

第37回 研究助成

B 実践部門・報告 V・英語能力向上をめざす教育実践

高等学校のライティング指導における 生成 AI の活用とその効果 — 機能的適切さ(Functional Adequacy)に着目して —

研究者: 渡邊 大志 神奈川県／神奈川県立総合教育センター 指導主事

(申請時 神奈川県／神奈川県立藤沢工科高等学校 教諭・横浜国立大学教職大学院 在籍)

《研究助言者: 小泉 利恵》

概要

本研究は、高等学校における英語ライティング指導において、生成 AI を活用した学習が学習者のパフォーマンス、とりわけ「機能的適切さ(Functional Adequacy)」に与える効果を検証することを目的とした。神奈川県内の公立高校1年生356名(分析対象214名)を対象に、事前・事後テスト、生成 AI を活用した練習課題、およびアンケート調査を実施した。評価は教師3名と Google Gemini により、機能的適切さの4観点(内容・タスク達成度・理解のしやすさ・一貫性と結束性)に基づき行った。分析の結果、教師評価と AI 評価の間には中程度の相関が認められた一方、AI は一貫して高めに評価する傾向があった。混合計画分散分析では「内容」で時間×練習の交互作用が有意となり、練習あり群のみで事前から事後に有意な向上が見られた。他の観点では時間効果は一部認められたものの、練習効果は限定的であった。アンケート調査では、効率性・即時性・気づきの促進・心理的安心感が肯定的に評価される一方、英語の基礎力不足や AI 依存への懸念も示された。以上より、生成 AI は機能的適切さに基づくライティング指導の補助的ツールとして有用であり、その効果を最大化するには、教師による足場かけと批判的活用を促す指導デザインが必要であることが示唆された。

1 はじめに

ChatGPT の2022年11月の公開以降、学校教育における生成 AI の利用可能性について議論が行われ、利活用に関する研究も進展している。文部科学省(2024)は、子どもたちが将来生成 AI のようなツールを使いこなせるようになることの重要性を指摘し、学校教育の中で適切な使用方法を体験的に学ぶことが有効であると述べている。同時に、「人間中心の原則」を強調し、生成 AI は「人間の能力を補助・拡張し、その可能性を広げる有用な道具」(p.7)と位置づけられている。つまり、生成 AI はあくまでも人間の目的を達成するための手段であり、活用自体を目的化すべきではない、ということである。

こうした背景を受け、英語教育の分野でも生成 AI を活用した指導法や学習法が検討されている(門田, 2024など)。特にスピーキングやライティングといった発信技能においては、教師に代わってフィードバックを提供できるという点で、時間的・空間的制約を受けずに自律学習を支援する有効なツールとして注目されている。また、学校現場ではライティング指導における添削の時間確保が困難であることも指摘されており(川村他, 2019)、ライティング指導の十分な実施が妨げられている現状がある。もし生徒自身が生成 AI を活用してライティングのパフォーマンスの改善を行うことができれば、教師の負担軽減と同時に生徒の学習機会の拡充にもつながるだろう。

さらに、現行の高等学校学習指導要領では「思考力・判断力・表現力等」の育成が重視されており、単なる言語形式の正確さだけでなく、目的や場面、状況に応じた適切な言語使用や内容面での表現力を高めることが求められている（文部科学省，2018）。したがって、学習者のパフォーマンスを内容面から評価できる枠組みの導入も併せて考えていく必要があり、この点において「機能的適切さ（Functional Adequacy）」が有効な構成概念となる可能性がある。

本研究では、高等学校におけるライティング指導に焦点をあて、機能的適切さの枠組みを用いた生成AIの活用可能性を実証的に検討する。

2 先行研究

2.1 生成AIによるライティングの評価とフィードバック

生成AIとライティングの評価についての研究で有名なものは Mizumoto and Eguchi (2023) が挙げられる。ChatGPT の GPT-3.5 による自動ライティング採点 (AWS) の信頼性と精度を、人間と機械学習とで比較した研究である。採点対象のライティングとしては、学習者コーパスである TOEFL 11 から英作文 12,100 本を使用した。これらはすでに評価者によって3段階の評価 (high / medium / low) がつけられており、その評価結果が ChatGPT による採点とどの程度一致するかを検討した。生成AIのプロンプトには、IELTS のライティングのルーブリック (10段階) に基づいて包括的に0～9点をつけるように入力した。結果としては、人間と ChatGPT による採点の一致度は比較的高く、生成AIを併用すれば、教師が内容面など必要な部分に時間を割いて集中できるのではないかと述べられている。

ChatGPT を学習者に使用させる際の手順について、水本 (2025) が先行研究を基に示している (表1)。

■表1: 英語ライティング指導においてChatGPTを学習者に使用させる際の手順

| ステップ | 具体的活動 |
|------------------------|---|
| (1) 事前準備 | <ul style="list-style-type: none"> ● 教師に ChatGPT の使い方を説明 ● 学習者に適切な使用方法を指導 |
| (2) 最初のドラフト作成 | <ul style="list-style-type: none"> ● 学習者が自力でドラフトを書く |
| (3) ChatGPT によるフィードバック | <ul style="list-style-type: none"> ● 学習者がドラフトを ChatGPT に入力 ● 文法・語彙・構成・内容についてアドバイスを受ける |
| (4) ドラフトの改訂 | <ul style="list-style-type: none"> ● 学習者が ChatGPT の助言を基に修正・改善 ● 必要に応じて ChatGPT に追加質問 |
| (5) 教師やピアによるレビュー | <ul style="list-style-type: none"> ● 改訂版に対してフィードバックを得る |
| (6) 最終版の作成 | <ul style="list-style-type: none"> ● 受けたフィードバックを踏まえて最終版を完成 |
| (7) 振り返り | <ul style="list-style-type: none"> ● ChatGPT 使用経験や学習効果について学習者が振り返りを行う |

注：水本 (2025) より引用

この手順の中では、学習者は教師の支援を受けながら、ChatGPT を活用して自身のライティングスキルを自律的に向上させる方法を体験的に学ぶことになる。重要なことは、教師の役割であり、教師は生成AIが学習者によって効果的に使用されるように指南するべきである。特に、生成AIからのフィードバックから「気づき (noticing)」が促進されるような手立てや授業デザインをしていくことに、留意すべきであると述べている。

生成 AI を活用したライティングの授業の実践例としては山下他(2024)がある。大学生が自身で使用するための AI を含む英文作成支援ツール Transable を開発し、それを実際に3回の授業で使用した実践報告である。Transable に搭載されているシステムは、以下の4つである。

- Grammarly: 文法や語法の修正提案ができる
- DeepL: 日本語訳の確認ができる(意図した内容になっているか)
- ChatGPT: 様々な表現を調べる(テンプレートでプロンプトが用意されている)
- TOEFL iBT, GTEC, CEFR, IELTS のルーブリック: 選ぶと即時的に採点とコメントを出力する

結果としては、生徒のライティング力の向上も見えて取れ、生徒の Transable に対する使用感は肯定的な意見が多かった。一方で、フィードバックを受けての生徒によるほとんどの修正が、単語レベルでの修正にとどまっている点が懸念点として挙げられていた。生成 AI が生徒の自律的な学習ツールとして機能するようになれば、教師がフィードバックする内容をより焦点化したり、個別最適化されたフィードバックを生徒に提供したりすることが可能になるだろうということが示唆された。

学習者による自動採点システムからのフィードバックを活用したライティングの修正に関しては、生徒が修正した量だけでなく、どのような修正を行ったのかという質にも気を配る必要がある。Liu and Yu(2022)は、参加者に Write & Improve (W & I) with Cambridge という自動採点システムから得られたフィードバックを基に、自分で書いたものの修正を求めた。W & I with Cambridge は直接語彙フィードバック(DWF)、間接語彙フィードバック(IWF)、間接文レベルフィードバック(ISF)の3つのフィードバックを行うことができるシステムである。学習者がそれらをどのように活用するかを、アイ・トラッキング、回顧的インタビュー、省察記録を基に分析した。結果は、フィードバックを用いて修正された率は75%と高かったが、表面的なミスの修正が多く、内容面(アイデアや論理構成)に関する修正は23%と少なめであった。直接的なフィードバックは修正につながりやすいが学びにはつながりにくい一方で、間接的なフィードバックは学習者のレベルが一定以上ないと修正につながりにくい可能性が示唆された。これは、W & I with Cambridge が DWF や IWF が中心であることや、ISF が単に「文の質が低い」ことをその部分をハイライトして示すだけである、といった自動採点システムの制約によるものが考えられる。内容面に関する直接的なフィードバックの効果の検討が必要であろう。

さらに近年では、生成 AI を含む自動フィードバックや自動評価(AWE/AWF)に関するレビューや研究が多数報告されている。Shi and Aryadoust(2024)は1993年以降の AI ベースの自動フィードバック研究を体系的に整理し、AWF は文法や語彙といった表層的誤り訂正には有効である一方、内容や構成など高次スキルへの効果は限定的であることを示した。また、Karatay and Karatay(2024)は AWE 研究の統合的レビューを行い、AWE が学習者の自己修正や自律性を促進することを認めつつも、教師フィードバックを完全に代替するものではなく、補完的に活用すべきだと論じている。また、Feng et al.(2025)は2014年から2024年までの112本の研究をレビューし、AI と L2 ライティング研究の多くが大学生を対象に短期的効果を検証するもので、理論的基盤が弱い点を指摘している。特に、ChatGPT の無料公開以降の研究では、学習者や教師が AI をどのように受け止めるかや、AI がアイデア生成や協働的学習をどのように支援するか、といった新たな課題が急速に注目されつつあることが報告されている。学習者の認識面については、Allen and Mizumoto(2024)が日本の大学生を対象に、ライティング活動におけるピアレビューと ChatGPT からのフィードバックを比較した研究を行い、多くの学習者が AI を「信頼できる権威的なフィードバック源」として評価していたことを明らかにした。同時に、学習者同士の協働的な学びにも一定の価値を感じていることもわかり、AI とピアフィードバックの補完的活用が望ましいと結論づけている。

また、AI 評価の信頼性に関しては、Tate et al.(2024)が ChatGPT によるエッセイスコアリングを分析し、人間評価者との一致はおおむね中程度であるが、内部的・一貫性(再現性)はむしろ AI の方が高い場合もあると報告している。これにより、形成的評価では十分に活用可能である一方で、ハイスタークスの試

験や総括的評価においては、まだ慎重な運用が必要であるとされる。評価の場面でどのように生成AIを導入すべきかどうかに関しては、Voss et al. (2023) が言語テストにおける生成AIの利用についての議論において、実際の言語使用を反映するためにはAI利用を許容すべきだという立場と、構成概念の歪みや公平性・倫理的リスクから制限すべきだという立場からの視点や考え方を整理している。そのまとめにおいて、言語テストとAIの開発者が協働し、人間の価値観に合致した適切な支援技術の利用を可能にする「ガードレール」を備えたアセスメントを設計していくことが求められると述べている。

2.2 機能的適切さ

学習者の言語パフォーマンスの内容面についての評価基準として、機能的適切さ (Functional Adequacy) がある。Kuiken and Vedder (2017) は、第二言語ライティング評価における従来の複雑さ (complexity)、正確さ (accuracy)、流暢さ (fluency) といった CAF と呼ばれる指標では、タスク達成度や意味伝達の適切さを十分に捉えられないと指摘し、新たに機能的適切さの評価枠組みを提案した。それまで内容を評価するための評価軸は様々な提唱されてきていたが、Kuiken and Vedder は①内容、②タスク達成度、③理解のしやすさ、④一貫性と結束性の4つの観点に整理し、6段階のリッカートスケールで学習者の産出を評価するルーブリックを作成した (表2)。

■表2: 機能的適切さ評価する4つの観点

| |
|--|
| 内容 (Content) : 文章に含まれる情報単位の数に適切で関連性があるか? |
| 1: アイデアの数がまったく適切でなく不十分であり、アイデア同士に関連性がない。 2: アイデアの数がほとんど適切ではなく、アイデアに一貫性がない。 3: アイデアの数はある程度適切だが、あまり一貫性がない。 4: アイデアの数は適切であり、十分に一貫している。 5: アイデアの数は非常に適切であり、非常に一貫している。 6: アイデアの数は極めて適切であり、非常に一貫している。 |
| タスク達成度 (Task requirements) : 課題の要件 (ジャンル、発話行為、レジスターなど) をどれだけ満たしているか? |
| 1: 課題の質問および要件のいずれにも答えていない。 2: 質問および要件のうち、半分未満しか答えていない。 3: 質問および要件のうち、約半分に答えている。 4: 質問および要件のうち、大半 (半分以上) に答えている。 5: はほとんどすべての質問および要件に答えている。 6: すべての質問および要件に答えている。 |
| 理解のしやすさ (Comprehensibility) : 読者が文章の目的や主張を理解するために必要な努力の程度はどれくらいか? |
| 1: 文章はまったく理解できない。アイデアや目的が明確に示されておらず、読者が文章を理解しようとしても効果がない。 2: 文章はほとんど理解できない。目的が明確でなく、読者は筆者のアイデアを理解するのに苦労する。読者はほとんどのアイデアや目的を推測する必要がある。 3: 文章はある程度理解できる。いくつかの文は一読では理解しにくい。再読によって目的やアイデアが明確になるが、いくつかの疑問は残る。 4: 文章は理解可能である。ごく少数の文が不明瞭だが、再読によって無理なく理解できる。 5: 文章は非常に理解しやすく、読みやすい。理解の問題はまったくない。 6: 文章は極めて理解しやすく、非常に読みやすい。アイデアと目的が明確に示されている。 |

一貫性と結束性 (Coherence and cohesion) : 文章は一貫性と結束性があるか? (結束装置やストラテジーの使用など)

- 1: 文章はまったく一貫性がない。無関係な文の流れや一貫性の破綻が非常に多い。筆者は照応表現をまったく使っておらず、文章はまったく結束していない。接続表現はほとんど使われず、アイデア同士に関連性がない。
- 2: 文章の一貫性は乏しい。筆者はしばしば無関係な文の流れを使っており、一貫性があってもそれは繰り返しのものによるものである。照応表現はごくわずかにしか使われていない。一貫性の破綻が見られる。文章の結束性も乏しく、アイデアは接続表現でうまくつながっておらず、接続表現自体がほとんど使われていない。
- 3: 文章にはある程度の一貫性があるが、無関係な文の流れや繰り返しが頻繁に見られる。同じ主語が(省略されていても)連続する文で使われることが多い。いくつかの照応表現は使われているが、一貫性の破綻もある。文章にはある程度の結束性があり、一部の接続表現は使われているが、主に接続詞に限られている。
- 4: 文章には一貫性がある。無関係な文の流れは稀にあるが、筆者は時折繰り返しに頼って一貫性を保っている。十分な数の照応表現が使われており、一貫性の破綻は一部あるかもしれないが、文章は結束しており、書き手は接続表現を上手に使っている。使われる接続表現は接続詞に限られないこともある。
- 5: 文章は非常に一貫している。筆者が新しいトピックを導入するとき、それは通常接続詞や接続句を使って行われる。繰り返しは非常に稀である。照応表現は多数使われており、一貫性の破綻はない。文章は非常に結束しており、アイデアは副詞的または動詞的な接続表現によってよく結びつけられている。
- 6: 筆者は、新しいアイデアを接続詞や接続句で統合することで、極めて一貫性を確保している。照応表現は適切に使用されている。無関係な文の流れはほとんどなく、一貫性の破綻はない。文章の構成は非常に結束しており、接続表現(特に接続句、動詞構文、副詞句)を巧みに使用して、アイデア間の関係を記述している。

注. Kuiken and Vedder (2017) より引用, 日本語訳は筆者による

Kuiken and Vedder による、オランダの学習者を対象としたライティング課題の分析では、この新しい尺度が評価者間で高い信頼性を示し、CAF だけでは捉えきれないタスク達成度などを測定できることが確認された。また、CAF とは部分的な関連が見られるものの、必ずしも言語的な正確さが高いことと機能的適切さが一致するわけではなく、誤りがあっても十分に伝達できれば、機能的適切さが高く評価され得ることも示された。さらに、学習者にフィードバックをする際、与えられたトピックに応じたフィードバックを与えることができることも、機能的適切さを評価軸として用いることの利点であると主張している。

機能的適切さが CAF とは異なる指標であるという研究は他にもある。例えば、Révész et al. (2016) は、機能的適切さに類似した指標として、コミュニケーションの適切さ (Communicative adequacy) と CAF の関連を検証し、CAF だけでは学習者の産出の全体的な質を説明するには不十分であると報告している。また、Koizumi and In'nami (2024) も CAF と機能的適切さの関連を調査し、機能的適切さが CAF からある程度予測できることは認めつつも、両者は異なる構成概念であるとし、CAF では説明しきれない学習者のパフォーマンスの側面を機能的適切さが補完すると論じている。De Jong et al. (2007) のタスクの複雑性が流暢性と機能的適切さに与える影響を調査した研究においても、機能的適切さは CAF とは異なる次元の評価であると主張している。以上を踏まえると、これらの先行研究では CAF と機能的適切さが部分的に関連はするものの、別の能力を評価するものであると位置づけている。一方、Lu and Suzukida (2025) では、スピーキングにおいて機能的適切さと複雑性や正確性との関係は見られなかったが、ライティングにおいては複雑性と正確性が高いと機能的適切さが高まったと述べている。また、熟達度の低い学習者は、ライティングにおいて正確性、特に語彙の正確性が機能的適切さと関係があることを報告している。

鈴木 (2023) は、スピーキング能力における機能的適切さについて自動採点の可能性に言及し、研究の必要性について述べているが、ライティング能力における機能的適切さについての自動採点、特に生成 AI を用いている研究はほとんどない。そこで、本研究では、高等学校における英語ライティング指導において、生成 AI から得た機能的適切さに関するフィードバックが学習者のライティングのパフォーマンス

にどのような影響をもたらすかを明らかにし、高等学校の英語ライティング指導における生成AIの効果的な活用法を検討することを目的とする。以下のリサーチクエスチョンを設定する。

RQ1 英語ライティングに対する機能的適切さについての教師の評価とAI評価はどの程度一致するか。

RQ2 生成AIによる機能的適切さに関するライティングへのフィードバックは、学習者のパフォーマンスを向上させるか。

RQ3 学習者は生成AIによる機能的適切さに関するライティングへのフィードバックについてどのように受け取り、修正に生かしているか。

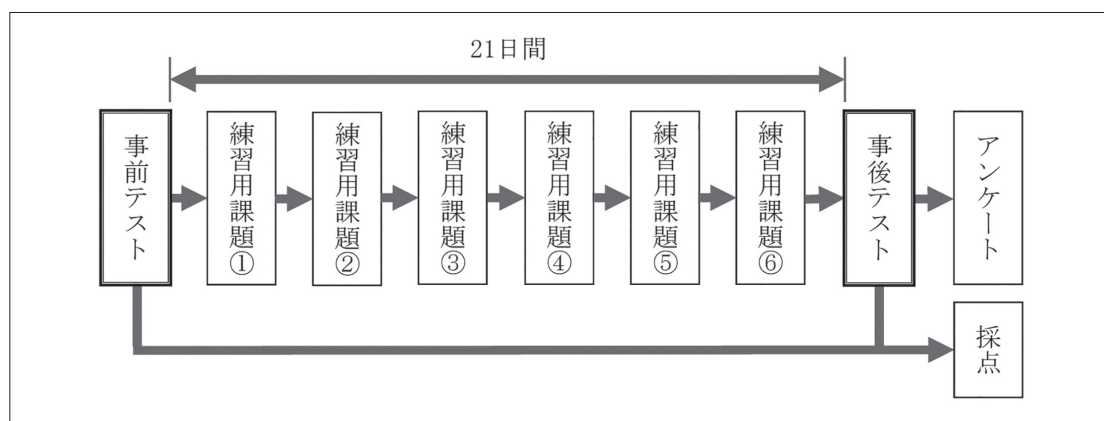
3 方法

3.1 研究対象者

本研究は、神奈川県内の公立高校1校の第1学年の356名の生徒を対象に実施した。対象者の英語レベルは、CEFRのB1相当である。ほぼ100%の生徒が大学入試の一般受験をするため、英語学習に対するモチベーションは高い。教師は日ごろから教科書で扱われている内容に関連したスピーキングやライティングを行っており、クラスによって多少のばらつきはあるようだが、英語を使って自身の考えを発信する活動に対して概して積極的に取り組んでいる。これまでの英語科の授業において、生成AIを活用したり、活用方法を指導したりしていないとのことだったため、本研究の協力を依頼する対象者として適切であると判断した。研究実施にあたり、まず当該校の校長に研究の目的や内容、データの活用の仕方について口頭と書面で説明をし、承認を得た。加えて、対象生徒に対しても、本研究の目的や方法、得られた情報の取り扱いについて十分な説明を行った。あくまでも自由意思での参加であり、参加しない場合にも不利益はないこと、本研究に係るテストや課題の提出状況などの情報は研究目的以外で使用されることなく、学校の成績には一切反映されないこと、研究実施期間中であっても参加を辞退できることなどを口頭で伝え、生徒から同意を得た上で実施をした。

3.2 データ収集

データ収集は、事前テスト、練習用課題、事後テスト、アンケート、採点という手順で行った(図1)。



■図1: データ収集の手順

3.2.1 事前テスト

事前テストは、英語コミュニケーション I の授業内で行われた。テストは、Google フォームで作成し、ライティングのトピックが提示されてから2分間の準備時間を設け、その後10分間でライティングを行う形式で実施した。この形式は、生徒たちが普段から授業内で行っている形式であり、本研究においてもそのまま採用することで、テスト形式に関しての混乱が起きないように配慮した。トピックは“A gift I got from someone”「誰かからもらった贈り物について友人に紹介する」という内容であった。このトピックは、当時行われていた授業の単元におけるゴールタスクのトピック「海外の食べ物をクラスメイトに紹介する」に関連させ、身近な話題に関するものを他人に紹介するというテーマ設定にすることで、授業との関連性を持たせた。なお、事後テストでも同じものを使用したため、事前テストで書いたものに関しては生徒に結果をフィードバックをしていない。

3.2.2 練習用課題

事前テストの後、21日間の間に3～4日おきに Google フォームで作成した練習用課題を配信し、授業外での取り組みを依頼した。練習用課題では、まず事前テストと同様に、トピック提示後に2分間の準備と10分間のライティングを行った。その後、書いた英文を、Kuiken and Vedder (2017)の機能的適切さの評価ルーブリックに基づいた評価とフィードバックを要求するプロンプト(筆者作成、図2を参照)と共に無料版の Google Gemini 2.5 Flash(以下、Gemini)に入力し、得られたフィードバックを基にリライトを行った。プロンプト内の指示に関しては、生徒や教師が使用しやすいように、なるべく短く簡潔な指示にとどめた。生徒に課題へ取り組ませる前に、筆者自身が試しに自作の英文を Gemini に入力し、フィードバックを求めてみた。すると、返ってきたコメントは観点別のものではなく全体的に曖昧であり、しかも高い確率で Gemini が英文全体を書き直して提示してしまうことがわかった。そこで、「それぞれの観点に関して改善のためのアドバイスをできるだけ具体的に個別に提示してください」という指示と、「書き直した全体の改善例は出さないでください」という指示を加えた。その結果、Gemini は機能的適切さの観点ごとに、生徒のライティング改善を促すような具体的なフィードバックを出力するようになった。筆者はこのプロンプトを練習用課題で用いるよう、生徒に伝えた。

あなたは英語ライティングの先生です。

生徒が“Something I Use Everyday”というトピックで、他人に紹介する内容のライティングをしました。これから入力されるその英文を、以下のルーブリックにしたがって観点別に評価をし、それぞれの観点に関して改善のためのアドバイスをできるだけ具体的に個別に提示してください。

※書き直した全体の改善例は出さないでください。

【表2の機能的適切さの評価基準をここに入力】

■図2: 練習用課題に使用したプロンプト

生徒は最初にした英文と、Gemini からの評価とフィードバック、リライトしたもの、気づいたことや学んだことや教師に質問したいことを、すべて Google フォームに入力し送信した。Gemini からのフィードバックを用いたこれらの手順については、水本 (2025) の提示した手順(表1)を参考にしたが、今回は授業外の自律的な学習として練習用課題を作成したため、(5)教師やピアによるレビューと(6)最終版の作成の段階は行わなかった。

練習課題のトピックについては、身近な話題に関するトピックを10個、当該生徒を指導している英語教師に提示して意見を仰ぎ、最終的に表3に示した6つに決定した。

■表3: 練習課題のトピック

| | |
|---|------------------------------------|
| ① | My favorite character |
| ② | Something I use every day |
| ③ | A convenient item for school life |
| ④ | Something that helps me relax |
| ⑤ | The best place I have ever visited |
| ⑥ | A book or movie that I like |

なお、当該学年の英語指導を担当している教師に、クラウドサービス上での定期的な課題の配布と、課題実施の生徒への声かけを、期間中毎日してもらった。生徒には、課題について配信された順番に取り組む必要はなく、取り組めるタイミングで自身で選んだ課題に取り組むように伝えた。

3.2.3 事後テスト

事後テストは、事前テストと同じ条件・トピックで行った。生徒は事前テストと全く同じものであることは知らされていなかった。採点は事前テストと同じ基準・方法で行った。事前テスト同様、事後テストにおいて書いたものについても、Geminiからのフィードバックは受けていない。

3.2.4 事後アンケート

事後テストの実施後に、対象者がAIと機能的適切さに関することをどのように受け取っているかを調査するため、Google フォームを用いてアンケートを実施した。アンケートの項目は、①AIを使ったライティングのフィードバックを使ってみてどう感じたか、②4つの観点(内容・理解のしやすさ・タスク達成度・一貫性と結束性)に基づいたフィードバックについてどのように感じたか、③その他感想やコメント、とした。実際にライティングの学習に使用した率直な感想を書いてもらえるよう、各項目の質問は大まかな内容にとどめ、自由記述形式での回答を求めた。授業時数の関係で、アンケートは授業外での実施となった。

3.2.5 採点

テストの採点は、著者と現職の高校教師1名、教職大学院在籍で英語科教師志望の大学院2年生1名の計3名と、無料版のGeminiで行った(採点時期は2025年7月下旬)。今回Geminiを採用した理由は、当該高等学校を所管する神奈川県がGoogleを学習のインフラとして採用しており、全ての生徒にGoogleのアカウントが配布されていることから、本研究実施後の使用継続の可能性も考慮したためである。また、無料版を選んだのも、すべての生徒が使用可能なツールを用いた場合の効果を検討するためである。

評価基準にはKuiken and Vedder(2017)の機能的適切さ(4観点×6段階)のスケールをそのまま採用した。Koizumi and In'nami(2022)は、教室内のスピーキングテストを用いて、機能的適切さの評価スケールを検討する研究を行ったが、評価スケールは全体的に機能していた一方で、評価者間の一致や、高いレベルの評価に少し問題があったため、1つの解決策として評価スケールを3段階にすることを提案している。しかし、Mizumoto and Eguchi(2023)はIELTSの10段階のルーブリックを用いてAIによるスコアリングを実施し、同じライティングについて2回評価をした際の一致率は52%、1ポイント差だったものは44%であったという点を踏まえ、生成AIによる評価に多少ブレはあるものの、この程度のばらつきは許容範囲であると判断し、評価のスケールを縮小せずそのまま用いた。

採点前には、事後テストが欠損している6人の生徒の事前テストのライティングを練習用として用いて、採点者3名で採点基準のすり合わせを行った。そこで合意した6つのライティングに対する評価と、機能的適切さのスケールをプロンプトとともに Gemini に入力し、採点するように指示をした(図3)。採点は、事後テストが終わってから行った。採点時には、個人が特定されないように処理をし、事前テストと事後テストの組み合わせもわからないように配慮した。

あなたは英語ライティングの先生です。

生徒が“My favorite character”というトピックで、他人に紹介する内容のライティングをしました。

これから入力されるその英文を、以下のルーブリックにしたがって観点別に評価をしてください。

【表2の機能的適切さの評価基準をここに入力】

お送りしたファイルに採点例がありますので、それを参考に採点をお願いします。

採点結果の数字のみ教えてください。

■ 図3: 採点に使用したプロンプト

3.3 データ分析

RQ1については、教師の評価と生成AIの評価の一致度を検討するために、教師と生成AIによる事前・事後テスト両方の得点を、機能的適切さの4つの観点別にそれぞれクロンバックの α 係数を算出した。その後、教師の評価と生成AIの評価について、スピアマンの相関分析を行い、相関係数を算出した。

RQ2に関しては、生成AIを用いたライティングの自律的学習による、生徒のパフォーマンスについての機能的適切さに対する効果を検討するために、機能的適切さの4つの観点別それぞれについて、事前・事後テストの得点を、練習あり・なしのグループに分けて比較、検討した。前提検討として、Levene 検定を行い、群間等分散性(内容:事前 $p=.11$, 事後 $p=.23$; タスク達成度:事前 $p=.13$, 事後 $p=.76$; 理解のしやすさ: 事前 $p=.07$, 事後 $p=.95$; 一貫性と結束性:事前 $p=.27$, 事後 $p=.10$)と、残差の Q-Q プロットをそれぞれ確認し正規性はおおむね良好であったため、2要因の混合計画分散分析を用いて分析を行った。

RQ3に関して、生成AIによる機能的適切さに関するライティングへのフィードバックを生徒がどのように受け取り、修正に活かしているかを見るために、生徒からのコメントとリライトされたライティングを質的に分析した。

なお、統計的な分析については JASP (Version 0.95) を使用した。

4 結果

授業のスケジュールの関係で、全9クラスのうち1クラスは練習のための期間が十分に取れず、もう1クラスは事後テストを実施する時間が確保することができなかったため、当該のクラスの生徒のデータは分析から除外し、対象者は276名になった。さらに、授業への欠席や遅刻等で事前・事後テストの両方、もしくはいずれかを受験できなかった生徒62名も分析から除外した。最終的に、214名分のデータを分析対象とした。授業外における練習課題に一度でも取り組んだ生徒を練習ありグループ($n = 62$)、一度も取り組まなかった生徒を練習なしグループ($n = 152$)としてグループ分けをし、結果を比較した。生徒の練習課題に対する取り組み状況は表4に示した。

■表4: 生徒の練習課題への取り組み状況

| 練習回数 | 生徒の人数 | 練習課題 | 取り組んだ生徒の人数 |
|------|-------|------|------------|
| 0 | 152 | ① | 58 |
| 1 | 37 | ② | 47 |
| 2 | 16 | ③ | 12 |
| 3 | 5 | ④ | 16 |
| 4 | 4 | ⑤ | 5 |
| 5 | 0 | ⑥ | 0 |

4.1 教師の評価と生成AIの評価の一致度 (RQ1)

まず、教師の3人の評価者の一致度を検討するために、事前・事後テストの結果における機能的適切さのそれぞれの観点について、3人のクロンバックの α 係数をそれぞれ算出した(表5)。

■表5: 評価者3人の記述統計と信頼性係数(クロンバック α)

| 観点 | テスト | N | 評価者A | | 評価者B | | 評価者C | | 評価者3人 | | |
|---------|-----|-----|------|------|------|------|------|------|-------|------|----------|
| | | | M | SD | M | SD | M | SD | M | SD | α |
| 内容 | 事前 | 214 | 3.71 | 1.02 | 3.32 | 0.88 | 3.72 | 0.81 | 3.56 | 0.75 | 0.76 |
| | 事後 | 214 | 4.14 | 0.96 | 3.62 | 0.77 | 3.79 | 0.61 | 3.85 | 0.62 | 0.68 |
| タスク達成度 | 事前 | 214 | 3.42 | 1.05 | 3.57 | 0.86 | 4.01 | 0.79 | 3.67 | 0.72 | 0.71 |
| | 事後 | 214 | 4.01 | 1.04 | 3.80 | 0.44 | 4.16 | 0.67 | 3.99 | 0.67 | 0.74 |
| 理解のしやすさ | 事前 | 214 | 3.57 | 0.99 | 3.76 | 0.60 | 3.73 | 0.81 | 3.69 | 0.62 | 0.64 |
| | 事後 | 214 | 4.16 | 1.02 | 3.86 | 0.44 | 3.72 | 0.80 | 3.92 | 0.56 | 0.50 |
| 一貫性と結束性 | 事前 | 214 | 3.15 | 1.07 | 3.13 | 0.79 | 3.43 | 0.76 | 3.23 | 0.70 | 0.71 |
| | 事後 | 214 | 3.77 | 1.06 | 3.24 | 0.54 | 3.65 | 0.60 | 3.23 | 0.70 | 0.73 |

評価者3名の内的一貫性は、内容(事前 $\alpha = 0.76$, 事後 $\alpha = 0.68$), タスク達成度(事前 $\alpha = 0.71$, 事後 $\alpha = 0.74$), 一貫性と結束性(事前 $\alpha = 0.71$, 事後 $\alpha = 0.73$)においては良好, 理解のしやすさ(事前 $\alpha = 0.64$, 事後 $\alpha = 0.50$)は中程度であった。観点によって多少ばらつきがあるものの、総じて一定の信頼性が確認されたため、以降の分析では3名の評価の平均値を教師評価として用いた。

次に、教師とGeminiによる採点の一致度を検討するために、教師の採点結果とGeminiによる採点結果を、機能的適切さの4つの観点別にスピアマンの相関分析を用いて比較をした(表6)。内容については、教師の採点結果($M = 3.69$, $SD = 0.83$)とGeminiの採点結果($M = 4.99$, $SD = 1.00$)には両者の間に中程度の正の相関が認められた($r = .58$, $p < .001$)。タスク達成度についても、教師の採点結果($M = 3.70$, $SD = 0.83$)とGeminiの採点結果($M = 5.14$, $SD = 0.94$)の間に中程度の正の相関が確認された($r = .54$, $p < .001$)。さらに、理解のしやすさでは教師の採点結果($M = 3.84$, $SD = 0.65$)とGeminiの採点結果(M

= 4.52, $SD = 0.76$)の間に中程度の正の相関が認められた($r = .48$, $p < .001$)。一貫性と結束性についても、教師の採点結果($M = 3.32$, $SD = 0.79$)と Gemini の採点結果($M = 4.94$, $SD = 1.10$)の間で中程度の正の相関が確認された($r = .51$, $p < .001$)。このように、すべての観点において、教師と Gemini の採点結果には中程度の相関が見られた。しかしながら、教師の採点結果に比べると、Gemini の採点結果の方が0.68～1.62点高くなっていた。有意差の有無を確認するために、まずシャピロ・ウィルク検定を行い、すべての観点において正規性が棄却($p < .001$)されたので、ウィルクソンの符号順位検定を用いて比較した。結果は、4観点すべてで有意差が確認された(内容: $z = 16.88$, $p < .001$, $r_{rb} = .98$; タスク達成度: $z = 17.34$, $p < .001$, $r_{rb} = .99$; 理解のしやすさ: $z = 14.48$, $p < .001$, $r_{rb} = .90$; 一貫性と結束性: $z = 17.09$, $p < .001$, $r_{rb} = .97$)

■表6: 教師と Gemini による評価の相関および差の検定

| 観点 | 教師 | | Gemini | | スピアマンの相関分析 | | ウィルクソンの符号順位検定 | | |
|---------|------|------|--------|------|------------|-----------|---------------|-----------|----------|
| | M | SD | M | SD | r | p | z | p | r_{rb} |
| 内容 | 3.69 | 0.83 | 4.99 | 1.00 | .58 | < .001*** | 16.88 | < .001*** | .98 |
| タスク達成度 | 3.70 | 0.83 | 5.14 | 0.94 | .54 | < .001*** | 17.34 | < .001*** | .99 |
| 理解のしやすさ | 3.84 | 0.65 | 4.52 | 0.76 | .48 | < .001*** | 14.48 | < .001*** | .90 |
| 一貫性と結束性 | 3.32 | 0.79 | 4.94 | 1.10 | .51 | < .001*** | 17.09 | < .001*** | .97 |

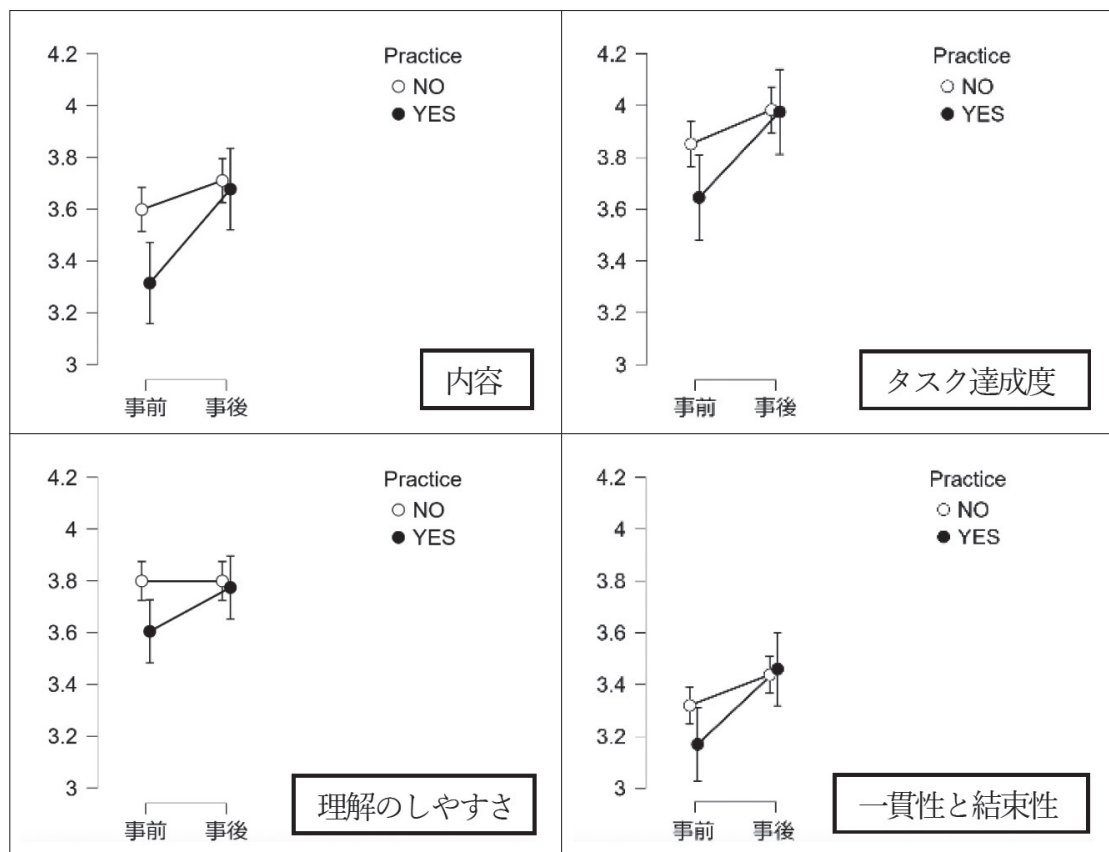
注: * $p < .05$, ** $p < .01$, *** $p < .001$

4.2 Gemini を活用した練習用課題の効果 (RQ2)

Gemini を活用した練習用課題が生徒の英語ライティングにおける機能的適切さに与える影響を検討するために、まず、機能的適切さの各観点ごとに、練習あり($n = 62$)と練習なし($n = 152$)のグループに分けて記述統計を確認した(表7, 図4)

■表7: グループ別の事前・事後テスト結果の記述統計

| 観点 | 時間 | 練習あり($n = 62$) | | | 練習なし($n = 152$) | | |
|---------|----|------------------|------|-------|-------------------|------|-------|
| | | M | SD | 事後－事前 | M | SD | 事後－事前 |
| 内容 | 事前 | 3.32 | 0.85 | 0.36 | 3.60 | 0.66 | 0.26 |
| | 事後 | 3.68 | 0.65 | | 3.86 | 0.53 | |
| タスク達成度 | 事前 | 3.65 | 0.87 | 0.33 | 3.85 | 0.62 | 0.13 |
| | 事後 | 3.98 | 0.66 | | 3.98 | 0.59 | |
| 理解のしやすさ | 事前 | 3.60 | 0.66 | 0.17 | 3.80 | 0.54 | 0.00 |
| | 事後 | 3.77 | 0.50 | | 3.80 | 0.51 | |
| 一貫性と結束性 | 事前 | 3.17 | 0.73 | 0.29 | 3.32 | 0.62 | 0.12 |
| | 事後 | 3.46 | 0.52 | | 3.44 | 0.44 | |



■図4: 機能的適切さの各観点におけるグループ別の事前・事後テスト結果の比較

両群とも平均はおおむね上昇したが、伸び幅は一貫して練習ありの方が大きかった。練習ありの事前と事後の差は、内容が0.36(練習なし0.26)、タスク達成度は0.33(練習なし0.13)、理解のしやすさは0.17(練習なし0.00)、一貫性と結束性は0.29(練習なし0.12)であった。

次に、練習の有無および事前・事後の差に対する統計的検定として、2要因の混合計画分散分析を実施し、時間の効果・練習の効果および交互作用の有無を検討した(表8)。

■表8: 観点別の2要因の混合計画分散分析の結果

| 観点 | 効果 | 項目 | 平方和 | df | 平均平方 | F | p | η^2p |
|--------|------|-----------|--------|-----|------|-------|----------|-----------|
| 内容 | 参加者内 | 時間 | 4.96 | 1 | 4.96 | 15.98 | <.001*** | .07 |
| | | 時間×練習 | 1.39 | 1 | 1.39 | 4.47 | .04* | .02 |
| | | Residuals | 65.84 | 212 | 0.31 | — | — | — |
| | 参加者間 | 練習 | 2.22 | 1 | 2.22 | 4.25 | .04* | .02 |
| | | Residuals | 110.61 | 212 | 0.52 | — | — | — |
| タスク達成度 | 参加者内 | 時間 | 4.70 | 1 | 4.70 | 13.96 | <.001*** | .06 |
| | | 時間×練習 | 0.87 | 1 | 0.87 | 2.59 | .11 | .01 |
| | | Residuals | 71.42 | 212 | 0.34 | — | — | — |

| 観点 | 効果 | 項目 | 平方和 | df | 平均平方 | F | p | η^2p |
|---------|------|-----------|--------|-----|------|-------|----------|-----------|
| タスク達成度 | 参加者間 | 練習 | 1.01 | 1 | 1.01 | 1.93 | .17 | .01 |
| | | Residuals | 111.62 | 212 | 0.53 | — | — | — |
| 理解のしやすさ | 参加者内 | 時間 | 0.63 | 1 | 0.63 | 2.83 | .09 | .01 |
| | | 時間×練習 | 0.63 | 1 | 0.63 | 2.83 | .09 | .01 |
| | | Residuals | 47.24 | 212 | 0.22 | — | — | — |
| | 参加者間 | 練習 | 1.06 | 1 | 1.06 | 2.91 | .09 | .01 |
| | | Residuals | 77.43 | 212 | 0.37 | — | — | — |
| | | | | | | | | |
| 一貫性と結束性 | 参加者内 | 時間 | 3.68 | 1 | 3.68 | 16.06 | <.001*** | .07 |
| | | 時間×練習 | 0.65 | 1 | 0.65 | 2.84 | .93 | .01 |
| | | Residuals | 48.57 | 212 | 0.23 | — | — | — |
| | 参加者間 | 練習 | 0.36 | 1 | 0.36 | 0.87 | .35 | .00 |
| | | Residuals | 87.48 | 212 | 0.41 | — | — | — |
| | | | | | | | | |

注: * $p < .05$, ** $p < .01$, *** $p < .001$

結果として、交互作用が有意だったのは「内容」のみであった(時間×練習: $F = 4.47$, $p = .04$, $\eta^2p = .02$)。その他の観点では交互作用は有意でなかった(タスク達成度: $F = 2.59$, $p = .11$, $\eta^2p = .01$; 理解のしやすさ: $F = 2.83$, $p = .09$, $\eta^2p = .01$; 一貫性と結束性: $F = 2.84$, $p = .93$, $\eta^2p = .01$)。時間の効果は、内容($F = 15.98$, $p < .001$, $\eta^2p = .07$)、タスク達成度($F = 13.96$, $p < .001$, $\eta^2p = .06$)、一貫性と結束性($F = 16.06$, $p < .001$, $\eta^2p = .07$)で有意、理解のしやすさでは有意ではなかった($F = 2.83$, $p = .09$, $\eta^2p = .01$)。練習の効果は、内容のみ有意($F = 4.25$, $p = .04$, $\eta^2p = .02$)で、その他は有意でなかった。

交互作用が有意であった内容について単純主効果をみると(表9)、時間の効果は練習ありで有意($F = 10.69$, $p = <.01$)だが、練習なしでは有意ではなかった($F = 3.37$, $p = .09$)。また、グループ間差は、事前で練習なし>練習ありで有意差が見られた($F = 6.91$, $p = <.01$)が、事後では有意差なしであった。($F = 0.15$, $p = .70$)。練習あり群は事後に得点が伸びたことで、事前に見られた有意差が解消されたことがわかる。

まとめると、Geminiを活用した練習課題は、機能的適切さにおける内容の観点に対して効果があった。他の観点に対しては、全体として時間経過による向上は確認できたものの、統計的に有意な差は確認できず、限定的であったといえる。

■表9: 内容の観点における単純主効果

| 時間 | | | | | |
|------|------|----|------|-------|--------|
| | 平方和 | df | 平均平方 | F | p |
| 練習あり | 4.08 | 1 | 4.08 | 10.69 | <.01** |
| 練習なし | 0.95 | 1 | 0.95 | 3.37 | .09 |
| 練習 | | | | | |
| | 平方和 | df | 平均平方 | F | p |
| 事前 | 3.56 | 1 | 3.56 | 6.91 | <.01** |
| 事後 | 0.05 | 1 | 0.05 | 0.15 | .70 |

注: * $p < .05$, ** $p < .01$, *** $p < .001$

4.3 学習者の Gemini からのフィードバックに対する認知と応答 (RQ3)

Gemini を活用した英語ライティングの練習課題に関するアンケートは146件の回答を得た(回収率68%, 146/214)。なお、練習なしのグループもアンケートに回答しており、その結果も含んでいる。

4.3.1 肯定的なコメント(76%, 111/146)

全体として、生徒はAIを活用した学習について、全体的に非常に肯定的な意見を寄せていた。特に、以下の点が特筆すべき傾向として見られる(表10)。

■表10: アンケートの肯定的なコメント

| 分類 | コメントの抜粋 | 分析 |
|------------------------|---|---|
| 効率性と即時性 (25%) | 「すぐに応答が返ってくるので、時短になる」 「集中力が続かないときでもすぐに使える」 「いつでもどこでも添削してもらえる」 | AIの即時的なフィードバックと迅速な対応を高く評価していた。従来の教師による添削では時間を要した部分が、AIの活用により効率化され、学習サイクルが速くなるというメリットを感じていることが窺える。 |
| 的確なフィードバックと学び (24%) | 「自身の文法ミスやスペルミスを正確に指摘してくれた」 「より自然な表現や語彙を教えてくれた」 「なぜ間違っているのか、どうすればよいのかを具体的に示してくれた」 | AIの具体的なフィードバックで弱点・改善点が明確になり、学習効果が高まり自律学習につながったとの声が多かった。さらに「新たな発見・気づき」が多く報告され、AIが新たな視点を与えていることが示された。 |
| ライティング力の向上 (15%) | 「以前より文章が長く書けるようになった」 「話のつながり方や、より適切な表現が深く理解できた」 「言いたいことをより正確に伝えられるようになった」 「構成を工夫するようになった」 | AIを活用した練習を繰り返すことで、自身のライティング能力に手応えを感じている生徒もいた。機能的適切さの観点に関わる部分で改善できたというコメントもあったことから、AIによるフィードバックが活かされていることが窺える。 |
| 心理的な安心感 (12%) | 「あまり先生に聞くことはなかったからいい体験だった」 「AIの添削で安心して学べた」 「気軽にアドバイスを求めることができるので、AIを参考にして英文を書くのもいいなと思った」 「スペルミスや文法の間違いを気にせずに作文をすることができた」 | 教師に直接質問することに抵抗を感じていた生徒にとって、AIは安心して学べる存在となっていたようである。些細な疑問でも、気軽にAIに尋ねられる環境が提供されていることがわかる。 |

4.3.2 批判的なコメント(24%, 35/146)

肯定的な意見が多数を占める一方で、批判的な意見も見られた(表11)。これらの意見は、英語ライティング指導に生成 AI を取り入れていく際に必要な学習者視点を提供している。

■表11: アンケートの批判的なコメント

| 分類 | コメントの抜粋 | 分析 |
|-----------------------|---|---|
| ツールとしての使いづらさ (10%) | 「評価基準をコピーする作業が手間」 「パソコンで打つのが大変だったり、時間がかかったりする」 「手で書く方がやりやすい」 | AI とのやり取り、そもそも PC でのタイピングに慣れていないことによる不便さを訴える意見があった。継続的に使用することで、解消できるか。 |
| 根本的な英語力の不足 (8%) | 「そもそも書きたい単語のスペルがわからなかった」 「文法が曖昧で混乱する」 「話の広げ方が難しく、トピックによってはあまり書けなかった」 「なぜその表現が良いのか、理由がわからないことがある」 | 根本的な語彙力や文法力に課題を感じている生徒もいた。AI はあくまでツールであり、自身の基礎力がなければ最大限に活用することは難しいと、生徒が感じていることが窺える。 |
| AI への依存に対する懸念 (6%) | 「AI に頼りすぎると、それに頼った文章しか作れなくなる」 「このままだと人間である意味がなくなる」 「自身の能力を潰されている気がする」 | AI の利便性からくる依存性や、学習効果に対する懸念も一部で見られた。これらの意見は、AI を何のためにどのように使うのか、という根本的な問題について考える必要性を示唆している。 |

4.3.3 フィードバックに基づいた修正

練習ありグループの中で、事前テストから事後テストにかけてライティングの評価が顕著に向上した2名の生徒の練習用課題におけるライティングについて、Gemini からのフィードバック(資料1)を受けてどのように修正が行われたかを分析した。一人目は、練習②の“Something I use every day”をトピックとして書いたものである。なお、生徒のライティングは、スペルミスや文と文の間のスペースが空いてないものなどもそのまま載せている。

【修正前】

Hello! I'm going to show you what I use every day.It's the smartphone. In modern Japan, the smartphone has become an indispensable tool for life because it plays various roles such as a game, a stopwatch, a timer, an alarm clock, a clock, and a dictionary.Smartphones can also serve as a substitute for telephones.That's why I'm always on my phone.Thank you for reading!

【修正後】※下線, 太字は筆者による。Gemini からのフィードバックを踏まえて変更された部分。

Hello! I'm going to introduce you to a tool I use all the time. It's a smartphone. Smartphones are essential to life in modern Japan because they allow you to play games, make phone calls, read e-books, and use the internet to do research. **I play a game called Blue Archives on my**

smartphone every day. This game is operated by a company called Yostar and is extremely popular. I also read light novels every day using an application called Kakuyomu. Light novels are stories with light content that are easy to read even for beginners. That's why I use my smartphone every day to play games and read e-books. Please tell me about the tools you use every day. Thank you for reading!

【Geminiからの改善のためのフィードバックの概要】

- ① スマートフォンの機能を具体的にどのように使っているかを説明する(内容)
- ② 読者に対して直接語りかけのような表現(例: "Imagine a world without it!" や "You might be surprised how much it helps me..." など)を加える(タスク達成度)
- ③ 文を主語から始めるだけでなく、時折、副詞句や前置詞句を文頭に持つことで、リズムが生まれより自然な英語になる(理解のしやすさ)
- ④ indispensableだけでなくessential, vital, crucialなど、文脈に応じて使い分けることで文章に深みが増す(理解のしやすさ)
- ⑤ 多機能性から「肌身離さず持っている」という結論に至るまでの論理的なつながりを、よりスムーズにする接続表現を検討(一貫性と結束性)
- ⑥ 具体例から抽象的な結論へ橋渡しとなるような文(例: All these features combined make my smartphone an indispensable companion for daily life.)の挿入を検討(一貫性と結束性)

この中で、修正する際に基になったと思われるフィードバックは①(内容)、②(タスク達成度)、④(理解のしやすさ)である。ゲームと小説という具体的なスマホの機能の使用が追加され、indispensableがessentialに変更されている。さらに、読者に問いかけるような1文を追加している。この生徒は、フィードバックの内容を自分なりに解釈して自分の言葉で修正を加えており、Geminiが提案している文をそのまま使うことはしていない、という点が興味深い点である。

二人目も、練習②の“Something I use everyday”をトピックとして書いたものである。

【修正前】

I'll introduce my eye mask.
My brother gave me it about half years ago, and I wear it every night now.
It's printed character's face who is a Waddle Dee.
I love it very much now. It's so cute and help my good sleep.
I can't imagine if I don't get it. How could I sleep without it?
If you have not ever use eye mask, you should try it.

【修正後】※下線, 太字は筆者による。Geminiからのフィードバックを踏まえて変更された部分。

I'll introduce my eye mask.
My brother gave it to me about half a year ago. Since then, I wear it every night now.
It's printed character's face who is a Waddle Dee. My brother and I like Kirby's Dream Land series. Especially, I like Waddle Dee. So, the face on it comforts me.
I love it very much now because it helps my good sleep.
Eye mask block rights, so I can focus on my sleep.
I can't imagine life without it. How could I sleep?
If you have not ever use eye mask and want to good sleep, you should try it.

【Gemini からの改善のためのフィードバックの概要】

- ①アイマスクの具体的な利点(なぜそれが良い睡眠につながるのか、など)やキャラクターの情報について、もう少し詳しく記述する(内容)
- ②「もしあなたが良い睡眠を探しているなら、ぜひアイマスクを試してみてください」のように、読者へのはたらきかけを意識する(タスク達成度)
- ③“gave me it”を“gave it to me”に、“half years”ではなく“half a year”、“I can't imagine if I don't get it”は、“I can't imagine not having it”や“I can't imagine life without it”とするとより自然である(理解のしやすさ)
- ④“and”よりも、時間の経過を示す副詞句(例:“Since then”または“Ever since then”)を使って、アイマスクをもらってから毎晩使っているという関係性をより明確にする(一貫性と結束性)
- ⑤理由を説明する際に“because”などの接続詞を使うと、より論理的なつながりになる。例:“I love it very much because it's so cute and helps me get good sleep.”(一貫性と結束性)
- ⑥“It's printed character's face who is a Waddle Dee.”の後に、“This design makes it even more special to me.”のように、補足的な情報を加えることで、前の文とのつながりを強めることができる。(一貫性と結束性)

この生徒は、Gemini からのフィードバックをすべて取り入れるように修正していた。③と④、⑤のように単語やフレーズレベルの修正に関しては、提案された案をそのまま採用している。しかし、その他の点に関しては、具体的な英語の修正案が提示されていなくとも、自分の言葉で表現をして修正を試みている。例えば、①については、“Eye mask block rights, so I can focus on my sleep.”や“My brother and I like Kirby's Dream Land series.”といったように、自身の持っている語彙や文法を駆使して表現を試みている。

このように、事前テストから事後テストにかけてライティングのパフォーマンスが向上した二人の生徒は、自分の言葉で表現をしようとする姿勢が共通していた。

5 考察

5.1 英語ライティング指導における生成 AI を活用した機能的適切さについてのフィードバック

本研究では、高等学校における英語ライティング指導の中で、生徒の自律的学習を促すため、機能的適切さという構成概念を用いた生成 AI によるフィードバックの効果を Gemini を用いて実証的に検証した。まず、教師評価と Gemini による機能的適切さの評価は中程度の一致であり、Gemini は教師よりも一貫して高めの評価をする傾向が見られた。以下は一人の生徒が書いた英文だが、教師の評価が(内容:1.33, タスク達成度:2, 理解のしやすさ:2.67, 一貫性と結束性:3)なのに対して、Gemini の評価は(内容:4, タスク達成度:4, 理解のしやすさ:4, 一貫性と結束性:4)であった。

My last year birthday, I got a gift which my mother gave me. I am using it everyday when I go to school. It is very convenient for me to go to school, because it is simple shape but it can have lot of things .

これは、Gemini が教師に比べ相対的に寛容な評価を行うことを意味しており、教師による最終的な

補正が必要であることを示している。英文によって、教師評価とAI評価が一致している観点も多くあるため、プロンプトを調整して一概に2つの評価を近づけようとするのは難しいように思われる。Xi(2023)は、AI評価が「ブラックボックス」であり、用いられているモデルやアルゴリズムが説明できない点は問題であるとしながらも、最終的な人間による評価の必要性和、対象がローステークスなものであるならば、評価が多少ブラックボックスでも問題ないだろうと述べている。つまり、生徒がライティングの練習用課題に取り組む際、形成的なフィードバックを得るためにGeminiを活用することは問題ないと考えられる。加えて、川村他(2019)でも述べられていたような教師の負担を軽減することにもつながるだろう。

練習用課題の効果に注目すると、全体としての時間の効果は内容・タスク達成度・一貫性と結束性の観点において有意であった(理解のしやすさについては非有意)。内容の観点では、練習×時間の交互作用が有意で、練習ありグループのみ、事前から事後へ評価が有意に向上し(Holm 補正後も有意)、練習なしグループの向上は有意ではなかった。さらに、事前では練習なしグループの評価が有意に高かったが、事後ではその差は消失した。したがって、Geminiを活用した練習用課題は、内容の観点に限って効果を示し、アイデアの追加や充実を促した可能性が高い。一方、タスク達成度と一貫性・結束性は時間の効果のみで、練習の主効果・交互作用はいずれも有意ではなかった。これらの観点は、タスクの文脈に応じて内容や表現を調整する力や、接続表現や論理展開に関わる高度な文章構成力を要するため、今回の短期間・自律型の練習ではAIフィードバックが十分に運用されにくかったと考えられる。先行研究でも、自動フィードバックは文法や語彙といった表層的側面の修正に偏りやすいことが指摘されており(Liu & Yu, 2022)、本研究の結果も同様の傾向を示したと考えられる。加えて、理解のしやすさは時間の効果も練習の効果も確認できず、この観点については教師による明示的指導や支援といったような足場かけが必要であることが示唆される。事後テストにおいてパフォーマンスが向上した二人の生徒のライティングの分析からは、Geminiからのフィードバックを参考にして、自分の伝えたい内容を自分の言葉で表現しようとする姿勢が見られた。AIから提案された修正案を、一旦自分の伝えたい内容と照らして考えてから書く、ということが必要だと考えられる。

実際の数字には表れなかったものの、アンケート結果において、生徒たちは生成AIが英語ライティングの学習において非常に有効なツールであり、特に効率的なフィードバックや学習へのモチベーション向上に大きく貢献している、と感じていたことがわかった。Geminiからのフィードバックに対して全体の76%が肯定的に受け止めており、効率性や即時性、心理的な安心感が高く評価されていた。特に、「AIに直されたところが新たな発見になった」との記述は、水本(2025)でも述べられているように、生成AIによるフィードバックが言語習得における学習者の気づきを促していることを示唆している。一方で、生成AIからのフィードバックから言語的な気づきを得るためには、生徒自身の基礎的な英語力が不可欠である可能性も見えてきた。プロンプトの中で具体的なフィードバックをするように指示をしても、機能的適切さに関するフィードバックは抽象的な指摘にとどまる場合も少なくない。山下(2024)も、学習者が自ら修正につなげられないタイプのフィードバックがあることを指摘している。そのような場合、生徒自ら言語的な気づきを得られる程度の基礎力がなければ、フィードバックを活かすことができない。実際、Lu and Suzukida(2025)でも、熟達度の低い学習者に対しては、機能的適切さについてではなく、語彙の正確性に焦点を当てた指導が効果的である可能性を指摘し、習熟度に応じた支援の必要性を述べている。つまり、生成AIを用いた自律的な学習ができるようになるまでの足場かけは、教師が必ず担い、指導・支援すべき部分であるといえる。また、生徒からの「便利すぎて人間の意味がなくなりそう」との意見は、生成AIへの過度な依存への懸念を反映しているといえる。生成AIを使うことを目的にせず、生成AIを使って自身の能力を高めていくという姿勢が重要であるというような指導も、生成AIを英語教育に取り入れる際には必要だろう。

5.2 教育的示唆

本研究の結果から、英語ライティング指導における機能的適切さに関する生成 AI の活用について、複数の教育的示唆を得られた。

第一に、生成 AI によるフィードバックの活用は教師の役割を不要にするのではなく、むしろ役割を明確にする、ということである。生成 AI の機能的適切さに関する評価は、教師の評価と一定の一致率は認められたものの、一貫性があるわけではなく、また学習者がその内容を必ず理解できるとは限らないということが、本研究でも明らかになった。そのため、教師は生成 AI からのフィードバックを鵜呑みにさせず、学習者が批判的に受け止め、「なぜこの表現が良いのか」を生徒自身が考えるような場を設定し、自らの表現に適用できるような思考力が育まれる授業をデザインすることが重要だといえる。Wang and Wang (2025) は、L2 学習者のライティングにおける批判的 AI リテラシー (Critical AI Literacy; CAIL) が「批判的気づき」「立場性」「対話方略」「アフォードランスの評価」という4つの要素から構成されているとし、それらを高めていくためには、プロンプト設計・出力の検証・AI と人間フィードバックの比較検討を授業に組み込むことで、CAIL を「追加スキル」としてではなく、従来のライティング教育に統合していく必要性を指摘している。具体的な活動としては、生成 AI に入力したプロンプトと出力されたフィードバックをクラス全体で共有し、それを基にどのように修正をするのが適切かを議論させるような活動や、学習者がお互いのライティングに対するフィードバックを作成し、生成 AI からのフィードバックと比較するようなタスクを行うことで、学習者の「気づき (noticing)」を促進できると思われる。さらに、生成 AI からのフィードバックが抽象的で理解が難しいときには、AI と対話することで理解を深めるような練習を行う場をデザインすることも考えられる。本研究においては、水本 (2025) の ChatGPT を学習者に使用させる際の手順のうち、「必要に応じて ChatGPT に追加質問」という部分についてはあまり強調しなかったため、その点に関しても授業内で丁寧に指導してから自律的学習につなげるなど時間をかけて行うことで、ライティングのパフォーマンスへの効果は違ったものになる可能性もある。このように、生成 AI をあくまでも1つの学習ツールとして明確に位置づけ、教材として扱うことで、生徒の CAIL 向上にもつながると考えられる。そして、そのような学びを教室内で臨機応変にファシリテートする役割こそ、教師が担うべきものとなるだろう。

第二に、生成 AI によるフィードバックは即時性と豊富な出力を特徴とし、学習者にとって「安全な練習の場」を提供できる点が挙げられる。アンケートにおいても学習者の多くが肯定的に評価しており、心理的負担を軽減しながら学習を継続できる可能性が示された。従来、教師による個別フィードバックは時間的・労力的に制約が大きく、十分な量を提供することが難しかった。しかし、本研究において、教師による機能的適切さに関する評価との中程度の相関を示すことがわかったため、生成 AI はその補完的役割を担うことが可能だろう。特に「内容」の改善に寄与したという結果は、学習者が新たなアイデアを得たり発想を広げたりすることができるという点で、教育的意義が大きいといえる。逆に、その他の観点に関しては、本研究における自律的学習では有意な向上は見られなかったため、練習用課題の実施前や実施中に、より丁寧な説明や指導を行い、3つの観点に関しても向上につながるようにする手立てが必要であるといえる。アンケートの中で、「あまり先生に聞くことはなかったからいい体験だった」と答えていた生徒がいたように、教師に質問するのが苦手な生徒もいる。そのような生徒たちが生成 AI に気軽に質問して学んでいけるように、教師が授業の中で効果的な活用法を積極的に指導していくことが求められるだろう。

第三に、AI の利用には過度の依存が伴う危険性がある点に留意すべきである。本研究においても「AI に頼りすぎると、それに頼った文章しか作れなくなる」という生徒の声があった。AI への過度な依存は学習者の自律的な学習態度を損なう危険があるということを、自認してコメントしていた生徒がいたことは大変興味深い。全員がそれを自ら感じているわけではないだろう。高校の教育現場では、生成 AI を使わずに書く活動を明確にゴールタスクに据えることで、生成 AI の活用が生成 AI を使わずに書くことがで

きるようになるための手段と捉えられるだろう。單元ごとの指導で様々なライティングを扱い、それぞれに対する生成AIを活用した練習を授業内外で設計して教師が指導していくことで、生徒自身の英語力伸長のための生成AIの補助的な活用方法について学んでいくことができる。機能的適切さを評価することの教育的意義も大きい。誤りが残っていても、目的達成に十分であれば高く評価されるという視点は、学習者の自信を高め、実用的な言語運用能力の育成につながる。AIフィードバックを機能的適切さの観点と組み合わせて導入することは、学習者に「何が伝わる文章なのか」を理解させる手助けとなり、単なる文法的正確さの追求にとどまらない実践的なライティング指導を可能にするだろう。

5.3 研究の限界と今後の展望

本研究における限界と今後考えられる研究について述べる。まず、対象となった学習者が限定的であった点である。本研究では特定の高校に通う生徒を対象としたため、学習者の英語力や学習背景が比較的均質であった。したがって、本研究で得られた知見を異なる学力層や多様な学習背景をもつ学習者へ一般化するには注意が必要である。例えば、英語力が初級段階の学習者にとって生成AIの機能的適切さに関するフィードバックが過剰な負担となる可能性や、逆に上級者にとっては既知の情報にとどまり十分な学習効果を生まない可能性もあり、さらには中学生や大学生にとっては効果が変わってくる可能性がある。今後の研究として、異なる熟達度の学習者における調査をすることで、より機能的適切さに関する生成AIのフィードバックの効果を明らかにしていくことができるだろう。また、参加者のアンケートの中に、フィードバックの量が多すぎるという内容のコメントが複数見られたため、フィードバックの個数を限定するようにプロンプトを調整して実施することも考えられる。あるいは、生成AIからのフィードバックを読み取る際のガイドラインを作成し、着目するポイントを示すことも効果的かもしれない。

また、データ収集の方法と量に制約があった点も限界点として挙げられる。本研究では、実践が授業外での実施にとどまったため、長期的な学習成果や学習者の態度変容を追跡することはできなかった。そのため、生成AIからのフィードバックが持続的にライティング能力の向上に寄与するのか、あるいは一時的な効果にとどまるのかは今後の検討課題である。生徒からのアンケートの記述の一部からは、「構成を工夫するようになった」や「話のつなげ方や、より適切な表現が深く理解できた」など、機能的適切さに関する学びが起っていたことが窺える。より長期的な縦断的研究や実験的デザインを取り入れることで、機能的適切さに関する生成AI活用についての中長期的効果をより明確に検証できるだろう。今後の研究として、間隔を空けて行ったり、学期中でなく長期休みの課題として行うことなども考えられる。今回は学期中で、他の教科の学習も忙しい中で行うには負担が大きすぎた可能性がある。また、事前テスト後に、1回目の練習は教師の指導の下、授業中に組みませるなど、手順が明確に伝わった状態で自律的学習を始める方が取り組みやすい可能性がある。加えて、実践的な授業内での研究もまた必要だろう。教師がAIフィードバックを授業の中でどのように組み込むのか、学習者がそれをどのように受け止め、自律的学習へとつなげるのかを具体的に検証することで、生成AIと人間の協働による新たなライティング教育のモデルを構築できる可能性がある。

ループリックや採点に関しても、さらなる研究の余地がある。今回は、Kuiken and Vedder(2017)の機能的適切さのループリックのみを使用した。語彙や文法といったような言語形式に関するループリックも加えて同時に評価することで、学習者のライティングパフォーマンスの伸びをより明確に把握することができるだろう。本研究では生成AIによる採点を行う際に、プロンプトに入力した採点例が6つのみであり、1から6の採点が包括的に入るような例になっていなかった。また、AIには採点スコアのみ提示し、なぜその採点になるかという説明をプロンプトに入れていなかった。プロンプトに採点基準についてのより詳細な情報を入力することで、AIによる評価を教師による評価により近づけられる可能性もあるかもしれない。

さらに、生成AIからのフィードバックの質に関する分析も不十分である。本研究では機能的適切さの

観点を導入したが、AIのフィードバックそのものの質や特徴についての検証や、それらに対して学習者がどのように受け止め修正に活かすのか、修正につながりやすいものとつながりにくいものの違いなどについて質的に詳らかにできれば、それらは生成AIから適切なフィードバックを出力させるためのプロンプト改善につながるだろう。また、異なるAIモデルを比較し、どのモデルがどのような観点で有効性を示すかを明らかにすることや、生成AIと教師のフィードバックを組み合わせる最適な方法を探ることも考えられる。

6 おわりに

本研究は、高等学校の英語ライティング指導において、機能的適切さに関する生成AIを用いたフィードバックの効果を検証した。21日間という限られた期間内での実施ではあったが、本研究で行った実践が生徒のライティングのパフォーマンスのうち、機能的適切さの内容に関わる部分の向上に寄与しているということが確認できた。また、教師の評価と生成AIの評価がある程度一致するということや、生徒は機能的適切さに基づいた生成AIからのフィードバックに関して肯定的に捉え、学びにつながるという実感をもっていたことがわかった。教師が行っていたライティング指導の一部を生成AIに任せられる可能性や、その際に使用できる評価の基準として、機能的適切さの有用性も見えてきた。本研究が、生徒が生成AIを活用して自律的に英語ライティングを学ぶための指導設計、およびその関連の研究の一助となれば幸いである。

謝辞

本研究を実行する機会を与えてくださった公益財団法人 日本英語検定協会、関係者の皆様、ならびに、選考委員の先生方に、心より御礼申し上げます。実施に当たって、小泉利恵先生からは、大変お忙しい中、本研究に関する多くのご助言、アイデアを賜りました。また、研究依頼校の先生方や生徒の皆さんからは、熱心な取り組みと貴重な振り返りのコメントを頂きました。加えて、2名の採点協力者は真摯に、そして迅速に採点に取り組んでくださいました。研究にご協力くださった方々に心よりお礼申し上げます。なお、本報告書に関して、開示すべき利益相反関連事項はありません。

引用文献

- Allen, T. J., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-foreign-language learners' preferences for editing and proofreading strategies. *RELC Journal*, 0(0). <https://doi.org/10.1177/00336882241262533>
- De Jong, N., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2007). The effect of task complexity on fluency and functional adequacy of speaking performance. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, I. Vedder, & K. Contactforum (Eds.), *Complexity, Accuracy and Fluency in Second Language Use, Learning & Teaching* (pp. 53-63). KVAB, Universa Press, Wetteren.
- Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology*, 29(1), 1-27. <https://doi.org/10.64152/10125/73629>
- Huang, J., & Mizumoto, A. (2025). The role of generative AI in mediating L2MSS and engagement with written feedback in EFL learning: A structural equation modeling approach. *Annual Review of Applied Linguistics*, 1-17. <https://doi.org/10.1017/S0267190525000029>
- 門田修平 (2024).『AIフル活用! 英語発信力トレーニング』コスモピア.
- Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 123, Article 103332. <https://doi.org/10.1016/j.system.2024.103332>
- 川村晃市・柏木治美・康敏 (2019).「教師の英作文添削を支援する誤り検出システム」『日本教育工学会研究報告集』19(1), 83-88. <https://hdl.handle.net/20.500.14094/90005737>
- Koizumi, R., & In'nami, Y. (2022). Assessing functional adequacy using picture description tasks in classroom-based L2 speaking assessment. *JLTA Journal*, 25, 60-79. https://doi.org/10.20622/jltajournal.25.0_60
- Koizumi, R., & In'nami, Y. (2024). Predicting functional adequacy from complexity, accuracy, and fluency of second-language picture-prompted speaking. *System*, 120, Article 103208. <https://doi.org/10.1016/j.system.2023.103208>
- Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing*, 34(3), 321-336. <https://doi.org/10.1177/0265532216663991>
- Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78-105. <https://doi.org/10.125/73480>
- Lu, X., & Suzukida, Y. (2025). Relationships of linguistic complexity and accuracy to functional adequacy across L2 English speaking and writing: Roles of L2 proficiency and task type. *International Journal of Applied Linguistics*. Advance online publication. <https://doi.org/10.1111/ijal.12775>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), Article 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- 水本 篤 (2025).「AIとライティング教育—英語ライティング教育における生成AIの活用と課題—」. 李 在鎬・青山玲二郎(編著)『AIで言語教育は終わるのか? 深まる外国語の教え方と学び方』(第8章) くろしお出版.
- 文部科学省 (2018).「高等学校学習指導要領(平成30年告示)解説 外国語編 英語編」https://www.mext.go.jp/a_menu/shotou/new-cs/1407074.htm
- 文部科学省 (2024).『初等中等教育段階における生成AIの利活用に関するガイドライン Ver. 2.0』https://www.mext.go.jp/a_menu/other/mext_02412.html
- Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828-848. <https://doi.org/10.1093/applin/amu069>
- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, 36(2), 187-209. <https://doi.org/10.1017/S0958344023000265>
- 鈴木駿吾 (2023).「第二言語スピーキング能力の機能的達成度は自動判定できるか?」『日本音響学会誌』79(3), 177-183. https://doi.org/10.20697/jasj.79.3_177
- Tate, T. P., Steiss, J., Bailey, D., Graham, S., Moon, Y., Ritchie, D., Tseng, W., & Warschauer, M. (2024). Can AI provide useful holistic essay scoring? *Computers and Education: Artificial Intelligence*, 7, Article 100255. <https://doi.org/10.1016/j.caeai.2024.100255>
- Voss, E., Cushing, I., Ockey, G., & Yan, D. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4-5), 514-533. <https://doi.org/10.1080/15434303.2023.2273799>
- Xi, X. (2023). Advancing language assessment with AI and ML—leaning into AI is inevitable, but can theory keep up? *Language Assessment Quarterly*, 20(4-5), 357-376. <https://doi.org/10.1080/15434303.2023.2291488>
- Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*, 67, 101187. <https://doi.org/10.1016/j.jslw.2025.101187>
- 山下美朋 (2024).「激震の英語ライティング教育」山中司 [編] (2024)『AI・機械翻訳と英語学習 教育実践から見えてきた未来』(pp.35-57) 朝日出版社.
- 山下美朋・山中司・杉山滉平 (2024).「AIツールを活かした英語ライティング授業—英文作成支援ツールTransableを導入して—」『立命館高等教育研究』24, 75-87. <https://ritsumei.repo.nii.ac.jp/records/2000853>

資料1-1: Geminiからのフィードバックの例(一人目)..... (事前テストから事後テストにかけてライティングの評価が顕著に向上した2名)

| | 内容 | タスク達成度 | 理解のしやすさ | 一貫性と結束性 |
|-------|------|--------|---------|---------|
| 事前テスト | 2.67 | 2.67 | 3 | 2.33 |
| 事後テスト | 4.67 | 4.67 | 4.67 | 4.33 |

評価と具体的な改善アドバイス

内容 (Content)

評価: 3

アイデアの数はある程度適切ですが、あまり一貫性がありません。スマートフォンが「日用品」であること、そしてその多機能性については述べられていますが、具体的な「あなたが」どのように使っているかという情報が不足しており、読者に紹介するという目的からすると、個人的な視点やエピソードがほとんどありません。

改善のためのアドバイス:

具体的な使用例の追加: スマートフォンが「ゲーム、ストップウォッチ、タイマー、目覚まし時計、時計、辞書」として機能するとありますが、あなたがそれらを具体的にどのように使っているのかを説明しましょう。「ゲーム」を例にとるなら、「私は朝の通学中に〇〇というゲームをして気分転換をしています」のように、具体的なゲーム名や、なぜそのゲームをするのか、それがあなたにとってどのような意味を持つのかを付け加えると良いでしょう。

個人的なエピソードの導入: 「なぜそれがあなたにとって欠かせないのか」をより深く掘り下げてください。例えば、「一度スマートフォンを家に忘れて出かけた時、どれほど不便を感じたか」といった個人的な経験を短く加えることで、読者はあなたの主張に共感しやすくなります。

タスク達成度 (Task requirements)

評価: 4

「Something I Use Every Day」というトピックに対して、スマートフォンを選び、その多機能性について述べているため、大半の要件は満たしています。しかし、「他人に紹介する内容」という点では、個人的なつながりや魅力が十分に伝わっていません。

改善のためのアドバイス:

「紹介」の視点を意識する: 読者が「へえ、そんな使い方もあるんだ!」と思ったり、「私も使ってみようかな」と感じたりするような「紹介」の要素を強化しましょう。単に機能の羅列ではなく、その機能があなたの日常生活をどのように豊かにしているのか、便利にしているのかを具体的に伝えることが重要です。

読者への語りかけ: 読者に対して直接語りかけるような表現 (例: "Imagine a world without it!" や "You might be surprised how much it helps me..." など) を少し加えることで、「紹介」のニュアンスが強まります。

理解のしやすさ (Comprehensibility)

評価: 5

文章は非常に理解しやすく、読みやすいです。文章構造がシンプルで、語彙も難しくないため、一読で内容を理解できます。

改善のためのアドバイス:

表現の多様化: 理解しやすいことは素晴らしいですが、さらに読者の興味を引きつけるために、少し表現の幅を広げることを検討しても良いでしょう。例えば、同じ「とても大切」を意味する言葉でも、"indispensable" だけでなく "essential", "vital", "crucial" など、文脈に応じて使い分けることで、文章に深みが増します。

文頭のバリエーション: すべての文が「It's the smartphone.」「In modern Japan...」「Smartphones can also...」「That's why...」のように、主語から始まることが多いです。時折、副詞句や前置詞句を文頭に持つてくることで、リズムが生まれ、より自然な英語に近づきます。

一貫性と結束性 (Coherence and cohesion)

評価: 4

文章には一貫性があり、無関係な文の流れは稀です。スマートフォンという共通のトピックを中心に記述されており、接続詞も良好に使われています。

改善のためのアドバイス:

接続表現の洗練: 「That's why I'm always on my phone.」という締めくくりの文は、それまでの多機能性の羅列に対する結論としては少し唐突に感じるかもしれません。多機能性から「肌身離さず持っている」という結論に至るまでの論理的なつながりを、よりスムーズにする接続表現を検討しましょう。

例: 「These diverse functions make my smartphone an extension of myself, so I truly can't imagine a day without it.」のように、多機能性が「肌身離せない」理由であることをより明確に示すと良いでしょう。具体例と抽象化のバランス: 「ゲーム、ストップウォッチ ...」と具体例が列挙された後、「That's why I'm always on my phone.」と抽象的な結論につながっています。この間のギャップを埋めるために、具体例から抽象的な結論へ橋渡しとなるような文（例: 「All these features combined make my smartphone an indispensable companion for daily life.」）を挿入することを検討すると、より結束性が高まります。

資料1-2: Geminiからのフィードバックの例(二人目)
(事前テストから事後テストにかけてライティングの評価が顕著に向上した2名)

| | 内容 | タスク達成度 | 理解のしやすさ | 一貫性と結束性 |
|-------|------|--------|---------|---------|
| 事前テスト | 3.67 | 3.33 | 3 | 3.67 |
| 事後テスト | 5 | 5.33 | 4.67 | 5 |

「Something I Use Every Day」というトピックで書かれた英文、拝見しました。よく書けていますね。
 あなたのアイマスクへの愛情が伝わってきます。

それでは、ルーブリックに沿って、各観点から評価と具体的な改善アドバイスをさせていただきます。

内容 (Content)

評価：4

アイデアの数は適切で、アイマスクを紹介するという点において十分に一貫しています。アイマスクをもたらした経緯、キャラクターのデザイン、そしてそれが睡眠に役立っていることなど、必要な情報が盛り込まれています。

改善のためのアドバイス：

アイマスクの具体的な利点について、もう少し詳しく記述すると、読者にとってさらに興味深い内容になります。

例えば、**「なぜそれが『良い睡眠』につながるのか」**について、具体的に説明を加えてみましょう。「光を完全に遮断してくれるから」「目の疲れが取れる感じがするから」など、具体的な感覚や効果を記述することで、読者はその利点をより鮮明に想像できます。

また、「Waddle Dee」のキャラクターについて、もう少し情報を加えることもできます。例えば、「Waddle Dee は有名なゲームのキャラクターで、そのとぼけた表情がかわいくて癒される」といった補足があれば、Waddle Dee を知らない読者にもその魅力を伝えられます。

タスク達成度 (Task requirements)

評価：5

「他人に紹介する内容」という課題の要件をほとんど満たしています。自分の使っているもの(アイマスク)を紹介し、その理由や効果、そして読者への勧めまで書かれています。

改善のためのアドバイス：

全体的に課題は達成できていますが、読者への働きかけをもう少し意識すると、さらに魅力的な紹介文になります。

例えば、「How could I sleep without it?」という問いかけは良いですが、その後に読者に具体的な行動を促すような一文を加えることもできます。「もしあなたが良い睡眠を探しているなら、ぜひアイマスクを試してみてください」のように、読者に直接語りかける形にしてみましょう。

また、「おすすめポイント」を箇条書きでまとめるなど、紹介文としての体裁をさらに整えることも検討してみてください。

理解のしやすさ (Comprehensibility)

評価：5

文章は非常に理解しやすく、読みやすいです。書かれている内容も明確で、読者が文章の目的やアイデアを理解する上で、特に努力を必要としません。

改善のためのアドバイス：

ほとんど問題ありませんが、さらに自然な英語表現にすることで、より滑らかに読めるようになります。

例えば、「My brother gave me it about half years ago」は、「My brother gave it to me about half a year ago」とする方が自然です。「half years」ではなく「half a year」が正しい表現です。

「I can't imagine if I don't get it」は、「I can't imagine not having it」や「I can't imagine life without it」とすると、より自然な英語表現になります。「get it」は「手に入れる」という意味合いが強く、この文脈には少し不自然です。

一貫性と結束性 (Coherence and cohesion)

評価：4

文章には一貫性があり、アイマスクについての内容で統一されています。接続詞もいくつか使われており、アイデア間のつながりは概ね良好です。

改善のためのアドバイス：

アイデア間のつながりをより明確にするために、接続詞や接続表現をさらに効果的に使うことで、文章全体の結束力を高めることができます。

文頭で**「My brother gave me it about half years ago, and I wear it every night now.」**のように、「and」で文をつなぐのは良いですが、時間の経過を示す副詞句（例："Since then, " または "Ever since then, "）を使って、アイマスクをもらってから毎晩使っているという関係性をより明確にすることができます。

「I love it very much now. It's so cute and help my good sleep.」のように、理由を説明する際に**「because」**などの接続詞を使うと、より論理的なつながりになります。例：「I love it very much because it's so cute and helps me get good sleep.」

文と文の接続に変化を持たせることで、単調さを避け、より洗練された印象を与えることができます。例えば、「It's printed character's face who is a Waddle Dee.」の後に、「This design makes it even more special to me.」のように、補足的な情報を加えることで、前の文との繋がりを強めることができます。

このフィードバックが、あなたのライティングスキル向上の一助となれば幸いです。もし他に何か質問があれば、遠慮なく聞いてくださいね