

## 生成系 AI を用いた学習成果の評価

## Learning Outcomes Assessment Using Generative AI

染谷 銀志<sup>\*1</sup>, 上野 春毅<sup>\*2</sup>, 小松川 浩<sup>\*1</sup>Ginji SOMEYA<sup>\*1</sup>, Haruki UENO<sup>\*2</sup>, Hiroshi KOMATSUGAWA<sup>\*1</sup><sup>\*1</sup> 公立千歳科学技術大学大学院 理工学研究科<sup>\*1</sup> Graduate School of Science and Technology, Chitose Institute of Science and Technology<sup>\*2</sup> 公立千歳科学技術大学 理工学部<sup>\*2</sup> Faculty of Science and Technology, Chitose Institute of Science and Technology

Email: m2240320@photon.chitose.ac.jp

あらまし：著者らは、学習者のレポートや授業の目標設定といった学習成果を、生成系 AI を用いて自動評価し、その結果を学習者にフィードバックした。自動評価の結果と、教員による手動評価の結果とを比較し、どの程度自動評価が可能か検証するとともに、AI による評価を学習者にフィードバックすることの有用性を評価した。

キーワード：生成系 AI, 学習成果, 主体性, 自動評価, フィードバック

## 1. はじめに

学習者本位の教育が求められる高等教育では、自律的に学習をする力が必要で、基本的な学習態度としての主体性を育むことが重要となる。このためには、教師が学習者の学習成果を評価し、それを学習者に戻す中で、学習者のメタ認知を高める取組が求められる。しかし、こうした取組を実効化するには、教員の定期的な評価に対する人的な負担を考慮しつつ、学習者の自律的な活動に繋げていくための有効な方策の検討が必要となる。

本研究では、生成系 AI を活用した学習成果の自動評価と、これを用いた学生へのフィードバック方策の有用性を調べ、主体性誘引の可能性を探ることを目的とする。この際、溝上が定義する主体性の3層モデル<sup>(1)</sup>のうち、短期的主体性と中期的主体性に着目する。そして、授業課題に対する評価と振り返り支援（短期的主体性）と、授業全体の目標設定と振り返り支援（中期的主体性）それぞれを生成系 AI によってどこまで自動評価できるかを検証する。その上で、AI 評価を学習者にフィードバックする方策の有用性を評価する。

## 2. 生成系 AI を用いた自動評価

本研究では、短期的主体性をレポート課題から評価し、中期的主体性を目標設定から評価した。目標設定は半学期（15週）分の授業を見据えて記述するためである。本章では、短期的主体性と中期的主体性を、どのような評価方法によって、どの程度生成系 AI が自動評価できたのか記述する。

## 2.1 短期的主体性の自動評価

## 2.1.1 レポート課題の評価基準

本項では、短期的主体性の評価対象であるレポート課題の評価基準について説明する。本研究では、公立千歳科学技術大学秋学期1年の授業内でレポートを課した。レポートの設問を基に評価観点を6つ定め、各観点の記述の有無を評価基準とし、二値分

類の6点満点とした。

## 2.1.2 自動評価方法

本項ではレポートの自動評価の流れ（図1）について説明する。まず、レポート（学習成果）の文章データと教員の定めた評価観点の情報を、生成系 AI である ChatGPT のモデル 4o のプロンプト部分に組み込んだ。ChatGPT によって自動評価（AI 評価）された結果と、教員による手動評価の結果を比較し、どの程度自動評価が可能か検証した。

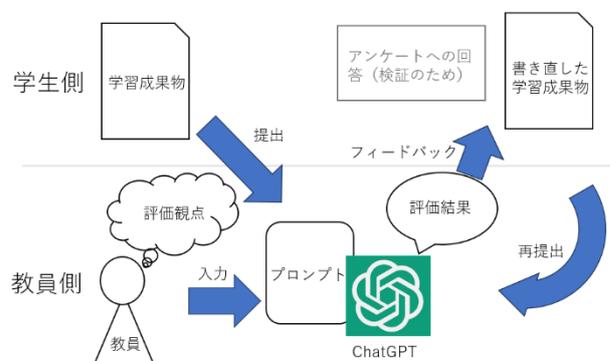


図1 自動評価の流れ

## 2.1.3 自動評価結果

レポートについて、教員による評価結果の平均値と AI 評価の平均値（n=226）を表1に示す。

表1 短期的主体性の評価結果（6点満点）

	評価点数
教員評価の平均値	4.75
AI 評価の平均値	5.47

平均値には 0.7 点程の差があるが、6つの評価観点別で評価した場合、一致率が約9割以上の観点4つと、一致率が8割を下回る観点2つに分かれた。

## 2.2 中期的主体性の自動評価

### 2.2.1 目標設定の評価基準

本研究では、公立千歳科学技術大学秋学期1年の授業内で目標・計画・方略の3つの設問を設定した。目標では「授業を通してどのような力を身につけるか」を記述し、計画では「設定した目標を達成するためにどのように取り組むか」を記述し、方略では「興味を持って取り組むことや挑戦したいこと」を記述する。これらをまとめて目標設定とする。

目標設定の評価基準は、まず、初回授業で学生が記述した内容を教員が確認したのちに、目標・計画・方略のそれぞれで5~6つの評価観点を定めた。レポートと同様に各観点の記述の有無を評価基準とし、二値分類としたが、目標設定については5個以上の観点を記述した場合も最大4点とした。

### 2.2.2 自動評価方法

目標設定の自動評価の流れは、レポートの自動評価の流れと同様である。目標・計画・方略のそれぞれについて、文章と評価観点の情報をChatGPTのプロンプトに与えた。

### 2.2.3 自動評価結果

目標設定 (n=237) の評価結果を表2に示す。

表2 中期的主体性の評価結果 (各4点満点)

	目標	計画	方略
教員評価の平均値	2.72	2.91	2.82
AI評価の平均値	2.62	2.38	2.92

教員評価の平均値とAI評価の平均値は、計画の項目において差が見られるが、他の2項目では近い値であった。

## 3. AI評価のフィードバック方策の有用性

自動評価した結果を、学生にフィードバックする方策の有用性を評価した。

### 3.1 レポートにおけるフィードバック方策の有用性

#### 3.1.1 授業設計とフィードバック方策の構築

自動評価により出力される評価結果を、個々の学生にフィードバックし、その結果を受けた上で学生にレポートの再提出を可能とする授業設計とした(図1)。具体的には、学生自身が記述したレポートに対するAI評価を、各学生がメールで確認できるようにした。評価結果には、自身のレポートに各評価観点の記述が含まれているか否かの○×、そのように判断した理由、合計の評価点数が示されている。以上をフィードバックした上で、学生にアンケートを課し、評価観点の情報を共有した。この評価観点の情報を基に、AI評価の結果が正しいと思うか否か(AI評価妥当性)を回答させた。また、フィードバックの結果を受け、学生が自身の記述内容に不足があると判断した場合は、評価結果に関わらず再提出可能とした。再提出した場合は、改めて評価された結果が再びメールから確認可能である。学習者の学

びを支援するこの一連の処理を、本研究ではフィードバック方策と定義する。

#### 3.1.2 フィードバック方策の有用性評価

AI評価妥当性、再提出した学生の割合、再提出後のAI評価の平均値を表3に示す。

表3 レポートのフィードバック結果 (6点満点)

	評価点数
学生視点のAI評価妥当性	98.9%
再提出割合	11.1%
再提出後のAI評価の平均値	5.63

学生視点では、AI評価が正しいと判断した学生がほとんどであり、再提出する学生が約1割と少ない結果になった。しかし、表1のAI評価の平均値より評価点数が0.16点ほど上昇した。

## 3.2 目標設定におけるフィードバック方策の有用性

### 3.2.1 授業設計とフィードバック方策の構築

レポート同様の授業設計とフィードバック方策のもと、フィードバックやアンケートを実施した。

### 3.2.2 フィードバック方策の有用性評価

AI評価妥当性、再提出した学生の割合、再提出後のAI評価の平均値を表4に示す。

表4 目標設定のフィードバック結果(各4点満点)

	目標	計画	方略
学生視点のAI評価妥当性	90.2%	82.8%	93.1%
再提出割合	70.7%	71.8%	63.2%
再提出後のAI評価の平均値	3.49	3.30	3.57

学生視点でAI評価が妥当と判断した学生が約8~9割であるにも関わらず、再提出する学生の割合が7割前後と高い。そして、再提出後のAI評価では平均値が0.65~0.95程度高い。

## 4. おわりに

本研究では、学習成果の自動評価による結果と、教員評価による結果を比較するとともに、フィードバック方策を構築し、その有用性を評価した。自動評価については教員評価とある程度一致することが認められ、生成系AIによって学生の能力を評価できる可能性があることが示唆された。また、AI評価を学生にフィードバックすることで、学習者の自律的な活動に繋げることができる可能性が示された。

今後は、本研究で用いたフィードバック方策の仕組みを、他大学でも活用可能な汎用性のある仕組みに改善することで、他の教育機関においても能力評価が可能であるか検証する予定である。

### 参考文献

- (1) 溝上慎一: “社会に生きる個性-自己と他者・拡張的パーソナリティ・エージェンシー”, 東信堂, 東京 (2020)