

ループリックの効果的な使い方の検討¹

岩 本 祐 樹

キーワード〈ループリック、パフォーマンス評価、パフォーマンス課題、形成的評価〉

0. はじめに

従来、学習とは、分離可能で普遍的なものであるという知識注入主義が前提となっており（山田・森・毛利・岩崎・田中，2015）、与えられた知識をいかに暗記し、出力できるかが能力として捉えられてきた（北澤，2019）。しかしながら、知識とは学習者が環境と関わりながら主体的に構成していくものであるとする「構成主義」という学習理論がジャン・ピアジェ（Piaget, J.）によって創始された。構成主義の考え方では、学び手は、環境との相互作用をしながら、学んだ知識を自分の中に取り入れるために自分の考え方を変えたり（調節）、自分が理解できるように知識を変形させたり（同化）、していくものだとなっている（北澤，2019）。

また、近年では、知識を習得するだけでなく、使いこなしたり創造したりする力を育成すべきとの風潮もある（松下，2010）。知識を活用するといった場合、能力を階層的に据えることが多い。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。

このように、構成主義による学習理論と「使える」レベルの能力の重要性は、中等教育でも強調されており、こうした変化にともない、学習成果に関する評価方法も転換する必要がある。つまり、従来、評価においては、「知識・理解」をどう客観的に評価するかという点に重きが置かれていたが、パラダイムの変遷により、以下の2点が課題となっている。

1. 知識の構成過程や「使える」レベルの到達度合いをどのように問うか
2. またそれをどのように評価するか

この2つの問いに対する答えは様々あるだろうが、1に関する1つの方法が、パフォーマンス

1 本稿は、第32回「英検助成研究」に掲載された筆者の論文をもとに作成した。もとの論文は、ループリックの効果を実証することが目的であるが、本稿は先行研究をもとにループリックの使い方を検討することに目的がある。一部内容が重なる部分もあるが、本稿の目的に合致するように、大幅に加筆、修正を加えた。

ンス課題であり、2のその評価方法として注目を集めているのが、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²だとしている（図1）。

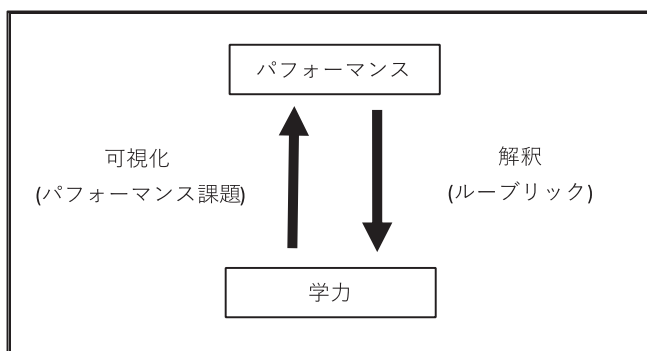


図 1: パフォーマンス評価の構図（松下, 2007, p.11）

沖（2014）が、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言い難い」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。単純に質的なパフォーマンス課題に対して、ルーブリックを作成し、評価、フィードバックするだけでは、その効果は望めないと考えられる。本稿では、このルーブリックに焦点を当て、その理論的枠組みを概観し、実証研究を紹介する。その上で、主に中等教育における教科学習でのその効果的な使用方法を検討していく。

1. ルーブリックの理論的枠組み

1-1. ルーブリックとは？

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。例えば、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

2 パフォーマンス評価と類似する評価に「真正の評価」がある。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。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。

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。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。

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り、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うことができる、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。第1に、ルーブリックの構造の点から、「分析的ルーブリック」と「全体的ルーブリック」に分けることができる。分析的ルーブリックは、ある課題を評価する際に、複数の評価規準³を設定し、それぞれについてレベル分けをして（尺度）、評価基準を記述しているものである。一方、全体的ルーブリックは、規準を複数に分けずに、全ての規準を1つにして評価基準を記述し、全体的に評価するルーブリックである。具体的には、英語のライティングの評価であれば、「論理性」「創造性」「文法」などの複数の「評価基準」を設定し、それぞれについて、3, 2, 1のような尺度に分け、各尺度に対応する「評価基準」の記述をするのが分析的ルーブリックである（図2）。一方、全体的ルーブリックは、評価規準を設けず、評価の尺度のみを設定し、その尺度に対応する評価基準を記述する形である。したがって、1つの尺度の記述の中に、複数のポイントが含まれていることが想定される。第2に、ルーブリックは、評価する対象によって、「一般的ルーブリック」と「課題特殊的ルーブリック」に分類することができる。前者は、ある領域において、一般的に適用できるルーブリックで、例えば、「批判的思考」のようなルーブリックは、特定の課題だけではなく、広く利用することができる。後者は、当該課題だけに適用されるルーブリックのことを指す。第3に、ルーブリックを使用する期間からその種類を分類することもできる。短期間、あるいは1度限りの評価のために用いられる採点ルーブリックと、繰り返し類似した課題を与え、長期的に成長のレベルを捉えたい場合に用いられる「長期的ルーブリック」の2つに分けることができる。長期的ルーブリックは、特定の課題のルーブリックの記述語の抽象度をあげることで作成することができるとされている（西岡・田中, 2009）。

また、ルーブリックには、以上に記した評価規準と評価基準、尺度に加え、各尺度に対応

3 「評価規準」は、目標に準拠した評価を行うことを意味し、その目標のことを指す。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）。

する典型的なパフォーマンスの事例が添付されることもある。この事例は「アンカー作品」と呼ばれ、学習者が各尺度における評価基準を具体的に理解し、課題が求めていることを明確に把握する手助けとなりうる。ルーブリックの作成においては、その記述をわかりやすく、また具体的にすることが効果的な使用のために重要であり、アンカーはその補助として有効であると言える。

以上のように、ルーブリックには複数の種類があり、その目的に応じて適切なタイプのルーブリックを選択する必要がある。

尺度	評価基準		
	論理性	独創性	文法
	3 …できる。 …している。	3 …できる。 …している。	3 …できる。 …している。
	2		
1			

記述語

図2: 分析的ルーブリックのイメージ

1.2. ルーブリックの利点

ルーブリックはパフォーマンス評価の際に使用されるが、様々な面で利点があると指摘されている。

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、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がなされるかに

についての記載があることが有効であるとされており、ループリックはこの両方を満たしていると言える (Stevens & Levi, 2005)。また、後述するように形成的に使うことで多くの利点がある。

1-3. ループリックの信頼性と妥当性

ループリックが評価である以上、そこには信頼性と妥当性を確保することが求められるが、信頼性と妥当性を備えたループリックを開発するには大きな労力と時間がかかる (松下・小野・高橋, 2013)。ループリックによるパフォーマンス評価は、高次の能力について広い範囲をカバーし、評価場面と実際に能力の発揮が求められる場面が近いという点で妥当性は高いとされている (松下, 2012)。しかし、山田ら (2015) が指摘するように、パフォーマンス自体が、要素に分断してそれを寄せ集めただけでは再現できない性質を持つため、規準の選択や尺度の設定に関して、必ずしも規準と尺度のマトリックスでは表現しきれない、という原理的な問題を抱えている。

また、信頼性については、客観テストにおいては評価が2回実施されても、同じかまたは似たような結果を示すかどうかが重要であり、課題の示し方や測定手続き、学習者の状態などをコントロールすることで誤差が小さくなる。一方、パフォーマンス評価の信頼性においては、評価者が異なっても、同じかもしくは類似した採点が出るかという「テストの採点の一貫性」⁴に焦点が置かれる。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ループリックを使用するのだが、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。同じループ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。よって、ループ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、共同でループリックの作成・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。

1-4. 形成的評価としてのループリック

ここまでで述べてきたように、ループリックは、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るので、質を量に変換する装置とい

4 複数の評価者間で評価基準を共通理解し、同じ採点規則に従うことなどによって、評価の一貫性が確保されているかを検討する視点のことを「比較可能性」と呼ぶ (田中, 2010)。従来の信頼性の概念を発展させたものとしてギップス (Gipps, C.) によって提唱された。

う一面を持つ（松下，2012）。したがって、学習者の作文や論述、英語などのライティングにおいても、成績評価としてルーブリックによる評価を学習者に渡して終わり、という使われ方がなされることも多いように感じる。しかしながら、安藤（2014）は、ルーブリックの機能は、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、ルーブリックはその採点機能ばかりに重点が置かれると、学習者が評価される点のみに傾注することになり、結果として質の高いパフォーマンスが生まれづらくなったり、より高次の段階のパフォーマンスは評価対象から外されてしまったりするケースも考えられる（山田ら、2015）。安藤（2014）は、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。また、Stiggins（2001）も、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。

また、形成的なルーブリックの使用により、指導→パフォーマンス課題の実施→ルーブリックによる採点→フィードバック→指導の改善、というサイクルが生まれるため、形成的な評価としてのルーブリックの使用は教員側から見たとき、「構成主義的統合」を促進する意味合いを持つ（Panadro & Jonsson, 2013）。

1-5. ルーブリックとメタ認知能力

形成的にルーブリックが使用されるのは、学習者がルーブリックの評価基準を理解し、フィードバックされた際には、自分のパフォーマンスがなぜ与えられた評価になるのか気づき、それを次の課題に活用することが大事であるからと考えられる。ゆえに、その「気づき」をより促すために、メタ認知活動と併用して、ルーブリック評価がなされることが多い。

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自己効力感を高めること、④不安の低減、の4点を挙げている。つまり、ルーブリックによって、課題が求めていることや課題の構成要素が学習者に明確になり、またそれが課題に対する不安の低減や自己効力感の向上につながり、その結果、パフォーマンスの向上が期待できる。そして、ルーブリックによるフィードバックも直接的にパフォーマンスの向上につながると考えられている。また、Panadro & Jonsson は、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、メタ認知活動（自己評価）、ルーブリックの使用期間、ジェンダー、トピック、その他の要因⁵の5つを

5 例えば、「課題の成績における重要度」、「ライティングにおける推敲にかかる時間」などが挙げられる。Panadro & Jonsson は、こうした要因の研究は少なく、パターンを導き出すことは難しいとしている。

挙げている。

特に形成的評価としてルーブリックが使用される時は、その評価は「採点」という側面よりも、「フィードバックとしての評価」と捉えられるべきである。したがって、評価基準に関して、自分自身で省察することが必要であるから、Panadro & Jonsson が挙げているパフォーマンスに寄与する要因の中でも、特に自己評価や他者評価などのメタ認知活動がルーブリックの使用と共に導入されることが多い。山田ら（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⁶、ルーブリックを用いると可能になる。

2. 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のいくつかを本節では紹介する。

寺嶋・林（2006）は、大学の演習型の授業⁷において、学習活動におけるプロセス評価にルーブリックを使用し、その効果を受講者への質問紙の結果から分析した。ルーブリックは、研究実施者である授業担当教員と観察者で作成したが、受講者にも検討をしてもらい、修正を行った。そして、受講者はこのルーブリックをもとに、学習の途中や終了時に自身の学習活動を自己評価した。特に授業の最後に提出する自己評価レポートでは、ルーブリックの各項目のどの段階に当たるのかを検討させた上で、その理由を記述させた。以上の実践を踏まえ、各授業後に自由記述によるアンケートと全ての授業終了後に4件法によるアンケート調査を実施した。以上の調査結果を総合すると、ルーブ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。また同時に、評価の軸となるルーブリックの修正にも自分たちが加わり、都度自己評価を行うことに対して、困難さを感じたり、慣れていなかったりすることも調査によって、明らかになった。寺嶋と林は、学生は「成績は教員から与えられるものである」という認識を持っているので、自己評価をして評価視点の内容を検討させることは難しく、継続した実践が必要であるだろう、としている。

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、ルーブ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評価の仕方（インフォームド

6 評価活動への参加は、自己評価や他者評価のみならず、寺嶋・林（2006）や松井（2015）が実践しているように（後述）、ルーブリックの評価基準の作成自体に生徒を関与させることなども含む。

7 長崎大学教育学部の「情報メディア論」の講義で実施。受講生は62名。

アセスメント)が、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、ルーブ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、ルーブ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、ルーブ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際された総合テストでは、5日間の授業で扱った問題では、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。

Arter, Spandel, Culham & Polland (1994) は、「書き方の特性 (writing traits)」⁸ 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なった実験群(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事の重要性を指摘している。

Goodrich (2001) は、形成的評価としてルーブ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年生を対象にルーブ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、記述式の質問においては、ルーブ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・理解が見られた。具体的には、統制群においては、優れたライティングの基準について抽象的に説明を行う傾向があったが、実験群の学習者は、ルーブリックに記述されている内容を含む、多様な説明をする傾向が見られた。しかしなが

8 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。

ら、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、ループリックの効果が十分に現れなかった理由として、教員側及び学習者がループリックに慣れていなかったこと、ループリックの記述が学習者には馴染みがなかったこと、推敲に当てる時間が十分でなかったことを挙げている。

Goodrich & Boulay (2003)⁹ は、中学1年生と2年生に対し、ライティングにおけるループリックを用いた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、2回推敲する時間を設けた。その中で、実験群の1回目の推敲では、ループリックの7つの項目のうちの3つに関して、自己評価について指導する時間を設けた。具体的には、ループリックと自分自身のライティングを比較し、評価基準を満たす箇所が自分のライティングにあればその部分に色をつけて印をつけ、基準を満たしていないようであれば、その旨をメモさせた。その上で、ライティングを推敲させ、2回目の授業に持参するよう指示した。2回目の授業では、ループリックの項目の後半の4つについて、1回目と同様に自己評価を促す授業を行い、2度目の推敲を行い、最終的なライティングを完成させた。一方、統制群に対しては単に推敲の時間を設けるのみにした。以上の条件のもと、両群の最終的なライティングを採点したが、スコアに有意な差は見られなかった。Goodrich & 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。また、自己評価が効果を持つためには、自己評価中に特定された問題点に対する指導を行うこと、自己評価の時間を長くする、また自己評価の正確さについてのフィードバックを与えることなど、自己評価に関するさらなる指導の必要性について言及している。

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おこなった。その実践から、教員が作成したルー

9 Goodrich & Boulay (2003) では、ループリックの効果を「実施した学校の特徴」「エッセイのトピック」「ジェンダー」の観点からも分析しているが、これらを対象とした研究は十分行われてはいないため、本稿ではこれらの点は分析の対象としなかった。

ブ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ブ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プ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。

岩本 (2021) は、ループブリックを事前提示することが日本語母語英語学習者の英語ライティングに効果があるかどうかを、中学1年生を対象に実験群と統制群に分け調査した。両群は約3週間ごとに異なるテーマのエッセイライティング¹⁰を3回行った。実験群はまず、内容面と言語面の2つの評価規準からなるループブリックと内容面に関してそれぞれの尺度に対応するアンカーを共に提示され、評価基準についての説明を受け、ライティングをおこなった。1回目のライティングは、ループブリックに採点されたスコアを表記し、2回目のライティング時に答案とともに返却された。その上で、2回目のライティングを実施したが、その際、評価やループブリックの説明はしなかった。3回目のライティングも2回目と同様に実施された。一方で、統制群は、何も提示されずに1回目のライティングをおこなった。統制群については、採点者が実験群で使用しているループブリックを使い評価をおこない、そのスコアのみを2回目のライティング時に生徒に返却し、その後2回目のライティングをおこなった。3回目のライティングも2回目同様に実施した。

このような実験計画であるので、ループブ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おこなった。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、ループブ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ブリックの事前提示効果と解釈し、その効果を検証した。

ループブリックを提示することで課題が求められていることが明確になり結果としてパフォーマンスも向上すると推測されたが、実験群と統制群の間で、ライティングのスコアに有意な違いはなかった。その理由として、1. ループブリックの記述が曖昧であり実験群の生徒は必ずしもライティングで求められていることを明確に理解できなかった、2. 求められていることは理解できたが、それを満たす知識を十分に持ち合わせていなかった、3. ループブリックを提示されなかった統制群の生徒もライティングで求められていることを既に知っていた、4. ループブリックを事前提示することに効果がそもそもない、の4つが挙げられる。これらの理由を考えると、ループブリックを作成する際の記述の明確さや学習者のレベルを考慮した課題や評価基準の設定について検証することが必要であると言える。

10 ライティングのテーマは、「どの月が好きですか」「どのTV番組が好きですか」「日本の中でどの都市が好きですか」の3つ。テーマの難易度や順序が結果に影響を与えないよう、両群をそれぞれ3つのグループに分け、カウンターバランスをとった。

3. 考察

以上に示したように、ループリックスの実証研究は多いが、結果は様々である。各研究で実施されているアンケート調査においては、概ね生徒はループリックスが課題の遂行に役立つものと思っていることが示されているが、パフォーマンス課題において有意に効果を持つかどうかについて一致した結果はでていない。また、多くの実践が自己評価や他者評価、ループリックスの明示的な指導など、その他の教育実践と併用してその効果を検証しようとしていることもループリックスの性質を表している。これらの実証研究と本稿の前半で紹介した理論的枠組みを踏まえると以下のことが言える。

1. ループリックスはパフォーマンス課題などの質的評価を可能にするが、単に採点のために使用するのではなく、形成的評価として用いることでより大きな教育的効果が期待できる。
2. 形成的評価に用いることは、生徒がループリックスの評価に主体的に関わることも意味する。それは単にループリックスによるフィードバックを受けるというだけではなく、ループリックスの作成に関わること、自己評価や他者評価をすることを含む。ゆえに、そうした活動と共にループリックスを使用することで、メタ認知能力が向上し、その結果として、パフォーマンスに良い影響を及ぼすのかもしれない。
3. 一方で、そのような活動がパフォーマンスに効果的だったことを示す実証研究は多くはなく、ループリックスを有効に使用することは思いのほか難しいと言える。ループリックスの明示的な指導やループリックスの作成への関与、メタ認知活動などを併用することでその効果は現れやすくなると推測できるが、例えば、メタ認知能力の向上とパフォーマンスの向上が同時に起こるわけではないので、長期的な視野が必要である。またそうした活動に学習者が十分に慣れるための時間も必要と言える。
4. ループリックスの効果を高めるためには、ループリックスの記述を具体的にわかりやすく記載することは必須である。Alter and Chappuis (2006) は、ループリックス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すること」が大事であるとしている。また、学習者がよりループリックスを理解するためには、アンカーを添付することも有効であると考えられる。
5. 特に形成的にループリックスを使用する場合、その「効果」をどう捉えるのか、という点

にも教員は注意する必要がある。「効果」をどう捉えるかという問題は、言語習得におけるフィードバック研究の例を考えるとわかりやすい。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）戦略の発達過程にその焦点がある（Ferris, 2010）。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 で定義される意味合いでの「効果」があらわれるのかもしれない。

このように、近年注目を集めているループリックではあるが、その効果的な使用について考えると、様々な要因が関与し、その使い方は多様であり、今度のさらなる研究も必要である。

最後に、本稿の主要なテーマではないが、入試などのハイ・ステイクスな評価¹¹におけるループリックを用いたパフォーマンス評価の可能性に触れる。

ループリックを用いたパフォーマンス評価は、本稿の 1-3 でも触れたように信頼性や妥当性に課題があるため、例えば大学入試のような選抜試験で使用することは難しいと考えられる。しかしながら、ニューヨーク・パフォーマンス・スタンダード・コンソーシアム（New York Performance Standard Consortium: 以下、NYPSC）では、加盟校に求めているパフォーマンス評価の信頼性が認められ、州統一テスト受験義務免除に至ったという事例がある（遠藤，2012）。ニューヨーク州は、高校の卒業要件として、英語、数学、社会、理科といった科目に州の統一試験を求めている。しかし、NYPSC は、加盟校に課しているパフォーマンス評価課題を確実に実施するために、州統一試験のうち英語以外の科目の受験免除が認められた。これは、NYPSC が実施しているパフォーマンス評価の信頼性が確保されているためであり、それを可能にしているのがモデレーションである。遠藤（2012）によれば、このモデレーションでは、まずコンソーシアム加盟校の教員の代表 150 名が、提出された生徒の作品を読み、ループリックで評価を行う。その後、学校ごとに教員が、同じ作品群から選ばれたものを読んで評価するという 2 段階の体制をとっている。また、外部の教育者も同様に評価活動を行う。このモデレーションが教員研修の一環として年間のサイクルに位置付けられており、このモデレーション研究により、NYPSC のパフォーマンス評価が州統一の標準テストよりも

11「ハイ・ステイクスな評価」とは、入試などの選抜に関わる評価、採用や昇進など人事に関わる評価などの被評価者にとって利害関係の大きい評価のことを指す（田中，2010）。

妥当性が高いことが認められ、統一テストの受験義務免除に至ったという。

これは、パフォーマンス課題が統一のテストの代替となることが認められた例であり、ハイ・ステイクスな場面におけるルーブリックの使用の可能性を示している。例えば、筆者が勤務する中央大学附属高等学校においては、大学への内部推薦は、主に校内で実施している定期試験の結果で選抜されるが、試験においては客観的な評価が求められる。このような状況においても、教員間でのルーブリックの開発とモデレーション研究により、パフォーマンス評価を取り入れる余地は十分にあるだろう。また、中央大学は、その他にも3つの附属校があり、各校でそれぞれ大学への推薦者を決定しているが、学校をまたいで統一したパフォーマンス課題を課し、ルーブリック評価をすることで、大学推薦を決定することも可能であろう。その際には、上述のNYPSCが実施するような教員間でのモデレーション研究が十分にできるかどうかは鍵となると言える。

参考文献

- 安藤輝次 (2014)。「ルーブ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, a 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.
- 遠藤貴広 (2012)「州テスト政策に対抗する草の根の教育評価改革 - New York Performance Standard consortium を事例に」北野秋男・吉良直・大桃敏行編『アメリカ教育改革の最前線 - 頂点への競争』学術出版会 pp.231-243.
- 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.
- 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.
- 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」(pp.77-111)・有斐閣コンパクト
- 岩本祐樹 (2021)。「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」EIKEN BULLETIN, vol.32, 84-107.
- 北澤武 (2019)。「構成主義と構築主義」大島純・千代西尾祐司 (編) 益川弘如・河崎美保・山口悦司・大浦弘樹・望月俊男・北澤武・大島律子・河野麻沙美・大崎理乃『主体的・対話的で深い学びに導くー学習科学ガイドブック』北大路書房
- 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 (編著) (2010)。「〈新しい能力〉は教育を変えるか—学力・リテラシー・コンピテンシー」ミネルヴァ書房
- 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.
- 松下佳代・小野和宏・高橋雄介 (2013)。「レポート評価におけるルーブリックの開発とその信頼性の検討」『大学教育学会誌』32 (1) : 107-115.

- 西岡加名恵・田中耕治編著 (2009)『「活用する力」を育てる授業と評価 中学校一パフォーマンス課題とルーブリックの提案』学事出版
- 沖裕貴 (2014) .「大学におけるルーブ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) .「ルーブ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) .
- 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
- 寺嶋浩介・林朋美 (2006) .「ルーブ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) .「ルーブ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ブ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.