

# ChatGPT を用いた記述式問題の採点

## Scoring Descriptive Questions Using ChatGPT

川本 響生<sup>\*1</sup>, 藤堂 健世<sup>\*1\*2</sup>, 高橋 聡<sup>\*1</sup>

Hibiki KAWAMOTO<sup>\*1</sup>, Kense TODO<sup>\*1\*2</sup>, Satoshi TAKAHASHI<sup>\*1</sup>

<sup>\*1</sup> 関東学院大学 理工学部

<sup>\*1</sup> College of Science and Engineering, Kanto Gakuin university

<sup>\*2</sup> 東京工業大学 情報理工部院

<sup>\*2</sup> School of Computing, Tokyo Institute of Technology

Email:f20K6034@kanto-gakuin.ac.jp

**あらまし**：近年、教員の不足が問題となっている。不足している原因として、学校行事や地域貢献など、多岐に渡る膨大な業務量が一つの原因だと考えられる。業務量を改善するために、AI の活用が挙げられている。しかし、AI が教師の多岐に渡る業務に対応できるのか活用事例が少なく、検証が不足している。本研究では採点業務に着目し ChatGPT を用いて国語の自由記述式問題の採点を行えるか調査する。ChatGPT には解答が違う内容を 5 件入力し、採点を行った。

**キーワード**：ChatGPT, 自由記述式, 記述式問題, 全国学力・学習状況調査

### 1. はじめに

2021 年 1 月末に文部科学省から公開された「教師不足」による実態調査<sup>(1)</sup>では、2021 年 5 月の時点で全国 2056 人の教員が不足していることが示された。教師が不足している原因の一つとして、その膨大な業務量の多さが負担になっていると考えられる。教師の業務は多岐に渡っており、授業だけではなく、学校行事や地域貢献など様々な業務が挙げられる。

教師の業務量を減らすための解決策として、AI の活用が挙げられる。文部科学省が取り組んでいるデジタル教育において、AI の活用に取り組んでいるが、AI が本当に教師の業務を遂行できるのか現場での活用事例が少なく、検証が不足している。そこで、本研究では採点業務に着目し ChatGPT<sup>(2)</sup>を用いて記述式問題の採点が行えるかどうか、調査を行う。

本研究では、国語の自由記述式問題を採用し採点を行う。国語の自由記述式問題に着目した理由として、ChatGPT は対話型 AI のため文章の理解することができ、国語の自由記述式問題の採点が可能ではないかと考えた。ChatGPT はユーザーが入力した内容に対して会話形式で文章を表示する機能がある。この機能を利用して、国語の自由記述式問題を採点する。また、国語の自由記述式問題は評価基準や文字数の多さから採点に時間が掛かり、採点者の負担が大きい。そのため、もし ChatGPT での採点が成功すると、大幅な採点業務の改善に繋がると考えている。

### 2. 提案手法

#### 2.1 利用する記述式問題

ChatGPT で採点する記述式問題は国語の自由記述式問題は、平成 20 年度学生全国学力・学習状況調査、調査問題小学生国語 B<sup>(3)</sup> (以下国語 B とする)の問題

を採用した。

国語 B の問題を採用する理由は、記述式形式であり、本文、問題、評価基準、模範解答が明確に表記されているためである。また、ChatGPT は特性上、明確な情報を多く入力しなければならない。明確な情報を入力しなければ誤った結果を出力してしまう。そこで、明確な情報が多い記述式形式の国語 B の問題を利用する。採点に利用した本文、問題、模範解答を図 1、図 2、図 3 に示す<sup>(3)</sup>。

中田さん 先生、こんにちは。インタビューを引き受けて頂いてありがとうございます。  
 それでは、よろしくお願ひします。先生は、スポーツがお好きだと聞いております。  
 そこで、まずスポーツに関する質問からお聞きします。

高山先生 いいですよ。なんでも聞いてください。

中田さん 先生の得意なスポーツは何ですか。また、いつから始めました。

高山先生 得意なスポーツは、野球、水泳です。特に野球が得意で、小学校四年生から始めました。  
 高校三年生のときに、全国高等学校野球選手権大会に出場して、甲子園に行くことができたんですよ。あの時の感動は忘れられないなあ。

中田さん 次の質問です。先生は、どのような時にスポーツの楽しさを感じますか。

高山先生 体を動かし、汗をいっぱいかいて、清々しい気持ちになる時ですね。

中田さん スポーツを通して、どんなことを学んでこられましたか。

高山先生 より高い目標に向かってあきらめずに努力すること、そして、チームワークの大切さですね。また、ルールを守ることの大切さを学びましたよ。

中田さん 先生が子供の頃からスポーツを楽しみ、色々なことを学んでこられたことがわかりました。  
 それでは、スポーツ以外のお話を伺いたいと思います。

図 1 本文<sup>(3)</sup>

先生は、スポーツがお好きだと聞いております。そこで、まずスポーツに関する質問からお聞きします。  
 この質問に対して「良かった」という評価になった。そのような評価になった理由を四十字以内で説明してください。

図 2 問題文<sup>(3)</sup>

前もって聞いていたことを元に、先生が答えやすい内容を取り上げているから

図 3 模範解答<sup>(3)</sup>

#### 2.2 ChatGPT に入力する内容

ChatGPT には本番の始まりに「あなたには国語の記述式問題の採点を行って貰います。」と入力し、小

学生全国学力・学習状況調査の調査問題で使用されている評価基準を参考とし、入力する。図1、図2、図3で示した本文、問題文、模範解答を入力する。その後、図4に示すA、B、C、D、Eの5件の解答を入力する。本番の終わりに「解答者A、B、C、D、Eの採点結果を表示してください」と入力することでA、B、C、D、Eの採点結果を出力する。

A、B、C、D、Eの解答は採点者の観点から筆者が作成した。採点時に問題となるのが、正誤の判断よりも正解ではあるが完全正解とは評価できない解答に対しての採点である。そのためA、Eの解答は正解と不正解が明確な解答を作成し、B、C、Dの3件の解答は正解ではあるが完全正解とは評価できない解答を作成した。本研究では◎、○、×の3段階で評価を行うため、Aが◎、B、C、Dが○、Eが×の評価となる。

評価基準は小学生全国学力・学習状況調査の調査問題の評価基準を参考するため、解答が文章の趣旨に合っている場合の評価を◎とする。文章の趣旨に合っているが言葉や内容が足りない解答の評価は○、文章の趣旨に合っていないければ×と評価する。模範解答<sup>(3)</sup>を参考にA、B、C、D、Eの解答を作成し、小学生全国学力・学習状況調査の調査問題の評価基準に合わせて採点を行う。

A	事前に聞いていないことを元に、先生が答えやすい内容を選んでいるから
B	先生が答えやすい内容を選択している
C	先生の得意なことを聞いている
D	質問に対して的確に答えられるような内容を選択している
E	前もって聞いていないことを、取り上げているから

図4 A、B、C、D、Eの解答

### 2.3 ChatGPTによる採点

ChatGPTに5件の解答を10回採点させ、A、B、C、D、Eの採点結果を出力させる。出力した採点結果をもとに記述式問題の採点を行うことができるか調査を行った。

### 3. 採点結果

表1はChatGPTを用いて採点を行った採点結果である。10回の採点を通して、A、Bは◎の評価を10回受け、Eは×の評価を10回受ける結果になった。Cは○の評価を9回受け、×の評価を1回受けた。Dは◎の評価を6回、○の評価を4回受ける結果となった。

表1 ChatGPTを用いた採点結果

評価	A	B	C	D	E
期待結果	◎	○	○	○	×
◎	10	10	0	6	0
○	0	0	9	4	0
×	0	0	1	0	10

### 3.1 考察

今回、筆者が作成した解答は、期待結果がAが◎、B、C、Dは○、Eは×の評価が期待される解答であった。ChatGPTの10回の採点を通して、Aは10回とも◎、Eは10回とも×と期待結果通りになった。しかし、Bは10回とも◎、Cは○が9回、×が1回という評価になりDは◎が6回、○が4回という結果になった。B、C、Dの解答は10回とも○の評価を受ける解答であり、期待結果とは違う採点結果になった。

今回、ChatGPTの採点ではAとEの採点結果から正誤の判断を正しく行えた。しかし、○と評価すべき解答に対して採点の評価を変えてしまっている。このことから正誤が明確な解答に対しては、評価を下すことができるが、文章の趣旨に合っているが言葉や内容が足りない解答に対しては正確な採点を行えない課題があることがわかった。

文章の趣旨に合っているが言葉や内容が足りない解答に対して正確な採点を行うことができない原因として、評価基準の伝え方に問題があると考えられる。本研究では、小学生全国学力・学習状況調査の調査問題で使用されている評価基準を参考とし、入力した。しかし、それではChatGPTにB、C、Dの解答に対して、正しい採点を行わせることができなかった。そのため、ChatGPTに合わせた評価基準を作成する必要がある。

### 4. まとめ

本研究ではChatGPTを用いて記述式問題の採点を行えるか検証した。文章の趣旨に合っている解答、文章の趣旨に合っていない解答の採点は可能であった。しかし、本研究では利用した手法では文章の趣旨に合っているが言葉や内容が足りない解答の採点を行うことはできない結果となった。文章の趣旨に合っているが言葉や内容が足りない解答に対して正しい採点ができない以上、学校現場で使用するには課題が残ることがわかった。

#### 参考文献

- (1) 文部科学省：“「教師不足」に関する実態調査”，[https://www.mext.go.jp/a\\_menu/shotou/kyoin/mext\\_00003.html](https://www.mext.go.jp/a_menu/shotou/kyoin/mext_00003.html)（参照 2023.5.30）
- (2) OpenAI：“ChatGPT”，<https://openai.com/product/chatgpt>（参照 2023.5.30）
- (3) 国立教育政策研究所：“平成20年度学生全国学力・学習状況について”，[www.nier.go.jp/08tyousa/08tyousa.htm](http://www.nier.go.jp/08tyousa/08tyousa.htm)（参照 2023.5.30）
- (4) 文部科学省：“文部科学省におけるデジタル推進人材育成の取り組みについて”，[https://www.mext.go.jp/content/20201223-mxt\\_kanseisk01](https://www.mext.go.jp/content/20201223-mxt_kanseisk01)（参照 2023.5.30）