



教室で英語のエッセイの評価をする生成AI
— powered by DALL · E3 —

生成AIは英文評 価者になりえる のか

河内健志
(前橋工科大学)

斎藤隆枝
(帝京大学)

英作文の評価・フィードバック

例えば、学生50人が書いたエッセイを採点し、フィードバックをするとする...



- 大変な仕事
- 多大な時間と労力
- 集中力を要する
- 公平かつ有益な評価とフィードバックは容易でない
- スキル向上に貢献できるような具体的なアドバイスを提供する必要性
- フィードバックを通じて学生のモチベーションを維持または向上させる必要性

K You

あなたは大学の英語教師です。これから学生50人が書いたエッセイを採点し、フィードバックをします。あなたは、この仕事を大変な仕事だと思えますか。

ChatGPT

はい、大学の英語教師として50人の学生が書いたエッセイを採点し、フィードバックをする任務は、**大変な仕事だと思えます**。この作業には**多大な時間と労力、集中力が必要**であり、**それぞれの学生に公平かつ有益な評価とフィードバックを提供することは簡単なことではありません**。教師は、**学生の努力を正しく評価し、彼らのスキル向上に貢献**できるような**具体的なアドバイスを提供**する必要があります。

それに加えて、教師は各エッセイの内容を深く理解し、学生がどのように課題に取り組んだかを評価しなければなりません。また、**個々のフィードバックを通じて学生のモチベーションを維持または向上させる役割も担っています**。さらに、このプロセスは教師自身の専門知識と教育スキルを反映するものでもあるため、教師にとっても成長の機会となります。

- 習熟度に応じて**明示的なフィードバック**をする方法が効果的である

(Lee 2008, Bitchener & Ferris 2012)。

- フィードバック後、学習者が見直しをするとは限らず、学習者の主体的なライティングスキル・意欲の向上につながることができているのか不明
- **Criterion**は明示的なフィードバックを行い、ライティングスキルの向上に導く。しかし、**高額!**

目的

生成AI (GPT4/GPT3.5/Gemini)は
Criterionの代替の英文評価ツール
になりえるかを検証する。

方法

英文評価ツール

- Criterion (有料, ETS社)
- GPT3.5 (無料, Open AI社)
- GPT4 (有料, Open AI社)
- Gemini (無料, Google社)

検証方法

- 学生30名（2023年当時1年生：英語を専攻しない）のエッセイをCriterion/GPT4/GPT3.5/Geminiで評価し、比較する。
- Criterionの評価値を基準とする。
- 評価項目：
 - ・ Grammar (8小項目) / Usage (10小項目) / Mechanics (12小項目) / Style (6小項目) の間違いの数 (Criterionの評価項目を生成AI (GPT4/GPT3.5/Gemini) にも適用)
 - ・ エッセイのスコア
- GPT4/GPT3.5/Geminiの評価はエッセイ1編につき3回評価する。
 - ※ Criterionは複数回評価させても数値が変わらないため、エッセイ1編につき1回評価する。
- GPT4/GPT3.5/Geminiの評価値は変動係数とする。

検証期間

- 2024年2月15日 ～ 2月27日

エッセイ

インストラクション

1. ① トピックセンテンス
② テーマを決める
③ 3つのキーワード
④ トピックセンテンスの完成
⑤ 『英文エッセイ 基本の「き」』を視聴
(<https://www.youtube.com/watch?v=tT3FJR8LAHA>)
2. エッセイの基本構成
 - ① Introduction (導入)
 - ② Body (本論) :
 - ③ Conclusion (結論)
3. 実際のエッセイを分析
(テーマ、トピックセンテンス、3つのキーワード)

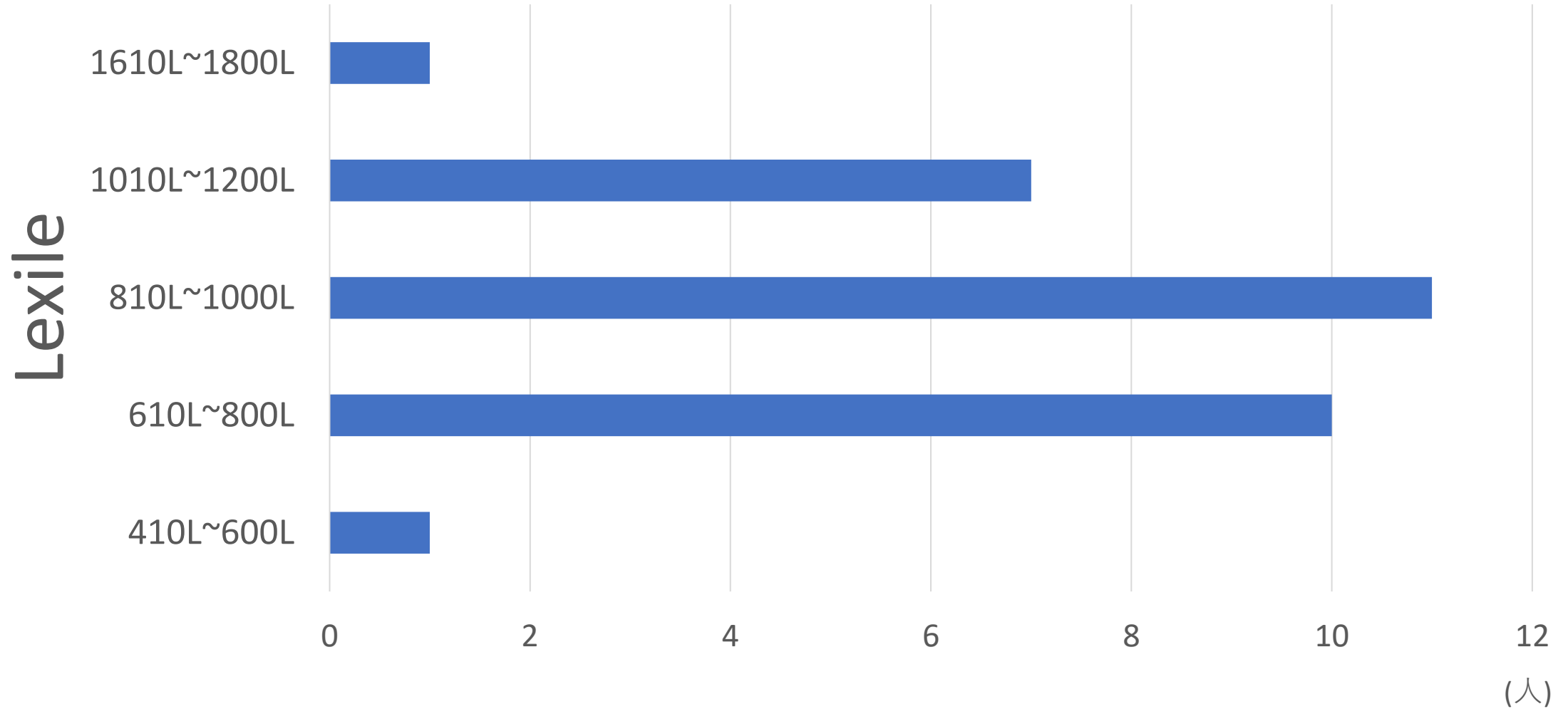
テーマ: Music

Music not only entertains people but also influences their thinking and behavior. How do you think music has the power to influence as well as to entertain people? Explain your views with reasons or examples from your reading, observations or experience. (450~500 words)

音楽は人々を楽しませるだけでなく、彼らの思考や行動にも影響を与える。音楽は人々を楽しませるだけでなく、どのように影響を与える力があると思いますか？読書や観察、経験から得た理由や例を挙げて、あなたの考えを説明してください。

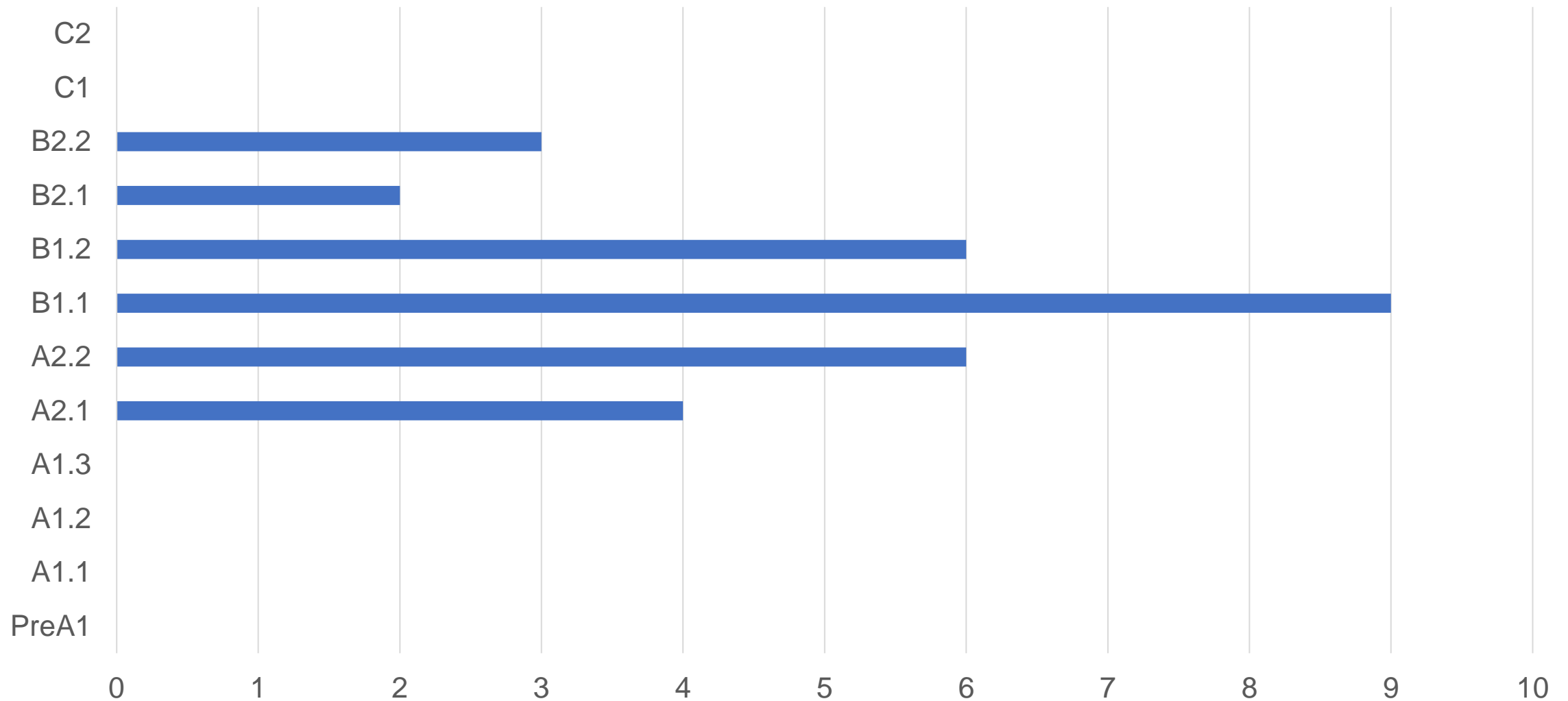
(DeepL翻訳: <https://www.deepl.com/translator>)

Lexile指数*



*文章の単語数や難易度、構文の複雑さ、長さなどを総合的に数値化したもの。

CEFR-J*



*CEFR-J Wordlistをもとにして、CVLA: CEFR-based Vocabulary Level Analyzer (ver. 2.0) にて算出

Essayの評価方法

Essay

評価①

評価①

- ・ Criterion
- ・ 採点 (1回)

評価②

評価②

- ・ GPT4 (有料)
- ・ 採点 (3回)

評価③

評価③

- ・ GPT3.5 (無料)
- ・ 採点 (3回)

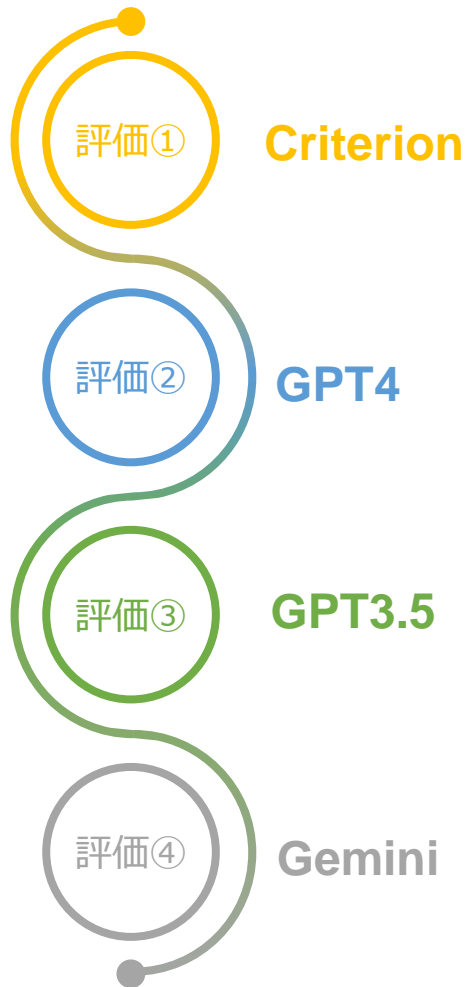
評価④

評価④

- ・ Gemini (無料)
- ・ 採点 (3回)

Essay1~30

Essay 1



Essay 2



Essay 30



Criterion

📄 エッセイは、自動採点エンジン **e-rater** によって自動採点される。

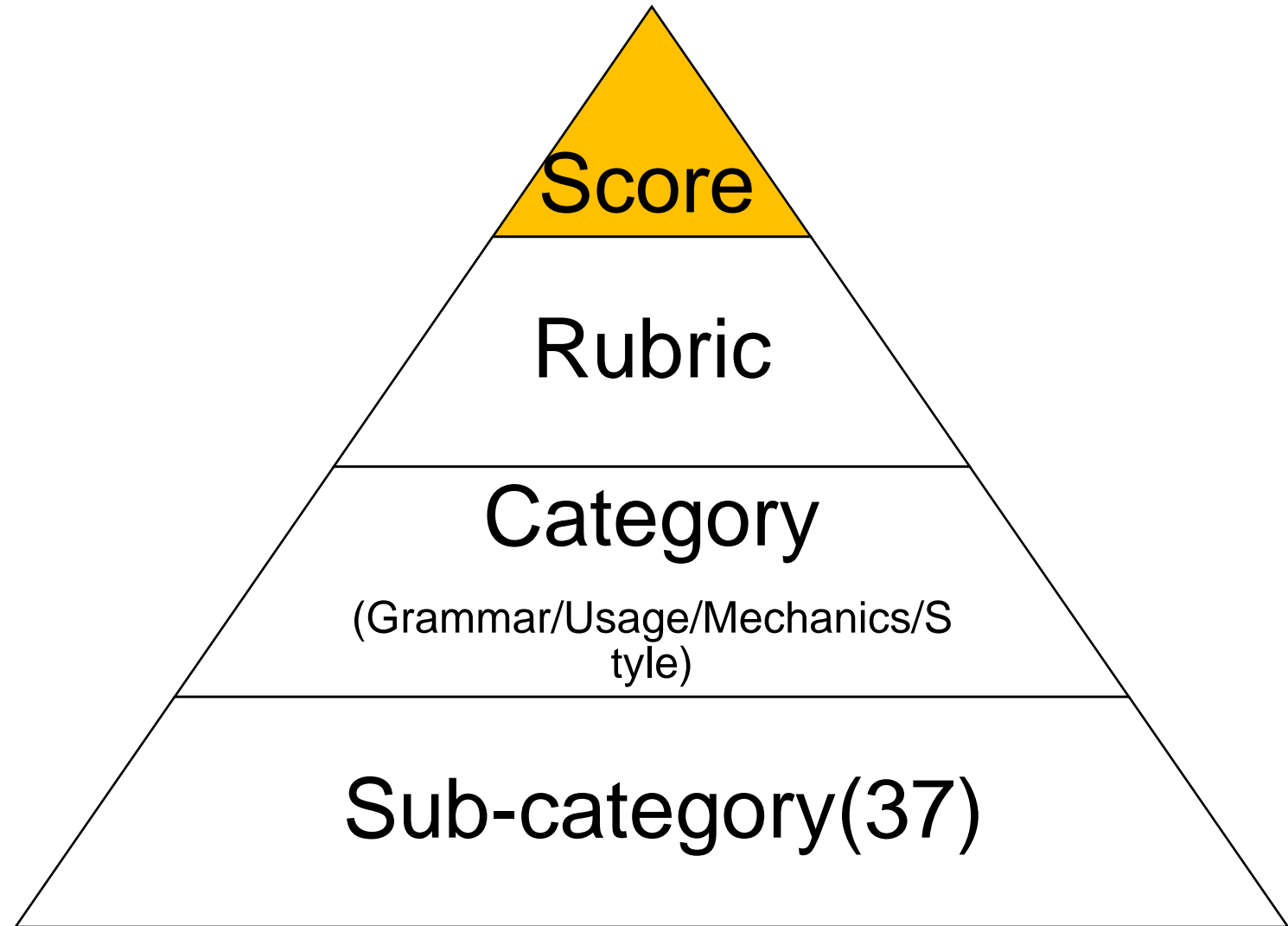
※ TOEFL iBT[®] テストで採用

📄 人間の採点者とのスコアの一致率が**97%**
(Attali & Burstein 2006)

(<https://etsjapan.jp/criterion/scoring.html>)

採点の行程イメージ

Score & Feedback



4/6

Your essay needs some help.

1) Look at the Organization and Development screen in the response tab. Are you missing anything you should add? If you have a clear outline, maybe there are some points you need to expand on.

2) Check how well your essay is organized. Make sure you provide paragraph breaks when you switch from one major idea to the next. Make sure you have provided transition words that make the relationship between the parts of your essay clear.

3) Once you have made sure that your essay is well-organized and well-developed, look at the grammar, usage, mechanics, and style tabs. Criterion has highlighted words and phrases that you should check. Sometimes you won't need to make any changes, but if you take a close look at everything that is highlighted, you will be able to eliminate a lot of problems that your teacher shouldn't have to correct for you.

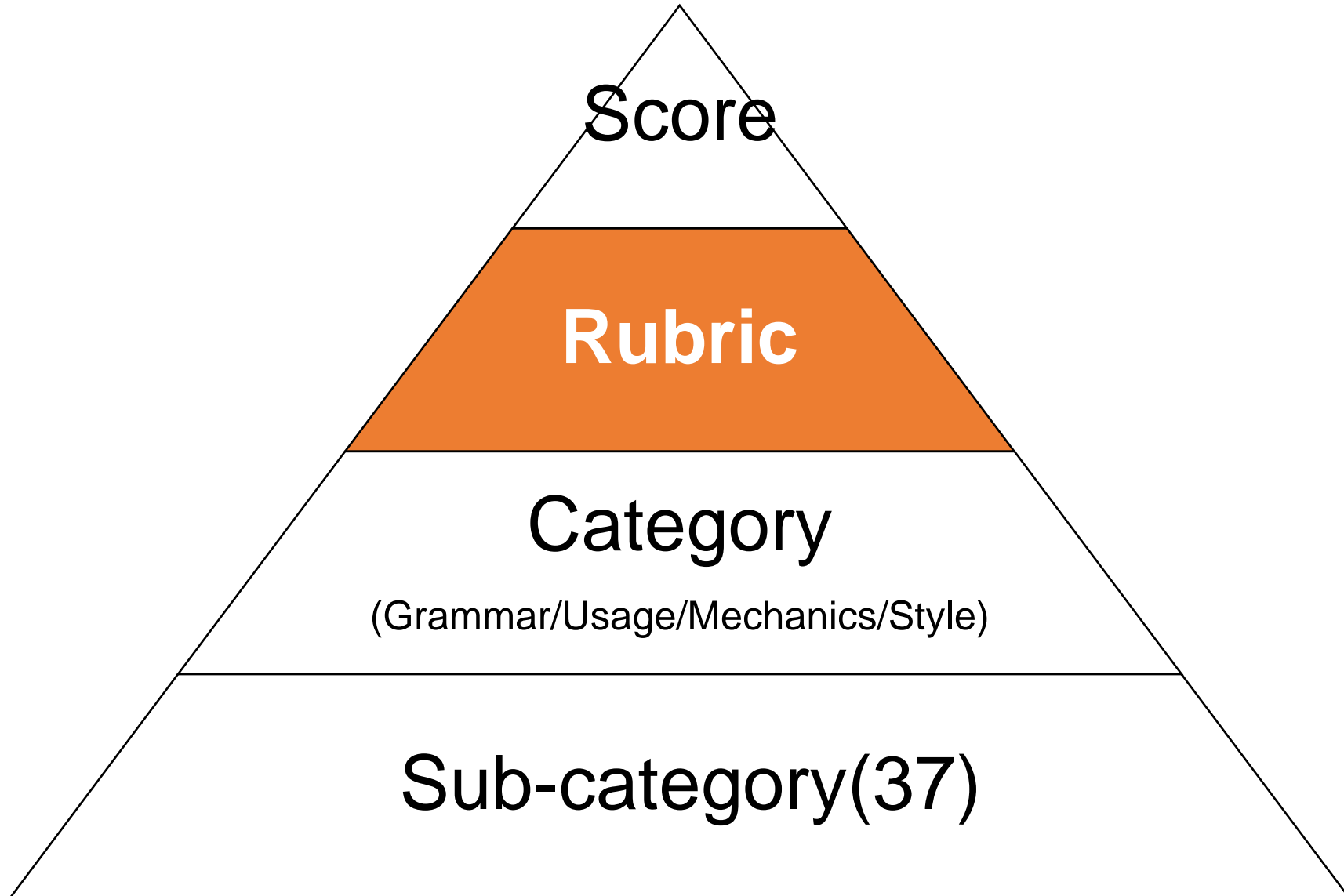
あなたのエッセイには助けが必要です。

1) 「構成と展開」画面を見てください。を見てください。あなたが追加すべき加えるべきものはありますか？アウトラインがはっきりしているのであればもっと詳しく説明すべき点があるかもしれません。

2) 自分のエッセイがどの程度整理されているかをチェックする。段落
が切り替わるときに段落の区切りがあることを確認する。主要なアイ
デアから次のアイデアに切り替わるときには、段落区切りをして
ください。また移行語（transition words）を使って、エッセイの
エッセイの各パート間の関係を明確にします。

3) 自分のエッセイがよく整理され、よく展開されていることを確認し
たら構成され、よく練られていることを確認したら、文法、用法、
仕組み、文体のタブを見ます、用法、仕組み、スタイルのタブを見
ます。クライテリアンはチェックすべき語句が強調表示されていま
す。変更する必要がない場合もありますがハイライトされた箇所を
すべてよく見れば強調表示されているすべての箇所をよく見れば、
教師が指摘すべきでない多くの問題を取り除くことができます。教
師が修正する必要のない多くの問題を排除することができます。あ
なたのために。

採点の行程イメージ



| Score | Rubric | 日本語訳 |
|-------|--|---|
| 6 | <p>You have put together a convincing argument. Here are some of the strengths evident in your writing: Your essay:</p> <ul style="list-style-type: none"> • Looks at the topic from a number of angles and responds to all aspects of what you were asked to do • Responds thoughtfully and insightfully to the issues in the topic • Develops with a superior structure and apt reasons or examples (each one adding significantly to the reader's understanding of your view) • Uses sentence styles and language that have impact and energy and keep the reader with you • Demonstrates that you know the mechanics of correct sentence structure, and American English usage virtually free of errors | <p>あなたは説得力のある議論をまとめた。あなたの文章に見られる長所をいくつか挙げてみましょう：</p> <p>あなたのエッセイ</p> <ul style="list-style-type: none"> • トピックを多面的にとらえ、依頼されたことのあらゆる側面に対応している。 • トピックの問題点に対して、思慮深く洞察に満ちた回答をしている。 • 優れた構成と的確な理由や例（一つひとつが読み手の理解を大きく深める）で展開している。 • インパクトとエネルギーがあり、読み手を引きつける文体や言葉を使っている。 • 正しい文章構成とアメリカ英語の使い方を知っていることを示す。 |

| Score | Rubric | 日本語訳 |
|-------|---|---|
| 5 | <p>You have solid writing skills and something interesting to say. Your essay:</p> <ul style="list-style-type: none"> • Responds more effectively to some parts of the topic or task than to other parts • Shows some depth and complexity in your thinking • Organizes and develops your ideas with reasons and examples that are appropriate • Uses the range of language and syntax available to you • Uses grammar, mechanics, or sentence structure with hardly any error | <p>あなたには確かな文章力があり、何か面白いことが言える。 あなたのエッセイは</p> <ul style="list-style-type: none"> • トピックや課題のある部分により効果的に反応している。 • 他の部分よりも効果的に反応している思考に深みと複雑さがある。 • 適切な理由と例を用いて自分の考えを整理し、発展させている。 • 使用可能な言語や構文を使いこなしている。文法や構文をほとんど間違えずに使っている。 • 文法、構文、文型を使用している。 |

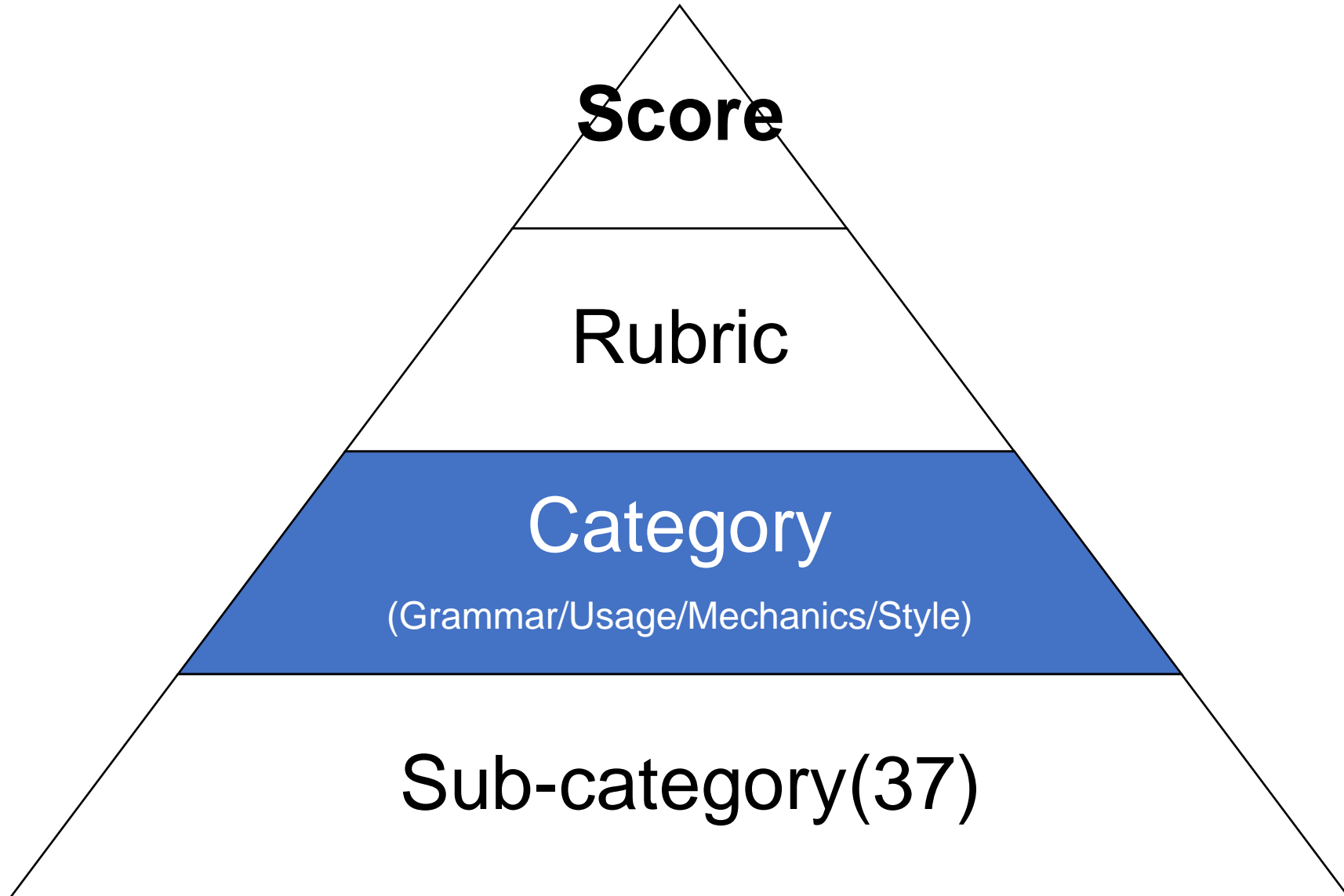
| Score | Rubric | 日本語訳 |
|-------|---|---|
| 4 | <p>Your writing is good, but you need to know how to be more persuasive and more skillful at communicating your ideas. Your essay:</p> <ul style="list-style-type: none"> • Slight some parts of the task • Treats the topic simplistically or repetitively • Is organized adequately, but you need more fully to support your position with discussion, reasons, or examples • Shows that you can say what you mean, but you could use language more precisely or vigorously • Demonstrates control in terms of grammar, usage, or sentence structure, but you may have some errors | <p>あなたの文章は良いが、もっと説得力を持たせ、自分の考えを上手に伝える方法を知る必要がある。 あなたのエッセイ</p> <ul style="list-style-type: none"> • 課題の一部を軽視している • トピックを単純に扱っているか、反復している • 十分に構成されているが、議論、理由、例で自分の立場をもっと十分にサポートする必要がある。 • 言いたいことは言えるが、もっと正確に、あるいは力強く言葉を使う必要がある。 • 文法、用法、文型の面で統制がとれているが、若干の誤りがある |

| Score | Rubric | 日本語訳 |
|-------|--|---|
| 3 | <p>Your writing is a mix of strengths and weaknesses. Working to improve your writing will definitely earn you more satisfactory results because your writing shows promise. In one or more of the following areas, your essay needs improvement. Your essay:</p> <ul style="list-style-type: none"> • Neglects or misinterprets important parts of the topic or task • Lacks focus or is simplistic or confused in interpretation • Is not organized or developed carefully from point to point • Provides examples without explanation, or generalizations without completely supporting them • Uses mostly simple sentences or language that does not serve your meaning • Demonstrates errors in grammar, usage, or sentence structure | <p>文章に長所と短所が混在しています。あなたのライティングは将来性があるため、改善に取り組むことで、より満足のいく結果が得られることは間違いありません。以下の部分で1つ以上改善が必要です。</p> <p>あなたのエッセイは、</p> <ul style="list-style-type: none"> • トピックや課題の重要な部分を無視している、または誤解している。 • 焦点が定まっていない、または解釈が単純または混乱している。 • 整理されていない、またはポイントからポイントへ注意深く展開されていない。 • 説明のない例や、完全な裏付けのない一般論を提示している。 • ほとんど単純な文章、または意味を成さない言葉を使っている • 文法、用法、文構造に誤りがある |

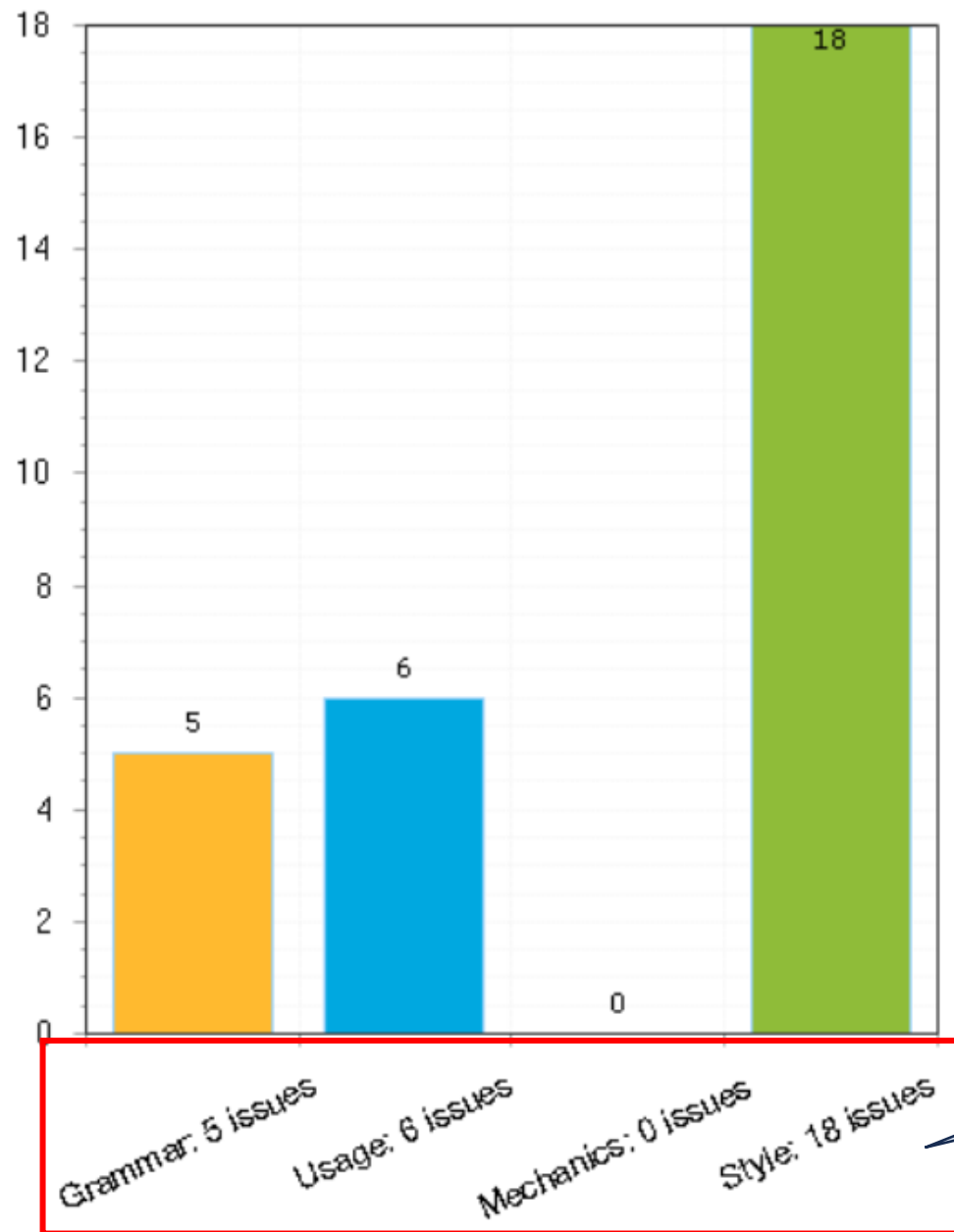
| Score | Rubric | 日本語訳 |
|-------|--|---|
| 2 | <p>You have work to do to improve your writing skills. You probably have not addressed the topic or communicated your ideas effectively. Your writing may be difficult to understand. In one or more of the following areas, your essay:</p> <ul style="list-style-type: none"> • Misunderstands the topic or neglects important parts of the task • Does not coherently focus or communicate your ideas • Is organized very weakly or doesn't develop ideas enough • Generalizes and does not provide examples or support to make your points clear • Uses sentences and vocabulary without control, which sometimes confuses rather than clarifies your meaning | <p>文章力を向上させるための課題があります。トピックを取り上げていないか、自分の考えを効果的に伝えていない可能性があります。あなたの文章はわかりにくいかもしれません。あなたのエッセイには次のような点があります：</p> <ul style="list-style-type: none"> • トピックを誤解しているか、課題の重要な部分を無視している。 • 首尾一貫して焦点を絞らず、自分の考えを伝えていない。 • 構成が非常に弱い、またはアイデアを十分に展開していない • 一般化しており、論点を明確にするための例や裏付けがない。 • 文や語彙をコントロールできず、意味を明確にするよりもむしろ混乱させることがある。 |

| Score | Rubric | 日本語訳 |
|-------|--|---|
| 1 | <p>You have much work to do in order to improve your writing skills. You are not writing with complete understanding of the task, or you do not have much of a sense of what you need to do to write better. You need advice from a writing instructor and lots of practice. In one or more of the following areas, your essay:</p> <ul style="list-style-type: none"> • Misunderstands the topic or doesn't show that you comprehend the task fully • Lacks focus, logic, or coherence • Is undeveloped--there is no elaboration of your position • Lacks support that is relevant • Shows poor choices in language, mechanics, usage, or sentence structure which make your writing confusing | <p>ライティングスキルを向上させるためにやるべきことはたくさんあります。課題を完全に理解した上で書いていない、あるいは、より良い文章を書くために何をすべきかをあまり理解していない。ライティングのインストラクターからのアドバイスと、たくさんの練習が必要です。あなたのエッセイには次のような点があります：</p> <ul style="list-style-type: none"> • トピックを誤解している。 課題を十分に理解していない。 • 焦点、論理、一貫性が欠けている。 • 自分の立場を詳しく説明していない。 • 適切なサポートがない言語、構成、用法、文の構成に不適切な選択がある。 • 文章の構成が悪く、文章がわかりにくい。 |

採点の行程イメージ



Trait Feedback Analysis



Average Number of Words per Sentence: 13.2

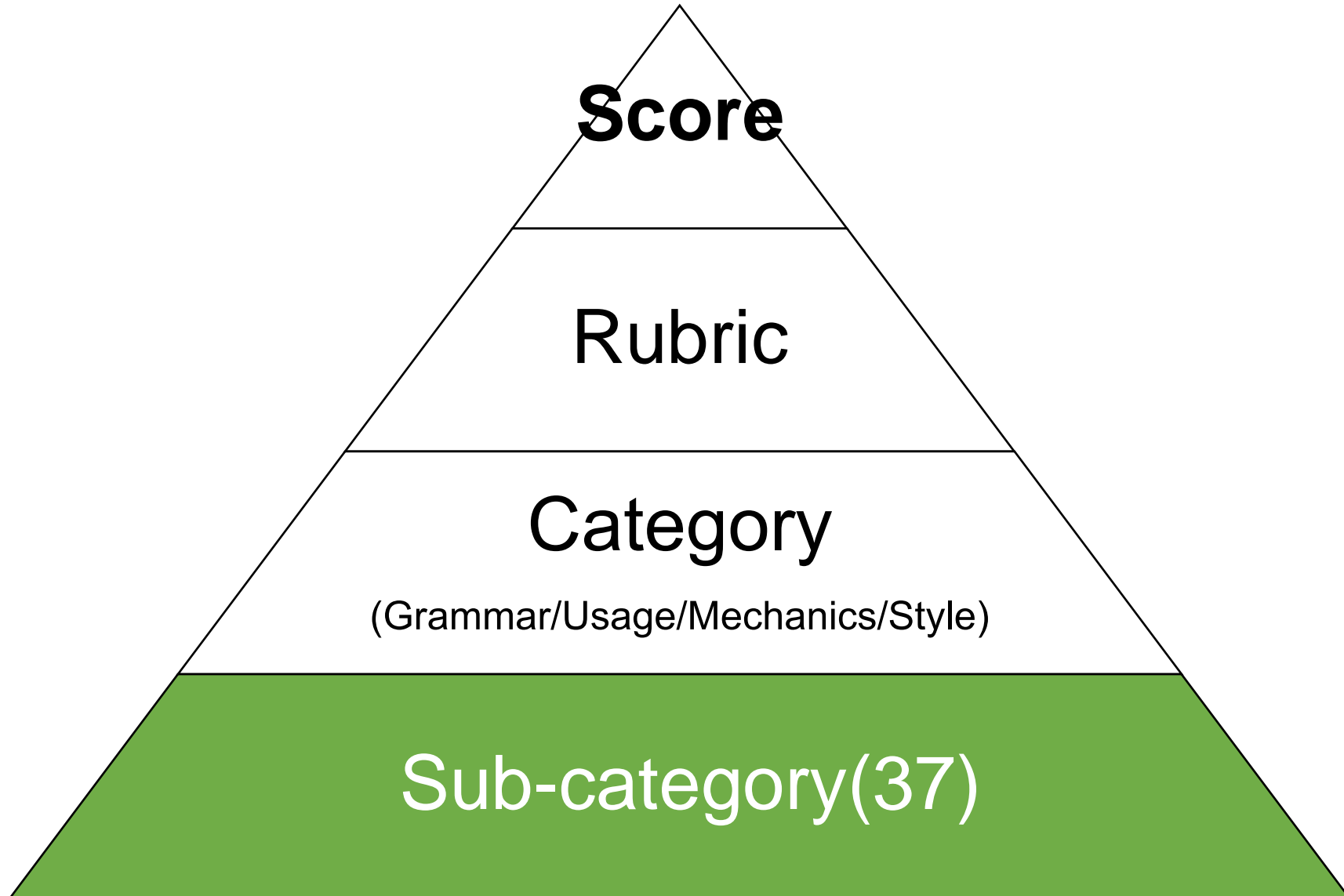
Total Number of Sentences: 34

Unique Words Count: 129

Total Number of Words: 449

各カテゴリーのエラー総数

採点の行程イメージ



| Grammar | Usage | Mechanics | Style |
|--|--|---|--|
| Fragments (断片的な文) | Determiner Noun Agreement (限定詞と名詞の一致) | Spelling (綴りの誤り) | Repetition of Words (頻発している言葉) |
| Run-on Sentences (無終止文) | Missing of Extra Article (冠詞の誤り) | Missing Capitalization of Proper Nouns (固有名詞が小文字で始まる) | Inappropriate Words or Phrases (不適切な語・句) |
| Garbled Sentences (混乱した文) | Confused Words (言葉の取り違え) | Missing Initial Capital Letter in a Sentence (文の始めが小文字) | Sentences Beginning with Coordinating Conjunctions (接続詞で始まる文) |
| Subject-Verb Agreement Errors (主語と動詞の一致) | Wrong Form of Word (誤った形態) | Missing Question Mark (疑問符の欠落) | Short Sentences (短い文) |
| Verb Form Errors (動詞の活用形ミス) | Faulty Comparison (比較の誤り) | Missing Final Punctuation (文末句読点の欠落) | Long Sentences (長い文) |
| Pronoun Errors (代名詞の誤り) | Preposition Error (前置詞の誤り) | Missing Apostrophe (アポストロフィの欠落) | Sentences with Passive Voice (受動態の適切な使用) |
| Possessive Errors (所有格の誤り) | Nonstandard Word Form (口語／砕けた形態) | Missing Comma (コンマの欠落) | |
| Wrong or Missing Word (誤った単語の使用、単語が抜けている) | Negation Error (否定の誤用 (二重否定)) | Missing Hyphen (ハイフンの欠落) | |
| | Wrong Part of Speech (品詞の誤り) | Fused Words (融合語) | |
| | Wrong Article (冠詞の誤り) | Compound Words (複合語) | |
| | | Duplicate Words (重複している言葉) | |
| | | Extra Comma (余分なコンマ) | |

GPT4
GPT3.5
Gemini

生成AIへの指示①

You are the rater of an essay.

Please count the number of mistakes in the essay for the following 4 categories and their each category. Then, make a table for 4 categories and their each category.

[Grammar]

- Fragments
- Run-on Sentences
- Garbled Sentences
- Subject-Verb Agreement Errors
- Verb Form Errors
- Pronoun Errors
- Possessive Errors
- Wrong or Missing Word

[Usage]

- Determiner Noun Agreement
- Missing of Extra Article
- Confused Words
- Wrong Form of Word
- Faulty Comparison
- Preposition Error
- Nonstandard Word Form
- Negation Error
- Wrong Part of Speech
- Wrong Article

[Mechanics]

- Spelling
- Missing Capitalization of Proper Nouns
- Missing Initial Capital Letter in a Sentence
- Missing Question Mark
- Missing Final Punctuation
- Missing Apostrophe
- Missing Comma
- Missing Hyphen
- Fused Words
- Compound Words
- Duplicate Words
- Extra Comma

[Style]

- Repetition of Words
- Inappropriate Words or Phrases
- Sentences Beginning with Coordinating Conjunctions
- Short Sentences
- Long Sentences
- Sentences with Passive Voice



ChatGPT

Here's the table summarizing the mistakes in each category:

| Category | Count |
|---------------------------------|-------|
| Grammar | |
| - Fragments | 0 |
| - Run-on Sentences | 0 |
| - Garbled Sentences | 0 |
| - Subject-Verb Agreement Errors | 0 |
| - Verb Form Errors | 1 |
| - Pronoun Errors | 2 |
| - Possessive Errors | 0 |
| - Wrong or Missing Word | 2 |
| Usage | |
| - Determiner Noun Agreement | 0 |

生成AIへの指示②

Grade the Essay. Score range shall be 1~6. When scoring, you need the following rubric.

- Score 6: the essay have been put together a convincing argument.
Here are some of the strengths evident in the writing
- Score 5: the essay has solid writing skills and something interesting to say.
- Score 4: the writing is good, but you need to know how to be more persuasive and more skillful at communicating your ideas.
- Score 3: the writing is a mix of strengths and weaknesses. Working to improve the writing will definitely earn the essay more satisfactory results because the writing shows promise. In one or more of the following areas, the essay is needed improvement.
- Score 2: the essay have been worked to do to improve the writing skills. The essay probably have not addressed the topic or communicated your ideas effectively. The writing may be difficult to understand.
- Score 1: the esssay have been much worked to do in order to improve the writing skills. The essay is not writing with complete understanding of the task, or the wrriter do not have much of a sense of what the writer need to do to write better. The essay is needed to give advice from a writing instructor and lots of practice.



ChatGPT

Based on the provided rubric and the identified mistakes in the essay, here is the grading:

Overall, the essay demonstrates some strengths but also has several areas for improvement, particularly in grammar, mechanics, and style. The writing lacks clarity and precision in expression, with multiple errors affecting comprehension and flow. While there are interesting points made about the impact of music, the essay could benefit from stronger organization and more compelling arguments. Therefore, considering the criteria provided:

Score: 3

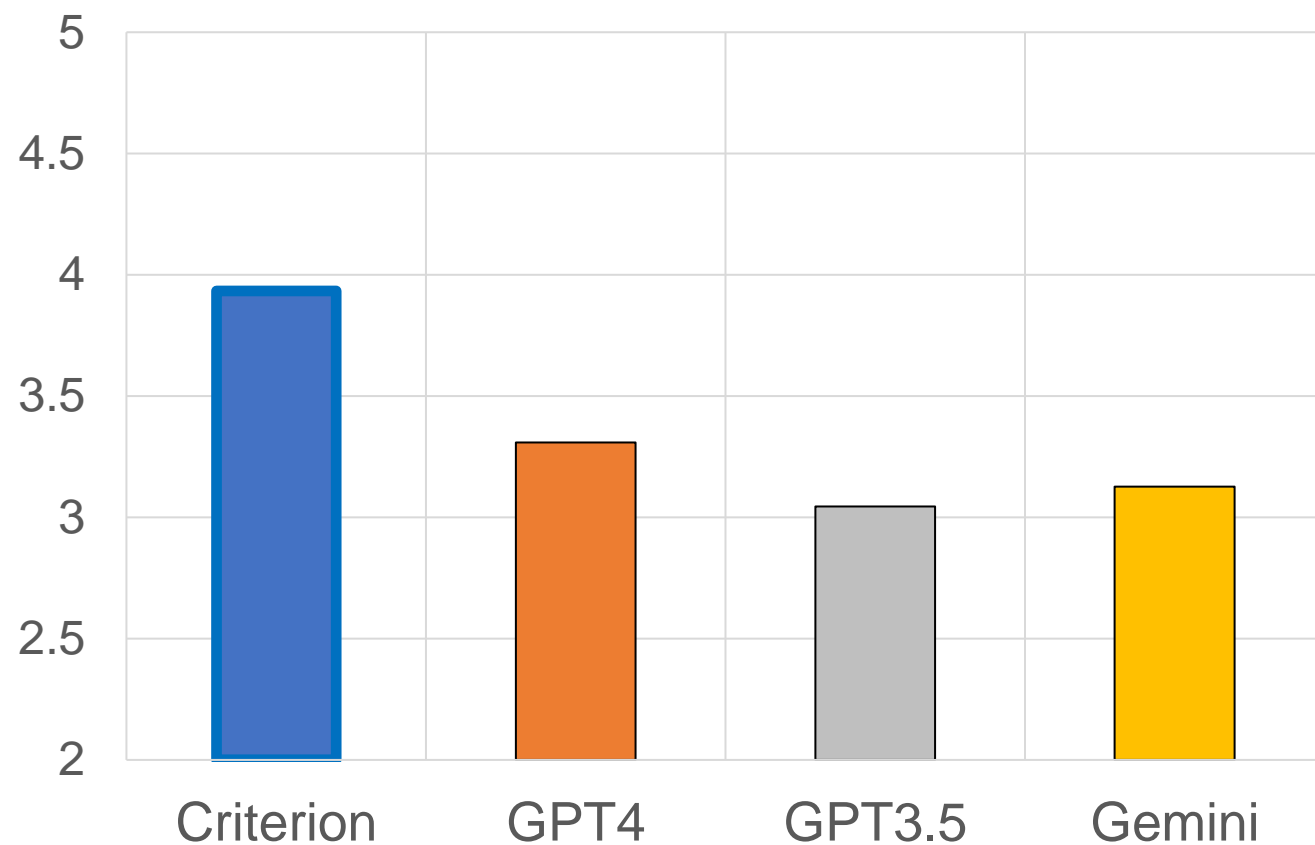
提供されたルーブリックとエッセイで確認されたミスに基づき、採点します：

全体的に、このエッセイはいくつかの長所を示していますが、特に文法、力学、文体において改善すべき点もいくつかあります。文章は明瞭さと表現の正確さに欠け、複数の誤りが理解力と流れに影響を与えている。音楽の影響について興味深い指摘がある一方で、エッセイはより強く構成され、より説得力のある論証の恩恵を受けることができるでしょう。したがって、提供された基準を考慮すると

検証 1

平均スコア

平均スコア n=30

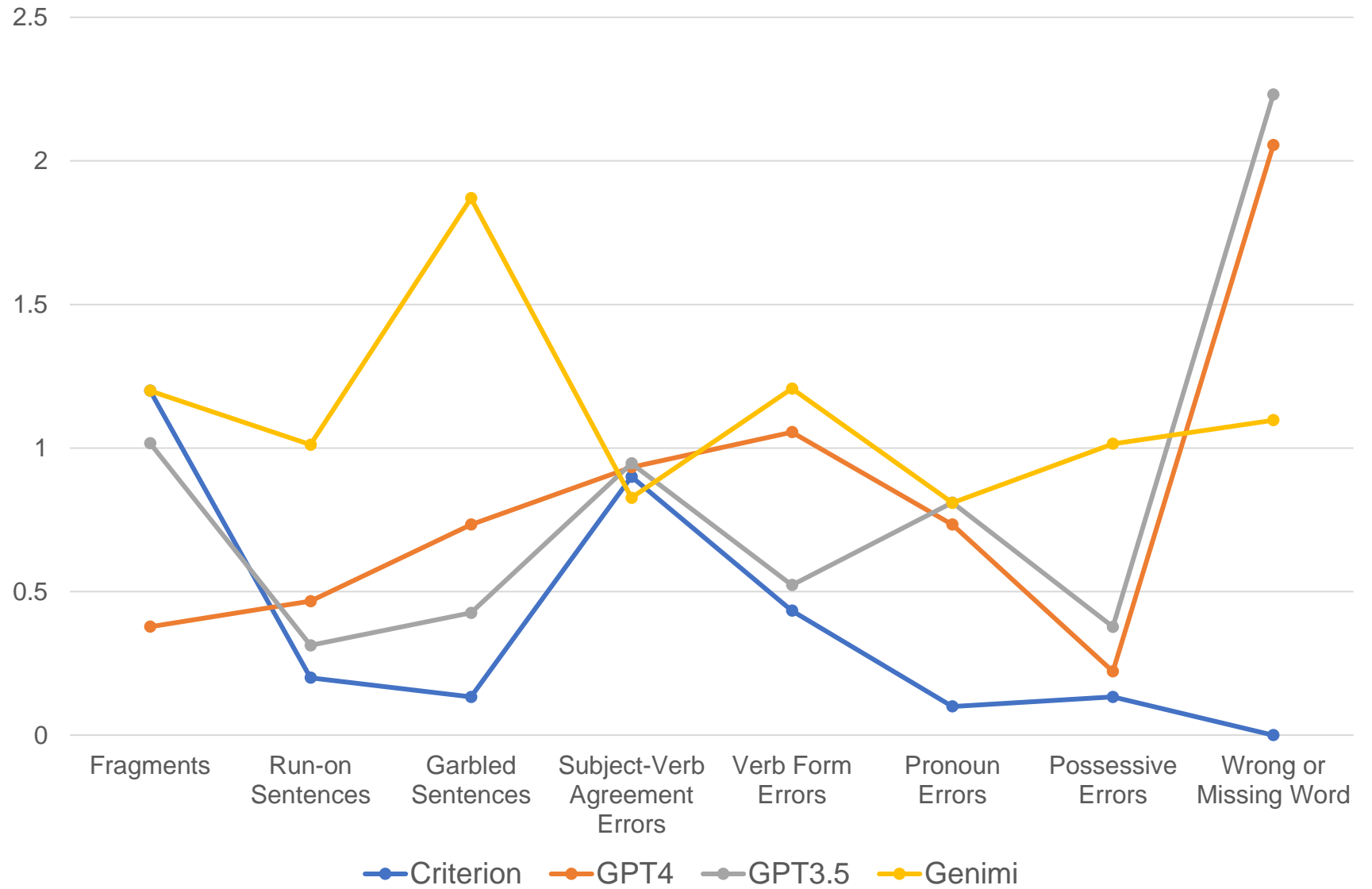


検証 2

平均エラー検出数

検証 2.1

平均エラー検出数

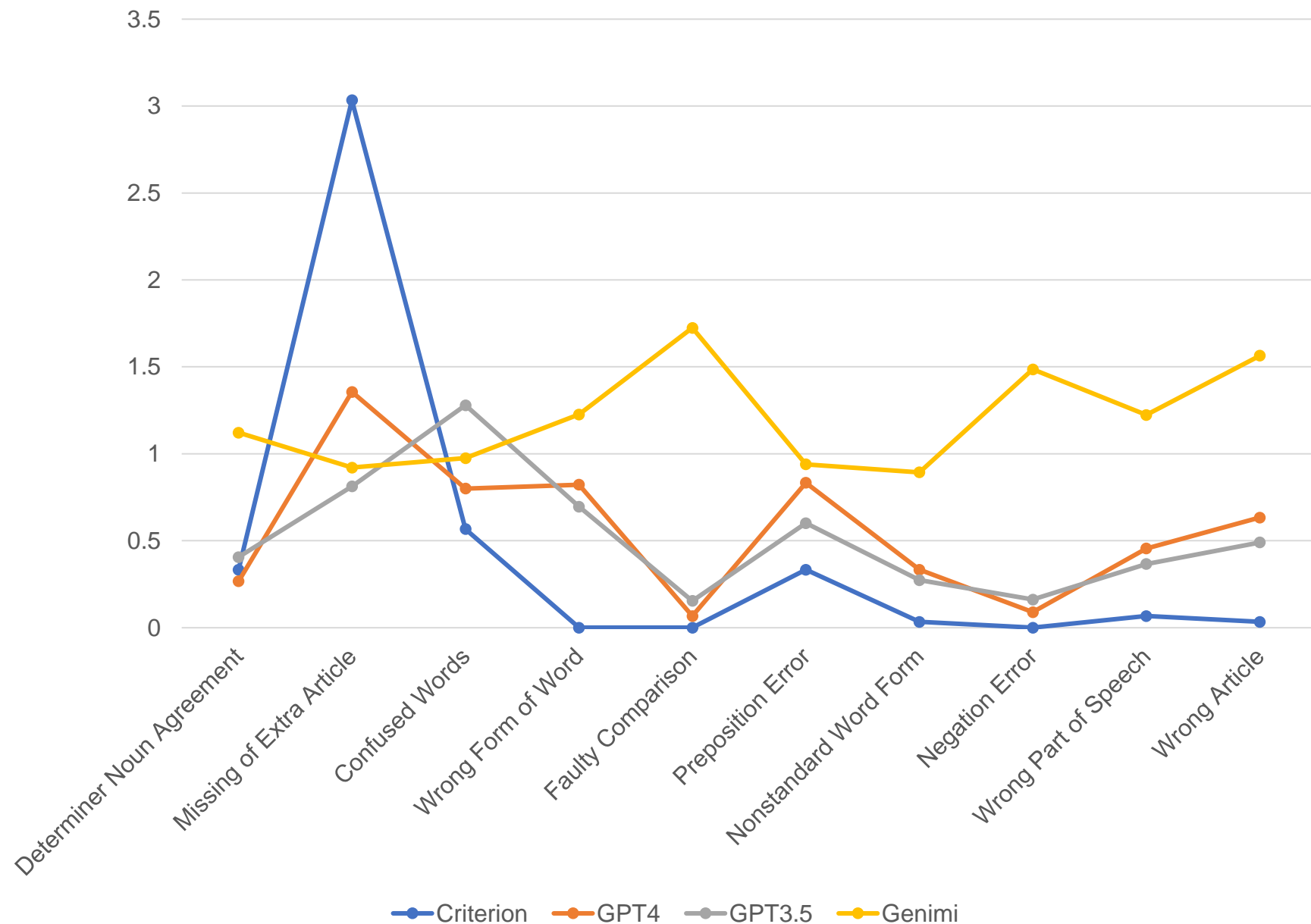


Grammar

検証 2.2

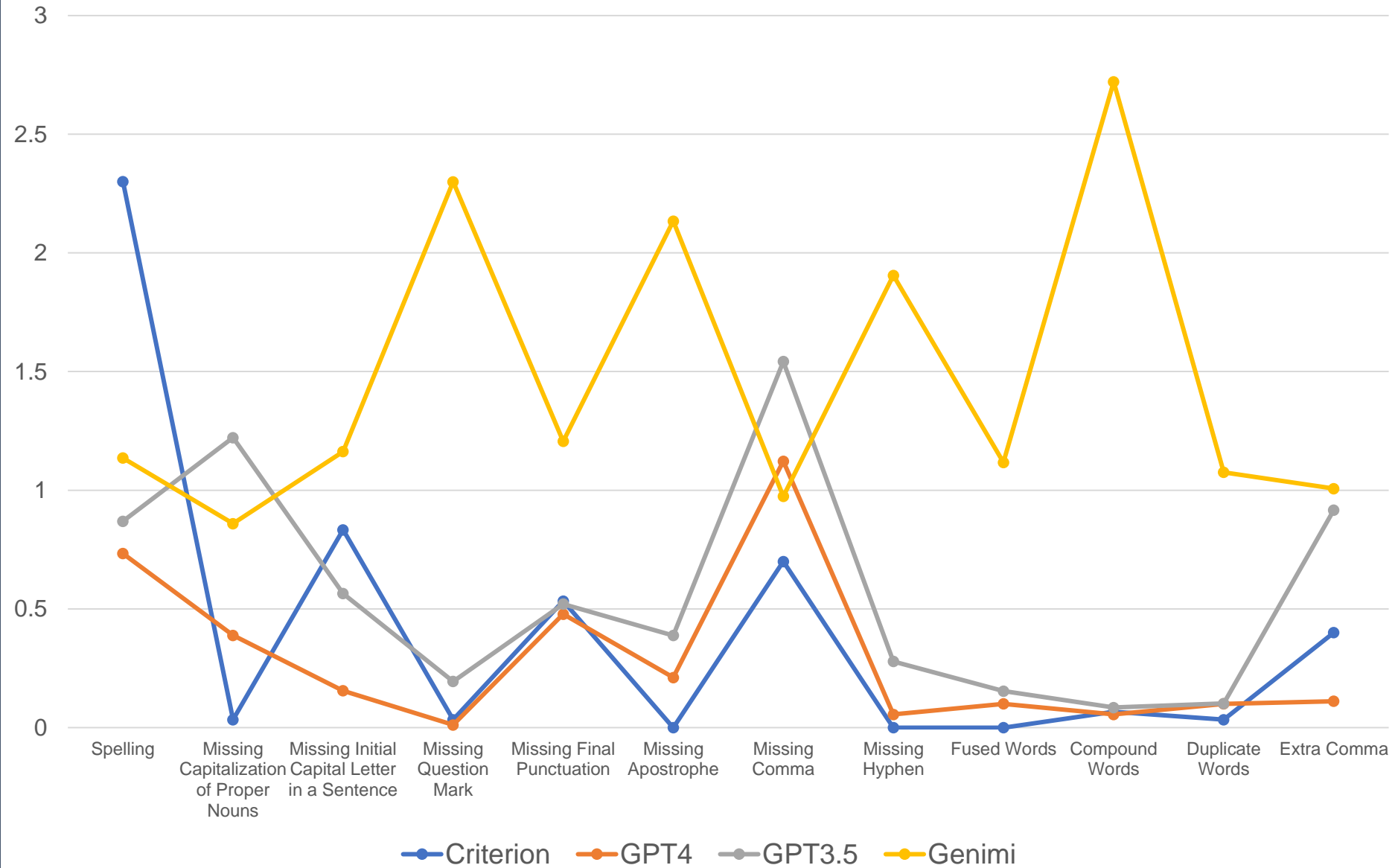
平均エラー検出数

Usage



検証 2.3

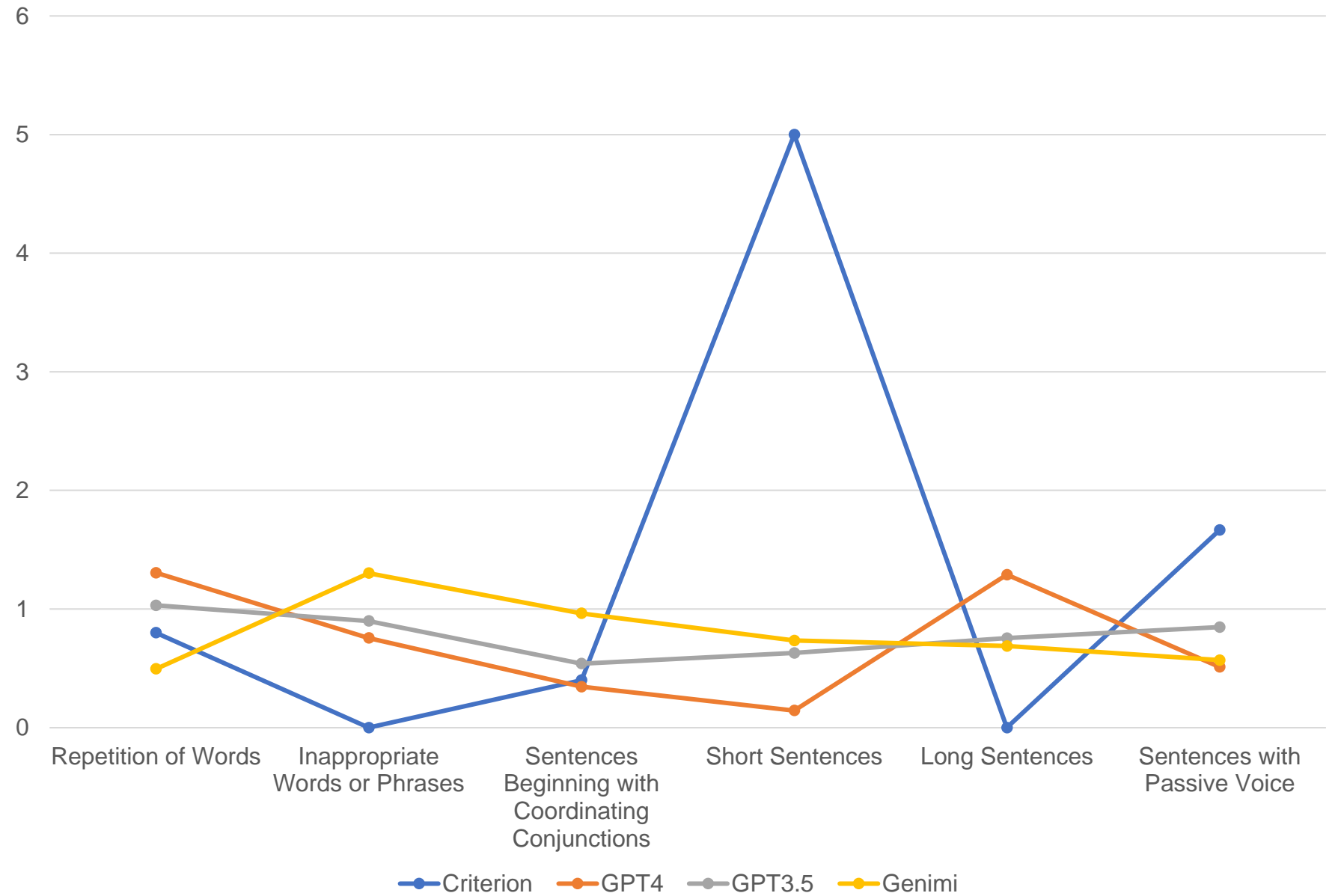
平均エラー検出数



Mechanics

検証 2.4

平均エラー検出数



Style

検証 3

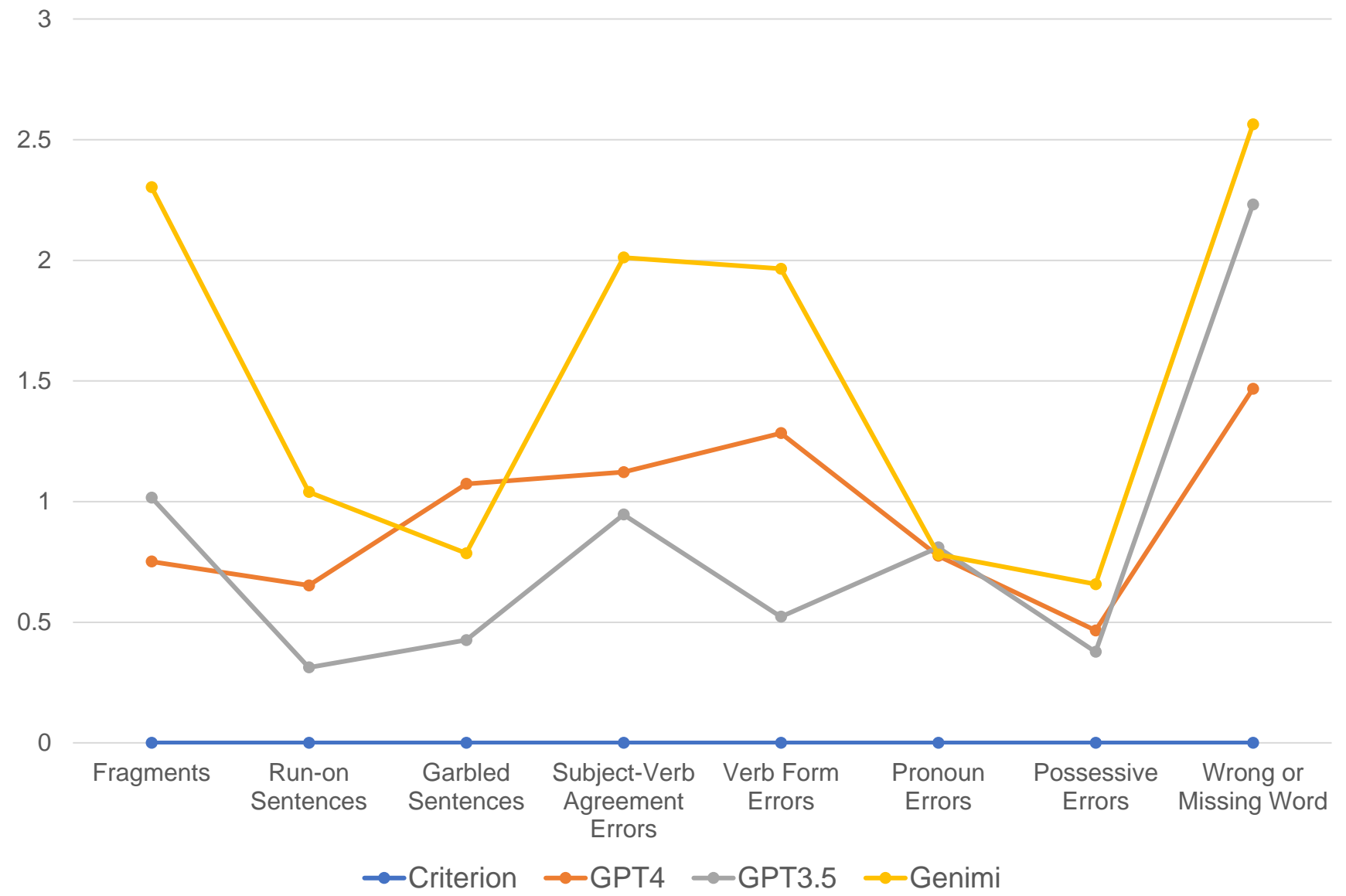
エラー検出数標準偏差

➤ 安定性の指標

検証 3.1

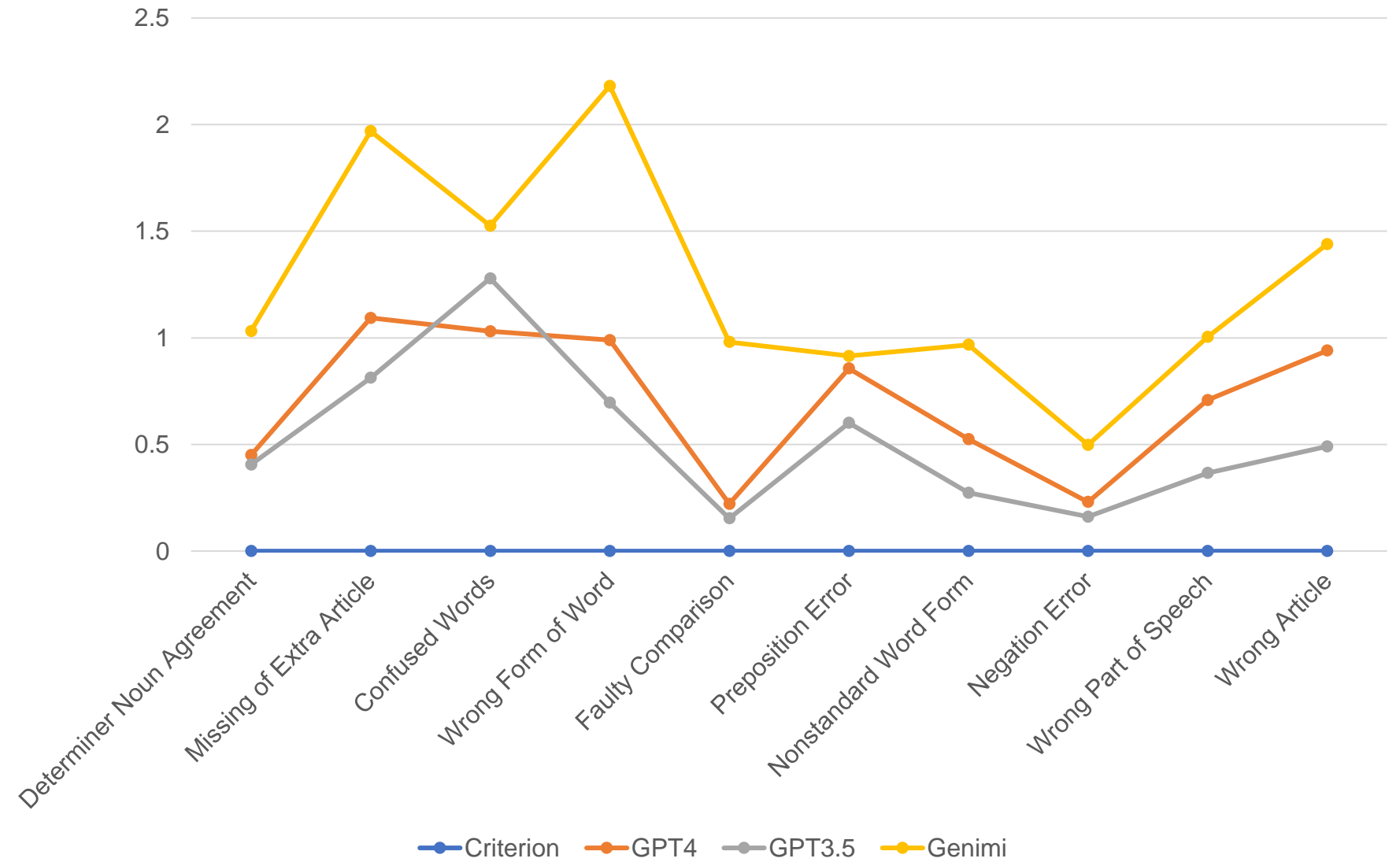
エラー検出数標準偏差

Grammar



検証 3.2

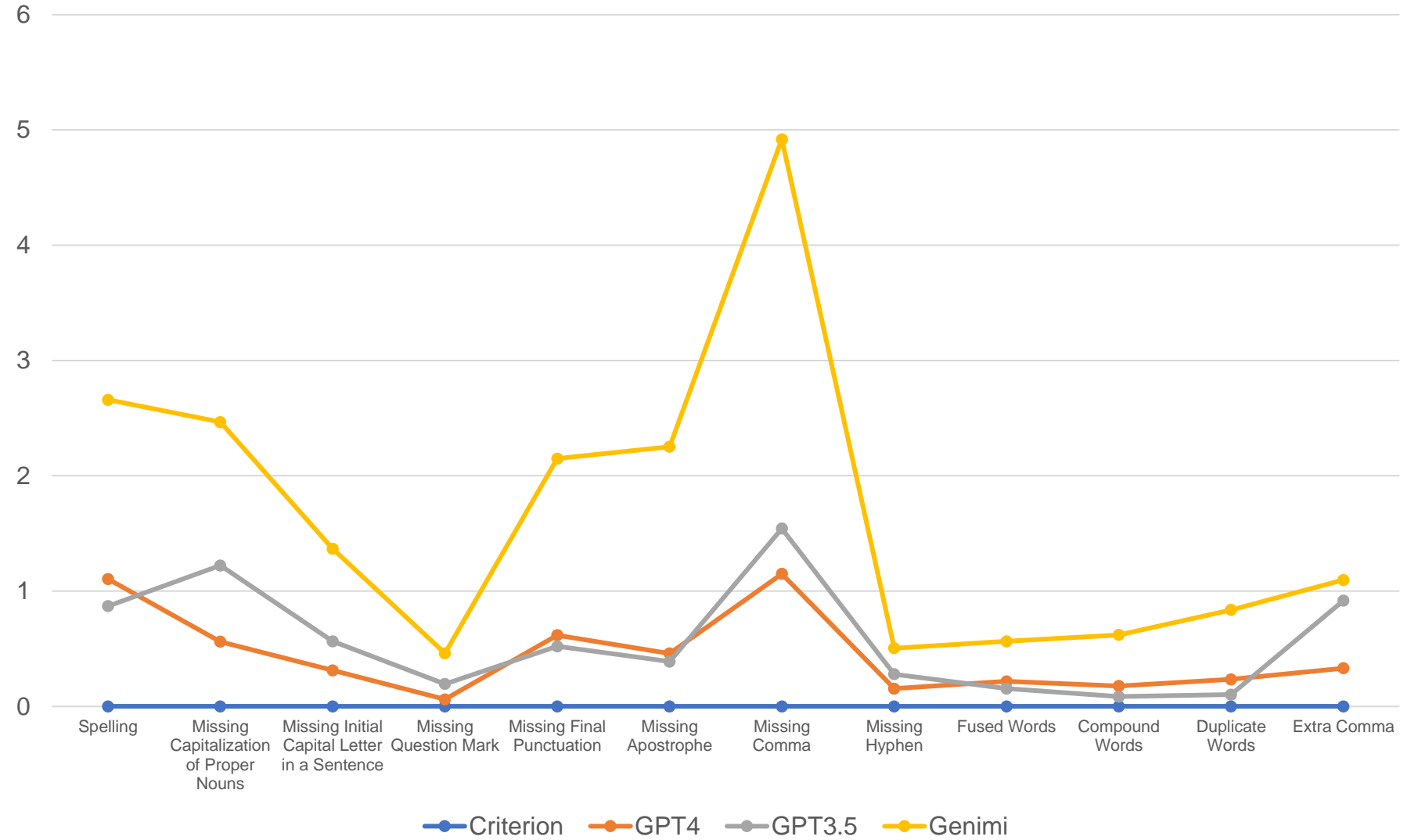
エラー検出数標準偏差



Usage

検証 3.3

エラー検出数標準偏差

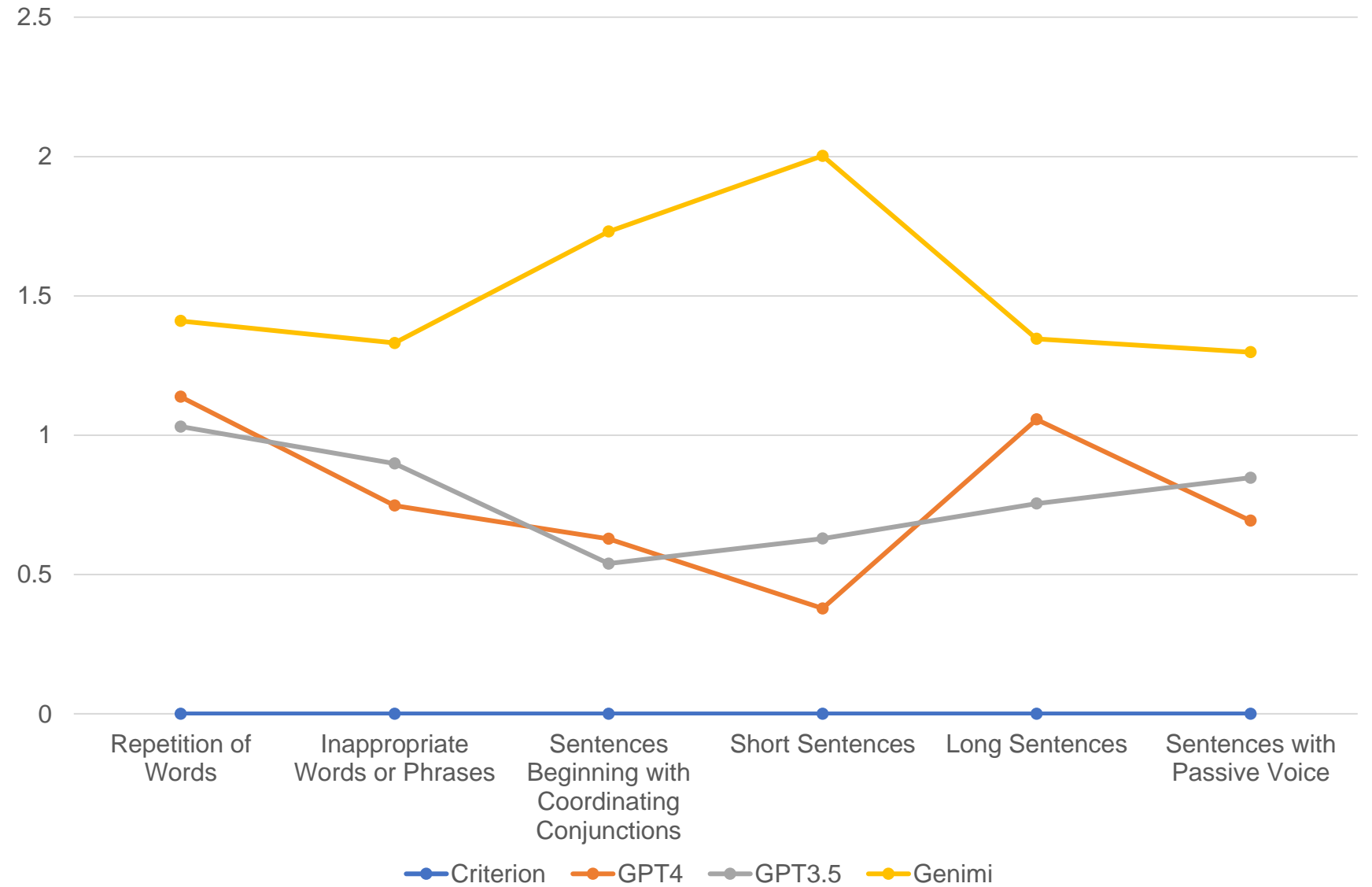


Mechanics

検証 3.4

エラー検出数標準偏差

Style



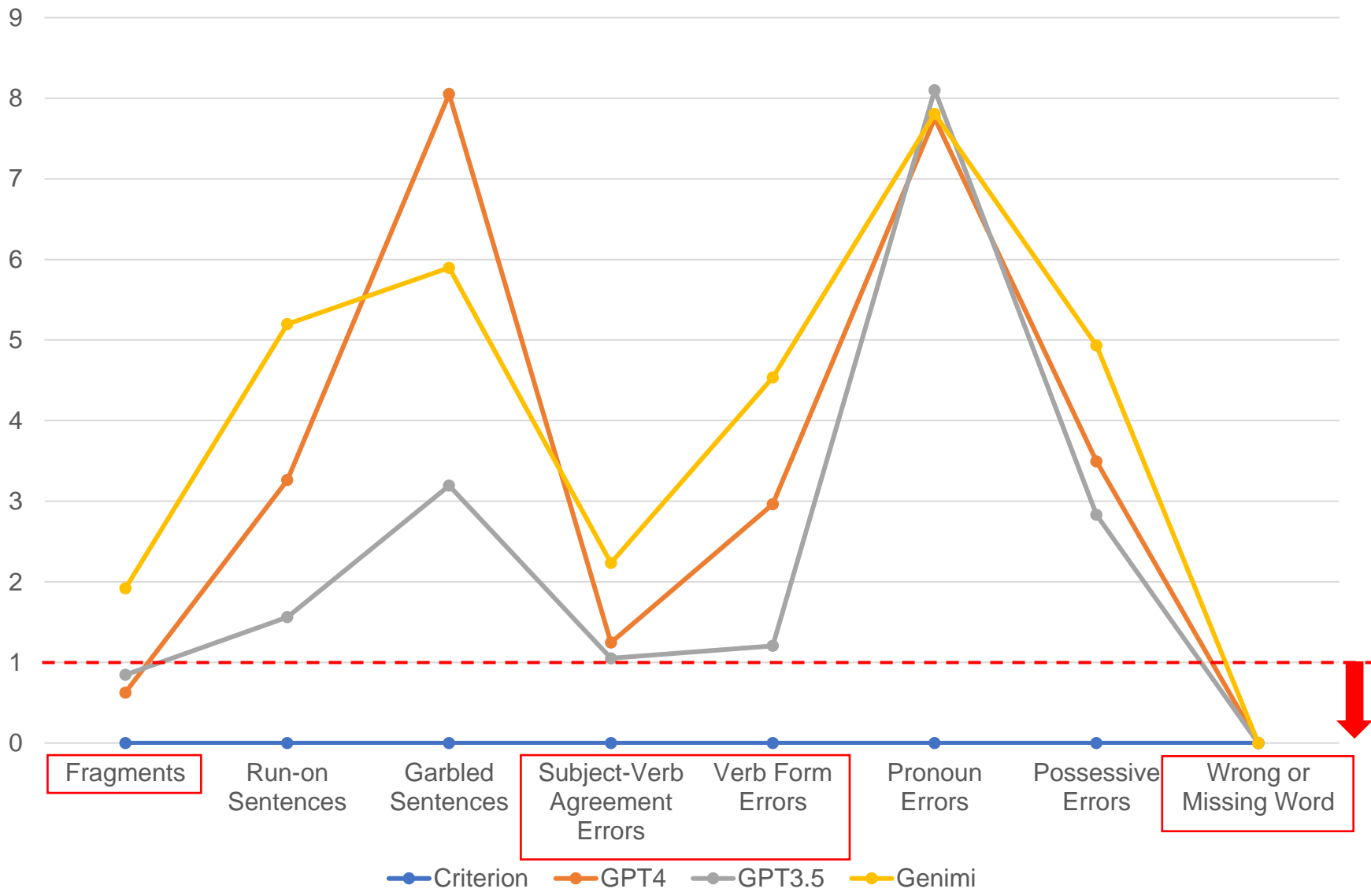
検証 4

エラー検出 変動係数

➤ 検出精度の指標

検証 4.1

変動係数

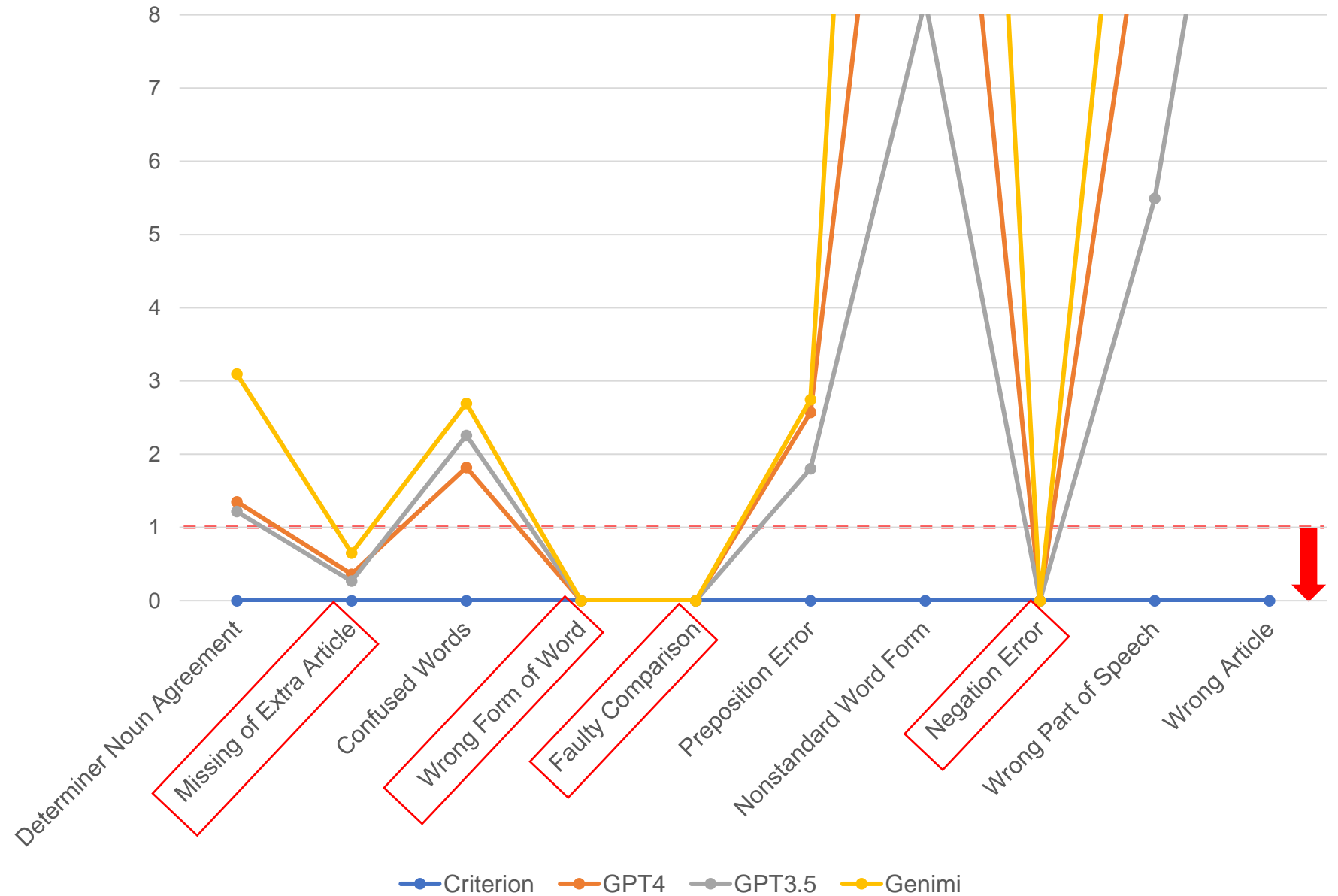


Grammar

検証 4.2

変動係数

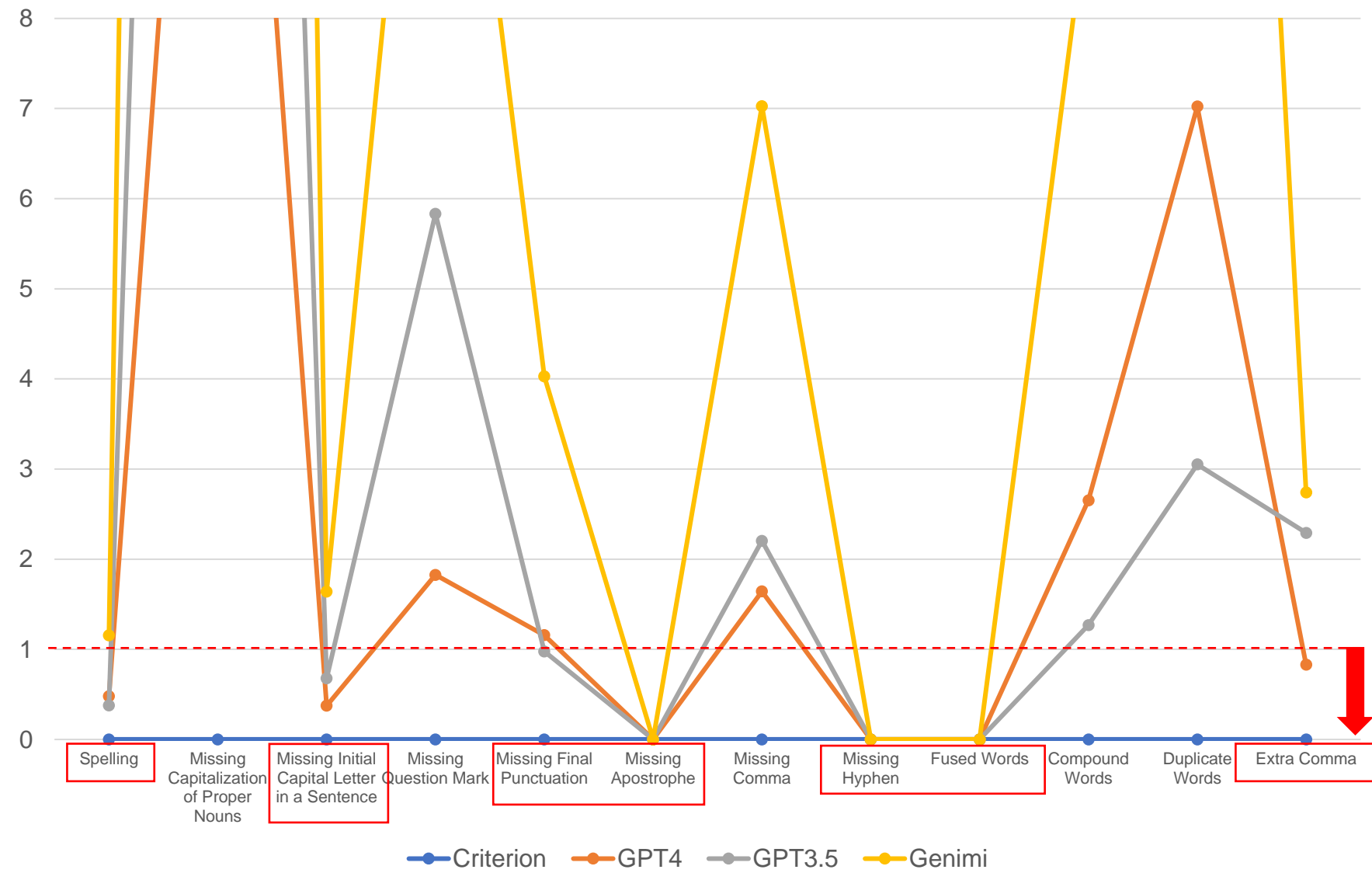
Usage



検証 4.3

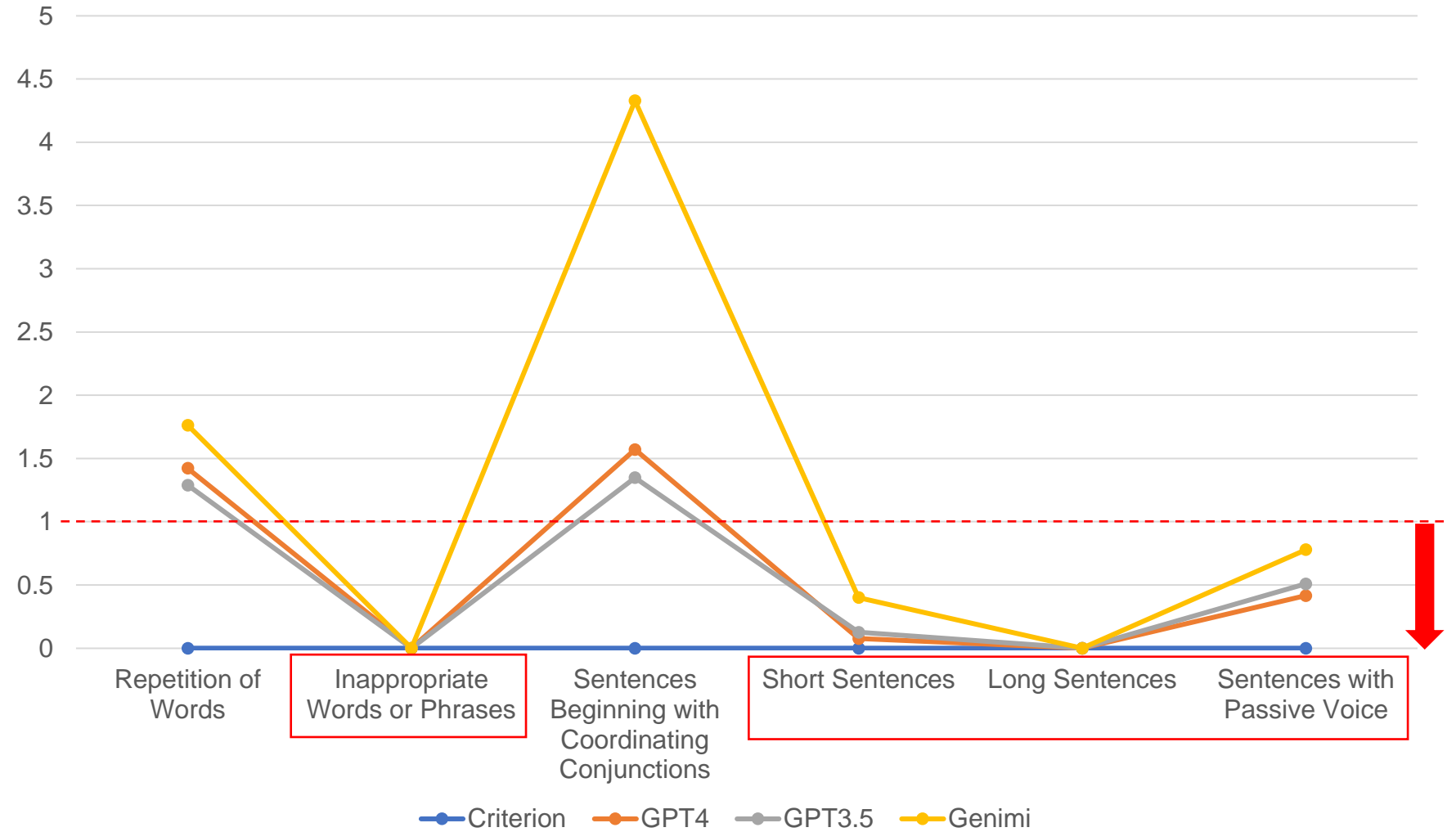
変動係数

Mechanics



検証 4.4

変動係数



Style

まとめ

生成AI (GPT4/GPT3.5/Gemini) の得意なこと

Grammar

- Fragments
- Subject-Verb Agreement Errors
- Wrong or Missing Word

Usage

- Determiner Noun Agreement
- Missing of Extra Article
- Faulty Comparison
- Negation Error

Mechanics

- Spelling
- Missing Initial Capital Letter in a Sentence
- Missing Final Punctuation
- Missing Apostrophe
- Missing Hyphen
- Fused Words
- Extra Comma

Style

- Inappropriate Words or Phrases
- (Sentences Beginning with Coordinating Conjunctions)
- Short Sentences
- Long Sentences
- Sentences with Passive Voice

Q.結局、どの生成AIが良いの？

A.Gemini (無料) < GPT3.5 (無料)

GPT4 (有料)

- 📄 生成AIで英文評価項目の**得手不得手**がある
- 📄 得手不得手を理解した上で**Criterion**の代替が可能である
- 📄 不得手の項目に特に**注意して校正**する必要がある
- 📄 生成AIを用いて英文評価することで**主体的な学習者の育成**可能性がある

参考文献

- Attali, Y. and Burstein, J. (2006) “Automated essay scoring with e-rater v.2,” *The Journal of Technology, Learning, and Assessment*, Vol. 4, No. 3, pp. 1–31.
- Bitchener, J. & Ferris, D. (2012) *Written Corrective Feedback in Second Language Acquisition and Writing*. Routledge.
- Lee, I. (2008) “Student Reactions to Teacher Feedback in Two Hong Kong Secondary Classrooms.” *Journal of Second Language Writing* 17(3), pp.144-164.

※本研究は文科省科研費基盤Cの助成を受けて実施された。
(課題番号23K00724)