先行研究によれば、ルーブリックを事前に提示することは評価の透明性をもたらし、その課題が求めていることが学習者に明確になることで、パフォーマンスの質も向上することが期待できる。また、ライティングを複数回実施し、ルーブリックを伴うフィードバックをした上で、新たに同じ種類のライティングを行えば、その効果は事前提示の効果に加えて大きくなると考えられる。しかし、事前提示に対しては、事後の質問紙などにより学習者から肯定的な認識はあるものの、その効果が実践において調査された研究は多くはない。そこで、本実践は日本語母語英語学習者である中学1年生を対象にライティング課題を実施し、その論理性や一貫性などの内容面と文法や語彙などの言語面の2つの側面<sup>4</sup>におけるルーブリックの事前提示とフィードバックの効果を検証する。以下の研究課題を設置し、3回のライティングの授業実践と事後のアンケートを実施する。

#### RQ1

ライティング課題において、ルーブリックを事前に提示した場合、そうでない場合と比較し内容面においてパフォーマンスが向上するか

#### RQ2

ライティング課題において、ルーブリックを事前に提示した場合、そうでない場合と比較し言語面においてパフォーマンスが向上するか

#### RQ3

ライティング課題において、ルーブリックを使用した場合、そうでない場合と比較し、課題に対する目標意識や、その後の学習への意欲が高くなる傾向が見られるか

### 3.2 参加者

本実践には、中央大学附属中学校の1年生の2クラスの69名が参加した。本実践は、ループリックの事前提示の効果を調査することが目的であるため、34名からなる1つのクラスにはループリックを提示し、フィードバックもループリックによって実施した(実験群)。そして、35名からなるもう一方のクラスは、ループリック自体を使用しない条件(統制群)のもとライティング課題を実施した。

クラス間の英語運用能力の均等性を可能な限り確保するために、英検の取得級を元に事前に分析の対象となる参加者を決定した。その結果、どちらのクラス(グループ)も、英検3級取得者が12名、4級取得者が18名となり、それぞれ計30名を分析の対象とした。

### 3.3 実践と調査の手順

#### 3.3.1 ライティング課題のテーマ

本実践は、上記のように、ループリックを提示する実験群と、ループリックを全く提示しない統制群に分けて実施した。どちらのグループにも、3回のライティングをおよそ3週間の間隔で課した。ライティングの内容は、①1年間でどの月が好きですか、②日本のどの都市が好きですか、③どのテレビ番組が好きですか、の3種類を用意した。それぞれのライティング課題は難易度も異なり、また順序による効果があることも予想されることから、表1に示すようにそれぞれのグループを10名(英検3級取得者4名、英検4級取得者6名)ずつの3つのグループに分け、カウンターバランスをとった。

#### 3.3.2 ループリック

本実践では、ライティング課題を、内容面と言語面の2つの側面で評価するために、主に論理的な主張ができていないかを評価する尺度(内容面)と文法や語彙の使用について評価する尺度(言語面)の2つの尺度が含まれるループリックを使用した(表2)。このループリックは、国際バカロレアの1つの科目であるEnglish Bのライティング評価で使われているものを中学1年生にも理解できる表現に訳して、作成したものである。

#### 3.3.3 実施手順

表3は、本実践の概要を示している。本実践は、実践実施者が勤務する中学校での週5時間ある英語の授業内に実施した。実験群には、1回目のライティングをする前に、ループリックを提示し、ループリックの概要説明と、それに従って採点がなされることを説明した。また、内容面のループリックにのみ、評価5、評価4、評価3、評価2、評価1に対応するライティングの例(アンカー)を添付した(資料1)。アンカーは、「どの教科が好きですか」というテーマのもとにそれぞれの尺度に当たる答案を作成し、生徒にその理由とともに提示した。特に、評価5に相当する答案の例に関しては、論理性と一貫性があることをループリックの記述とアンカーを参照しながら説明した。さらに、アンカーとは別に、評価2に相当する答案をもう1つ用意し、評価2であることを言わずに、それがどの評価に該当するかを生徒自身に考えてもらい、その後評価2であることとその理由を提示した。そして、それらの全ての説明が終了した後に、1回目のライティングに移行した。ライティングの際は、教材として生徒が共通に使用している紙の和英・英和辞書の使用を許可した。

■表1: カウンターバランスによるライティング課題の振り分け

	実験群 (30名)			統制群 (30名)		
	グループ1	グループ2	グループ3	グループ1	グループ2	グループ3
第1回	月	都市	TV番組	月	都市	TV番組
第2回	TV番組	月	都市	TV番組	月	都市
第3回	都市	TV番組	月	都市	TV番組	月

一方で、統制群に対しては、何も提示せずにライティング課題を実施した。統制群においても紙の辞書の使用を認めた。ライティングはどちらのグループも20分間で終わった。

2回目のライティングは、1回目からおよそ3週間後に実施された。実験群においては、ルーブリックに採点されたスコアを表記し、答案と共に返却した(採点については後述)。それ以外は、評価やルーブリックの説明はせず、2回目のライティン

グを行なった。統制群については、採点者が実験群で使用しているルーブリックを使い評価を行ない、そのスコアのみを生徒に返却した。よって、生徒は、内容面と言語面において、それぞれ0点から5点の点数がフィードバックされるだけであった。そして、その後、2回目のライティングに入った。

3回目のライティングも2回目からおよそ3週間後に実施した。3回目はどちらの群も2回目と同様の手続きで実施したが、ライティング後、それ

表2: 本実践で使用したルーブリック(International Baccalaureate, 2011, Language B guideより抜粋し筆者が修正。)

	内容面	言語面
5	理由は説得力がある。説明の流れも非常に良く、わかりやすい。 ・理由は、具体的に非常に良く表現されており、妥当である。 また、興味深いものである。 ・ライティングの構成は、首尾一貫してわかりやすく、整理されている。	文法・語彙、文の作り方が非常に良い。 ・語彙が豊富で、正確かつ効果的に使われている。 ・主語・動詞のある文をしっかり作ることができている。 ・間違いがほとんどない。
4	理由は明確で十分な説得力がある。説明の流れも良く、わかりやすい。 ・理由は、十分に表現されており、妥当である。 ・ライティングの構成は、わかりやすく、整理されている。	文法・語彙、文の作り方が良い。 ・語彙が豊富で、正確に使われている。 ・主語・動詞のある文をしっかり作ることができている。 ・細かい間違いは多少あるが、意味に関わる間違いはほとんどない。
3	理由はある程度明確で、部分的に説得力がある。 説明の流れは、良い箇所がある。 ・理由は、おおむね妥当である。 ・ライティングの構成は、明確である。	文法・語彙、文の作り方は適切であるが、誤りがいくつかある。 ・最低限の語彙を用い表現できている。 ・主語・動詞のある文をおおむね作ることができている。 ・意味に関わるものも含め、いくつか間違いがある。
2	理由はほとんど明確ではなく、説得力もほとんどない。 説明の流れもあまり良くない。 ・理由は、妥当である箇所がある。 ・ライティングの構成は、明確である部分もある。	文法・語彙、文の作り方はおおむね正しいが、誤りが多くある。 ・語彙の種類が限られている。 ・主語や動詞のない文が多く見られる。 ・意味に関わるものも含め、間違いが多くある。
1	理由は明確ではなく、説得力がない。説明の流れも悪い。 ・理由は、見当外れである。 ・ライティングの構成が曖昧で、わかりにくい。	文法・語彙、文の作り方が限られており、誤りが多くある。 ・語彙の種類がかなり限られている。 ・主語や動詞がある文がほとんどない。 ・意味に関わるものも含め、間違いがかなり多くある。
0	上記に記されている基準に達していない。	上記に記されている基準に達していない。



それぞれのグループに今回のライティング課題に関するアンケートを行った(後述)。

このような実験計画であるので、ループリックの事前提示による純粋な効果は1回目のスコアを比較することで行う。2回目以降は、その前のライティングのフィードバックがなされるため、ループリックの事前提示による効果とフィードバックの効果が混合したものになる。したがって、2回目と3回目のライティングにおいては、その前の自分自身のライティングのフィードバックも含めた上でのループリックの事前提示効果と解釈し、その効果を検証した。なお、ループリックによる評価を提示せずに、3回書いてもらうことも考えたが、採点后に点数の返却は通常行なうため、実際の指導に近い形の実験計画にした。また、この計画では、1回目では、実験群におけるループリックの提示や例示、解説の有無が統制群と異なる点であり、2回目、3回目では、ループリックの提示の有無がグループでの相違点となる。しかし、2回目、3回目での実験群へのループリック提示は、1回目での例示・解説を含むループリックの事前提示を思い出してもらう程度とも考えられる。

### 3.3.4 採点とフィードバック

回収された答案は、中央大学附属中学校・高等

学校に所属する英語科教員2名に採点を依頼した。採点者は、採点のブレが大きくなるように、筆者と共に、ループリックと生徒にも提示したアンカーをもとに、採点の基準について話し合った。採点する際には、採点者同士のすり合わせは行わず、それぞれがループリックをもとに評価を行なった。フィードバックには、2名の採点者が評価したスコアの平均をとり、小数点以下は切り捨てた数値を採用した。これは、例えば評価が3.5になった場合、その内容は、ループリックの評価4には達していない、3の範疇の中にあると判断したためである。統制群においても、同じループリックをもとに採点を行った。なお、採点者には、実践の内容は明らかにせず、実験群と統制群が存在することは知らされず、60名分の答案をランダムにしたのちに、採点をしてもらった。

### 3.3.5 事後アンケート

表4に示すのが、3回目のライティング課題が終わった後に、実験群と統制群に行なったアンケートである。質問の(6)～(16)は、両グループに共通した質問項目であり、ライティング中に各要素(一貫性、構成、文法、語彙、評価へのモチベーション)を意識したかどうか、フィードバックに対する感想、今回のライティングは英語学習に有益であったか、また今後のモチベーショ

■表3: 本実践の概要

	グループ		
	内容		
第1回	ライティング①の実施	<ul style="list-style-type: none"> <li>ループリックの提示</li> <li>ループリックの説明</li> <li>各評価尺度の例を提示、その理由を説明</li> <li>ライティング①を実施</li> </ul>	<ul style="list-style-type: none"> <li>ライティング①を実施</li> <li>※特に指示は無し</li> </ul>
第2回	ライティング①のフィードバック	ループリック評価を返却	ループリックによる評価の点数のみを返却
	ライティング②の実施	ループリックを提示し、ライティング②を実施	ライティング②を実施
第3回	ライティング②のフィードバック	ループリック評価を返却	ループリックによる評価の点数のみを返却
	ライティング③の実施	ループリックを提示し、ライティング③を実施 ・アンケートを実施	<ul style="list-style-type: none"> <li>ライティング③を実施</li> <li>アンケートを実施</li> </ul>
第4回	ライティング③のフィードバック	ループリック評価を返却	ループリックによる評価の点数のみを返却

ンとなったかの主に3要素から成る。また、実験群のみに対して、ルーブリックの理解度と使用を確認するために、(1)～(5)の質問を聞いた。また、どちらのグループに対しても、最後に自由記述欄を設けた。

このようなアンケートを実施することで、両グループの生徒が今回のライティング学習に対して持った認識において、グループ間にどのような傾向があるのかを調査した。

■表4: 事後アンケートの項目

(1) 授業中に行ったルーブリックの説明は理解できましたか。
(2) その際に、例として提示されたWritingは、なぜその評価になるのか理解できましたか。
(3) ルーブリックの評価の記述はわかりやすかったですか。
(4) ルーブリックはWritingを書く上で、役に立ちましたか。
(5) Writing中に、ルーブリックの評価基準を確認したり、意識したりしましたか。
(6) Writing中に、内容が筋の通ったWritingにしようと思われましたか。
(7) Writing中に、全体の構成を良くしようと努力しましたか。
(8) Writing中に、文法のミスを少なくしようと努力しましたか。
(9) Writingの中で、いろいろな語彙を使おうと努力しましたか。
(10) Writingをしている際に、良い評価を取りたいと思われましたか。
(11) 返却された答案が、その評価になる理由が理解できましたか。
(12) 返却された1回目の評価は、2回目のWritingに、役に立ちましたか。
(13) 返却された2回目の評価は、3回目のWritingに、役に立ちましたか。
(14) 計3回のWritingは英語の勉強になったと思えますか。
(15) 3回分のWritingを書いたことで、今後もっと英語の勉強をしようと思われましたか。
(16) 今回のようなWritingを繰り返し行ったら、Writingが向上すると思えますか。
※生徒は、1. 全然そうではない 2. ほとんどそうではない 3. どちらとも言えない 4. 少しそうだ 5. とてもそうだ の5段階で回答。 ※(1)～(5)の質問は実験群のみ。(6)～(16)の質問は2群共通。

## 4 検証方法と結果

ライティングのスコアは、分析においては2人の評価者が採点したものの平均をとり、その数値をスコアとして採用した。したがって、生徒にフィードバックした評価とは異なり、小数点以下も含む数値で分析にかけた。また、本実践で行う統計処理は、全て有意水準を5%に設定し、分析を行った。

2人の採点者の評価に一貫性があるかどうかを相関関係の度合いで調査した。その結果、1回目、2回目、3回目の内容面のスコアはそれぞれ、 $r = .761$  ( $p < .001$ )、 $r = .671$  ( $p < .001$ )、 $r = .536$  ( $p < .01$ ) となり、1回目は強い相関、2回目、3回目は、中程度の相関が確認された。また、一致度を検定するカッパ係数は、 $k = .682$ 、 $k = .460$ 、 $k = .467$  となり、1回目がかなりの一致、2回目、3回目で中程度の一致となった。

言語面のスコアは、1回目、2回目、3回目でそれぞれ、 $r = .499$  ( $p < .001$ )、 $r = .693$  ( $p < .001$ )、 $r = .236$  ( $p = .208$ ) で、1回目、2回目で中程度の相関が見られたが、3回目は弱い相関はあるものの、有意な相関がなかった。カッパ係数は、 $k = .300$ 、 $k = .557$ 、 $k = .194$  であり、1回目がまずまずの一致、2回目が中程度の一致が確認されたが、3回目は、わずかに一致するにとどまった。

言語面の3回目の相関には課題があるものの、それ以外は、中程度の相関以上が確認できたので、これらのスコアを用いて、それぞれのライティングのスコアの平均を比較した。

なお、今後の実践や研究への参考のために、内容面と言語面が共に評価5の生徒のライティングと、共に評価3の生徒のライティングを資料として添付する(資料2)。

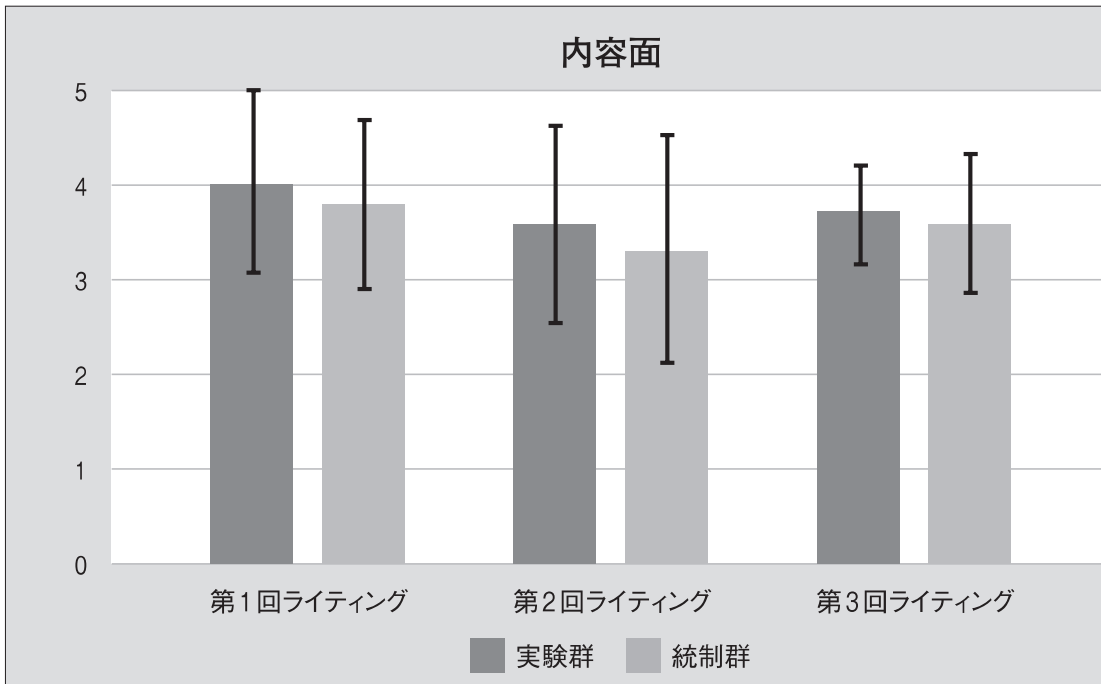
### 4.1 内容面のライティングスコアの結果

表5に内容面のライティングスコアを示した。図3と図4は、スコアのグラフとスコアの人数分布をそれぞれ表している。スコアを比較するために、「事前提示の有無」とライティングの「回数」の2つを要因とし、事前提示の有無は実験群と統制群の2水準(対応なし)、回数は1回から3回の3水準(対応あり)に設定して、混合計画による分散分析を行った。分散分析を行うための前提である等分散性を確認するために、各ライティングのグループ間において、Leveneの検定を行った。その結果、 $F(1, 58) = 0.041$  ( $p = .841$ )、 $F(1, 58) = 1.229$  ( $p = .272$ )、 $F(1, 58) = 3.115$  ( $p = .083$ ) となり、それぞれのグループ間で、分散が異なるとは言えないことが確認されたので、分散分析を実施した。

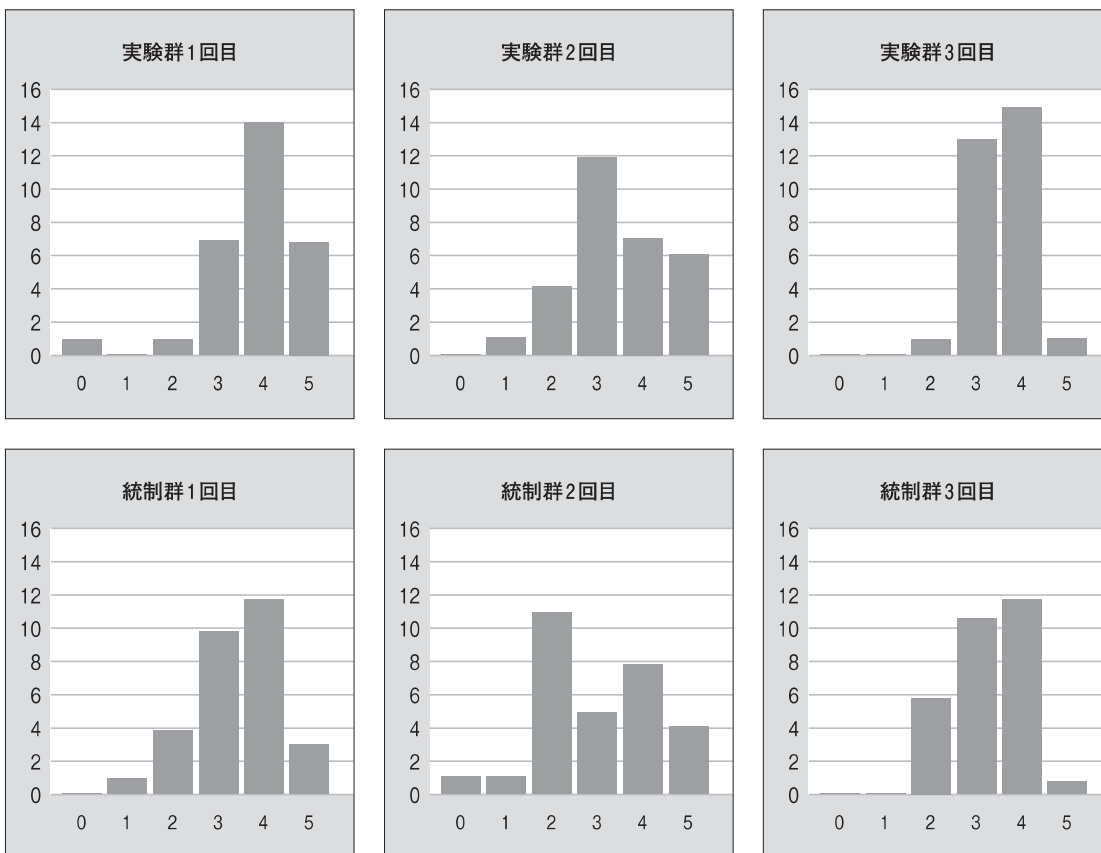
表6は、内容面のスコアの分散分析の結果である。分散分析の結果、事前提示の有無と回数の交互作用は、 $F(2, 116) = 0.159$  ( $p = .854$ ) となり有意でなかった。事前提示の有無についても、 $F(1, 58) = 1.219$  ( $p = .274$ ) で有意な差は見られなかった。また、回数に関しては、 $F(2, 116) = 5.715$  ( $p = .004$ ) で有意な差が見られた。そこで、ホルムの方法による多重比較を行なった結果、1回目と2回目に有意な差があり、両群とも1回目から2回目にかけてスコアが有意に下がっていた( $t(58) = 3.110$ 、 $p = .009$ 、 $d$ (効果量) = .430、 $1-\beta$ (検定力) = .374)。

■表5: 内容面のライティングスコア

	第1回ライティング		第2回ライティング		第3回ライティング	
	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差
実験群	4.050	0.958	3.600	1.030	3.717	0.510
統制群	3.800	0.900	3.358	1.190	3.600	0.723



■図3: 各回の内容面のライティングスコアのグラフ(Y軸はスコア。エラーバーは標準偏差。)



■図4: 内容面のスコアの人数分布 (上部が実験群。下部が統制群。X軸はスコア。Y軸は頻度。)

■表6: 内容面のスコアの分散分析表

	SS	df	MS	F	p	$\eta^2$	1- $\beta$ (検定力)
事前提示	1.850	1	1.850	1.219	.274	.020	.197
事前提示の誤差	88.009	58	1.517				
回数	6.039	2	3.019	5.715	.004	.090	.552
交互作用	0.167	2	0.084	0.159	.854	.003	.061
交互作用の誤差	0.1674	116	0.528				
全体	61.2931	179					

#### 4.2 言語面のライティングスコアの比較

表7に、言語面のライティングスコアの結果を示す。図5と図6は、スコアのグラフとスコアの人数分布をそれぞれ表している。また、内容面と同様に、言語面のスコアも分散分析を実施するた

めに、Leveneの検定を行った。その結果、各ライティングにおけるグループ間の比較において、 $F(1, 58) = 0.116$  ( $p = .734$ ),  $F(1, 58) = 2.563$  ( $p = .115$ ),  $F(1, 58) = 1.087$  ( $p = .302$ ) となり、分散が異なるとは言えないことを確認した。

■表7: 言語面のライティングのスコア

	第1回ライティング		第2回ライティング		第3回ライティング	
	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差
実験群	3.575	0.869	3.417	0.797	3.667	0.467
統制群	3.675	0.728	2.983	0.962	3.692	0.531

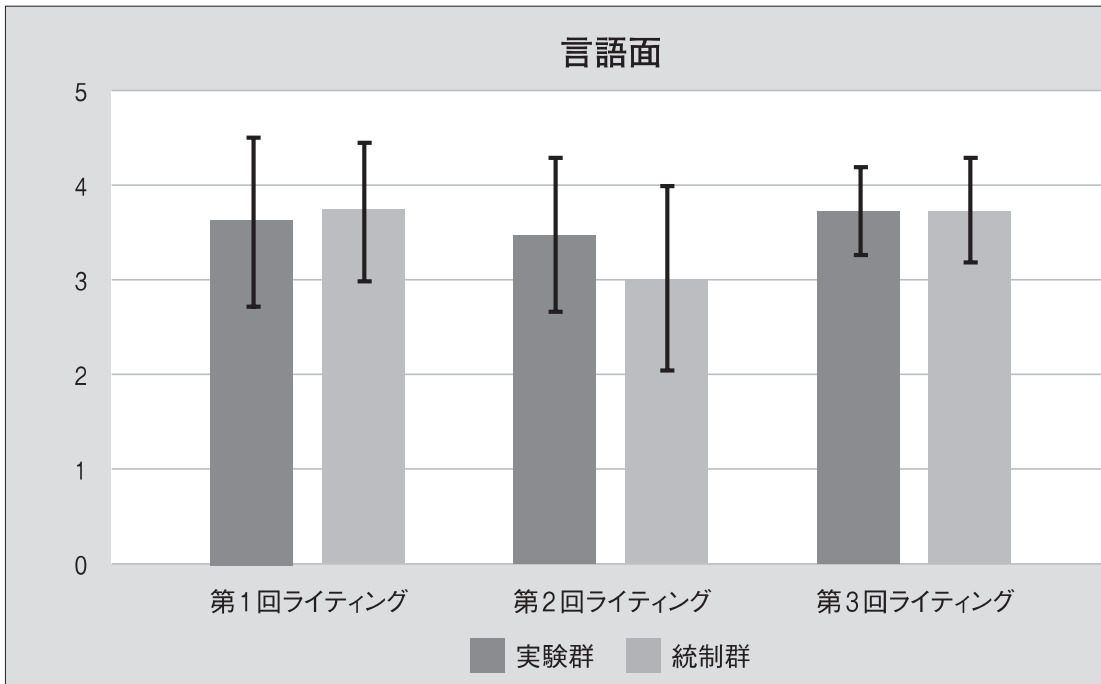
表8は、言語面のスコアの分散分析表である。分散分析の結果、事前提示の有無と回数の交互作用については、 $F(2, 116) = 3.054$  ( $p = .051$ ) で、有意な差は見られなかった。事前提示の有無の要因も、 $F(1, 58) = 0.519$  ( $p = .474$ ) で有意な差は見られなかった。また、回数の要因は、 $F(2, 116) = 10.093$  ( $p < .001$ ) で有意差が見られた。ホルム

の方法で多重比較を行った結果、1回目と2回目 ( $t(58) = 3.171$ ,  $p = .005$ ,  $d$ (効果量) = .500, 1- $\beta$ (検定力) = .479), 2回目と3回目 ( $t(58) = 4.428$ ,  $p < .001$ ,  $d$ (効果量) = .650, 1- $\beta$ (検定力) = .697), で、有意な差が見られた。

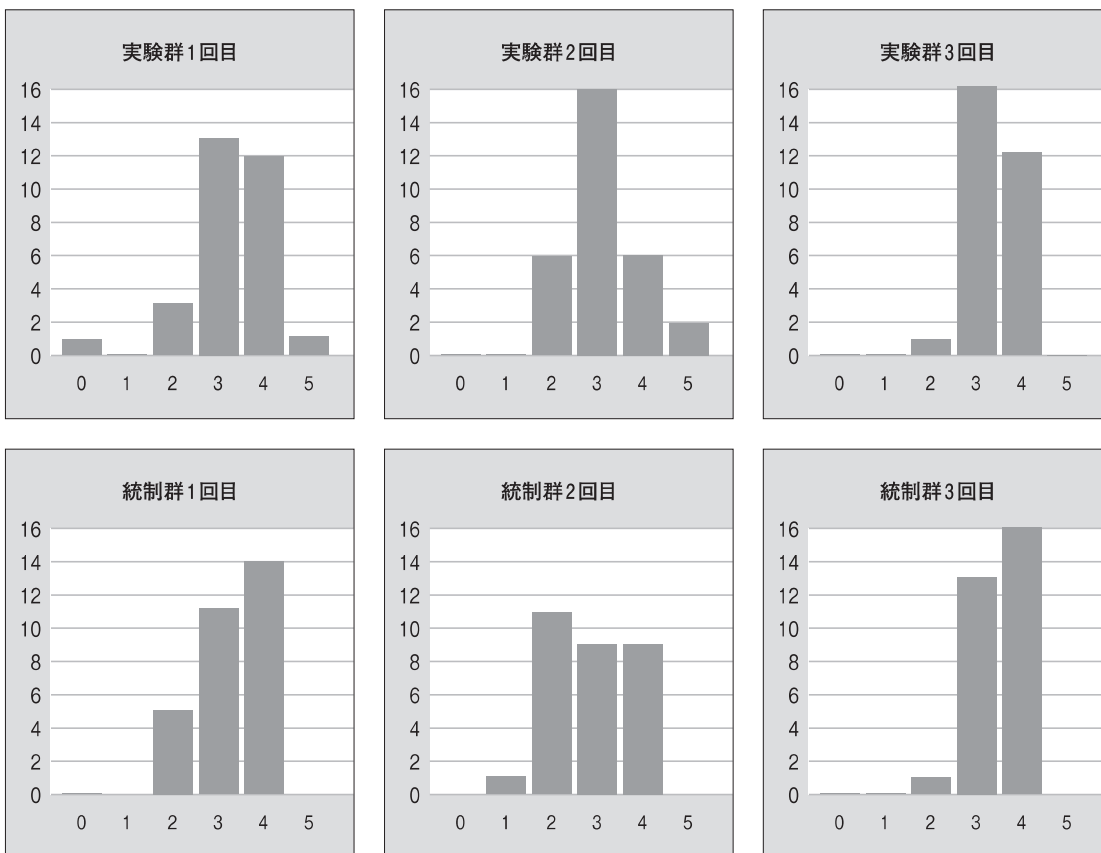
■表8: 言語面のスコアの分散分析表

	SS	df	MS	F	p	$\eta^2$	1- $\beta$ (検定力)
事前提示	0.475	1	0.475	0.519	.474	.009	.113
事前提示の誤差	53.097	58	0.915				
回数	8.263	2	4.131	10.093	<.001	.148	.810
交互作用	2.501	2	1.250	3.054	.051	.050	.320
交互作用の誤差	47.486	116	0.409				
全体	111.812	179					





■図5: 各回の言語面のライティングスコアの比較 (Y軸はスコア。エラーバーは標準偏差。)



■図6: 言語面のスコアの人数分布 (上部が実験群。下部が統制群。X軸はスコア。Y軸は頻度。)

### 4.3 アンケート結果

表9にアンケートの結果を示す。前節でも述べたようにループリックに関する質問を含んだ項目の(1)～(16)を実験群に、また(6)～(16)を統制群

に回答してもらった。すべての質問に生徒は5段階で回答した。その結果、フィードバックについて聞いた項目(12)と(13)を統合して考えると実験群の方が高い数値を示していると言える。しかし、その他の項目で顕著な違いは見られなかった。

■表9: 事後アンケートの項目と記述統計

	実験群		統制群	
	平均	標準偏差	平均	標準偏差
(1) 授業中に行ったループリックの説明は理解できましたか。	4.53	0.62		
(2) その際に、例として提示されたWritingは、なぜその評価になるのか理解できましたか。	4.30	0.59		
(3) ループリックの評価の記述はわかりやすかったですか。	4.30	0.78		
(4) ループリックはWritingを書く上で、役に立ちましたか。	4.27	0.85		
(5) Writing中に、ループリックの評価基準を確認したり、意識したりしましたか。	4.33	0.65		
(6) Writing中に、内容が筋の通ったWritingにしようと思意識しましたか。	4.67	0.47	4.47	0.72
(7) Writing中に、全体の構成を良くしようと努力しましたか。	4.47	0.56	4.50	0.56
(8) Writing中に、文法のミスを少なくしようと努力しましたか。	4.48	0.56	4.47	0.88
(9) Writingの中で、いろいろな語彙を使おうと努力しましたか。	4.31	0.79	4.13	0.92
(10) Writingをしている際に、良い評価を取りたいと思いましたか。	4.41	0.72	4.50	0.67
(11) 返却された答案が、その評価になる理由が理解できましたか。	4.14	0.78	4.30	0.90
(12) 返却された1回目の評価は、2回目のWritingに、役に立ちましたか。	3.97	0.76	3.77	0.84
(13) 返却された2回目の評価は、3回目のWritingに、役に立ちましたか。	4.17	0.70	3.79	0.76
(14) 計3回のWritingは英語の勉強になったと思いませんか。	4.55	0.56	4.67	0.54
(15) 3回分のWritingを書いたことで、今後もっと英語の勉強をしようと思いませんか。	4.17	0.79	4.17	0.86
(16) 今回のようなWritingを繰り返し行ったら、Writingが向上すると思いませんか。	4.41	0.62	4.50	0.62

## 5 考察

本実践研究は、ループリックの事前提示により、課題が求めている評価規準および評価基準が生徒に透明になることで、パフォーマンスも向上するという予測のもと、英語ライティング課題を実施した。しかしながら、内容面、言語面ともにループリックの事前提示の有意な効果は示されなかった。以下に、ループリックをどのように使えば良いかという観点から、本実践で掲げた研究課題について、結果を踏まえて考察する。

### 5.1 RQ1: ライティング課題において、ループリックを事前に提示した場合、そうでない場合と比較し内容面においてパフォーマンスが向上するか

3回のライティングの内容面において、1回目のスコアは2回目より有意に高かったが、これは2つの解釈ができる。第1に、採点に関して、特に回数間で評価にぶれが生じた可能性は排除できない。今回の実践では、ライティングの間の期間が3週間ほどあるため、ループリックを用いても回数により評価の判断が異なった可能性がある。実際に、2人の採点者の一致度は、1回目は、 $k = .682$ であったが、2回目、3回目は $k = .460$ ,  $k = .467$ となり、かなりの一致から、中程度の一致となっている。

それが1回目のライティングが2回目よりも有意に高い結果となった1つの原因となったことはありえる。第2に、生徒の観点から見ると、1回目のフィードバックを踏まえて、さらに上の評価を目指そうと書いたゆえに、向上の過程として2回目のスコアが下がったのかもしれない。表10は、3回のライティングの内容面のスコアの推移を表している。この表を見ると、2回目でスコアが下

がったが、3回目でその2回目より上がった生徒は両群あわせて20名おり、他のケースよりも比較的人数が多い。このことは、実験群、統制群ともに良いものを書こうとして、2回目でスコアを下げた可能性を間接的にはあるが支持している。今後、生徒のライティングの詳細な分析により、この過程を分析したい。

■表10: 3回のライティングの内容面のスコアの推移

	統制群	実験群
3回を通して下がらず上がった	4人	4人
2回目で下がり、3回目で2回目より上がった	11人	9人
3回を通して上がらず下がった	6人	10人
2回目で上がって、3回目で2回目より下がった	8人	5人
3回全てで変わらない	1人	2人

本実践の研究課題の1つ目である内容面のルーブリックの事前提示の効果については、実験群と統制群の比較においては、有意な差が見られず、その効果は確認できなかった。この点に関しては、以下の3点の可能性を検討する。

第1に、今回のルーブリックの記述は、本実践に参加した生徒にとって、具体性が弱く、必ずしも明確な基準となり得なかったことが考えられる。Alter and Chappuis (2006) は、ルーブリックが満たすべき要件として、内容や組織と、明瞭性と

いう2つを挙げており、明瞭性については、「特定化しにくい言葉を使わないようにしているか」としている。ルーブリックの基準の理解について質問をしているアンケート項目の(1)～(5)への実験群の生徒の回答は、4.27～4.53の範囲にあり、全体としては基準を理解できたものと解釈できる。また、自由解答でも「ルーブリックを意識できた」「評価が分かりやすかった」と記述している生徒も7名いた。一方で、以下のようにルーブリックの曖昧さを指摘する生徒もいた。

- 評価4と評価5の違いがわかりにくかった (2名)
- 1段階ごとの基準の違いがわかりにくかった
- 具体的にどのように直せば良いのか、例などが欲しかった
- ダメな例があると良かった

このような意見があったことを踏まえると、一部の生徒には明瞭性が不十分だったために、事前提示が十分に機能しなかったのかもしれない。例えば、今回使用した内容面のルーブリックの「理由」に関する記述は、評価5が「具体的によく表現されており妥当」、評価4が「十分に表現されており妥当」、評価3が「概ね妥当」というように曖昧さは否めない。また、「どの教科が好きですか」というテーマのもとに書かれたアンカーについて

の説明では、「英語が好きな理由として、『海外に行きたいから』というものは、海外の全てで英語が通じるわけではないから、理由としては妥当ではない」というような指導をしたが、このような説明だけでは、論理性や妥当な理由づけが何であるのかを十分に理解できなかったかもしれない。このような点を考えると、「ルーブリックの評価が分かりやすかった」と答えた生徒においても、具体的な修正点など改善点が必ずしも明確であっ

たかはわからず、パフォーマンスに影響を与えなかったことは考えられる。

第2に、ルーブリックを提示されなかった統制群の生徒は、すでに今回のライティングで何が求められているのかを理解しており、ルーブリックを提示された実験群と同じ状況にあった可能性もある。アンケートの結果を見ると、「内容を筋の通ったものにしようと努力したか」、「構成をよくしようと努力したか」といった内容面に関わる質問のスコアは、前者は0.2ポイントほどの違いはあるものの統制群においても4.47という比較的高い結果であり、両群で大きな違いがあるとは言えない。統制群においては、「筋の通った」「論理的な」「構成」といった言葉は、実践中には使用せず、アンケートで初めて質問した要素である。それに関わらず、こうした項目への回答が、実験群と同様の傾向を示すことは、今回のようなエッセイライティングでは、理由づけとその一貫性が大事である、ということを経験者の提示がなくても意識できていたかもしれない。また、同様に、統制群に対するフィードバックでは、点数のみが提示されるだけであるが、アンケート調査では、「返却された答案が、その評価になる理由が理解できましたか」という質問に実験群よりもやや高い回答をしている。こうした点からも、ルーブリックを提示されなかったものの、評価として何が求められているのかを理解していた可能性を考慮することができる。今後、学習者がどのように事前提示やフィードバックを解釈し、パフォーマンスに活用しているのかをより深く理解するために、本実践で実施したような選択肢式のアンケートを用いて調査することが必要であろう。

第3に、そもそも事前提示だけでは、パフォーマンスに影響を与えるほどの効果を得られない可能性もある。実験群は、初回のライティングの授業において、ルーブリックと各尺度におけるアンカーを提示され、評価基準の説明を受けた。また、1つの答案について、どの評価に当てはまるかを検討してもらった。しかし、そうした事前提示だけではルーブリックや、評価基準に対する理解は十分得られず、例えば、フィードバックの際に、再度ルーブリックの記述について説明を加えたり、生徒の答案をアンカーとして説明をしたりす

ることで、パフォーマンスに影響を与えるくらいの効果が表れるのかもしれない。また、Goodrich Andrade (2001) が指摘するように、ルーブリックにより、良いライティングに関する理解が高まったとしても、ただちにスコアに反映されないのかもしれない。一方、Goodrich Andrade & Boulay (2003) は、ルーブリックの提示の効果が大きい可能性に言及しているが、2回の推敲後のスコアを分析しているため本研究と単純に比較することはできない。さらに、寺嶋、林(2006)など多くの研究がなされているように、自己評価が有効であるかもしれない。事前提示に加え、下書きの時点や、課題を完了した時点、またフィードバックを踏まえて、自己評価をすることで、ルーブリックの記述と自分が足りていないもののギャップに気づき、それが次のパフォーマンスに活かされる可能性は今後さらに検討されるべきである。相互評価に関しても同様の検証が必要である。また、ライティングに関しては、その向上にはもともと時間がかかるため、今回の実践のように2ヶ月弱での実践ではなく、長期的に実施することで、事前提示も意味があるものになってくるのかもしれない。形成的な評価という観点からも、事前提示が長期的にライティングの向上に及ぼす影響も、今後検証する必要があると言える。

## 5.2 RQ2: ライティング課題において、ルーブリックを事前に提示した場合、そうでない場合を比較し言語面においてパフォーマンスが向上するか

言語面においても、2回目のスコアが1回目と3回目と比較して、有意に高い結果を示した。内容面でも述べたように、これには採点の信頼性と生徒のライティングの向上の過程という2つの可能性が考えられる。言語面の評価の信頼性に関しては、1回目は一致度が「まずまずの一致」で、3回目に関しては採点者の相関関係は、 $r = .236$ 、一致度も「わずかな一致」とどまった。このように、2人の採点者のずれが2回目のスコアが有意に低くなったことに影響していることはあり得るだろう。また、ライティング向上の過程に関して、内容面同様に表11に、3回の言語面のスコアの推移を示した。2回目で下がり、3回目はその2回目より上がった生徒は両群で24名おり、内容面同様の傾向を示した。

■表11: 3回のライティングの言語面のスコアの推移

	統制群	実験群
3回を通して下がらず上がった	8人	6人
2回目で下がり, 3回目で2回目より上がった	14人	10人
3回を通して上がらず下がった	6人	7人
2回目で上がって, 3回目で2回目より下がった	0人	5人
3回全てで変わらない	2人	2人

実験群と統制群の比較には有意差は見られず、ルーブリックの提示に有意な効果は見られなかったが、言語面に関しては、その要因として以下の4点が考えられる。

第1に、内容面と同様に、ルーブリックの記述が明瞭ではなかったのかもしれない。特に、言語面に関しては、実験群の初回のライティング時に、どのようなライティングが言語面において評価されるのかについて詳しくは説明をしなかった。また、言語面に焦点をおいたアンカーも提示しなかった。このため、例えば、言語面のルーブリックの中の「文の作り方」「語彙の豊富さ」と言った記述の具体的な内容がわからず、パフォーマンスに活かされなかったのかもしれない。

第2に、中学1年生にとって、ライティング課題そのものが負荷の高い課題であり、言語面を意識することができなかったのかもしれない。本実践研究に参加した生徒は、中学校1年生から英語を本格的に学習した者がほとんどであった。対象となった生徒は、入学以降、比較的早い段階から、英語の授業の中で、自己紹介や自分の好きなものの紹介といったライティングをしてきた。そのため、書くことにはある程度抵抗のない状態であったと推測できるが、自分の意見を理由と共に主張するタイプのライティングの経験は浅く、20分で自分の主張をする課題は負荷が高かった可能性がある。このために、意識はしていたとしても、言語の形式や語彙の種類、文法などの正確さに注意を払えなかったのかもしれない。中学1年生に対しては、パフォーマンス課題の難易度が低い自己紹介などのテーマであれば、言語知識にも注意が向けることが可能となり、言語面への事前提示効果が見られるのかもしれない。

また、第3に、ルーブリックの言語面の評価で求められていることを実践しようとしたが、その基準を満たす言語知識を持ち合わせていなかった可能性もある。本実践は、1回目のライティングから3回目のライティングまでの期間は、6週間しかなく、言語面に注意を払ったものの評価に見合う言語知識がまだ獲得されておらず、パフォーマンスに反映できなかったのかもしれない。例えば、辞書で単語を調べて使おうとしても、品詞の知識はまだ浅いため、その語の使い方がわからず、使用を躊躇ってしまう、というケースなども想定される。学年が上がった中学2年生や中学3年生に対して同様のライティングを課した場合、事前提示が、蓄積された言語知識や、調べた語彙や表現などを幅広く使用することを活性化させるかもしれない。

以上のように、第1～3で述べた可能性は、パフォーマンス評価を実施する際は、課題のレベル、ルーブリックの記述、学習者の習熟度の3点をバランスよく検討する必要性を示唆している。

第4に、事前提示は言語面のパフォーマンスの向上に効果を与えるほどの効果がない可能性もある。ルーブリックの機能の1つは、評価を透明にすることであるが、言語面の評価の透明性においては、ルーブリックの記述をより具体的にしたり、フィードバックも個人個人に合わせたコメントを付与したりするなどの形で実施しなければ、パフォーマンスに効果が見られないのかもしれない。特に、ライティングのフィードバックにおいては、どのようなフィードバックが効果的か、そもそも言語面のフィードバックは有効なのか、について研究結果は一致していない(田中, 2015)ため、さらに研究を重ねる必要がある。



### 5.3 RQ3: ライティング課題において、ループリックを使用した場合と、そうでない場合を比較し、課題に対する目標意識や、その後の学習への意欲が高くなる傾向が見られるか

3回のライティングが終了した後、それぞれのグループにアンケートを実施した。実験群と統制群の共通の項目において、フィードバックの有効性について聞いた項目(12)、(13)で若干の傾向の違いが表れた。フィードバックに関しては、統制群はスコアのみの開示であるので、フィードバックとして不十分だったと感じる生徒は実験群より多くなることは予想されたことである。

しかしながら、その他については、実験群と統制群で大きな違いが見られなかった。アンケートによる調査方法は、当然ながら解釈には制限や限界があり、実際の生徒の意識を反映しているとは限らないが、事前提示に効果があれば、それぞれの項目に関しても、回答の傾向が異なることが予測される。今回は、ライティングのスコアの比較に有意な差が表れなかったが、今回実施したアンケート項目で聞いているような内容において、少なくとも生徒の意識の点で違いがより明確に見られてくると、パフォーマンスへも影響してくる可能性もある。そういった意味では、より生徒が具体的に理解し意識できるループリック作りと例を伴う説明が、ループリックを効果的に使用する上で、重要な鍵の1つとなってくると考えられる。

## 6 まとめ

本実践は、中学1年生のライティング活動において、ループリックの事前提示がパフォーマンスに有効かどうかを統制群とスコアを比較することで実証的に調査することを目的としたが、その効果は示されなかった。ループリックの事前提示においては、記述の明確さ、ループリックの説明の仕方、アンカーの提示などが効果的な提示として必要であり、またそれを複数回行うことで、より評価基準を理解することができるかもしれない。また、本実践がそうであったように、事前

提示やフィードバック、自己評価や相互評価自体のそれぞれの有効性を検証することも大事であるが、Panadro & Jonsson(2013)がループリックの使用を効果的にする要因として挙げている自己評価やメタ認知活動と組み合わせることで、ライティング力が向上するかどうかを検証することも必要である。ループリックを自己評価や相互評価とともに使うことは、「学びとしての評価」という観点からも重要であると考えられる。

また、ループリックによる「効果」をどう捉えるか、という観点も非常に重要な意味を持っている。例えば、本実践でもフィードバックについて言及し、ループリックによるフィードバックの有効性を調査したが、第2言語習得研究(SLA)と第2言語ライティング研究では、フィードバックの効果をどう捉えるかが異なっている(田中, 2015)。前者は、フィードバックを受けて、そのフィードバックが新たなライティングをする際に効果を持っているかを検証するのに対し、後者は、1つのライティングのドラフトの段階でなされたフィードバックが書き直し後に効果を持っているかどうかを対象としており、ライティング全体に対する自己編集(self-editing)ストラテジーの発達過程にその焦点がある(Ferris, 2010)。この点を踏まえると、本実践では、SLAの観点からループリックの事前提示とフィードバックの効果を検証したといえるが、ループリックの性質を考えると、ライティング研究が対象にするドラフト段階での発達過程にこそループリックの提示やフィードバックが効果を持つことはありそうである。そして、その積み重ねの先にSLAで定義される「効果」が表れるのかもしれない。今後は、この「効果」をどう捉えるかという点も踏まえて、ループリックをどう使うかについて検証し、実践していくことが必要だと思われる。

注

(1) 実社会を想定して課される課題を用いて活用する力を評価することを真正の評価と呼ぶ。Wiggins(1998)は、真正を「大人が仕事場、市民生活、私生活の場で『試されている』、その文脈を模写したりシミュレートしたりすること」(p. 24)と定義している。後述のパフォーマンス評価も実際の社会での知識の活用が問

われている課題の場合、真正な評価の対象となり、両者はオーバーラップする場面が多い。

- (2) 「評価規準」は、目標に準拠した評価を行う際の目標を指す。これに対し、「評価基準」は、「規準」(目標)に従って、教師が実際に評価をおこなうときに、指標として用いるもので、例えば、「こういうものができていれば5」で、「ここまでであれば4」というように具体化された評価の指標のことである(田中, 2010)。
- (3) アメリカの教員が北西地区研究実験所(Northwest Regional Educational Laboratory)の支援のもと開発された小学校3年以上を対象にした書き方のためのルーブリック。
- (4) ライティングのフィードバックには、内容や構成についての「内容面」のフィードバック(content-based feedback)と、文法や語彙などの「形式面」のフィードバック(grammar-based feedback)の2つがあるとされている(田中, 2015)。

## 謝辞

本研究への機会とご支援をくださいました、公益財団法人 日本英語検定協会の皆様、ならびに選考委員の先生方に心より御礼申し上げます。「研究はひっそりやるもの」という小池先生の言葉を胸に、日頃の指導と共に本研究を進めて参りました。そしてその中で、常に的確にかつ今後の指導や研究にもつながるような助言を下さりました指導助言者である小泉利恵先生には深く感謝申し上げます。ありがとうございました。また、多忙な中、本研究の採点を快く引き受けていただいた中央大学附属中学校・高等学校の園田恒子教諭および小堀英男教諭に改めて御礼申し上げます。ありがとうございました。

## 参考文献(\*は引用文献)

- \* 安藤輝次 (2014). 「ループリックの学習促進機能」『関西大学文学論集』, 64(3), 1-26.
- \* Arter, J., Spandel, V., Culham, R., & Polland, J. (1994). The impact of training students to Be self-assessors of writing · paper presented at the Annual Meeting of American Educational Research Association (New Orleans, LA, April 4-8), 2-20.
- \* Arter, J., & Chappuis, J. (2006) *Creating & recognizing quality rubrics*, Educational Testing Service.
- \* Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, 32, 181-201.
- \* 藤森千尋・小泉利恵 (2011). 「スピーチ発表におけるALTとJTEによる全体的評価: 正確さ・流暢さ・複雑さの客観的測定と生徒による相互評価と比較して」 *KATE (Kanto-koshinetsu Association of Teachers of English) Journal*, 25, 21-31.
- \* Goodrich Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education* 4(4).
- \* Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, 97 (1), 21-34.
- \* International Baccalaureate (2011). *Language B guide*. Cardiff: International Baccalaureate Organization (UK) Ltd.
- \* 石井英真 (2015). 「教育目標と評価」西岡加名恵・石井英真・田中耕治 (編). 『新しい教育評価入門一人を育てる評価のために一』(pp.77-111). 有斐閣コンパクト
- \* 石井英真 (2016). 「資質・能力ベースのカリキュラムの危険性と可能性」『カリキュラム研究』25, 83-89.
- \* 松井市子 (2015). 「評価rubricを活用した英語ライティング力と自己評価力の育成をねらった実践」EIKEN BULLETIN, vol.27, 146-164.
- \* 松下佳代 (2007). 「パフォーマンス評価—子供の思考と表現を評価する—」日本基準ブックレット No.7.
- \* 松下佳代 (2012). 「パフォーマンス評価による学習の質の評価—学習評価の構図の分析にもとづいて—」『京都大学高等教育研究』18, 75-114.
- \* 沖裕貴 (2014). 「大学におけるループリック評価導入の実際—公平で客観的かつ厳格な成績評価を目指して—」, 『立命館高等教育研究』14, 71-90.
- \* Panadro, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144.
- \* Stevens, D. D. & Levi, A. J. (2005). *Introduction to rubrics*. Sterling, VA: Stylus Publishing.
- \* Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.) Upper Saddle River, NJ: Merrill/Prentice-Hall.
- \* 鈴木雅之 (2011). 「ループリックの提示が学習者に及ぼす影響のメカニズムと具体的事例の効果の検討」『日本教育工学会論文誌』35(3), 131-143.
- \* 田中耕治 (編) (2010). 『よくわかる教育評価』〔第2版〕 ミネルヴァ書房
- \* 田中真理 (2015). 「ライティング研究とフィードバック」大関浩美 (編・著), 名部井敏代・森博英・田中真理・原田三千代 (著) 『フィードバック研究への招待—第二言語習得とフィードバック—』, (pp. 107-138). くろしお出版
- \* 寺嶋浩介・林朋美 (2006). 「ループリックの構築により自己評価を促す問題解決学習の開発」『京都大学高等教育研究』12, 63-71.
- \* 鶴田京子 (2018). 「ループリックを用いて段階を踏んだ、継続的なスピーキング活動の実践」EIKEN BULLETIN, vol.30, 161-176.
- \* Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve Student performance*. Jossey-Bass.
- \* Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development. ウィキンズ, G. & マクタイ, J. (2012). 『理解をもたらすカリキュラム設計—「逆向き設計」の理論と方法—』(西岡加名恵訳) 日本標準
- \* 山田嘉徳・森朋子・毛利美穂・岩崎千晶・田中俊也 (2015). 「学びに活用するループリックの評価に関する方法論の検討」『関西大学高等教育研究』6, 21-30.

## Model Writing

### テーマ：Which subject do you like?

#### Model Writing① 【内容面：評価5】

My favorite subject is English. First, I want to talk with foreign people in English. I want to understand foreign culture. Second, I like an English book. I often read one. It is interesting. Therefore, I want to study English.

#### Model Writing② 【内容面：評価4】

My favorite subject is English. First, I want to talk with foreign people in English. Second, I like an English book. I often read one. It is interesting. So, I want to study English.

#### Model Writing③ 【内容面：評価3】

I like English. First, I want to talk with foreign people in English. Second, I like an English book. So, I like English.

#### Model Writing④ 【内容面：評価2】

My favorite subject is English. First, I want to talk with foreign people in English. Second, I like foreign movies. So, I want to study English.

#### Model Writing⑤ 【内容面：評価1】

I like English. I study English every day. English is difficult. An English class is fun.

#### Model Writing⑥ 【内容面：評価 ???】

I like English. First, I like foreign countries. I want to visit many countries. Second, English is interesting. I want to speak English. So, I like English.

資料2: 生徒が書いた実際の答案(上: 内容・言語とも評価5。下: 内容・言語ともに評価3).....

### Writing Practice ③

Topic: Which city in Japan do you like? (日本のどの都市が好きですか)

I like Hokkaido. I have three reasons. The first reason is it is very rich in nature. There are many tourist spot. It is very beautiful. The second reason is the food is very delicious and there are many types of food. The third reason is there are many valuable animals in Hokkaido. You can see animals in the city. It is really different from Tokyo. So I like Hokkaido.

### Writing Practice ③

Topic: Which month do you like? (1年間でどの月が好きですか)

I like May. Because, I have birthday. On birthday, I can eat delicious food and receive presents. I don't like Hot season. So, I like the temperature in May. May is very refreshing season.