

ChatGPT を利用した英検 2 級のライティング分析

竹ノ内朋子
四天王寺大学

An Analysis of the Writing Section of EIKEN Grade 2 Using ChatGPT

TAKENOUCHI, Tomoko
Shitennoji University

Abstract

英検 2 級のライティングの解答例と学習者の解答を CEFR に基づいて評価分析するために ChatGPT を利用する。併せて語彙レベルチェックを行い、評価の妥当性を考察する。評価点数は期待できないものの、概ね評価は妥当でライティング指導にも有効な面があると考えられる。

Keywords: ChatGPT, writing, CEFR, パフォーマンス評価

1. はじめに

本稿では ChatGPT を英検 2 級のライティング分析に利用し、評価、採点、書き直しの提案ができるかどうかを検証する。英検のライティングは 1 次試験において課され、2 級に関しては、650 点満点、80 語から 100 語程度、所用時間は 20 分から 25 分程度で解答する。質問内容は社会の事象に関することが多く、2 つの理由とそれをサポートする論拠や説明を述べて説得力のあるもの、つまり、*argumentative essay* (論証型のエッセイ) にすることが求められる。(新谷, 2022. pp. 137–138)。CEFR では B1 に該当するレベルである。しかしながら、学習者に一番必要な評価基準 (点数) が公開されておらず、学習者への指導も指標がなく、指導者も迷うところが多い。英検 HP (https://www.eiken.or.jp/eiken/exam/2016scoring_w_info.html) にはライティングテストの採点に関して、観点として 4 項目 (内容、構成、語彙、文法: 各 4 点満点)、TOPIC に示された問いに答えているかどうか、などの注意点が表記されている。本稿では、これらの基準と対応するかどうかについて解答例を ChatGPT に分析させ、その評価傾向を明らかにしたい。これは、ライティング評価をルーブリックのような形で学習者に事前に提示することによって、ライティングのパフォーマンスを高めることができるからであり (Goodwin, 2023)、今後の学

習・指導の指針とすることができるからである。また、学習者の解答とも比較し、評価基準を推測する一助とする。さらに、語彙レベルのチェックを合わせて行うことで、その評価基準の妥当性を考察する。

2. 方法

2.1 使用ツール

- ChatGPT (<https://chat.openai.com/>)
- New Word Level Checker (<https://nwlc.pythonanywhere.com/>)
- Voyant Tools (<https://voyant-tools.org/>)

2.2 分析素材

- 英検 HP に掲載してある 2020 年度から 2022 年度の解答例を 6 本分析する。本会場実施のものと同曜日開催の準会場実施のものを検証する。
- 学習者解答は解答例と同時期のもので、2 例（合格 1 例・不合格 1 例）を分析する。これらは受験後に学習者に解答を思い出して書き直しを依頼した。

2.3 解答例の評価

英検 HP に記載されていた 2020 年度から 2022 年度の英検 2 級の解答例 6 本を ChatGPT にて分析した。ChatGPT はエッセイの評価がある程度の精度でできるということが報告されているため (Mizumoto & Eguchi, 2023), CEFR に基づいて評価するよう指示し、その評価内容を箇条書きにて求めた。なお、ChatGPT にアクセスした期間は 2023 年 1 月 13 日から 17 日までである。質問は下記の日本語で行い、ChatGPT の回答は英語であったため DeepL にて和訳し、筆者が内容を確認した。

英検の HP 掲載の出題と解答例と評価観点は以下の通りである。

2021 年第 3 回(土)

TOPIC

Today, some local governments in Japan use social networking services (SNS) to send emergency information about natural disasters. Do you think more local governments will do this in the future?

POINTS

Safety, Speed, User-friendliness

解答例: I think that more local governments in Japan will use social networking services to send emergency information about natural disasters. To begin with,

by using social networking services, local governments can quickly provide the latest information to people. This information is necessary for people's safety when natural disasters happen. In addition, the government will be able to share emergency information with large numbers of people. Today, more and more people are using social networking services. For these two reasons, I think more governments will do this in the future. (89 words)

英検による評価観点（各観点4点満点）

1. 語彙：課題にふさわしい語彙を正しく使えているか
2. 内容：課題で求められている内容が含まれるか
3. 構成：英文の構成や流れがわかりやすく論理的であるか
4. 文法：文構造にバリエーションやそれを正しく使えているか

3. 結果

3.1 ChatGPT による分析

上記で示した解答例と以下の質問（Q）を ChatGPT に入力したところ、A のような回答が得られた。CFER ではこの解答例は B1 と評価された。観点としては英検のものと同様に評価しており、概ね同様と見なせる。但し、文法に関しては、文脈上間違いではない箇所を間違いと指摘している。

Q: この英文を CEFR に基づいて評価して下さい。根拠を箇条書きで教えて下さい。

A: CEFR B1

- 文章はシンプルでわかりやすい言葉で書かれている。
- 主旨は明確に表現されているが、文法や語彙に若干の誤りがある。
- 主な考えを裏付ける基本的な理由や根拠が含まれている。
- 簡単な文型と語彙を使用しているが、より複雑な文型や語彙が含まれている。

Q: 文法の間違いを指摘して下さい。（下線部筆者）

A: For these two reasons, I think more governments will do this in the future.

↓

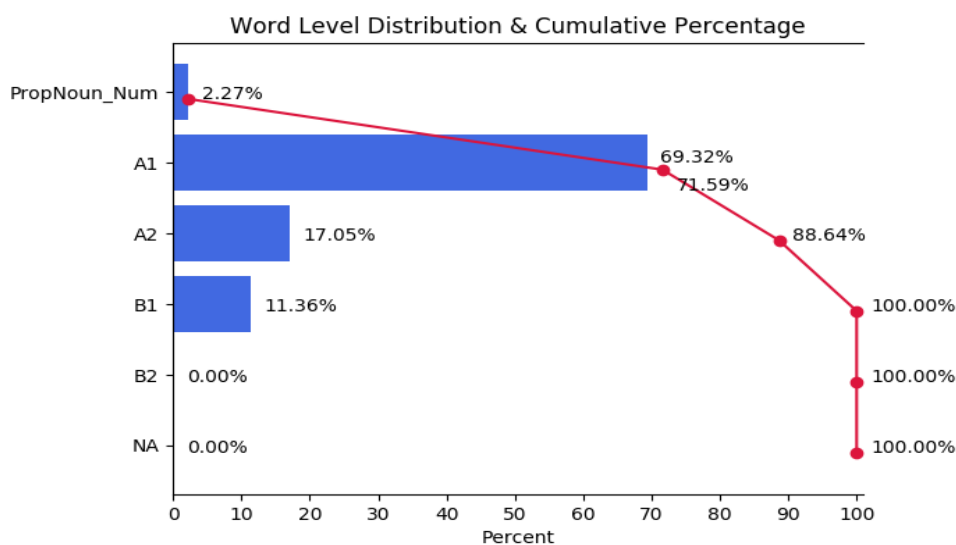
For these two reasons, I think more local governments in Japan will use social networking services in the future.

3.2 New Word Level Checker と Voyant Tools による分析

この解答例を New Word Level Checker（水本, 2022）に掲載されている語彙リストである CEFR-J で分析したところ、B1 レベルの単語が 11.36% 占めていると分析された（social networking services, disasters, safety, able など）。また、Voyant Tools にて分析した結果、1 文の平均語数は 14.8 語であった。解答例としては that を含む複文や関係詞等を含む比較的長い文が解答例として書かれていることがわかった。

図 1

New Word Level Checker による語彙レベル分析



3.3. 学習者の解答の分析

次に学習者の英文を 2 例、以下の点数を求める指示で ChatGPT に評価させた。学習者 A（英検 1 次試験不合格）の解答に対して、評価点は得られず、A2 レベルと評価された。ChatGPT ではプロンプトを工夫しない限り、数値的な評価はできないということがわかる。

この解答は B1 レベルの単語を 12% 使用していたにもかかわらず、A2 レベルと評価されたのは、文法的に誤りが若干あったためと、1 文の平均語数が 12 語と短いことが影響しているのではないかとと思われる。また、この解答を文字数指定して書き直しを指示したが、文字数を大幅に越えた解答しか得られなかった。文字数指定した書き直しはできないようである。

Q: 解答例の英文を 650 点として、次の英文を評価して下さい。

A: CEFR A2 (点数は報告されない)

- 文章は簡単な言葉で書かれており、主旨は明確であるが、文法、語彙、文構造に誤りがある可能性がある。
- 主な考えを裏付ける基本的な理由や証拠が含まれているが、詳細さや複雑さに欠ける。
- 語彙は基本的で反復的。
- 文法的な誤りがいくつかある。例えば、“Knowing natural disasters information is as possible as fast, it is safer.” は “Knowing information about natural disasters as soon as possible is safer.” とすべきであろう。

(下線部筆者)

同様に学習者 S (英検 1 次試験合格) の評価は B1 レベルであった。B1 レベルの単語の使用は 5% と少なかったが、文法ミスはほぼなく、1 文の平均語数が 14.2 語であったためにこの評価が算出されたと考えられる。

表 1 に他の解答例と学習者の解答を同様に分析したものを示す。解答例の 1 文の平均語数は 14.6 語であり、使用語彙に B2 レベルが含まれていれば B2 レベルと評価する傾向があることが示唆された。但し、解答例には文法の間違ひはないが、学習者の解答には文法の間違ひがあるため、語彙レベルのみでは評価できない可能性もあることに注意が必要である。

表 1

解答例と学習者の比較一覧表

No.	実施年度	出題内容 (日本語要約)	CEFR レベル	1文の平 均語数	Word level				
					A1	A2	B1	B2	TTL (%)
1	2020-3	室内ペット飼育	B1	14.0	76	13	11	0	100
2	2021-2	美術館の入場料無料	B2	15.7	82	9	3	3	97
3	2021-3	光のデコレーション	B1	14.8	70	16	12	0	98
4	2021-3土	非常時のSNS利用	B1	14.3	78	11	10	0	99
5	2022-1	史跡に直接行くか	B2	13.7	76	14	6	1	97
6	2022-1土	市中の美術品設置	B1	14.8	79	12	9	0	100
				14.6	76.8	12.5	8.5	0.67	98.5
学習者									
A	2021-3土	非常時のSNS利用	A2	12.0	63	21	12	0	96
S	2022-1土	市中の美術品設置	B1	14.2	85	9	5	0	99

4. 考察

以上の分析と結果により、ChatGPTによる解答例のCEFRに基づく評価は概ね妥当と考えられるが、ChatGPTのみによる評価では根拠が乏しいと考えられる。特に学習者の解答を評価する場合は、他のデータ分析ツールと組み合わせ、語彙レベルの分析など裏付けになる根拠を提示することが指導上重要であると考えられる。もちろん、指導者など人間が学習者の解答をしっかりと読むことが重要であるということも言うまでもない。文法の間違いについては、文脈を考慮すると必ずしも妥当なものとは限らないこともあるが、ChatGPTを使用することによって、明らかな間違いは指摘してくれるので有効であろう。また、一番欲しい情報としての評価点に関しては、今回は算出されなかった。指定した文字数での解答例の作成や書き直しも今のところできない。但し、パラフレーズや構文など、ChatGPTの出力には参考になる部分はあるので適宜指導に活かすと良いだろう。

後日、同様の操作を試みたが、評価が一定しなかった。ChatGPTをライティング指導に活かすには、プロンプトの出し方にも工夫が必要であるうえ、結果を検証する教師側の英語力も必須である。ChatGPT自体も日々進化しており、さらに多くの検証が必要である。

また、ChatGPTに利用上の注意点を聞いてみたところ以下のような回答があった。

But keep in mind that I am a language model and my responses are generated based on patterns and examples in the data I was trained on, so please consider using the information in the context of your research or application.

ただし、私は言語モデルであり、私の回答は私が学習したデータのパターンや事例に基づいて生成されることを念頭に置いて、研究やアプリケーションの文脈で情報を利用することを検討してください。(DeepLによる和訳)

教員のみならず、学習者もAIツールを利用し様々な課題に活用していることを考慮しつつ、利用する際はそのような点に留意し、有効に活用していくことがこれからの英語学習・指導において必要であるといえるだろう。

参考文献

- Goodwin, R. (2023). Using rubrics to improve writing skills: A study in Kuwait. *Language Testing in Asia*, 13. <https://doi.org/10.1186/s40468-023-00224-6>
- 水本 篤 (2022). 「New Word Level Checker の概要」『外国語教育メディア学会 (LET) 関西支部メソドロジー研究部会報告論集』12, 1-24. <https://doi.org/10.31219/osf.io/whr9a>

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*.

<https://doi.org/10.31219/osf.io/2uahv>

日本英語検定協会 (n.d.). 「ライティングテストの採点に関する観点および注意点」

Retrieved January 10, 2023, from

https://www.eiken.or.jp/eiken/exam/2016scoring_w_info.html

日本英語検定協会 (n.d.). 「2級の過去問・対策」 Retrieved January 31, 2021, from

https://www.eiken.or.jp/eiken/exam/grade_2/solutions.html

新谷奈津子 (2022). 「ライティングの学習」 中田達也・鈴木祐一 (編著) 『英語学習の科学』 (pp. 131–147). 研究社