

フィードバックの視点にもとづいた 児童の作文活動における自己評価の検討

長崎 耕作*・深見 俊崇**

Kosaku NAGASAKI・Toshitaka FUKAMI
A Consideration of Self-Assessment in Students' Writing Activities
Based on a Feedback Perspective

ABSTRACT

本研究では、児童の作文活動における自己評価のフィードバックのタイミングが自己評価の正確さにどのような効果や影響を与えるのかを明らかにすることを目的とした。初稿と完成稿で得た児童の記述と自己評価の数値を、自己評価の正確さの定義にもとづいた「自己評価-教師評価」の式によって分析し、文章作成時にフィードバックを受ける（即時フィードバック）群と文章作成時にフィードバックを受けない（遅延フィードバック）群で対応のないt検定を行った。その結果、フィードバックのタイミングの違いによって児童の自己評価の正確さに差異がないことが確認された。しかしながら、作文の完成稿についての自己評価の正確さには両者の間に小さな効果量（ $d=0.26$ ）が認められたことから、即時フィードバックを実施することが作文活動における児童の自己評価の正確さを高めるための一要因となる可能性があることが示唆された。

【キーワード：自己評価、遅延フィードバック、即時フィードバック、自己評価の正確さ、書くこと】

1. 問題と目的

近年、諸外国において自己評価は、学業成績との関係性で注目されており、自己評価が学習の成果にポジティブな影響を与えるという認識は一般的なものとなっている（Brown & Harris, 2013）。

こうした、自らの学習プロセスや成果物を調整・修正していく学習は、とりわけ作文（Writing）においてポジティブな結果が得られている（e.g., Andrade et al., 2010）。例えば、物語文と自己評価に関して調査を行ったRoss et al. (1999) は、物語文の自己評価の練習をした児童の方が、同程度の作文力を有する児童の成績を上回ったことを報告している。またAndrade et al. (2008) は、初稿段階の文章をループリックで自己評価することによって、その後の作文の質が向上したことを示している。

以上のような作文活動における自己評価の営みは、日本の文脈でも重要視されている（e.g., 富士原, 2015）。しかしながら、作文活動における自己評価の研究は、諸外国におけるその研究の蓄積と比較して極めて少ない。したがって、本研究では、諸外国における自己評価に関する研究の動向を整理したうえで、自己評価をめぐる課題を明らかにする。それをふまえ、日本の作文活動における自己評価の実施方法について検討していく。

2. 諸外国における自己評価の研究動向の整理

自己評価は、「あらかじめ設定された規準（Criteria）にもとづいて行われる、学習プロセスおよび最終成果物

の質的評価」(Panadero, 2011, p.78)と定義される。また、「児童・生徒が自分自身の学習プロセスや成果物の質について説明（例えば、アセスメント（Assess）し、場合によってはメリットや価値を割り当てる（例えば、エバリュエートする（evaluate））多種多様なメカニズムや技法」（Panadero et al., 2016, p.804）である。その自己評価の実施ならびに自己評価をとおして児童・生徒・学生が有益な結果を得るにあたっては様々な要素について考慮する必要がある。例えば、自己評価の有用性の認識を促すこと（e.g., Yan et al., 2023）や、自己評価を実施するための技法を具体的に構想すること（e.g., Panadero et al., 2016）、自己評価の正確さを高めること（e.g., Brown & Harris, 2013）などである。

そうした様々な要素が絡み合う自己評価について、Andrade (2019; 2010) は「自己評価とはフィードバックである」（p.2）と自己評価の目的を端的な言葉で表している。自己評価の目的はフィードバックの生成であり、学習のプロセスと成果物の調整・修正をするための情報を提供するものでなければならない（Andrade, 2019）。

Andradeが提示したこれらの視点は、他の先行研究にも一貫して確認できる。「学習のプロセスと成果物の調整・修正」について、Panadero & Alonso-Tapia (2013) は、ミスの発見と修正によってパフォーマンスの向上を目指すことが自己評価の目的であるとしている。また、Boud & Falchikov (1989) も、自己評価において最終的成果物のみに着目するのではなく、学習のプロセスにも着目することを重要視している。「フィードバックの

* 島根大学大学院教育学研究科教育実践開発専攻

** 島根大学教育学部小学校教育専攻

2024年8月30日受付

2025年2月19日受理

生成」に関して、Yan et al. (2023) は、タイムリーなフィードバックが自己評価の有用性の認識を促すための要素として重要であることを示している。また、Panadero et al. (2020) は、フィードバックを実施するタイミングによって生徒が用いる自己評価技法のレパートリーが変化するとしている。

自己評価研究においては、上述した学習のプロセスと成果物の両者の調整・修正やフィードバックについての議論と同様に、自己評価の正確さについての議論が活発である(Andrade, 2019)。自己評価における正確さとは、「児童・生徒・学生の自己評価と評価基準の近さ」(León et al., 2023, p.3)と定義される。すなわち、児童・生徒・学生が自らの成果物を評価したものと、教師等の他者が特定の評価基準にもとづいて成果物を評価したものととの差がどれだけ小さいかによって確認される。Dochy et al. (1999) は、自己評価のポイントの1つを、誤りを修正するために何が間違っていたかを適切に判断し、何が正しかったかを特定し、それらの行動を繰り返すことであると述べている。また、Panadero et al. (2016) は、自らの成果を正確に把握して、修正のための筋道を明らかにすることが自己評価において重要であるとしている。この点に関してはBoud & Falchikov (1989) も主張していることである。すなわち、テストの点数などの総合的な得点や評定を正確に予測するよりも、自分の成果物について何が間違っていて何が正しいのかを正確に把握できることの方が重要なのである。これらの主張からわかるとおり、自己評価における正確さは自己評価の目的である学習の調整・修正に必要不可欠となる。また、自己評価の正確さについてのメタ分析を行ったBrown & Harris (2013) は、自己評価の正確さと学業成績との間に中程度の相関関係があることを示した。このことから、自己評価の正確さの如何が、学業成績と大きく関わってくることが示唆される。

以上までに確認してきたように、自己評価におけるフィードバックと自己評価の正確さは、いずれも自己評価の根本的な目的である、学習プロセスや学習における成果物を調整・修正するために重要な役割を果たすものである。この2つの自己評価における要素は、自己評価の正確さについてメタ分析を行ったLeón et al. (2023)によって、より強固な関係性であることが明らかになった。彼らは、自己評価の正確さに影響を与え得る要因について分析したところ、自己評価の正確さを高める有意な要因は「フィードバック」と「教育レベル」のみであることを明らかにした。ここでの「教育レベル」とは、初等・中等・高等教育として表される学年段階のことであり、その段階が進むことによって自己評価の正確さが高まるという結果が得られている。しかしながら、「教育レベル」は自己評価の正確さにポジティブに影響し得る要因であるが、それをコントロールすることはできないものである。したがって、現在の知見では、「フィードバック」の実施方法について検討することが最も自己評価の正確さを高めるために必要な手立てであると言え

る。同様に、自己評価の正確さを高めるためのフィードバックの実施方法を検討することが自己評価の目的を遂行するために非常に重要な役割を果たすと言える。

その一方で、フィードバックを行うタイミングについての議論が盛んであり多様な見解も確認される。例えば、Panadero et al. (2020) は、即時フィードバック (immediate feedback) を行うことによって、自己評価者が用いる自己評価技法のレパートリーが限定的になるため、遅延フィードバック (delayed feedback) を行うことでその限定性が少なくなると主張している。一方、即時フィードバックを行うことで、修正情報が早く提供され、効率的な定着がもたらされる可能性が高いという主張もある (Phye & Andre, 1989)。

Shute (2008) は、こうした背反した議論を多面的な視点から考察し、次のような視座を得た。遅延フィードバックは、児童・生徒・学生の能動的な認知的・メタ認知的処理への関与を促すことや自律性・自己効力感の向上を生み出すことができる。その一方、学習に苦戦している児童・生徒・学生には学習の方向性を即座に示すことができないという点で有益な結果を生み出さないとしている。即時フィードバックは、結果と原因を明示的に示したり、学習への動機づけや学習活動の決定を促進したりすることができるとしている。しかしながら、児童・生徒・学生が単一の情報へ依存したり、それによって慎重さに欠ける判断を引き起こしたりしてしまう危険性もある。以上をふまえて、学業成績の高い児童・生徒・学生にとっては遅延フィードバックの方が有益であり、学業成績の低い児童・生徒・学生にとっては即時フィードバックの方が有益であると結論づけている。

これまで整理した諸外国の自己評価の研究動向にもとづいて、自己評価における正確さを高めるためのフィードバックの実施方法について検討することを本研究の主題とする。なお、本研究で扱うフィードバックの実施方法を検討する視点としては、Shute (2008) による議論をふまえてフィードバックを行うタイミングを設定する。この視点にもとづいてフィードバックのタイミングの違いが自己評価の正確さに与える影響について考察する。以下では、フィードバックのタイミングの違いによって生じる、自己評価の正確さへの影響の差異を調査するための実践を構想し、その結果と考察を述べる。

3. ルーブリックを用いた文章作成における自己評価を取り入れた実践の構想

3.1. 自己評価の段階性

Andrade (2010) は自己評価の段階性を「評価基準の明確化」「基準に対しての成果の批評」「復習 (フィードバックにもとづく改訂)」の3つのステップとして理解している。

最初のステップである「評価基準の明確化」では、児童・生徒・学生に学習において期待することを明確にすることが目的となる。課題またはパフォーマンスに対する期待を、教師、児童・生徒・学生、またはその両方が

ともに、模範課題を見直したり、ループリックを共同作成したりして、明確にする営みである。

第2ステップである「基準に対する成果の批評」では、期待される点（評価基準）から自らの作品を批評することが求められる。児童・生徒・学生は、エッセイ、単語問題、実験レポート、スピーチなど、課題のラフまたは初稿を作成する。初稿作成後、進行中のパフォーマンスを期待と比較することで、課題の進捗状況を確認し、ゴールへと向かう段階である。例えば、ループリックを使用して学習を進めたAndrade et al.(2008)では、生徒がループリックにもとづいて下書きを改変していくための材料を集める段階が、この第2ステップにあたる。色鉛筆を使い、生徒はループリックの重要なフレーズに一色でアンダーラインを引き（例えば、説得力のあるエッセイのループリックで「意見を明確に述べる」に青でアンダーラインを引く）、そのフレーズで明示されている基準を満たした証拠に下線を引いたり、下書きに丸をつけたりする（例えば、説得力のあるエッセイの下書きで自分の意見に青でアンダーラインを引く）。もし基準を満たしていないことがわかったら、完成稿を書くときに改善できるようにチェックする。

3つ目のステップは「復習（フィードバックにもとづく改訂）」である。このステップでは、児童・生徒・学生は自己評価からのフィードバックを改訂の指針にしてゴールへ向かう。ここではフィードバックにもとづいて自らの成果を改変していくことが最も重要である。

3.2. 自己評価を構成するための要素

先にも示したとおり、自己評価の実施段階にあたっては、様々な要素をふまえて実施内容を構成していくことが求められている。本研究で調査対象としているのは、自己評価におけるフィードバックのタイミングの違いによる、自己評価の正確さの差異であるが、先行研究で明らかになっている自己評価を構成していくために必要な要素、また、考慮すべき要素についてふまえたうえで実践を構想した。

まず、児童・生徒・学生の自己評価の実施に影響する要素について「評価ツールの使用」と「自己評価の練習の機会の確保」を取り入れた。Yan et al. (2023) は、その要素の最も重要なものとして自己評価の有用性に対する認識を挙げている。Panadero & Alonso-Tapia (2013) では、自己評価に対する有用性を認識されない限り自己評価は実施されないとされる。また、Yan et al. (2020) も自己評価の有用性の認識は、学習者が自己評価を実施する背景にある最も大きな因子であると結論づけている。以上から、自己評価の有用性に対する認識を促進していく要素を検討していくことが自己評価の構築の段階において重要であると言える。この点をふまえて、本実践では、自己評価の有用性の認識を高めるために、上記の要素を取り入れた。「自己評価の練習の機会の確保」をするために、実践の中では自己評価を行う手順を学習する機会やその練習をする機会を設けた。なお、「評価ツールの使用」の要素は、2点目に説明する技法

とも重なるものである。

そして、自己評価の技法として「ループリック」を採用した。自己評価の技法は、先に指摘した有用性への認識と正確さに関する諸研究において両者に関係する要素として示されている。特にYan et al. (2023) では、自己評価の実施を促す足場かけとして、自己評価の技法について検討することの重要性を示唆している。本研究では、様々な自己評価技法の中から、現時点で確実な有意性が認められている (Panadero et al., 2012) 「ループリック」を活用した自己評価を実施した。

3.3. ループリックの作成

上述したループリック (図1) の作成にあたっては、Andrade (2010) がループリック作成時に使用したSpandel & Stiggins (2001) を参照した。また、実践対象の学級担任を務めるA教諭とB教諭もループリックの作成に参加し、児童らの実態に合った語彙・文言の選定やループリックに示す分量について相談した。ループリックとして児童に示した文言については図1に示す。

4. 研究の方法

4.1. 実践の内容

実践の構想にあたっては、当該小学校で使用されている、東京書籍「地域のみりよくを伝えよう」にもとづいて学習内容を決定した。児童は、上記課題に関する初稿と完成稿を書き、その学習プロセスにおいて自己評価を行う。調査で収集したデータは、①文章作成課題のワークシート(初稿・完成稿)、②①について児童がループリックをもとに自己評価を実施したワークシート(初稿・完成稿)である。

自己評価を行うにあたっては、先述したAndrade (2010) による自己評価の3ステップの流れを中心に自己評価が行える環境を整えた。ステップ1では、児童が参照するためのループリックを作成する。この段階では、第1にループリックを示すことによって児童が参照すべき基準を明確にすることが目的である。また児童の自己評価の有用性への認識を促すことが第2の目的となる。

ステップ2では、成果に対する批評を行い、修正箇所を見出すことが目的となる。そのために本研究ではAndrade (2010) の方法を適用した。児童にループリックにおける重要な文章をチェック(下線を引く・丸で囲むなど)させ、そのチェックに合うものが自分の書いた文章の中にあるかを確認させる。該当する文章が書かれていない場合は完成稿において修正できるようにメモをするよう指示する。

ステップ3では、フィードバックを行い、自身の成果を修正する方向性を決定することが目的となる。本研究では、遅延フィードバック群(以下、DF群)と、即時フィードバック群(以下、IF群)に分けてフィードバックを行う。なお、分ける単位は学級である。遅延フィードバックの定義は、テスト終了時(Dihoff et al., 2003, 2004)から7日間の遅延(Bruning et al., 1999; Robin, 1978)までである。即時フィードバックの定義は、応答後数秒

①アイデア
レベル3：自分の経験をもとにした内容であり，自分ならではの思いや感じ方が書かれている。
レベル2：自分の経験をもとにした内容であるが，自分ならではの思いや感じ方が書かれていない。
レベル1：何を伝えたいのかがわかりにくい。
②文章の組み立て
レベル3：「始め」「中」「終わり」があり，それぞれが工夫されている。
レベル2：「始め」「中」「終わり」がある。
レベル1：「始め」「中」「終わり」がなく，読みづらい。
③主張
レベル3：自分の気持ちや考えを，理由をつけて表している。
レベル2：自分の気持ちや考えを書いている。
レベル1：自分の気持ちや考えを伝えようとしていない。
④言葉選び
レベル3：だれにでも伝わる言葉が使われていて，自分の気持ちや考えにぴったりくる表し方をしている。
レベル2：だれにでも伝わる言葉が使われている。
レベル1：言葉が読む人に伝わらない。
⑤文章の読みやすさ
レベル3：読む人が一度読んだだけで伝えたいことがわかる。
レベル2：読む人が何度か読むと伝えたいことがわかる。
レベル1：読む人に伝わりにくい文章になっている。
⑥文章のマナー
レベル3：文字や「、」「。」、かい行のまちがいが1～2こであり，漢字も使っている。
レベル2：文字や「、」「。」、かい行のまちがいが3～5こであり，漢字も使っている。
レベル1：文字や「、」「。」、かい行のまちがいが6こ以上である。

図1 作成したループリックの文言¹

(Epstein, Epstein, & Brosvic, 2001) から，次の週 (Robin, 1978) まで幅広い。このように，遅延フィードバックと即時フィードバックのそれぞれの定義については，重複する部分も多くある。したがって，本研究では，これらの定義を加味して，遅延フィードバックを授業（文章作成）後にフィードバックを行うこと，即時フィードバ

ックを授業（文章作成）中にフィードバックを行うことと定義した。

また，フィードバックの方法と正確さの関連性について調査するために，ステップ2・3では，ループリックを参照して自分が基準に対してどの位置にいるのかを3段階（レベル1～3）の中の数値として示す活動を設ける。初稿の段階と完成稿の段階において数値として自らの成果を評価する機会を合計2回設ける。

4.2. 実践の対象

実践はX県Y小学校第5学年の2学級（合計71名）を対象に2回（合計4日間）行った。DF群の実践は，2024年9月2日（月）と9月3日（火）の日程で，男児17名と女児18名の合計35名が在籍する学級で実施した。IF群の実践は，2024年9月12日（木）と9月13日（金）の日程で，男児17名と女児19名の合計36名が在籍する学級で実施した。

4.3. 実践の流れ

本実践は国語科における4時間を用いて主著者が行った。以下に実践の流れをまとめた表を示す（表1）。

1時間目は児童がループリックの使い方を把握することによって参照すべき基準を明確化できるようにすること，自己評価の有用性への認識を高めることが目標となる。したがって，ループリックの使い方の説明や練習，ステップ2における手続きの理解を促すための学習となる。教師（主著者）が作成したスライド資料にもとづいてループリックに示される文言を理解しながら学習を進めた。授業の最後には，3.2.で示した自己評価の有用性の認識を促すために自己評価の練習ができる機会を確保するために，自己評価の練習として教師が作成した文章を評価する活動を行った。評価にあたっては，その後の授業における自己評価の手続きと同じく，ループリックの3段階を用いて評価した。

2時間目は児童が初稿作成することが目標となる。初稿作成時，DF群ではフィードバックを与えず，自力で文章作成に取り組んだ。なお，ループリックの文言についての質問等には対応した。IF群では教師が机間巡視をしながら児童らの進捗状況を観察しつつ，ループリックにもとづいてその後の文章作成についてフィードバックを与えた。なお，IF群の授業時には，主著者の他に当該小学校の教諭Aと教諭B，ならびに教員養成課程4年生の学部生2名が協力し，即時フィードバックが行えるようにした。協力者は全員，事前にループリックについての説明を受け，文章作成にあたってのフィードバックの質を統一できるようにした。初稿作成後，児童らは自分の初稿に対して3段階で評価を行い，完成稿に向けた見通しを立てた。授業後，主著者は両群の初稿に対しての遅延フィードバックを行った。フィードバックの内容としては，初稿をループリックの6つの視点にもとづいて評価し，評価内容を文章として当該ワークシートに記入した。なお，初稿に対して主著者が3段階の内どの

¹⑥における「かい行のまちがい」とは，改行時に新しい行の1マス目を空けることができていないことを示している。

表1 実践の流れ

	1時間目	2時間目	3時間目	4時間目
IF群	・ループリックの説明 ・例文提示 ・自己評価練習	・初稿作成 (IFあり) ・初稿への自己評価 ・初稿へのDF記述	・初稿へのDF説明	・完成稿作成 (IFあり) ・完成稿への自己評価
DF群	・ループリックの説明 ・例文提示 ・自己評価練習	・初稿作成 (IFなし) ・初稿への自己評価 ・初稿へのDF記述	・初稿へのDF説明	・完成稿作成 (IFなし) ・完成稿への自己評価

評価をつけたかは児童らに明らかにしていない。

3・4時間目は完成稿作成に取り組み、ステップ2・3の段階を行き来して、よりよい成果物を目指すことが目標となる。いずれの群も3時間目に初稿に対する遅延フィードバックが返却され、それにもとづいて完成稿を作成した。なお、IF群は授業時間中にフィードバックが与えられた。また、いずれの群も完成稿作成後にはループリックを用いて3段階評価を行った。

5. 結果と考察

5.1. 自己評価と教師評価の結果と考察

初稿と完成稿についての教師評価とループリックに示された自己評価の結果を実数値及び割合として表すことから、表2・3・4・5を得た。なお、表中に示される数値は、白塗りの部分に記された数値が実数値で、網かけ部分に記された数値が割合を示している。

5.1.1. 遅延フィードバック群についての結果

初稿作成後における自己評価では、どの項目にも共通してレベル2に約半数の児童が該当していた。項目①ではレベル2として評価する児童が15名(50.0%)確認された。項目②は、レベル1に該当する児童が最も多く、8名(26.7%)であった。項目③と④はほぼ同程度の割合が結果として確認され、レベル1と評価する児童は項目②に次いで多かった。(なお、項目③の実数値合計が1不足しているのは、その項目のみ評価を行っていなかった児童がいたためである)。項目⑤は、レベル2で最も割合が多い項目であった(17名:56.7%)。項目⑥は、全項目中レベル3で最も高い割合が確認された(12名:40.0%)。

初稿作成後における教師評価では、自己評価と比較して項目③と④の結果が大きく異なるものとなった。項目①では、概ね自己評価と同様の結果が得られた。項目②では、自己評価では最も多かったレベル1に該当する児童に差異が見られ、半分以下となった(3名:10.0%)。項目③と④では、自己評価の際はレベル3に該当していた児童が半数以上レベル2以下となった。特に、項目④では3分の1以上の差異が確認された(3名:10.0%)。項目⑤と⑥では、約20%以上の児童がレベル2から3と評価された。

完成稿作成後における自己評価では、いずれの項目のレベル3でも初稿作成後と比較して2倍以上割合が増加した。項目①では、レベル1と評価する児童が最も少なく(1名:2.9%)、反対に、レベル3と評価した児童は項

目⑥に次いで2番目に多かった(27名:79.4%)。項目②は、レベル3に該当する児童が最も少なかった(20名:58.8%)。項目③・④・⑤では、初稿作成後の結果と比較しても全てのレベルで同様の結果が得られた。いずれの項目もレベル3に該当する児童が2倍以上増加した。項目⑥では、レベル3に該当する児童が最も多く確認された(28名:82.4%)。

完成稿作成後における教師評価では、自己評価と比べて項目③と④における差が顕著であった。また、初稿作成後における教師評価と比較すると、どの項目においてもレベル1と評価される児童が減少し、レベル3に該当する割合はどの項目でも増加した。項目①では、レベル3に該当する児童が最も多く確認され26名(76.5%)であった。また、項目④と同様にレベル1に該当する児童は確認されなかった。項目②では、項目①と同様の結果が得られた。レベル3に該当する児童の割合は項目⑥と同様であった(24名:70.6%)。項目③と④ではレベル3に該当する児童が他の項目の3分の1の割合であった。また、自己評価と比較しても自己評価ではレベル3に該当していた児童の内の約4分の3がレベル2以下へ移行していた。項目⑤と⑥では、同様の傾向が確認された。自己評価と比較してもあまり大きな変化はなかった。

5.1.2. 即時フィードバック群についての結果

初稿作成後における自己評価では、どの項目においても30名以上(89%以上)の児童がレベル2以上の評価をしている。項目①では、レベル3と2が同程度確認された。項目②では、群の55.9%(19名)がレベル2と評価をした。項目③は、レベル3の割合が最も少なく、8名(23.5%)が確認された。また、レベル2の割合が最も多く確認された項目でもあった(23名:67.6%)。項目④では、項目②と同様の結果が得られた。項目⑤では、項目①と全く同じ結果が得られた。項目⑥においては、レベル3と評価する児童が最も多く確認された(22名:64.7%)。

初稿作成後における教師評価では、どの項目においても85%以上の児童がレベル2以上に該当した。項目①は、レベル1の割合が最も多く、5名(14.7%)が確認された。項目②は項目①に次いでレベル1の割合が多かった(3名:8.8%)。項目③はレベル3の割合が最も少なく、2名(5.9%)であった。項目④も項目③と同様にレベル3に該当する割合が低かった(4名:11.8%)。項目⑤では、レベル2と3に該当する児童が同程度確認された。これは、同項目ならびに項目①の自己評価で確認できた傾向と同じ結果である。項目⑥は、レベル3に該当する児童が最も多く、

表2 DF群における初稿作成時の自己評価と教師評価の結果

	自己評価						教師評価					
	3		2		1		3		2		1	
①	11	36.7	15	50.0	4	13.3	9	30.0	17	56.7	4	13.3
②	8	26.7	14	46.7	8	26.7	11	36.7	16	53.3	3	10.0
③	11	36.7	12	40.0	6	20.0	5	16.7	23	76.7	2	6.7
④	11	36.7	13	43.3	6	20.0	3	10.0	26	86.7	1	3.3
⑤	8	26.7	17	56.7	5	16.7	14	46.7	11	36.7	5	16.7
⑥	12	40.0	14	46.7	4	13.3	17	56.7	11	36.7	2	6.7

表3 DF群における完成稿作成時の自己評価と教師評価の結果

	自己評価						教師評価					
	3		2		1		3		2		1	
①	27	79.4	6	17.6	1	2.9	26	76.5	8	23.5	0	0.0
②	20	58.8	10	29.4	4	11.8	24	70.6	8	23.5	2	5.9
③	24	70.6	8	23.5	2	5.9	7	20.6	26	76.5	1	2.9
④	23	67.6	9	26.5	2	5.9	6	17.6	28	82.4	0	0.0
⑤	23	67.6	8	23.5	3	8.8	25	73.5	7	20.6	2	5.9
⑥	28	82.4	4	11.8	2	5.9	24	70.6	9	26.5	1	2.9

表4 IF群における初稿作成時の自己評価と教師評価の結果

	自己評価						教師評価					
	3		2		1		3		2		1	
①	16	47.1	15	44.1	3	8.8	11	32.4	18	52.9	5	14.7
②	11	32.4	19	55.9	4	11.8	14	41.2	17	50.0	3	8.8
③	8	23.5	23	67.6	3	8.8	2	5.9	30	88.2	2	5.9
④	12	35.3	19	55.9	3	8.8	4	11.8	30	88.2	0	0.0
⑤	16	47.1	15	44.1	3	8.8	17	50.0	16	47.1	1	2.9
⑥	22	64.7	9	26.5	0	0.0	23	67.6	10	29.4	1	2.9

表5 IF群における完成稿作成時の自己評価と教師評価の結果

	自己評価						教師評価					
	3		2		1		3		2		1	
①	31	91.2	3	8.8	0	0.0	32	94.1	2	5.9	0	0.0
②	22	64.7	12	35.3	0	0.0	28	82.4	6	17.6	0	0.0
③	24	70.6	10	29.4	0	0.0	8	23.5	26	76.5	0	0.0
④	23	67.6	10	29.4	1	2.9	6	17.6	28	82.4	0	0.0
⑤	23	67.6	10	29.4	1	2.9	27	79.4	7	20.6	0	0.0
⑥	30	88.2	3	8.8	1	2.9	25	73.5	9	26.5	0	0.0

23名(67.6%)であった。これも、同項目の自己評価における傾向と同様の結果が得られている。完成稿作成後の自己評価では、初稿作成後よりさらにレベル1と評価した児童が減少した。項目①では、レベル3と評価した児童数が31名(91.2%)となり、各項目の中でも最も多い割合となった。項目②では、初稿作成後に比べてレベル3を選択する児童が2倍に増加した(22名:64.7%)。項目③では初稿作成後と比較して3倍の児童がレベル3として評価をしている(24名:70.6%)。項目④では、項目③とほぼ同程度の結果が得られており、レベル3に該当する児童が最も多かった(23名:67.6%)。項目⑤は項目④と全く同じ結果となった。項目⑥では、項目①に次いでレベル3と評価する児童が多く確認された(30名:88.2%)。

完成稿作成後における教師評価では、レベル1に該当する児童が確認されず、項目③と④以外ではレベル3に該当する児童の割合が最も多くなった。項目①では自己評価時とほぼ同じ結果が得られ、32名(94.1%)の児童がレベル3に該当した。初稿作成後と比較するとその割合は約3倍に増加した。項目②では、自己評価ではレベル2であった児童の半数がレベル3と評価され、レベル3における割合が約20%増加した。初稿作成後ではレベル3の該当者が14名であったのに対して、この時点では28名に倍増した。項目③では、自己評価でレベル3と評価していた児童の3分の2がレベル2または1と評価され、レベル3の割合は自己評価と比較して3分の1になった(8名:23.5%)。初稿作成後(2名)と比較してレベル3に該当する児童は4倍になった。項目④では、レベル3に該当する割合が最も少なくなった(6名:17.6%)。初稿作成後とも大きな変動は見られない。項目⑤では、自己評価とほぼ同じ結果が得られ、27名(79.4%)がレベル3に該当した。初稿作成後と比べると約30%の割合でレベル3に該当する児童が増えている。項目⑥では、初稿作成後におけるレベル3に該当した割合と同程度の結果が得られた(25名:73.5%)。

5.1.3. 自己評価と教師評価の結果にもとづく考察

以上の結果における実数値や割合の比較をとおして得られた結果の傾向を以下の2点にまとめ、その傾向にもとづいて考察を行った。

1点目に、フィードバックのタイミングの違いによって書くことの学習における成果(教師評価)には大きな違いが見られた項目が、項目①のみであったことである。項目①に関しては、DF群よりもIF群の方がより高い成果を得られていた(DF完成稿:26名76.5%, IF完成稿:32名94.1%)。Shute(2008)は、遅延フィードバックと即時フィードバックの効果の差異に、課題の難易度と学業成績の高低が関係していると説明している。遅延フィードバックは、より簡単な課題や学習の転移を促進することが目的であるような課題には有効であるとされ、より学業成績が高い児童・生徒・学生に有効である。一方、即時フィードバックは、より難易度の高い課題に取り組む際に有効であり、より学業成績が低い児童・生

徒・学生に対して有効であるとされる。本研究で扱った学習課題は、児童らに馴染みのないループブックを使用した文章作成であり、児童らにとってみれば高難度の課題であった。その点ではIF群における成果(教師評価)がDF群のものよりも高くなるのが仮説として設定され、この結果からはその仮説を強めることができたと言える。

2点目に、自己評価と教師評価の乖離が大きいのは項目③「主張」と項目④「言葉選び」であることである。項目③では、教師評価においてレベル2の要件である「自分の気持ちや考えを書いている」を満たしている児童は多く確認された(DF初稿:23名76.7%, DF完成稿:26名76.5%, IF初稿:30名88.2%, IF完成稿:26名76.5%)が、その上位であるレベル3の要件の「自分の気持ちや考えを、理由をつけて表せている。」を満たす児童は少なかった(DF初稿:5名16.7%, DF完成稿:7名20.6%, IF初稿:2名5.9%, IF完成稿:8名23.5%)。一方で、自己評価においては、項目③でレベル3と評価する児童が多く確認された。初稿作成後はレベル3に到達していたと評価した児童の割合は少ない(DF:11名36.7%, IF:8名23.5%)にもかかわらず、完成稿作成後にはその割合が著しく増大した(DF:24名70.6%, IF:24名70.6%)。この結果が導かれた要因としては、文章全体をとおして伝えたい気持ちや考えについて一貫した理由を書くことができていることが考えられる。初稿作成後に項目③でレベル2に該当した児童へは、初稿作成後に与えるフィードバックにおいて、自分の気持ちや考えに理由を書くことを促す旨の添削を入れた。その結果、図2に見られるような、自分の気持ちや考えについての断片的な理由を書くことはできているが、文章全体をとおして伝えたい気持ちや考えについて一貫した理由を書くことができている文章が多く確認された。

この文章における自分の気持ちや考えは、「地域の人が親切である」ということであると読み取れる。そのため、主張がコミュニティセンターの人たちとのかかわりにおける具体的なエピソードにもとづいて展開されている。しかしながら、文章内で強調した部分である「勉強以外にも…」という文では、主題が「親切」から「楽しい」へとずれており、本来の自分の気持ちや考えを説明する理由となっていないと指摘できる。また、文章の最後には「このように…」の文で締めくくられているが、「親切である」という単一であった魅力が「いろいろな」魅力へと変わっている。この点においても、文章全体をとおして伝えたい気持ちや考えについて一貫した理由を書くことができていると言える。

項目④では、「言葉選び」として比喩表現を使うことをレベル3の要件の一例として例文に示していたが、その要件を満たしている児童は多く確認されなかった(DF初稿:3名10%, DF完成稿:6名17.6%, IF初稿:4名11.8%, IF完成稿:6名17.6%)。初稿作成後には、自己評価で項目④をレベル3とした児童は多くなかった(DF:11名36.7%, IF:12名35.3%)。一方で、完成稿において

「こんにちは！」と言うと、みんなあいさつをかえしてくれます。

みんな親切な地元の地いきがわたしは大好きです。

たとえば、コミュニティセンターの勉強会では、わからない問題があったら一人一人にヒントをくれたり、理かいするまでおしえてくれます。

そして、きゅうけい時間にジェンガをする時も、先生がいっしょにしてくれて、とっても親切だと思います。勉強以外にも、パステルアートや電気工作もあります。いろいろなことができるので毎回行くのがたのしみになります。ほかに、ウォークラリーやりんごがり、スキー研修などもあり、いろいろなき画をしてきている地いきの人にほんとに感しゃの気持ちでいっぱいです。

このように、地元の地いきにはいろいろなみ力があります。なのでわたしはこれからも地元の地いきを大事にしていきたいです。

(強調は筆者による。)

図2 IF群における児童の完成稿

自己評価時にレベル3と評価している児童は多く確認された(DF:23名67.6%, IF:23名67.6%)。この要因として、初稿作成後にレベル2の該当者に与えたフィードバックにおいて「誰にも伝わりやすい文章になっています」という肯定的な意味しか見出せない文言を使用したために、自らの文章のレベルが3であると誤認識してしまったことが考えられる。このことから、文言としてのフィードバックだけではなく、数値としてのフィードバックを与える必要性が示唆された。

5.2. フィードバックのタイミングの違いによる

自己評価の正確さの差異

5.1.1.において、表にまとめた数値にもとづいて正確さの差異を確認した。自己評価の正確さの数値を求めるにあたっては、自己評価の正確さの定義である「児童・生徒・学生の自己評価と評価基準の近さ」(León et al., 2023, p.3)にもとづいて得た「|自己評価-教師評価|」の式を用いた。正確さの数値に関しては、DF群とIF群との正確さの差異を確認するため、初稿作成後と完成稿作成後の時点で対応のないt検定をかけ、フィードバックのタイミングの違いによって自己評価の正確さにどのような影響をもたらすのかを分析した。なお、t検定においてその効果量も分析した。効果量の分析にあたっては、水本・竹内(2008)にもとづいて結果を導いた。

初稿作成後における自己評価の正確さにおいて、遅延フィードバックのみが行われる場合($M=2.93, SD=1.57$)と即時フィードバックが行われる場合($M=2.5, SD=1.85$)との自己評価の正確さには、有意な差が示されなかった($t(62) = 0.99, ns$)。また、効果量($d=0.20$)はほとんど認められなかった。同様に、完成稿作成後における自己評価の正確さについても、遅延フィードバックのみが行われる場合($M=2.68, SD=2.01$)と即時フィードバックが行われる場合($M=2.24, SD=1.35$)と

の自己評価の正確さには、有意な差が示されなかった($t(66) = 1.04, ns$)が、小さな効果量($d=0.26$)が認められた。

以上の結果から、フィードバックのタイミングの違いによって自己評価の正確さに差異がないことが確認された。しかしながら、完成稿作成後には即時フィードバックを行うことによって自己評価の正確さが高まる効果が認められた。このことから、即時フィードバックを実施することは作文活動における自己評価の正確さを高めるための一要因となる可能性があることが示唆された。

6. 結論

本研究では、諸外国の自己評価に関する研究の動向から、作文活動における自己評価の実施方法について検討した。まず、諸外国の自己評価に関する研究動向を整理することで、自己評価の正確さを高めるためのフィードバックの実施方法についての検討を行うことが、自己評価の目的を達成できる学習を構想するために必要であることを明らかにした。次に、その結果にもとづいて、フィードバックのタイミングの違いによって自己評価の正確さに差異が表れるのかを分析するための実践を開発・実施・評価した。その結果、フィードバックのタイミングの違いによって有意な差異は確認されなかったが、即時フィードバックの実施が、作文活動における自己評価の正確さを高める一要因となる可能性が示唆された。

以上の結果の一方で、即時フィードバックの実施方法について検討することが今後の課題として挙げられる。本研究における即時フィードバックは、授業内におけるフィードバックと定義して実施してきたが、その即時フィードバックは真に「即時」と言えるのか懐疑的である。文章作成過程において、真に文章作成後即時でフィードバックを与えるためには、教師対児童・生徒・学生をマンツーマンの環境で文章作成できるようにするか、AI活用のもとに実施することを考えていく必要があると考えられる。数名の教師に対して多くの児童・生徒・学生が向き合うという構造が一般的である日本の文脈を鑑みれば、後者のAI活用について具体的な実践を構想していくことの方が優先的であることは明確である。つまり、文章作成過程におけるAIを活用した即時フィードバックの実践としての研究の蓄積が必要になると言える。

文章作成過程におけるAIを活用した即時フィードバックとしては、AWE (Automated Writing Evaluation) や AES (Automated Essay Scoring) を活用した事例が多く見られる。こうした文章作成過程におけるAIを活用した即時フィードバックには、活用するAIの採点の正確さ(Liu & Kunnan, 2016)や採点の客観性(Shermis, 2014)について議論がなされている。Liu & Kunnan (2016)は、人間の採点の方がAESよりも正確な評価を与えることができたという結果を示している。そうしたネガティブな報告の一方でShermis (2014)

は、AESの評価の客観性に注目し、AESを活用することで人間より客観性の高い評価を行うことができると報告している。文章作成過程におけるAIを活用した即時フィードバックの実践を考えていくにあたっては、その即時性のよさばかりに着目するのではなく、上記のようなポジティブな面とネガティブな面について吟味する必要がある。今後は、このような多面的な視点から文章作成過程におけるAIを活用した即時フィードバックについて検討を重ねていきたい。

謝辞

本研究を実施するにあたっては、X県Y小学校の教職員ならびに児童の皆様にご協力を賜りました。また、同校のA教諭ならびにB教諭には授業計画および、授業実践にご協力いただきました。厚く御礼申し上げます。

参考文献一覧

- Andrade, H. L. (2010). Students as the Definitive Source of Formative Assessment: Academic Self-Assessment & the Self-Regulation of Learning. *In Handbook of Formative Assessment*. 90-105. Routledge.
- Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-Referenced Self-Assessment & Middle School Students' Writing. *Assessment in Education*, 17 (2), 199-214.
- Andrade, H., Y. Du, & X. Wang. (2008). Putting Rubrics to the Test: The Effect of a Model, Criteria Generation, & Rubric-Referenced Self-Assessment on Elementary School Students' Writing. *Educational Measurement: Issues & Practices*, 27 (2), 3-13.
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *In Frontiers in Education*, (4) ,87. Frontiers Media SA.
- Boud, D., & Falchikov, N. (1989). Quantitative Studies of Student Self-Assessment in Higher Education: a Critical Analysis of Findings. *High Educ* 18, 529-549.
- Brown, G., & Harris, L. R. (2013). Student Self-Assessment. *SAGE Handbook of Research on Classroom Assessment*.
- Bruning, R., Schraw, G., & Ronning, R. (1999). *Cognitive Psychology & Instruction*. Columbus, OH: Merrill Prentice Hall.
- Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The Role of Feedback during Academic Testing: The Delay Retention Effect Revisited. *The Psychological Record*, 53, 533-548.
- Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2004). Provision of Feedback During Preparation for Academic Testing: Learning is Enhanced by Immediate but not Delayed Feedback. *The Psychological Record*, 54, 207-231.
- Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The Use of Self-, Peer & Co Assessment in Higher Education: A Review. *Studies in Higher Education*, 24 (3), 331-350.
- Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate Feedback during Academic Testing. *Psychological Reports*, 88, 889-894.
- 富士原紀絵. (2015). 国語教育の作文指導過程における「文を見直す」行為に関する一考察：指導に用いられる用語の整理を通して. お茶の水女子大学人文科学研究, 11, 99-111.
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition & Communication*, 32 (4), 365-387.
- León, S. P., Panadero, E., & García-Martínez, I. (2023). How Accurate Are Our Students? A Meta-analytic Systematic Review on Self-Assessment Scoring Accuracy. *Educational Psychology Review*, 35 (4), 106.
- Liu, S., & Kunnan, A. (2016). Investigating the Application of Automated Writing Evaluation to Chinese Undergraduate English Majors: A Case Study of WriteToLearn. *Calico Journal*, 33 (1), 71-91.
- 水本篤・竹内理. (2008). 研究論文における効果量の報告のために. 基礎的概念と注意点. 英語教育研究, 31, 57-66.
- Panadero, E. (2011). Instructional Help for Self-Assessment and Self-Regulation: Evaluation of the Efficacy of Self-Assessment Scripts vs. Rubrics. *Unpublished Doctoral Dissertation*. Universidad Autónoma de Madrid, Madrid, España.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics & Self-Assessment Scripts Effects on Self-Regulation, Learning & Self-Efficacy in Secondary Education. *Learning & Individual Differences*, 22 (6), 806-813.
- Panadero, E., & Tapia, J. A. (2013). Self-Assessment: Theoretical & Practical Connotations. When It Happens, How is It Acquired & What to Do to Develop It in our Students. *Electronic Journal of Research in Educational Psychology*, 11 (2), 551-576.
- Panadero, E., Brown, G. L., & Strijbos, J.-W. (2016). The Future of Student Self-Assessment: A Review of Known Unknowns & Potential Directions. *Educational Psychology Review*. 28, 803-830. 10.
- Panadero, E., Ruiz, F., J. & Iglesias, S., I. (2020). Secondary Education Students' Self-Assessment: The Effects of Feedback, Subject Matter, Year Level, & Gender. *Assessment in Education: Principles, Policy & Practice*, (Online first).
- Phye, G. D., & Andre, T. (1989). Delayed Retention Effect: Attention, Perseveration, or Both? *Contemporary Educational Psychology*, 14 (2), 173-185.
- Robin, A. L. (1978). The Timing of Feedback in Per-

- sonalized Instruction. *Journal of Personalized Instruction*, 3, 81-87.
- Ross, J., C. Rolheiser, & A. Hogaboam-Gray. 1999. Effects of Self-Evaluation Training on Narrative Writing. *Assessing Writing* 6, 1, 107-132.
- Shermis, M. D. (2014). State-of-the-Art Automated Essay Scoring: Competition, Results, and Future Directions from a United States demonstration. *Assessing Writing*, 20, 53-76.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78 (1), 153-189.
- Spandel, V., & R. J. Stiggins. (2001). *Creating Writers Through 6-Trait Writing Assessment & Instruction (3rd Edition)*. Addison Wesley Longman.
- Yan, Z., Brown, G. T. L., Lee, C. K. J., & Qiu, X. L. (2020). Student Self-Assessment: Why do They Do It? *Educational Psychology*, 40 (4), 509-532.
- Yan, Z., Panadero, E., Wang, X., & Zhan, Y. (2023). A Systematic Review on Students' Perceptions of Self-Assessment: Usefulness & Factors Influencing Implementation. *Educational Psychology*, 35 (3), 1-28.