

The Influence of Evaluation Criteria Descriptions on Peer Evaluation Activity : The Relevance of Between Evaluation Forms to Advice Given

メタデータ	言語: jpn 出版者: 公開日: 2023-03-15 キーワード (Ja): キーワード (En): 作成者: 藤浦, 五月, 宇野, 聖子 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/1991

〔調査報告〕

評価項目の記述が他者評価活動に与える影響

—評価表とアドバイスコメントの関連性—

The Influence of Evaluation Criteria Descriptions on Peer Evaluation Activity:

The Relevance of Between Evaluation Forms to Advice Given

藤浦五月・宇野聖子

キーワード：評価、ルーブリック、レポート・ライティング、初年次教育、アドバイス

1. はじめに

多くの大学が、初年次教育においてレポート・ライティングの授業を取り入れているが、その評価方法については改善・検討が続いている。レポート・ライティングという行為自体が、テーマ設定、資料収集、読解、論理的な展開に必要な思考力や構成力など、表面的な執筆ルール以外にも多くの要素が関わっており、そのパフォーマンス全体を捉えて評価することが非常に難しいためである。近年、このような場面で多く用いられているのがルーブリック評価である。ルーブリックとは、成功の度合いを示す数値的な尺度と、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からなる評価基準表（田中 2003）である。定義にもあるように、ルーブリックでは、パフォーマンスの特徴を示す記述語という質的な側面に数値的尺度がついていることでパフォーマンスの質が量的に表現されている。松下（2012）は、脱文脈化できない、量的に測れないものを評価するためのパフォーマンス評価が、ルーブリックのこのような特徴により、再度脱文脈化や標準化を志向する契機となっていることを指摘している。レポートというパフォーマンスを測る手段がルーブリック評価表であれ、他の形式の評価表であれ、数値的尺度で示される場合、カリキュラム設計者や評価者はこうした「標準化への志向」に自覚的である必要があるだろう。

本来、パフォーマンス評価が、学習者が自身の学びの段階に気づき次のステップへの手がかりを得ることを目的とするのであれば、最も大切なのは点数を出すのではなく、学びのなかでその評価行為をどのように位置付けるかということである。ゆえに、パフォーマンス評価は、多くがルーブリック等を用いた評価行為（活動）のあとに、改善の機会を与えられるタイミングで行われる形成的評価であり、教員からの評価だけではなく学習者同士の学びあいに活用する事例も多い。筆者らの実践においても、アウトラインや第一稿を学習者が互いに読み合い、アドバイスする機会を設けている。

一方で、パフォーマンス評価に用いられる評価表に関する研究は、教員—学習者、学習者相互の評価がどの程度一致しているかという信頼性に関するものやライトの程度などの効果に関するものは多いが、その評価表を用いた後に想定されているアドバイス活動と関連させたものは少ない。パフォーマンス評価で出された評価を、学びのなかでより効果的に活用するためには、評価表単体ではなく、その後の学習者同士の活動との関連性を探る調査も必要であろう。

そこで本調査では、実際に授業で用いられたことのある2つの評価表（1つは非ルーブリック形式、もう1つはルーブリック形式）を用いて、評価の傾向とアドバイスコメントの傾向を分析する。2つの異なる評価表を用いることで、評価表それ自体が学習者の活動にどのような影響を与えるかを推測することができる。また、評価傾向やアドバイス傾向から、評価活動から次の活動に移る際に教員からどのような働きかけが必要かを探ることも可能である。本調査を、パフォーマンス評価を学びのなかでどのように位置付け、クラス内で活用すべきか検討するための一助としたい。

2. 先行研究

2.1 教員評価と学習者評価

パフォーマンス評価で課題として多く取り上げられるのは、評価の信頼性についてである。教員間、教員と学習者、学習者同士の評価について、さまざまな研究が試みられている。ここでは、教員と学習者評価の関係について整理したい。

Cho, Schunn, & Wilson (2006) は、3年間で16コース、708名を対象にライティングのピアレビューについて検証し、一つのレポートに対して少なくとも4名の学習者が評価した場合、その総合評価は教員の評価と同様に高い信頼性と妥当性を持つことを明らかにしている。一方で、教員自身が学習者による相互評価・アドバイス活動を有用であると判断し、相互評価やアドバイス活動をライティングの授業に取り入れたとしても、それだけでは学習者は教員と同様に有用であると感じるとは限らないことも指摘している。この調査の結果では、高いレベルの信頼性を確保するためには複数のピアが必要であるが、個人の評価のバラつきが大きいと信頼性認識が低くなり、学習者が相互評価結果を受容する度合いが低下することがわかった。Choらの結果から、相互評価の受容度を低下させるリスクを軽減するためには、学習者の評価のバラつきができるだけ少なくなるような評価表を用いる必要があることがわかる。

一方で、Choらは、信頼性は多様性とも表裏一体であることに注意を促しており、評価を統一することに注意を向けすぎて、一つのレポートに対する多様な反応について学ぶことを犠牲にすべきではないと述べている。つまり、評価活動自体の信頼性を損なうような評価のバラつきは回避しつつも、自身のパフォーマンスに対する多様な意見があることも学びとして価値づけることが大切である。

石川・阿部・川崎 (2022) は、実験レポートの再提出後、ピア・レスポンスのみ、ピア・レスポンスと教員によるフィードバックを行い、修正して最終稿として再提出させた。その後の質問紙調査の結果、教員によるフィードバックと組み合わせた場合、ピア・レスポンスの評価がややネガティブになることがわかった。学習者は、教員によるフィードバックを学習者同士

のフィードバックより価値あるものとして捉えていることが窺える。石川らの調査の結果は、Choら(2006)も指摘していた「多様性への価値づけ」が必要であることを示唆しているといえる。

これらの結果から見ると、学習者自身も「評価しやすい」と認識できる、評価のバラつきの少ない明快な評価表を開発することは、互いの評価活動の信頼性を向上させることが期待できる。一方で、評価活動の目的が学習者の成長である以上、相互評価の目的は「評価の点数をできるだけ揃えること」ではなく、多少評価のバラつきはあっても「多様な観点からアドバイスを受け、互いのレポートについて改善のヒントを得る機会として捉えること」とすべきだと考えられるが、そのような認識を持たせることは難しい。このような観点で評価機会を捉えられるようになるには、まずは評価活動において点数だけでなくその根拠となる理由や改善点を相互に伝えること、また伝え方を工夫することが大切であると考えられる。

2.2 教員の指示と内省記述の変化

上田・半田(2022)は、プロジェクト学習における内省支援の実践と評価において、振り返り課題を評価するルーブリック評価表をもとに解釈と分析を促す記述指示を準備し、受講者からの評価・記述量・抽象的概念化の観点から検証している。その結果、改善した記述指示を取り入れた実践において記述量に有意な差が見られ、抽象的概念化に該当する記述については「解釈」に該当する記述が増加する傾向にあることが示された。また、「分析」においては、分析を促す記述指示の要素を取り入れても「分析」に関する記述は増加しなかったことから、内省記述において、その内容には難易度の高低が存在することが示された。上田・半田(2022)の実践は、イベントを企画するプロジェクトであり、レポートに関する実践ではないこと、また内省についての研究であり相互アドバイスに関する研究ではないことから、本調査と異なる点も多いが、本研究結果は、ルーブリックを用いた内省を行う際に教員が与える指示の差が学習者の行動に影響を及ぼすことを示しており、本調査目的と関連性がある。また、学習者の記述内容をコーディングして分析した結果を指示内容に反映させたとしても、一律に効果があるわけではなく、項目により難易度が確認された点については、本調査で分析したアドバイス記述の分類で得られた課題にも同様のことが考えられる。今回の調査結果で得られた傾向と課題をアドバイス支援につなげる際には、項目ごとに難易度が存在することが想定されるため、本調査結果を実践に取り入れる際のヒントとなるであろう。

2.3 ピア指向性の変化と会話内容

富永(2011)は、文章表現授業のピア活動におけるピア指向性の変化と会話内容の関連性について分析している。ピア指向性は、3つの指向性から分析された。3つのピア指向性は、「意見歓迎指向(問題点の指摘を役立つものと感じ、メンバーからの意見を積極的に受け入れようとする指向)」、「フレンドリー指向(文章改善とは関係なく、メンバーとの楽しい会話を重視する指向)」、「高能力要求指向(文章力が低いメンバーとのピア活動は役に立たないと感じ、高いメンバーとのピア活動を望む指向)」である。そのうち、ピア活動後に意見歓迎指向とフレンドリー指向が高くなり、高能力要求指向が低くなった人物の会話構成は、【進行】【問題点】【修正

案】【気づき】で構成されていたという。一方、意見歓迎指向とフレンドリー指向が低くなり、高能力要求指向が高くなった学習者のやり取りの特徴では、【問題点】は指摘されるが、具体的な【修正案】が検討されないまま次に進んでしまうことが多かったという。このようなやりとりがピアの指向性に影響を与える一因であるとしている。また、富永（2012）では、ピア活動に関する意見受け入れ不愉快感の高いグループの自由記述では、e ラーニングや教師フィードバックを評価するコメントが見受けられたという。そのようなケースの場合は、ピア活動で解決できなかったことや気が付かなかったことを教師のフィードバックによりフォローされたことに満足感を感じていたようだとしている。

学習者同士で評価を行ったりアドバイスをしたりしても、その評価やアドバイスを学習者自身が価値あるものとして学びに紐づけなければレポートを吟味する契機にならないだろう。富永の研究からは、「何を改善すべき課題として伝えるか」だけではなく、「課題をどのように伝えと相手に受け入れられるか」という点について検討する必要があることがわかる。

本調査では、評価記述に差がある2つの評価表とその評価表に書かれたアドバイス記述を調査することで、評価記述の差とアドバイス記述が関連しているか、また関連している場合はどのような点が影響しているかを探る。また、本調査をもとに学習者自身がお互いの相互評価活動やアドバイスを価値あるものと感じ、能動的に取り組むためには、カリキュラムにおいてどのように評価活動を位置づけ工夫すべきかを考える。

3. 評価表の検討と開発のプロセス

本調査では、2種類の評価表を用いて調査を行った。1つ目の評価表A（表1）は、宇野・藤浦（2016）が開発した教材に掲載されている評価表である（宇野・藤浦, 2016: 40）。全15項目で、◎○△×を用いて評価をする形式である。本教材開発時は、全学に向けた大型のカリキュラムで使用することは想定していなかったため、評価表Bほど緻密な記述にはなっていない。

表1 評価表A（宇野・藤浦, 2016: 40）

	項目	自己評価 ◎○△×
1	テーマ設定の動機がはっきりしており、議論すべき必要性を感じるか	
2	基本用語について、本レポートでの意味（使用法）まで書かれているか	
3	議論をする上で必要な背景（基本的情報）が整理できているか	
4	一度読めば言いたいことがわかるか	
5	それぞれの問題と関係のある資料を選び、資料を写すだけでなく自分の言葉でまとめられているか	
6	今後の課題・提案を示せたか	
7	全体を通して（テーマを選んだ動機・基本的背景から導き出した問題点・資料を調べた上でわかった主張・今後の課題&提案）一貫性があるか	
8	全体を通して、一般的によく知られていることだけでなく、書き手の考えが伝わるように書けているか（よくニュースで聞き、読むまでもなくわかること・	

	資料の羅列のみ→x)	
9	全体に使用されている資料は信頼できる資料か	
10	タイトルを見て内容がわかるように工夫されているか	
11	全体を通してレポート用の言葉遣い(である・だ)になっているか	
12	授業で学んだ構成・表現を積極的に使用しているか	
13	読み手にわかるような言葉を選択して書いているか	
14	引用表現が正しく使えているか	
15	参考文献が正しく書けているか	
計		

2つ目の評価表Bは、武蔵野大学の初年次教育のレポート指導時に用いられている評価表である。こちらの評価表はルーブリックになっており、段階ごとの記述がある。2つ目の評価表は、全学初年次レポートライティングクラスを担当するにあたり、関わる教員数や学生数が増え(学生数約2,000名)、より公平かつ評価理由がより詳細にわかる評価表が現場で求められたため開発したものである。この評価表は、全学の取り組みが始まってから一度改訂しており、今回使用したものは改良後のルーブリックである。改訂にあたっては、教員間の採点調査と意見交換の結果を参考にしている。2つ目の評価表であるルーブリック全体記述は末尾に掲載するが、開発経緯と全文は藤浦(2021)に掲載している。ここでは、対応する記述の違いを紹介したい。特に、資料解釈も含めた「内容」に関する評価観点と「表現・形式」に関する評価観点の違いをそれぞれ紹介する。評価項目については、ルーブリックを作成する段階で項目を作成し直したため評価表Aと評価表Bの全ての項目が何かしらの項目に対応しているわけではない。

資料解釈も含めた「内容」に関する評価観点の違い、資料とまとめに関する項目を取り上げる。評価表Aでは「それぞれの問題と関係のある資料を選び、資料を移すだけでなく自分の言葉でまとめられているか(評価表Aの5)」という一文の記述であったが、評価表Bの記述は以下、表2の通りである。

表2 評価表Bの6「本論のまとめ」(表1評価表Aの5に対応)

6	第六段落	本論のまとめ	本論(1点目と2点目)からわかったこと・言いたいことを自分の言葉でまとめているか			
			a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント			□本論で資料を引用するだけではなく自分の考えを自分の言葉でまとめている。一点目の資料と二点目の資料から言える	本論で資料を引用するだけではなく自分の考えを自分の言葉でまとめているが、以下の点において不備がある	□引用表現の繰り返りで自分の言葉でまとめられていない □資料と解釈がほとんど合っていない (一箇所でもチェック)	□本論のまとめ(これらの資料からわかることは…)が書かれていない

	ことを総合的に自分の言葉でまとめている。	<input type="checkbox"/> 一点目と二点目の資料を総合的にまとめられていない（どちらか片方になっている） <input type="checkbox"/> 資料とまとめの解釈にややズレがある	クがあればこちら)	
--	----------------------	---	-----------	--

本学の初年次教育で行われている授業では、見本レポートを全員で読解・理解確認をしたあとに各自設定したテーマについて同じ構成で書くことを求められる。左上にある「第六段落」とは見本レポートの段落に該当する。その上で、4段階の評価ポイントについてそれぞれ評価記述がある。

続いて、表現・形式に関する記述の例として「参考文献」の評価記述を取り上げる。評価表Aでは、「参考文献は正しく書かれているか」という記述だが、評価表Bの記述は表3の通りである。

表3 評価表Bの12「参考文献」(表1 評価表Aの15に対応)

12	参考文献	参考文献は、必要な情報が指示通りの方法で統一されて書かれている			
		a. 素晴らしい (3)	b. まあまあ (2)	c. 発展途上 (1)	d. 点数なし (0)
評価ポイント	<input type="checkbox"/> 参考文献は、必要な情報が指示通りの方法で統一されて書かれている	<input type="checkbox"/> 参考文献の情報が足りない部分がある (筆者・組織名、出版年、資料名、出版社、URL、参照日など) <input type="checkbox"/> 参考文献の書き方の不備がある(書式不統一、並べ順など) (不備: <input type="checkbox"/> のチェック数に関わらず全体で1-2箇所程度)	<input type="checkbox"/> 参考文献の情報が足りない部分がある (筆者・組織名、出版年、資料名、出版社、URL、参照日など) <input type="checkbox"/> 参考文献の書き方の不備がある(書式不統一、並べ順、アウトラインの表をコピーして貼るだけなど) (不備: <input type="checkbox"/> のチェック数に関わらず全体で3箇所以上。アウトラインコピーの場合はチェック数関係なく「発展途上」)	<input type="checkbox"/> 参考文献が書かれていない	

以上の表記は、ルーブリック開発時に、授業を担当する複数教員から評価表の評価のしやすさ、負荷等をヒアリングし、実際に採点調査を行って改善したものである。つまり、このルーブリック開発は2つのステップを踏んでいる。まず、学生のレポート(成果物)から躓き点を整理して文言を作成した。その後、教員間の評価差異の調査や、採点時の困難点についてヒアリングした結果を踏まえて文言を改訂した。このように、多くの学生の評価を行うにあたり、担当者による評価差を最小限に抑えることを目指した。藤浦(2021)では、評価記述を検討す

これらのプロセスを具体的に示すことで評価表の客観性について検討している。

パフォーマンス評価においてあえて点数に注目し、バラつきの軽減を目指す文脈は2つある。一つは、複数評価者からの評価のバラつきが、評価を受ける側の評価行為に対する信頼を低下させるため、評価行為自体への信頼低下リスクを避ける文脈である。もう一つは、初年次教育などで複数教員がカリキュラムに関わり多くの学習者の評価を行うなど、公平性や実現可能性がより強く求められる文脈である。筆者らが評価表Bの開発に至ったのは後者であるが、学習者同士の評価活動の捉え方に関連するのは前者であろう。

ここで強調しておきたいのが、本調査はどちらかの評価表がより優れているかを決定するためのものではないという点である。パフォーマンス評価は、評価行為自体が文脈から切り離せないものである。例えば、同じ教材を扱い同じ課題を義務付けたとしても、学年、授業時間、クラスサイズ、それまで受けてきた関連講義、評価表以外のフィードバックの有無等により、評価表の記述は調整され得る。本調査では、詳述度の大きく異なる評価表を用いて調査を行うが、それはどちらかをよりよいものとして提示するのではなく、この調査結果をもとに、各現場の文脈に合わせた評価記述の改善・調整に役立てることを目的としている。また、アドバイス活動についても同様で、評価活動とアドバイス活動をつなげる取り組みを考える際の一助となることを目指したものである。

4. 調査課題

本調査の調査課題は、以下の通りである。

課題1 詳述度の異なる評価表で採点した際、採点傾向にどのような差異が生じるか。

課題2 詳述度の異なる評価表で採点した際、アドバイス傾向にどのような差異が生じるか。

5. 調査概要

5.1 対象者

本調査の協力者は、本学に所属し、かつ、初年次のレポート執筆のための授業(科目名「日本語リテラシー」)に参加した経験がある者とした。本調査は、授業とは別に協力機会を設けて行われ、謝金が支払われた。「日本語リテラシー」履修経験者とした理由は、本調査で使用する評価表はいずれも本実践の成果物を想定して設計されているためである。成果物としてどのようなレポートが求められているかは、授業内において見本レポートの読解から始まり、構成を意識したアウトライン執筆など段階を踏んで学ばれる。このようなプロセスを経て到達点(見本レポートのようなレポートを自らのテーマで執筆すること)は段階的に共有されていく。よって、評価の指標となる到達点の共有については調査単体(短時間)では難しいため、本講義を受講した経験のある学生を対象者とした。今回の評価調査の対象となるレポートは初見であるが、評価表A、Bについてはテキストに掲載されていたりルーブリックを使用していたりするため初見ではない。本調査の目的が評価記述の差異とアドバイスの傾向を調査することであることを考えると初見かどうかは結果の分析に大きな影響を与えないと判断した。

1グループ50名程度募集した。参加希望日時を選択する形式で行ったところ、グループ1・

グループ2も46名ずつ集まった。データ収集後のデータを確認したところ、グループ2の協力者2名の評価表および手続きに不備があったため、対象データは、グループ1は46名分、グループ2は44名分となった。

5.2 調査手順

調査協力者には、本調査がレポートを評価する調査であることを伝え、事前準備として評価対象となるレポートを読んできてもらった。対象となるレポートは両グループ同じもので、調査協力者全員が初見である。評価対象レポートは、本学の初年次のレポート執筆の授業で求められている課題と同じ課題が想定されていることを伝えた。採点対象となったレポートは、同じテキストを用いた実践において過去に提出された学生の成果物であり、提供許可を得て匿名で調査に使用した。藤浦ほか(2018)で行った教員間の評価差異を調査した際にも使用したものであり、教員間でも評価差異が大きく、かつ、資料や形式といった多様なポイントで評価差異が発生していたレポートである。字数は2,020字である。

グループごとにオンライン上で集まり評価表A(グループ1)または評価表B(グループ2)を用いて評価・コメントをしてもらった。評価表Aは、4段階(3点、2点、1点、0点)をチェックするものである。評価表Bも4段階であるが、ループリック形式であり、評価だけでなくその理由もチェックする。どちらの評価表もコメント欄がついており、「Aさんのレポートについて、読んでいて感じた疑問点やもっと知りたいと感じたことを書いてください」という指示と、「Aさんのレポートをより良くするためにアドバイスをお願いします」という指示の欄を設けた。今回の調査報告では、後者の改善アドバイスの傾向について報告することとする。

オンライン上での採点の取り組みの様子がわかるようにカメラオンでの参加を依頼し、その様子は調査アシスタント(2名)がチェックを行った。カメラオンについては、調査協力募集時に明記し、事前に到達している。両グループとも、全体の協力時間は60分で依頼した。オンライン調査のためのURL(Zoom)にアクセス後、評価対象となるレポートを読んできたか、評価対象となるレポートが手元にあるか確認した。その後、評価表をオンラインで配布し、評価表への評価チェックの記入と、気がついた点および改善のためのアドバイス記入をお願いした。アドバイスの記入まで終わった時点で、時刻を報告してもらい評価表チェックとアドバイス記述が終了するまでのどの程度の時間がかかったかを調べた。

5.3 調査状況と実践との関係

本評価調査は、授業内で行われたものではない。初年次必修科目「日本語リテラシー」は約2,000名が受講する科目であり、それぞれ別の評価表を用いた場合に生じうる学習機会の差についてフォローアップを行う時間がなく、教育および評価の公平性に関わると判断したため、授業外で参加協力者を集めて行った。このような条件から、カリキュラムの過程に組み込まれた評価活動を調査できているわけではない。報酬があることにより活動への取り組みの態度に差異が生じる可能性もある。また、協力用の被採点レポートがクラスメートのものではなく匿名であることから、相手のフェイスを侵害することについての心的負担も軽いと考えられる。実際の授業においては、一度回収したレポートを匿名化し番号などを振って割り当ててコメン

ト活動を行うことも、書き手本人を前にしてコメントすることもあり得る。本調査が全ての評価場面を想定しているわけではないことに留意されたい。このように、本調査条件は、実際の授業活動とはいくつか異なる点がある。また、前述したように、初年次のレポート授業を受けたことがある者を協力者としたが、協力者全員が同じクラスに所属していたわけではないため、テキスト、進度、課題、評価方法（評価表）、評価割合などが統一された状況であっても、その過程には若干の差異もあろう。そのため、本調査結果は、評価表による差異が、学習者のアドバイス記述にどのような影響を与えるかを予想するに留まっている。

しかしながら、本調査は、採点時間、採点のバラつき、採点結果の差を数値で、さらにアドバイスコメントについては記述内容の分類と行為を分けて分析しており、このように多様な観点から2つの評価表とアドバイス記述を調査した研究は管見の限り見当たらず、評価表の記述について検討するだけでなく、カリキュラム全体において評価活動をどのように位置づけ導入するかについて一定の示唆を与えようと考えられる。

6. 結果と考察

6.1 評価・アドバイス時間とアドバイス文字数

協力者が評価とアドバイスにかかった時間の差は表4の通りである。

表4 評価表のチェックとアドバイスにかかった時間

時間	評価表 A (N=46)		評価表 B (N=44)	
	人数	割合 (%)	人数	割合 (%)
20分以下	9	19.6	0	0
21-30分	23	50	8	18.2
31-40分	7	15.2	12	27.3
41-50分	6	13	19	43.2
51-60分	1	2.2	4	9.1
61分以上	0	0	1	2.3

評価表 A は、約7割 (69.6%) の協力者が30分以内に評価とアドバイスを終えているのに対し、評価表 B は30分以内に評価とアドバイスを終えた協力者は2割に満たない。41分以上かかった協力者が5割を超え、評価・アドバイス行為自体が時間を要する、負荷の高いものであるといえる。Wang (2014) は、EFL ライティングにおけるピアフィードバックの検証において、時間についても取り上げている。53名の学生のうち12名 (27%) の学習者が「時間的制約」をピアフィードバックの有用性に影響を与える要因として意識していることを指摘している。授業内に他者のレポートを読み、評価・コメントを書き、意見交換をするための十分な時間を確保できるか否かは、学生の相互評価・意見交換活動の満足度を左右する可能性がある。本結果を見ると、評価表のタイプにより必要な時間が異なることが予想されるため、授業活動を組み立てる際には用いる評価表の負荷を考え、時間を調整・計画する必要がある。評価とアドバイ

ス活動を分けることも一案として考えられる。

表5は、「Aさんのレポートをより良くするためにアドバイスをお願いします」という指示欄に書かれた自由記述アドバイスの文字数である。

表5 アドバイス文字数

文字数	評価表 A (N=46)		評価表 B (N=44)	
	人数	割合 (%)	人数	割合 (%)
100 以下	8	17.4	10	22.7
101-150	9	19.6	11	25
151-200	10	21.7	9	20.5
201-250	10	21.7	3	6.8
251-300	2	4.4	5	11.4
301-350	2	4.4	0	0
351-400	1	2.2	2	4.5
401 以上	4	8.7	4	9.1

上記のように、文字数については、評価表 A では 58.7%、評価表 B では 68.2%の協力者が 200 字以下となった。評価表 B の方が評価・アドバイス時間はかかっているが、アドバイス文字数については若干少ない傾向にある。

表1、表2・表3で示したように、評価表 A と評価表 B は評価記述に大きな差がある。評価表 A は、各項目について、「◎ (3点)」「○ (2点)」「△ (1点)」「× (0点)」をチェックするだけであるのに対し、評価表 B は、「すばらしい (3)」「まあまあ (2)」「発展途上 (1)」「点数なし (0)」を評価するだけではなく、更に下位の選択肢からそのように評価した理由をチェックするものとなっている。評価表のチェックとコメントの両方が終了したタイミングでの時間を申告してもらったため、申告時間は評価表のみにかかる時間ではないが、コメントの文字数傾向は両グループに大きな差はなかったことから、評価表のチェック自体にかかる時間の差が大きいものと思われる。アドバイス文字数が少なかった要因としては、一つ目は評価自体に時間がかかりアドバイスまで行う十分な時間がないと判断されたり、疲れが生じたりした可能性がある。また、評価表の下位分類の記述が細かくそちらにチェックを入れていることから、下位分類のチェックで被評価者が理解できるものとして多くのアドバイスは必要ないと判断されたことも考えられる。今回は記述形式でのアドバイスを求めたため、すでに評価表に書いてあるものとして省略された可能性も高く、口頭でのアドバイス活動では異なる結果が生じる可能性も考えられる。

6.2 点数傾向

以下、表6に評価表の平均スコアを示す。両評価表ともに満点を100とした場合の70%程度であった(表6の平均スコアの隣に割合で記載)。点数のバラつきに関しては、評価表 B の

ほうが少ない結果となった。複数の評価者が関わる場合、評価点のバラつきは評価活動そのものへの不信感を招く恐れがあるが、評価表 B のような評価観点の詳述はそのようなリスクを軽減させることが期待できる。項目ごとの採点者割合については後述する。

表 6 評価表総合点の平均スコアと標準偏差

	評価表 A (45 点満点)	評価表 B (36 点満点)
平均スコア	32.2 点 (72%)	26.2 点 (73%)
標準偏差	5.85	4.42

各項目点数の採点者割合を表 7 に示す。例えば、評価表 A を採点した協力者 46 名のうち、項目 1 で 3 点をつけた人物は 34 名で全体の 73.9%であった。

表 7 評価表 A 各項目点数の採点者割合

No.	項目 (表 1 に全文)	スコア平均	◎ (3 点)	○ (2 点)	△ (1 点)	× (0 点)
1	テーマと興味理由	2.74	73.9%	26.1%	0%	0%
2	定義	2.57	65.2%	26.1%	8.7%	0%
3	背景整理	2.11	32.6%	47.8%	17.4%	2.2%
4	一読理解	1.83	19.6%	43.5%	37.0%	0%
5	自分の言葉によるまとめ	2.13	34.8%	43.5%	21.7%	0%
6	課題と提案の提示	1.91	19.6%	52.2%	28.3%	0%
7	一貫性	1.98	32.6%	37.0%	26.1%	4.3%
8	書き手の考えがわかるか	2.17	37.0%	43.5%	19.6%	0%
9	資料の信頼性	2.33	45.7%	41.3%	13.0%	0%
10	タイトル	2.59	69.6%	19.6%	10.9%	0%
11	言葉遣い	2.5	60.9%	30.4%	6.5%	2.2%
12	学んだ表現の使用	2.02	30.4%	45.7%	19.6%	4.3%
13	読み手にわかるような言葉の選択	2.26	43.5%	39.1%	17.4%	0%
14	引用表現	1.8	28.3%	34.8%	26.1%	10.9%
15	参考文献	1.26	19.6%	21.7%	23.9%	34.8%

スコア平均 2.5 以上の項目は、項目 1「テーマ設定の動機がはっきりしており、議論すべき必要性を感じるか」と、項目 2「基本用語について、本レポートでの意味(使用法)まで書かれているか」であった。一方、スコア平均 1.5 以下の項目は、項目 15「参考文献が正しく書けているか」であった。

評価表 B の各項目点数の採点者割合結果は以下の通りである。

表8 評価表B 各項目点数の採点者割合

No.	項目（文末に全文）	スコア平均	すばらしい (3点)	まあまあ (2点)	発展途上 (1点)	点数なし (0点)
1	タイトル	2.5	56.8%	36.4%	6.8%	0%
2	テーマの関連性と目的	2.48	52.3%	43.2%	4.5%	0%
3	定義	2.48	56.8%	34.1%	9.1%	0%
4	情報の整理と課題の特定	2.2	27.3%	65.9%	6.8%	0%
5	本論の観点の適切さ	2.2	38.6%	43.2%	18.2%	0%
6	自分の言葉によるまとめ	1.82	34.1%	34.1%	11.4%	20.5%
7	提案・意見の提示	2.25	45.5%	34.1%	20.5%	0%
8	構成・段落・文の長さ	1.77	6.8%	63.6%	29.5%	0%
9	書式（フォント・文字数など）	2.02	6.8%	88.6%	4.5%	0%
10	資料の信頼性・数・適切な使用	2.25	38.6%	47.7%	13.6%	0%
11	引用表現	2.39	52.3%	34.1%	13.6%	0%
12	参考文献	1.82	15.9%	50%	34.1%	0%

スコア平均2.5以上の項目は、項目1「タイトルが議論点を明確に示しており、レポートの内容はタイトルと一致しているか」であった。一方、スコア平均1.5以下の項目はなかった。最もスコア平均の低い項目は、項目8の「見本レポート通りの構成・段落で、読みやすい文章の長さで書かれているか」で1.77であった。

評価表Aと評価表Bの点数の割合をみると、0%のチェック欄に大きな差がある。評価表Aでは、15項目のうち6項目に0点をつけた採点者が存在する。一方で、評価表Bは、12項目のうち1項目「本論（1点目と2点目）からわかったこと・言いたいことを自分の言葉でまとめているか」（表3参照）のみである。またその1項目も20.5%の評価者が0点をつけており、単なる厳しさの個人差ではないことが窺える。評価表Bのルーブリック項目6の点数なし（0点）は、「本論のまとめ（これらの資料からわかることは...）が書かれていない」がチェック項目になっている。約2割の「書かれてない」と判断した採点者は、「これらの資料からわかることは...」という、資料を自らの言葉でまとめ直す表現を手がかりに判断していた可能性が高いと考えられる。

その他項目において0%が選ばれなかった理由として、評価表Bの「参考文献」（表3参照）を例に挙げると、点数なし（0点）は「参考文献が書かれていない」であり、0点をつける基準が明確であったためであると考えられる。この0点に該当する記述は、ルーブリックを改善するにあたり、複数の科目担当者からの「そもそも授業の到達点となるレポート見本において、（授業で指導してきた）書くべきことが書かれていない場合の採点ができない」「不十分ではあるが自分なりに書いている学生と、全く書いていない学生とは区別したい」という声を受けて導入されたものであり、教員間の採点のズレも改善できた。一方で、評価表Bのように記述を具体的にしてもなお「すばらしい（3点）」は15.9%、「まあまあ（2点）」は50%、「発展途上（1

点)」は34.1%であり、評価点数の差が解消されるわけではない。このような場合、評価を行ったあとの評価者同士の話し合いにおいて、評価の根拠を話し合う機会を設けることが大切だろう。

次に、どのような項目において高い点数/低い点数がつけられる傾向にあったか、また傾向に差があったかを見ていく。評価表Aと評価表Bは、ルーブリックを作成する段階で項目を作成し直したため全ての項目が何かしらの項目に対応しているわけではない。今回の評価表Aと評価表Bの項目点数を比較するにあたり、第一筆者と第二筆者で対応する項目を話し合い、表9のようにまとめた。評価表Bでは、詳細記述に複数の内容が含まれているため、評価表Bの7に詳述されている観点は、評価表Aでは3つの項目(項目6、7、8)に関わっているということも生じている。また、評価表Bは、何年かの実践を経て改良されてきたものであるため新規項目が含まれる(項目5,9)。例えば、評価表Bの5については、より本論の構成を具体的にチェックできるような記述を加えて新規項目として設置されたものである。よって、単純に比較できるものではないが、評価表Bは、点数を選ぶだけでなく点数根拠となる記述にチェックを入れる形式になっていることから、より詳しいデータも参照しながら検討したい。

表9 評価表Aと評価表B 各項目点数の採点者割合

評価表A					評価表B						
No.	項目	3点	2点	1点	0点	No.	項目	3点	2点	1点	0点
1	テーマと興味理由	73.9%	26.1%	0%	0%	2	テーマの関連性と目的	52.3%	43.2%	4.5%	0%
2	定義	65.2%	26.1%	8.7%	0%	3	定義	56.8%	34.1%	9.1%	0%
3	背景整理	32.6%	47.8%	17.4%	2.2%	4	情報の整理と課題の特定	27.3%	65.9%	6.8%	0%
4	一読理解	19.6%	43.5%	37.0%	0%						
5	自分の言葉によるまとめ	34.8%	43.5%	21.7%	0%	6	自分の言葉によるまとめ	34.1%	34.1%	11.4%	20.5%
6	課題と提案の提示	19.6%	52.2%	28.3%	0%	7	提案・意見の提示	45.5%	34.1%	20.5%	0%
7	一貫性	32.6%	37.0%	26.1%	4.3%	7	提案・意見の提示	45.5%	34.1%	20.5%	0%
8	書き手の考え	37.0%	43.5%	19.6%	0%	7	提案・意見の提示	45.5%	34.1%	20.5%	0%
9	資料の信頼性	45.7%	41.3%	13.0%	0%	10	資料の信頼性・数・適切な使用	38.6%	47.7%	13.6%	0%
10	タイトル	69.6%	19.6%	10.9%	0%	1	タイトル	56.8%	36.4%	6.8%	0%
11	言葉遣い	60.9%	30.4%	6.5%	2.2%	8	構成・段落・文の長さ	6.8%	63.6%	29.5%	0%
12	学んだ表現の使用	30.4%	45.7%	19.6%	4.3%	8	構成・段落・文の長さ	6.8%	63.6%	29.5%	0%
13	読み手にわかる言葉の選択	43.5%	39.1%	17.4%	0%						

14	引用表現	28.3%	34.8%	26.1%	10.9%	11	引用表現	52.3%	34.1%	13.6%	0%
15	参考文献	19.6%	21.7%	23.9%	34.8%	12	参考文献	15.9%	50%	34.1%	0%

(ゴシック表記は最も割合の高い点)

前述した「参考文献」に似た傾向として「引用表現」も、評価表 B では 0 点の記述がはつきりしていることで低い点数の評価差は抑えられている。一方で、評価表 B では約半数が「素晴らしい (3 点)」をつけている。おそらくこれは「～によると、～という」という引用表現に関する記述が評価表内にあり、その表現をレポート内で見つけられれば「素晴らしい (3 点)」と判断している採点者が一定数いるためであると考えられる。評価表 A は、「引用表現が正しく使えているか」という記述のみで、具体的な引用表現は評価表内には登場しない。「引用表現」の採点データを詳しくみると、評価表 B において「まあまあ (2 点)」の採点者は、「一部引用表現 (～によると、～という) に不備がある」という観点にチェックを入れた人物は 1 名のみで、「出典情報が不十分だったり冗長だったりする (著者・組織名が冗長、出版年なし等)」が 14 名、「引用情報が冗長だったり、不十分だったりする (長すぎる、簡潔すぎる)」が 5 名であった。つまり、被採点レポートは、「～によると、～という」という表現は使用しており、表現のみを判断の手がかりにしている場合と、それ以外の項目を見ている場合とで判断が分かれたものと思われる。

また、評価表 A の 11 「全体を通してレポート用の言葉遣い (である・だ) になっているか」は、評価表 B の「見本レポート通りの構成・段落で、読みやすい文章の長さで書かれているか」という項目にあたるが、この項目では多くの下位記述がある。「構成バランスがやや悪い」「段落の一字下げ」「長すぎる文章/ねじれ文/話し言葉/目的語がない文がある」「文末の不統一箇所がある」「誤字/脱字がある」である。評価表 A の 11 は、「文末の不統一箇所がある」というチェック項目に該当すると考えられるが、この項目にチェックを入れた人物は 2 名しかおらず、逆に言えば、被採点レポートでは文末は統一できていたと考えられるため評価表 A の 11 の◎ (3 点) 割合が高いことと矛盾しない。評価表 B の同項目で多くチェックされていた下位記述は、「構成バランスがやや悪い (15)」、「段落の一字下げ (28)」、「長すぎる文章/ねじれ文/話し言葉/目的語がない文がある (16)」であった。このように記述詳細を見ていくと、両評価表で大きく採点傾向が異なるとは判断できなかった。全体として、タイトルの付け方やテーマに興味を持った理由などは、何かしらの理由が書かれていれば良い得点をつけやすい項目であることが窺えた。また、定義も、情報 (ことばの意味) を調べた上で定型表現が使用されていれば他の表現・形式よりは「できている」と判断されやすい。

一方で、表現・形式は、採点者が誤用やミスを発見すれば減点の根拠が明らかであるため厳しい点数をつけやすい。評価表 B には上述したような下位項目があるが、評価表 A には具体的な項目が挙げられていないため、減点の根拠となった箇所が評価表からはわからない。学習者自身が評価表を受け取った際に、どの記述が減点の根拠となっているか理解できているかを確認する必要があるだろう。

また、「自分のことばでまとめているか (評価表 A の 5、評価表 B の 6)」というような、解釈に関わる、表現・形式と内容の中間に位置するような項目の場合は、自分の言葉でまとめる

ための表現を評価基準にする採点者と、解釈を評価基準にする採点者として採点結果が分かれる傾向にあることが窺えた。表現や形式についてであれば、定型となっている表現を入れるべきだという気づきを与えるのは、評価表 B のように詳述してあれば可能かもしれない。しかしながら、解釈や内容（一貫性など）については、チェックする形式の評価表のみでは、評価者が被評価者にそのポイントを伝えることは難しい。そこで、本調査で収集した「レポートをより良くするためのアドバイス」の相違点を明らかにすることで、評価表とアドバイスの双方を授業活動に組み込む際のヒントを得たい。

6.3 アドバイス記述の分析

アドバイス記述には、内容や形式など多くの観点に関するコメントが含まれていたため、読点ごとに区切って一文単位で分類することにした。グループ 1（評価表 A）からは、139 文抽出できた。グループ 2（評価表 B）からは、189 文抽出できた。アドバイス文字数はグループ 2 の方が短い割合が高かった（表 5）ことから、グループ 2 のほうが抽出文が多いことは一見矛盾するが、401 文字以上のグループは文字数に幅があり、グループ 2 には、500 字後半～600 字半ばのアドバイスを行った協力者が 2 名いたためこのような文数となった。

一つの文は 2 つの観点からカテゴリー付けを行った。一つ目の観点は、表現・形式や内容など、何に対するアドバイスであるかという観点で、「分類」とした。もう一つの観点は、改善案を提示しているのか、問題を指摘しているのかなど、そのアドバイスで何をしているのかという観点で、「行為」とした。

コメント例をもとに説明すると、例えば、「内閣府ゲーム調査から自分の考えを書いているところで“私は”を文章の中で二回使用していた」という記述は、文章内の表現について述べているため「分類」は「表現・形式」とした。そして、この文は問題点について取り上げてはいるが、どのように改善すべきかについてまでは述べていないため「行為」は、「指摘」とした。「行為」の「具体提案」については、アドバイスをもらった相手が具体的にどのように改善するようにアドバイスされているのかがわかるようなもののみを対象とした。例えば、「話し言葉が混ざっているので直したほうが良い」という記述は、何が話し言葉に該当し、どのように直すべきかがすぐに認識できない可能性がある。このような記述は「指摘」に分類した。「～は話し言葉なので、...という表現に直したほうが良い」という記述は、次に取る行動が選択肢として具体的にイメージできるため「具体提案」とした。以下、アドバイス記述をカテゴリー別に分ける際の、「分類」と「行為」についてアドバイス例とともにまとめた。「分類」を表 10、「行為」を表 11 に示す。

表 10 アドバイス記述「分類」のカテゴリーと例

カテゴリー	カテゴリー内容	アドバイス例
資料	資料の信頼性や数、新旧に関するコメント	・ 信憑性のある機関をより使用した方が情報に確実性があると感じた
内容	一貫性や説得力、テーマに関するコメント	・ 本論の資料からまとめが飛躍しすぎていると感じた ・ 自分の考えをにかけているのは良いと思うが

		もう少し資料などからわかる事実を加えた方がよいと思う
表現・形式	一字下げ、フォント、参考文献、言い回しなどに関するコメント	・話し言葉を使わないように注意したほうが良い ・「考える」の多用が気になる
構成	構成や段落のバランスに関するコメント	・理由を述べる部分は段落を変えた方がわかりやすい ・最後の段落の自分の考えや提案を述べる部分が短い
その他	上記に該当しないコメント	・以下の点に気をつけるべきだと思う、など

表 11 アドバイス記述「行為」のカテゴリーと例

カテゴリー	カテゴリー内容	アドバイス例 (括弧内は表4の分類)
指摘	問題点を挙げているコメント	・論文では疑問文を使用するときに？を語尾につけないほうが良い (表現・形式)
箇所指定指摘	箇所を指定して問題点を挙げているコメント	・後半の自分の意見 (では、...を変えるには～の部分) が弱く、テーマとずれていると感じた (内容)
指摘理由	問題だと思ふ理由を述べて問題点を挙げているコメント	・「そう」といった指示語が多すぎるため何を指しているのかわからない (表現・形式)
箇所指定指摘理由	箇所を指定して問題だと思ふ理由を述べて問題点を挙げているコメント	・理由を述べていたはずであるにもかかわらず、まとめの段落では「一例」となっていた点が読みづらいと感じました (内容)
具体提案	どのように改善すべきか案を挙げているコメント	・「と私は考える」という形式が多すぎるため、だ・である調で断定していった方がよい (表現・形式)
箇所指定具体提案	箇所を指定してどのように改善すべきか案を挙げているコメント	・7段落目に、(例)「子供が自尊感情を満たすためにゲームをする」という記述があるが、(例)を例えば、などの言葉に直した方が適切 (表現・形式)
承認	良い点を述べたり、賛同・共感を示したりしているコメント	・気になる点は少しあるものの論点や意見が分かりやすく素晴らしいレポートでした (内容)
箇所指定承認	箇所を指定して良い点を述べたり、賛同・共感を示したりしているコメント	・第7段落の勉強をする時間を設定するという主張が良かった (内容)
異なる視点の提示	レポート内には書かれていない異なる視点を提示しているコメント	・たとえば制限される側はどう考えているのか、ゲームが与える悪影響は本当に一つもないのか、など思い込んでしまっている部分をあえて疑ってみたりするとより議論が深まるのではないかと思った (内容)
励まし	これからのことについて応援するコメント	・この点を改善したら、さらに輝くレポートになると思う (内容)
その他	上記に該当しないコメント	・以下の点に気をつけるべきだと思う、など

本分類をもとにまずは第一筆者が振り分け、ランダムに記述を抽出し、第二筆者が振り分けた。一致率は、87%であった。一致しなかった項目については、すり合わせを行った。その結果、参考文献に関する記述を第一筆者が「箇所指定」としてカウントしていなかったため、そ

の点は「箇所指定」として振り分け直した。若干の不一致箇所はあったものの、概ね一致していたことから、参考文献について訂正した第一筆者のデータを採用し分析対象とした。本分類をもとに振り分けた結果は、表 12 の通りである。

表 12 評価表 A と評価表 B のアドバイス記述における「分類」と「行為」の出現割合

評価表 A 対象 139 文：分析単位数 (1 文)

分類	資料	内容	表現・形式	構成	その他					
	8	58	66	4	3					
割合	5.8	41.7	47.5	2.9	2.2					
行為	指摘		指摘理由		具体提案		承認		異なる視点の提示	その他
	71		16		32		3		14	3
	51.1		11.5		23.0		2.2		10.1	2.2
	指摘	指定指摘	指摘理由	指定指摘理由	具体提案	指定提案	承認	指定承認		
	58	13	16	0	20	12	3	0		
各行為内割合	81.7	18.3	100	0	62.5	37.5	100	0		
全体割合	41.7	9.4	11.5	0	14.4	8.6	2.2	0		

※箇所指定は指定に省略 (例：箇所指定指摘→指定指摘)、網掛けのセルの単位は%

評価表 B 対象 187 文：分析単位数 (1 文)

分類	資料	内容	表現・形式	構成	その他						
	7	74	101	3	2						
割合 (%)	3.7	39.6	54.0	1.6	1.1						
行為	指摘		指摘理由		具体提案		承認		異なる視点の提示	励まし	その他
	86		28		48		12		8	3	2
	46.0		15.0		25.7		6.4		4.3	1.6	1.1
	指摘	指定指摘	指摘理由	指定指摘理由	具体提案	指定提案	承認	指定承認			
	61	25	19	9	28	20	11	1			
各行為内割合	70.9	29.1	67.9	32.1	58.3	41.7	91.7	8.3			
全体割合	32.6	13.4	10.2	4.8	15.0	10.7	5.9	0.5			

※箇所指定は指定に省略 (例：箇所指定指摘→指定指摘)、網掛けのセルの単位は%

表 4 で示した通り、評価表 B のグループは評価表 A のグループほどアドバイスに時間を割いていない可能性があるが、結果から見ると、アドバイスの傾向に両グループに大きな差異はなかった。まず、両グループともに、「表現・形式」「内容」に関する言及が多く、それぞれ合わせて 9 割を占める。「資料」や「構成」についての言及は少なかった。授業実践において、

「資料」や「構成」について学習者自身にそれらに気づく機会・改善できる力を得てほしいと考えている場合は、お互いに読ませアドバイスを行わせるだけでは不十分である可能性が高いといえる。多くの時間が「表現・形式」「内容」に関するアドバイスに割かれているだろう。例えば、「構成」であれば、「全体のバランス（極端に少ない段落がないかなど）や内容を提示する順番を見る」など、観点を絞ってチェックする機会を設けたりする必要があるが、上田・半田(2022)にもあったようにアドバイスに取り入れる難易度が高いゆえに言及されない場合は、難易度を軽減させるような工夫も求められる。

次に「行為」についての結果を見ると、両グループともに多いのは「指摘」であった。評価表 Bの方が、「指摘」「指摘理由」「具体提案」「承認」のいずれの行為においても「箇所指定」の割合が評価表 Aよりも高かった。評価表 Bでは、チェック項目が見本レポートの何段落目にあたるのかが明記されており箇所を明示しやすかった可能性もある。

富永(2011)では、ピア後に意見歓迎指向とフレンドリー指向が高くなり、高能力要求指向が低くなった人物の会話構成は、【進行】【問題点】【修正案】【気づき】で構成されており、一方、意見歓迎指向とフレンドリー指向が低くなり、高能力要求指向が高くなった学習者のやり取りの特徴では、【問題点】を指摘されるが、具体的な【修正案】が検討されないまま次に進んでしまうことが多かったことが報告されている。

今回の調査結果では、両グループ共に「指摘」の割合が最も多く、学習者は「問題点を指摘すること」は、アドバイスとして多く行っていることが窺えた。また、「指摘理由」の割合も両グループともに11.5%（評価表 A）、15%（評価表 B）と少なく、なぜその箇所が問題だと感じたかはあまり伝えられていない。富永(2011)の結果から見ると、【修正案】を相手に提示することは、ピア活動の雰囲気をより良いものにし意見を相互に受け止めやすくする際の鍵となりそうである。評価表 Bグループにより多く見られた「指摘箇所」の明示は、修正案の提示とまではいかないが、問題の箇所を特定できることはアドバイスを受けた人物が改善のアクションに入りやすい可能性がある。【修正案】にあたる、「具体提案」については、評価表 Aのグループが23.0%、評価表 Bのグループが25.7%であった。このことから「改善案の提示(具体提案)」はアドバイスに組み込む要素として十分に意識されているとは言い難い。そして、評価表 Aのグループと評価表 Bのグループにおける「具体提案」の出現割合に差がみられなかったところを見ると、相互アドバイスにおける「具体提案」の提示は評価表の詳述化のみでは促せない。評価表とアドバイス活動をどのように結びつけるかを指示する、あるいはどのような内容をどのように伝えるかなど「アドバイスをする際のポイント」を指示するなど新たな仕掛けが必要であるといえる。

7. まとめと今後の課題

本調査の課題は、①詳述度の異なる評価表で採点した際、採点傾向にどのような差異が生じるか、②詳述度の異なる評価表で採点した際、アドバイス傾向にどのような差異が生じるか、の2点であった。その結果明らかになったこと、そこから考えるクラス運営について記述したい。

まず、評価表の具体記述により、評価傾向のコントロールは可能であり、「点数なし」など、極端な点数差による評価のバラつきは軽減できると考えられる。評価表は、求める課題だけでなく、カリキュラムの形態（1人の教員が1つのクラスで完結するか、複数の教員が複数クラスを持つのかなど）にも関係しているため、必ずしも評価表Bほど細かく記述された評価表が求められる場面ばかりではない。しかしながら、評価表Aのような記述のものを扱う場合も、点数そのものを評価活動に利用する場合は、どのような場合が何点かを補足することは必要であると考えられる。今回の調査で評価表Bの0点の採点割合がコントロールされていたことを考えると、学習者と教員の極端な採点ギャップを埋める記述としても参考になるだろう。

一方で、評価表Aも評価表Bも、評価にバラつきは存在した。つまり、授業内で学生がそれぞれ行う評価のバラつきを評価記述により解消することは難しい。また、そうしたバラつきが学習者同士の評価活動自体への不信感を招くことも指摘されていることから、点数への意識を軽減させること（点数だけに注目しないよう伝える、チェックはするが点数は出さない・合計しない）ことや、点数を出す場合は、複数名で評価した平均点を被評価者にはフィードバックするなどの工夫も考えられる。

また、評価表に書いてある表現を採点の手がかりとしていることも窺えた。一方で、表現のみに手がかりを求める学習者は、表現のみで判断し高得点をつける可能性もあることから、表現以外の観点を意識させることも必要である。

次に、アドバイスについては、「表現・形式」や「内容」の「指摘」に、「箇所指定」や「具体提案」を追加する活動を組み込むと良いだろう。本調査から、「表現・形式」や「内容」についてのアドバイスは多くの学習者が目を向ける場所であることが示唆された。それらの問題点を「指摘」することは多く行なっているが、学習者同士のアドバイスにおいてアドバイスの受容度を高める「箇所指定」や「具体提案」は少なかった。アドバイスを受ける本人が気がつく能力も大切であるが、学習者同士が互いにチェックを行う際は、アドバイス行為そのものが円滑に進むような仕掛けも大切である。毎回全ての箇所ではなくとも、「3つ改善場所を指摘してもらって、あとは自分で探す」といったタスクも考えられる。また、アドバイスをする側が、「箇所指定」や「具体提案」を意識し言語化することで、観察眼が養われる可能性もある。また、「構成」や「資料」については言及されづらい傾向にあるため、観点を決めて全体で共有し、相互アドバイスを行う際にも言及されるよう、複数の観点から見る力を段階的に養っていく必要があるだろう。

評価活動で使用する評価表の複雑さやアドバイス活動は、全体の時間や負荷も考慮する必要がある。評価表Bのような評価表自体の評価負荷が高い場合は、まずは共通のレポートをクラス全体で評価してみて評価表そのものの理解をクラスで順を追って確認し、スムーズに評価できるようになってから（評価負荷を下げてから）評価活動に入るといった工夫も必要である。

最後に、本調査の課題を述べたい。今回の調査は、評価活動とアドバイスを同時に記述した際の結果である。評価活動に負荷がかかったり時間の制限がある場合、評価活動とアドバイスや意見交換のタイミングが異なることも考えられる。また、評価表は交換せず、アドバイスのみを口頭で行うような活動も考えられる。その際は、評価表Aと評価表Bの結果には今回に見られない差が生じることも予想される。評価表Bは評価表Aと比較し、「箇所指定」コメント

の割合が高かったことから、なぜその点数をつけたのかを具体的な箇所に結びつけて判断している可能性がある。一方で、評価表 A では時間が経つにつれてどのような理由でその箇所にチェックをしたのか理由が曖昧になってしまう可能性もある。このように、評価表を使用するタイミングと口頭でのアドバイスのタイミングが離れているときの結果も調査する必要があるだろう。また、評価表を見たり提示したりしながらアドバイスを行うか否かという条件の違いも調査すると、それぞれの状況下における評価活動・アドバイス活動の特徴が明らかになり、評価活動を組み立てる際に役立つであろう。

また、アドバイスには「異なる視点の提示」という、レポートには書かれていない視点についてのコメントが見られた。このようなレポート執筆者にない視点の提供は、レポートの内容そのものの吟味を促す可能性がある。本調査では、アドバイスのほかに、「気がついたこと、もっと知りたいと感じたこと」も別途記述してもらっている。そちらは内容を批判的に検討するためのコメントとして期待して設けたものである。このような指示でどのようなコメントが出てきたのか、改善のためのアドバイスとは別の特徴があるのかを調査し、レポートの内容を批判的に見る力の育成における課題も併せて探りたい。

付記

本研究は、科研費若手研究「初年次アカデミック・ライティングプログラムにおける効果的な評価システムの構築」（研究代表者：藤浦五月 研究課題番号 18K13070）の助成を受けて行われた。

注

1) 石川・阿部・川崎（2022）では、「ピアレスポンス」としている。

参考文献

- 石川奈保子・阿部真由美・川崎弥生（2022）「大学授業のレポート課題における再提出方式とフィードバック方法の検討」『日本教育工学会研究報告集』1, 127-134.
- 上田勇仁・半田純子（2022）「プロジェクト学習における内省支援の実践と評価—解釈と分析を促す記述指示が抽象的概念化に与える影響—」『日本教育工学会論文誌』46(3), 543-556.
- 宇野聖子・藤浦五月（2016）『大学生のための表現力トレーニング あしか アイデアをもって社会について考える（レポート・論文編）』ココ出版
- 田中耕治（2003）『教育評価の未来を拓く—目標に準拠した評価の現状・課題・展望—』ミネルヴァ書房
- 富永敦子（2011）「ピア・レスポンス指向性と会話との関連」『第53回日本教育心理学会総会発表論文集』263.
- 富永敦子（2012）「文章表現授業における大学生のピア・レスポンス指向性の変化と要因の分析」『日本教育工学会論文誌』36(3), 301-311.
- 藤浦五月（2021）「初年次レポート・ライティングプログラムにおけるルーブリックの改善」『Global studies』5, 125-146.
- 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.

Cho, K., Schunn, C. D., & Wilson, R. (2006) Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives. *Journal of Educational Psychology*, 98(4), 891-901.

Wang, W. (2014) Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing*, 19, 80-96.

藤浦 五月 武蔵野大学グローバル学部日本語コミュニケーション学科 准教授

宇野 聖子 京都大学国際高等教育院 非常勤講師

資料：評価表 B

名前()さん		
1	タイトル	タイトルが議論点を明確に示しており、レポートの内容はタイトルと一致しているか
評価ポイント	a. 素晴らしい(3)	<input type="checkbox"/> 議論点がタイトルから明確にわかり、レポート内容とも一致している
	b. まあまあ(2)	<input type="checkbox"/> 議論点がタイトルからわかりづらい <input type="checkbox"/> レポートとタイトルの内容が少しずれている
	c. 発展途上(1)	<input type="checkbox"/> タイトルからテーマや議論点がわからない <input type="checkbox"/> タイトルとレポートの内容が大きくずれている (一箇所でもチェックがあればこちら)
	d. 点数なし(0)	<input type="checkbox"/> タイトルが書かれていない
2	第一段落 テーマと目的	自分とテーマの関連性と、本レポートの目的がわかりやすく書かれているか
評価ポイント	a. 素晴らしい(3)	<input type="checkbox"/> テーマ設定が個人的な疑問・興味関心から出発しており、具体的に書かれている。よくある話題ではなく、少し深いテーマを掘り下げており、レポートの目的も一文で明示されている。
	b. まあまあ(2)	<input type="checkbox"/> テーマと自分との関連性がわかりづらい <input type="checkbox"/> テーマ設定がよくある話題・広すぎる話題でありもう少し掘り下げる必要がある <input type="checkbox"/> テーマ設定が狭すぎるため、もう少し関連領域や同じような事例にも広げて設定すると良い
	c. 発展途上(1)	<input type="checkbox"/> テーマ設定と自分との関連性がほとんどわからない <input type="checkbox"/> レポートの目的が明確に書かれていない。 (一箇所でもチェックがあればこちら)
	d. 点数なし(0)	<input type="checkbox"/> テーマを選んだ理由が書かれていない <input type="checkbox"/> レポートの目的が書かれていない (一箇所でもチェックがあればこちら)
3	第二段落 定義	レポートを読むために重要な用語について、信頼できる資料をもとに意味を整理し自分のレポートでの用い方を明示しているか
評価ポイント	a. 素晴らしい(3)	<input type="checkbox"/> レポートを読む上で読み手が理解した方が良い重要な用語を、信頼できる資料から調べられている。ただ意味を調べるだけでなく、自分のレポートでどのように用いるかを明示している。
	b. まあまあ(2)	<input type="checkbox"/> レポートを読む上で重要な用語を信頼できる資料をもとに調べられているが、以下の点においてやや不十分である <input type="checkbox"/> 他の用語も調べる・紹介する必要がある、あるいは用語の説明が不十分 <input type="checkbox"/> 重要な用語は調べられているが、不要な用語も書かれている <input type="checkbox"/> 提示された用語の場合、異なる資料(専門事典等)を用いた方が良い <input type="checkbox"/> 自分のレポートでどのような意味として用いるのかやや不明瞭である
	c. 発展途上(1)	<input type="checkbox"/> 以下の点において不十分である <input type="checkbox"/> 重要な用語とは言えない(この用語より説明すべき用語がある) <input type="checkbox"/> 信頼できる資料ではない <input type="checkbox"/> 自分のレポートでどのような意味として用いるのか示していない <input type="checkbox"/> 用語の定義・説明以外のものが書かれている (一箇所でもチェックがあればこちら)
	d. 点数なし(0)	<input type="checkbox"/> 定義も自分のレポートでの用語の用い方も書かれていない
4	第三段落 背景整理	議論をする上で必要な情報(数値等)を整理し、より深く調べるべき課題を特定できているか
評価ポイント	a. 素晴らしい(3)	<input type="checkbox"/> 議論をする上で必要な情報(数値・データなど)を調べられており、その情報からさらに課題を深く考え、レポートの目的とともに本論(一点目・二点目)につなげている。
	b. まあまあ(2)	<input type="checkbox"/> 背景説明に資料が使われているが、更に良い資料を探さなければならない <input type="checkbox"/> 背景の最後の論点提示で「何のために」「何について」のうち、どちらか一方しか書かれていない <input type="checkbox"/> 背景の最後に論点が提示されているが、背景から導き出された課題あるいは論点がわかりづらいなど、やや飛躍している
	c. 発展途上(1)	<input type="checkbox"/> 議論をする上で必要な情報(数値・データなど)を調べられていない。 <input type="checkbox"/> 背景説明に資料が使われていない <input type="checkbox"/> 背景資料から更なる課題が特定できていない <input type="checkbox"/> 本論に行く前に論点が提示されていない (一箇所でもチェックがあればこちら)
	d. 点数なし(0)	<input type="checkbox"/> 背景説明が書かれていない

5	第四・五 段落	本論と 主張	本論の1点目と2点目が主張を支えるための観点として適切か			
			a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント			<input type="checkbox"/> 論点の一点目と二点目が主張を支える観点として十分であり、資料もそれぞれの主張を支えるものとして適切である。	<input type="checkbox"/> 論点の一点目と二点目のどちらかが主張を支えるための観点としてやや説得力に欠ける <input type="checkbox"/> 論点と引用資料が合っていない、箇所がある <input type="checkbox"/> 一点目・二点目のどちらか一方しか資料解釈(自分の言葉による資料まとめ)ができていない <input type="checkbox"/> 論点の一点目と二点目が似通っている	<input type="checkbox"/> 本論が一点しかない <input type="checkbox"/> 論点を支えるための主張としてどちらも説得力に欠ける <input type="checkbox"/> 論点と引用資料が全く合っていない <input type="checkbox"/> 資料解釈(自分の言葉による資料まとめ)がない <input type="checkbox"/> 論点の一点目と二点目の観点が本論目的からズレている 例:一点目は理由・二点目は対策、一点目は日本の事例・二点目は海外事例(比較)になっている、リット・デリットなどを並べているなど (一箇所でもチェックがあればこちら)	<input type="checkbox"/> 本論が書かれていない
6	第六 段落	本論のま とめ	本論(1点目と2点目)からわかったこと・言いたいことを自分の言葉でまとめているか			
			a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント			<input type="checkbox"/> 本論で資料を引用するだけでなく自分の考えを自分の言葉でまとめている。一点目の資料と二点目の資料から言えることを総合的に自分の言葉でまとめている。	<input type="checkbox"/> 本論で資料を引用するだけでなく自分の考えを自分の言葉でまとめているが、以下の点において不備がある <input type="checkbox"/> 一点目と二点目の資料を総合的にまとめられていない(どちらか片方になっている) <input type="checkbox"/> 資料とまとめの解釈にややズレがある	<input type="checkbox"/> 引用表現の繰り返して自分の言葉でまとめられていない <input type="checkbox"/> 資料と解釈がほとんど合っていない (一箇所でもチェックがあればこちら)	<input type="checkbox"/> 本論のまとめ(これらの資料からわかることは…)が書かれていない
7	第七 段落	考察	レポートの目的および背景・本論1・本論2で書かれたことをふまえて、具体的な提案や意見が述べられているか			
			a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント			<input type="checkbox"/> 考察が、背景のまとめで提示された目的と一致している。また、本論1・本論2で述べたことを反映して、より詳細・具体的な意見や実現が見込める提案が提示されているため、全体を通して一貫性が感じられる。	<input type="checkbox"/> 考察は書かれているが、以下の点においてやや不十分である。 <input type="checkbox"/> レポートの目的と考察がズレている。 <input type="checkbox"/> 本論1・本論2で述べられた内容が一部しか反映されていない <input type="checkbox"/> 提案が一般的に既に知られている内容で、本レポートで調べなくても述べられるような内容になっている。具体性に欠ける(やや足りない)	<input type="checkbox"/> 考察は書かれているが、以下の点において不十分である。 <input type="checkbox"/> 提案や意見がレポートの目的と合っていない・大きくずれている <input type="checkbox"/> 本論1・本論2で述べられた内容が全く反映されていない (一箇所でもチェックがあればこちら)	<input type="checkbox"/> 考察が書かれていない
8	読みやす さ		見本レポート通りの構成・段落で、読みやすい文章の長さで書かれているか。			
			a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント			<input type="checkbox"/> 見本通りの構成・段落で書かれており、文章も簡潔で読みやすく書かれている	<input type="checkbox"/> 概ね見本通りだが、以下の点に修正が必要である。 <input type="checkbox"/> 構成バランスがやや悪い <input type="checkbox"/> 段落の一字下げ <input type="checkbox"/> 長すぎる文章/おじれ文/話し言葉/目的語がない文がある <input type="checkbox"/> 文末の不統一箇所がある <input type="checkbox"/> 誤字/脱字がある (不備:□のチェック数に関わらず全体で1-3箇所程度)	<input type="checkbox"/> 見本をよく見て下記の修正が必要である。 <input type="checkbox"/> 見本と大きく異なる構成、ナンバリングがない、段落バランスが悪い <input type="checkbox"/> 段落の一字下げ <input type="checkbox"/> 長すぎる文章/おじれ文/話し言葉/目的語がない文がある <input type="checkbox"/> 文末の不統一箇所がある <input type="checkbox"/> 誤字/脱字がある (不備:□のチェック数に関わらず全体で4箇所以上。見本と大きく異なる構成・ナンバリングがない場合は一箇所でもこちら)	<input type="checkbox"/> 段落・構成がない <input type="checkbox"/> アウトラインを貼り付ける、箇条書きになっているなど全体的にまとまった文章として書かれていない

9	書式ルール	指定された書式(フォント・文字の大きさ・行数・行間など)で、文字数を守って書かれているか			
		a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント		<input type="checkbox"/> 書式も文字数を守って書かれている(2,250-2,750)	<input type="checkbox"/> 書式に一部不備(1,2箇所)がある <input type="checkbox"/> 文字数がやや多い(2,750-3,000)、やや少ない(2,000-2,250)	<input type="checkbox"/> 書式に不備がある(3箇所以上) <input type="checkbox"/> 文字数が多い(3,000字以上)、少ない(1,999字以下)(一箇所でもチェックがあればこちら)	<input type="checkbox"/> 書式に重大な不備がある(コピーして貼り付けた形跡がある) <input type="checkbox"/> 文字数が著しく少ない(1,000字に満たない)・文字数不明(記載なし)
10	資料	信頼できる多様な資料を調べ、レポートの中で適切に使用できているか			
		a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0) / 不正行為(厳罰)
評価ポイント		<input type="checkbox"/> 複数資料を多様な観点から調べられており、主張と資料も一致し説得力がある	<input type="checkbox"/> 複数資料は調べられているが似たような資料(発信者)が多い <input type="checkbox"/> 資料は集められているが検索しやすい・アクセスしやすい資料のみ利用している <input type="checkbox"/> 毎年出される情報など、明らかに最新情報を調べた方が良い情報が最新のものではない <input type="checkbox"/> 資料は集められているが、主張と資料にズレがあるところがある	<input type="checkbox"/> 背景、本論ともに1つの資料や著者・発信元に頼っている(定義で1つ、背景・本論で1つの場合もこちら。背景・定義・本論1・本論2は別々の資料が望ましい) <input type="checkbox"/> 匿名の個人のブログなど信頼性に欠ける資料がある <input type="checkbox"/> 資料が少ない(1つの論点のみに資料を使うなども含む)(一箇所でもチェックがあればこちら)	<input type="checkbox"/> 背景・定義・本論1・本論2のいずれか1つでも資料を用いておらず、これまで得た情報について確認せずに感想文のように書いている(点数なし) <input type="checkbox"/> 引用表現を用いずに資料内容を書いている(不正行為)
11	引用表現と要約	引用表現を適切に使い、必要な情報を簡潔にまとめて引用しているか。			
		a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	不正行為(厳罰)
評価ポイント		<input type="checkbox"/> 引用表現を適切に使い、必要な情報を簡潔にまとめている	<input type="checkbox"/> 一部引用表現(～によると、～という)に不備がある <input type="checkbox"/> 出典情報が不十分だったり冗長だったりする(著者・組織名が冗長、出版年なし等) <input type="checkbox"/> 引用情報が冗長だったり、不十分だったりする(長すぎる、簡潔すぎる)(不備:□のチェック数に関わらず全体で2-3箇所)	<input type="checkbox"/> 引用表現(～によると、～という)に不備がある <input type="checkbox"/> 出典情報が不十分だったり冗長だったりするところが多い(著者・組織名が冗長、出版年なし等) <input type="checkbox"/> 引用情報が冗長だったり、不十分だったりする(長すぎる、簡潔すぎる)(不備:□のチェック数に関わらず全体で4箇所以上)	<input type="checkbox"/> 引用表現を用いずに資料内容を書いている <input type="checkbox"/> 引用表現を用いず部分的に変えたり切りはりして情報を使っている <input type="checkbox"/> 他者が書いたものを自分のものとして提出している
12	参考文献	参考文献は、必要な情報が指示通りの方法で統一されて書かれている			
		a. 素晴らしい(3)	b. まあまあ(2)	c. 発展途上(1)	d. 点数なし(0)
評価ポイント		<input type="checkbox"/> 参考文献は、必要な情報が指示通りの方法で統一されて書かれている	<input type="checkbox"/> 参考文献の情報に足りない部分がある(筆者・組織名、出版年、資料名、出版社、URL、参照日など) <input type="checkbox"/> 参考文献の書き方の不備がある(書式不統一、並べ順など)(不備:□のチェック数に関わらず全体で1-2箇所程度)	<input type="checkbox"/> 参考文献の情報に足りない部分がある(筆者・組織名、出版年、資料名、出版社、URL、参照日など) <input type="checkbox"/> 参考文献の書き方の不備がある(書式不統一、並べ順、アウトラインの表をコピーして貼るだけなど)(不備:□のチェック数に関わらず全体で3箇所以上。アウトラインコピーの場合はチェック数関係なく「発展途上」)	<input type="checkbox"/> 参考文献が書かれていない