

サイバー大学研究紀要

Research Bulletin of Cyber University

第3号
(2026年3月)



目 次

研究報告

マイクロクレデンシャルとオープンバッジによる継続学習への効果……川原 洋 …… 1

生成 AI 時代における大学教育の質向上

— 上位認知レベル育成と教育システム — …… 除村 健俊 …… 9

Operational Efficiency and Feasibility of Generative AI
in English Writing Instruction:

A Study on Teacher-led Assessment Model ……Kensaku Ishimaki …… 23

サイバー大学における実践的データサイエンス教育の統合設計

— イベント・自由参加コース・ハイブリッド型ハッカソン／アイデアソンの役割と効果 —
……………藤澤 弘美子・浅田 麻菜・安間 文彦 …… 43

短 報

Model Context Protocol による学習管理システム拡張の試み

— LTI との併用による生成 AI への接続 — …… 田中 頼人 …… 49

解 説

Evaluating MEXT's Plan to Replace Japanese University

English Entrance Exams with Commercial Assessments
…………… Jared R. Baierschmidt …… 55

外部資金による研究プロジェクト紹介

相関関係に基づくラーニングパス最適化による

定量化個人学習支援 …… 陳 健・池田 大樹 …… 69

AI 社会実装人材育成のための 3 段階プロトタイプ構築法の

開発と分野対応教育法 …… 大江 信宏・除村 健俊 …… 71

マイクロクレデンシャルと オープンバッジによる継続学習への効果

川原 洋¹

1. 概要

2024年度よりサイバー大学は従来の学位プログラムをコース・プログラム制（CP）の履修科目体系を維持しながら、マイクロクレデンシャル制（MC）へ全面的に移行した。MCの狙いは、卒業研究科目に至る体系的な専門分野の学力向上にある。さらに全く異なる他の専門分野、あるいは網羅的な同分野の学習意欲も助長するという、専門教育における専門性と多様性の両立である。MCの運用は開始からまだ2年経たところではあるが、MCの制度的効果が、一般および本学卒業生の再入学による科目履修を目的とする学習者数の著しい増加という形で現れ始めている。

本稿では、これらの学位の取得にとらわれない一般科目等履修生と卒業生による継続的学習のインセンティブと、これを助成するキャリア形成を目指す高等教育のあり方を提言する。

2. サイバー大学のマイクロクレデンシャルによる教育目的

2.1. 学習者主体の専門性の強化と多様性の両立

サイバー大学がMCを導入した第一の目的は、この学習者主体の専門性強化と多様性の両立である。これは、伝統的な大学教育における「専攻（Major）」と「副専攻（Minor）」のアナロジーを、デジタル時代の柔軟な単位構成によって再定義する試みである。

従来の学位プログラムでは、学生は入学時に選択した学科やコースの枠組みに縛られがちであり、卒業に必要な単位を揃えることが目的化する懸念があった。しかしMC制では、卒業単位修得に至る科目履修体系を細分化・テーマ別にクラスター化することで、学生は自らの主専攻（Platinum バッジを目指す経路）を維持しながら、並行して他の学修分野の上位MC（Gold バッジ）を獲得することを「副専攻」として定義できる。この仕組みにより、例えば「AIエンジニア」を目指す学生が、同時に「デジタルマーケティング

¹ サイバー大学 学長

グ」や「経営管理」の体系的な知識を証明可能な形で修得することができる。

2.2. オープンバッジの獲得欲を学修インセンティブに転化

教育工学的な観点から、学習の継続には適切なフィードバックとマイルストーンの設定が不可欠である。サイバー大学は、学修分野別の履修証明書をオープンバッジ¹⁾により付与することで、学習者の「獲得欲」を直接的な履修インセンティブへと転化させるシステムを構築した。

大学教育における「単位」は、学内での卒業要件としては機能するものの、学外、特に労働市場においてはその具体的な内容が不可視であることが多い。オープンバッジは、国際標準規格に基づいたメタデータを保持しているため、バッジ取得者がどのような知識を習得し、どのような評価基準をクリアしたかを、第三者が即座に検証することが可能である。この「可視化された資産」としてのバッジが、単なる「単位の積み上げ」という無機質な作業を、自身の学修ポートフォリオを構築するクリエイティブな活動へと変容させている。アンケート結果において、62.2%の学生が「科目を受講する意欲が向上する」と回答している事実は、このインセンティブ設計の有効性を如実に物語っている（後掲の図3参照）。

2.3. テクノロジーの進化と社会情勢の変化への即応

いわゆる情報化社会における情報技術の陳腐化は劇的に加速している。サイバー大学のMCは、包含される学修内容を随時更新し、バージョン管理を行う「ダイナミック（動的）」な設計を採用している点に大きな特徴がある。これは、一度取得すれば半永久的に効力を持ち内容が固定される「学位」や、静的な「マイクロディグリー」とは一線を画す概念である。

例えば、生成AIの急速な普及に伴い、2025年度からは「生成AI活用」分野の教育プログラムが追加され、客員教員としてシリコンバレーの起業家や技術者を招聘するなど、最先端の知見を即座にカリキュラムへと反映させた。しかし、学習者は、一度MCを取得した後も、社会情勢の変化に応じて更新される最新のMCを獲得するために、科目等履修生として学び続ける動機を持つことになる。このように、MCは単なる「修了証明」ではなく、学習者を生涯にわたる「知識やスキルのアップデート」へと誘う羅針盤として機能している。

3. サイバー大学のマイクロクレデンシャルの特徴と構成

3.1. MCの厳格な質保証とダイナミックな更新性

サイバー大学のMCは、その社会的価値を担保するために、厳格な質保証の枠組みに基づいている。MCの対象となるのは、大学または大学と共同で実施される学修成果の評価が行われた教育プログラムや科目のみに限定されている。これは、大学側が評価の実態

を把握できない外部の講座やセミナーへの参加、あるいは「学位」そのものの修了を MC の対象外とするという、明確な境界線に基づいている²⁾。

また、前述の通り、MC はバージョン管理が行われており、産業界のニーズや技術革新に合わせてその構成科目が動的に入れ替わる。この「ダイナミック MC」のモデルは、学修者が常に最新のスキルセットを証明できることを意味し、これが履歴書や SNS (LinkedIn 等) での学修歴証明としての実用性を高めている。2025 年 12 月現在、オープンバッジ 3.0³⁾ の普及により、分散型アイデンティティ (DID)⁴⁾ 技術を活用した真正性の証明が可能となり、サイバー大学の MC はこの最新の技術基盤に支えられている。

3.2. 階層化された積み上げ型構造

サイバー大学の MC は、学習の到達度と難易度を視覚的に明示するため、バッジの色による階層化 (Bronze、Silver、Gold、Platinum) を行っている。この積み上げ型 (スタッカブル) の構造は、学生が長期的な学習目標を見失わず、一歩ずつ専門性を深めていくためのガイドとなる (図 1、図 2)。



図 1 専門科目 MC とオープンバッジ構造

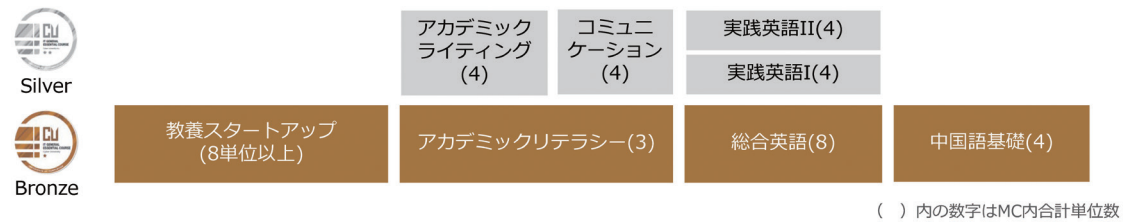


図 2 語学教養科目 MC とオープンバッジ構造

この構造において、各バッジの取得には指定されたすべての科目の単位修得が必須とされている。これにより、断片的なスキルの習得ではなく、大学教育としての体系性が維持されている。特に、卒業研究を Platinum バッジと位置づけることで、学位取得というマクロな目標と、個別のスキル習得というマイクロな目標が、一つの連続したパスとして統合されている。

3.3. リベラルアーツと専門教育の融合

サイバー大学の教育方針である「ビジネスのわかる IT エンジニア、IT のわかるビジネスパーソン」の育成を実現するため、MC はハードスキル（専門知識）だけでなく、ソフトスキル（リベラルアーツ）も包括している。

リベラルアーツ系 MC である「教養スタートアップ」は、人文科学、社会科学、自然科学、キャリアリテラシーの4分野、計69科目から構成される大規模なモジュールである。また、アカデミックリテラシー（スタディスキルやアカデミックライティングなど）や、グローバルコミュニケーションを支える英語・中国語のバッジも用意されている。これにより、学生はITやビジネスの専門性を深めると同時に、批判的思考力、創造力、コラボレーション能力といった、AI時代にこそ価値が増す汎用的な能力を証明可能な形で修得できる。

4. 学生による MC 制への反応

4.1. 在校生の意識変化と学習動機の変容

2024年6月に実施された在校生向けアンケート調査の結果は、MC制が学生の学修行動にポジティブな影響を与えていることを裏付けている（図3）。アンケートに回答したMC制が開始される以前に入学し、卒業までには時間がある在校生の74.4%はMCカリキュラムへ移行した理由として「オープンバッジを取得したかった」と回答している。またMC制への期待についても、62.2%の学生が「科目を受講する意欲が向上する」と前向きな姿勢の回答をしている。

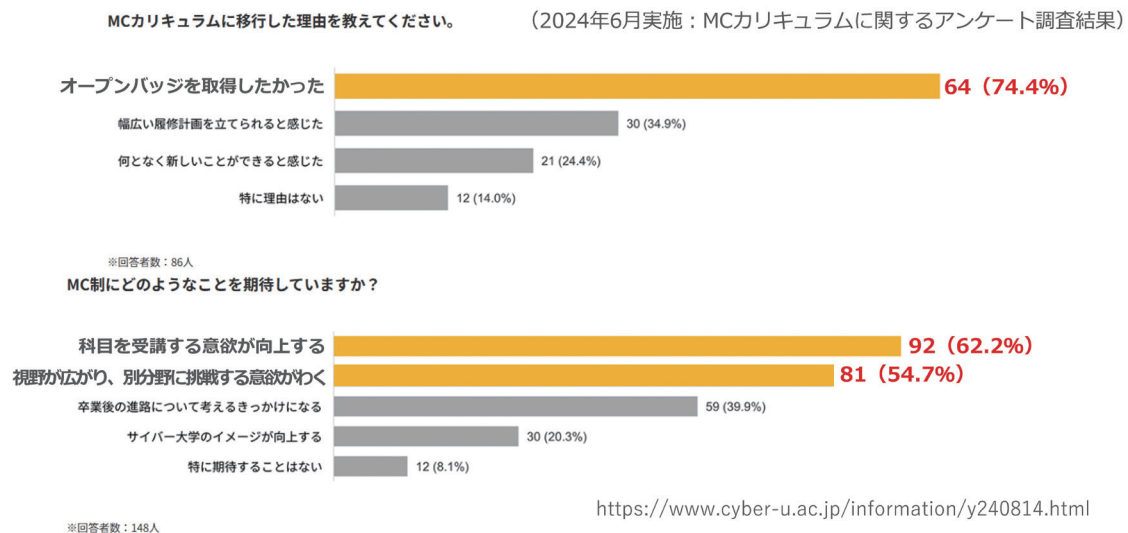


図3 MCに関する学生アンケート結果（一部）

これらのデータは、MCが単なる証明書の発行にとどまらず、学生の「探索的学習」を

促進していることを示唆している。特に 54.7% が「別分野への挑戦意欲」を挙げている点は、冒頭で述べた「専門性と多様性の両立」が、制度レベルだけでなく学生の意識レベルでも達成されつつあることを示している。

4.2. 科目等履修生および卒業生の再入学者数の著しい増加

MC 導入による最も顕著な定量的変化は、科目等履修生および卒業生の履修動向である。サイバー大学では、2024 年度の MC 全面移行に伴い、オープンバッジの獲得を目的とする在学中に未履修だった科目の単位を修得するために再入学する卒業生が急増した。以下の図 4 は、科目等履修生と卒業生科目等履修生の推移を示している。

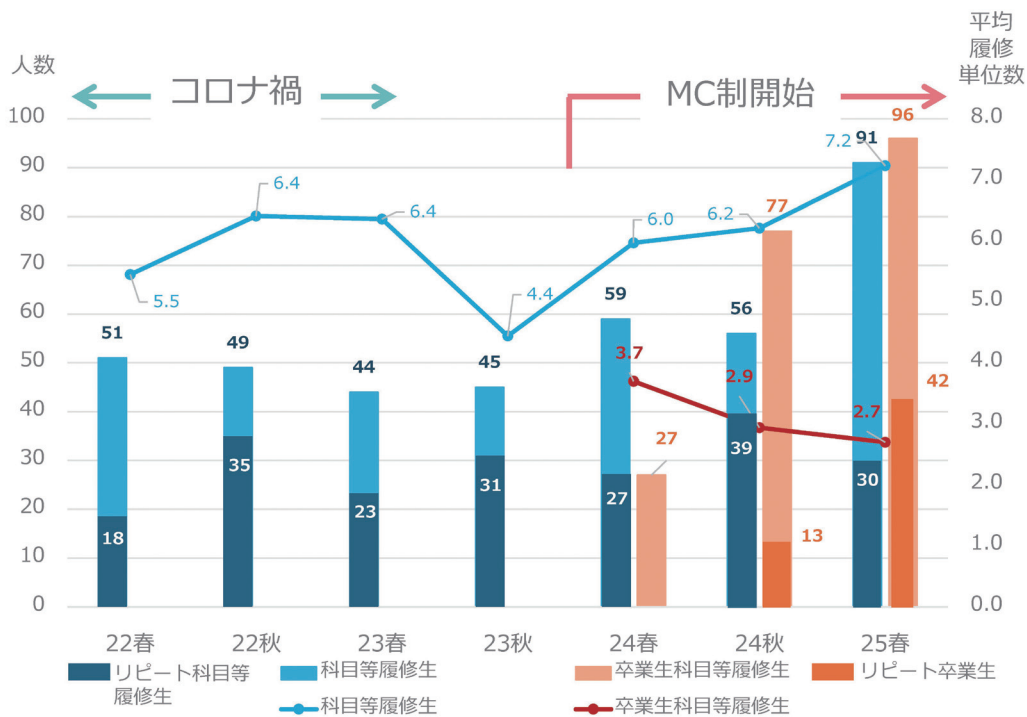


図 4 MC 導入前後の一般・卒業生科目等履修生数と平均履修単位数の推移

2024 年度 (MC 開始年度) より卒業生による履修者数が大幅に増加している

以前より卒業生が科目等履修生として再入学することはあったが、MC を開始した 2024 年度春学期より「卒業生向け科目等履修生受け入れ制度」⁵⁾ を開始したことで、2024 年度秋学期において、卒業生の再入学者数が一般の科目等履修生数を超え、2025 年度春学期には両者とも高い水準を維持している。特筆すべきは、科目等履修生 1 人あたりの平均履修単位数が 2022 年度春学期の 5.5 単位から 2025 年度春学期には 7.2 単位へと増加している。これは科目ごとの興味関心に基づく受講から、バッジ取得という体系的な目標に基づいた受講へと、社会人の意欲的な学習スタイルに変化したことを物語っている。

5. まとめと今後の挑戦

5.1. 教育モデルの根本的な転換：Degree から Competency へ

サイバー大学における MC 導入から2年間の軌跡は、高等教育が「学位 (Degree)」という一つの大きな証書を授与する場から、個々の「能力 (Competency)」を連続的に証明し、蓄積していくプラットフォームへと変容しつつあることを紹介した。特に卒業生の再入学の急増は、大学卒業資格が「4年後の到達点」ではなく、生涯にわたりスキルをアップデートし、その正当な証明 (オープンバッジ) を取得し続けるための「恒久的な学習拠点」となり得ることを示している。

オープンバッジは、学習者にとっての「収集の楽しみ」や「小刻みな達成感」という内発的動機付けを引き出すと同時に、労働市場における「スキルの可視化」という外発的価値を両立させる極めて有効なメディアとして機能している。これは、大学教育の質を卒業生の就職率というフローの指標だけでなく、卒業後の継続的な学習参加率というストックの指標で評価する新たな時代の幕開けと言える。

5.2. 今後の技術的・制度的課題

MC とオープンバッジの真の社会実装に向け、克服すべき課題も明確になりつつある。最も重要な課題は産業界による大学教育の評価である。MC が包含する学修歴が、企業の採用や人事評価、昇進において「共通の物差し」として機能するためには、大学側からの積極的な働きかけと、法人向け人材育成プログラムとしての認知獲得が不可欠である。

5.3. 未来への提言：継続学習のコミュニティとしての大学

サイバー大学の試みは、変化の激しい現代社会における「大学の在るべき姿」に対する一つのモデルを提示している。教育コンテンツの継続的な更新を行い、魅力的な MC を継続的に提案し続け、学生や卒業生からの学習要望を共有する「継続学習のコミュニティ」への移行である。

今後は、AI やデータサイエンスといった先端技術分野のみならず、人文知や社会実装力を含めた多様な MC の体系化をさらに進める必要がある。また、正科生、科目等履修生、卒業生、そして企業の従業員といった多様な属性の学習者が、それぞれのニーズに合わせて MC を組み合わせ、独自のキャリアを形成していく「スキル・エコシステム」の構築を目指すべきである。サイバー大学が先導したこのマイクロクレデンシャル・モデルが、日本の高等教育全体のデジタルトランスフォーメーション (DX) を加速させ、多くの国民が「継続学習」を日常的慣習として身に付け、その成果としての自己実現によって社会に寄与することを切望する。

注および参考文献

- 1) オープンバッジとは（一般社団法人 オープンバッジ・ネットワーク）
<https://www.openbadge.or.jp/about-ob/>（2026/2/1 参照）
- 2) サイバー大学における MC の定義
https://www.cyber-u.ac.jp/about/pdf/self-check/mc_basic_policy.pdf（2026/2/1 参照）
- 3) OpenBadge 3.0 <https://www.openbadge.or.jp/about-ob/index.html#ob3>（2026/2/1 参照）
- 4) 分散型アイデンティティ（DID: Decentralized Identity）とは、**特定の企業や政府に依存せず、個人が自身のデジタル ID（識別子）を主体的に作成・管理できる仕組み**で、ブロックチェーンなどの分散型技術で実現され、プライバシー保護とセキュリティ向上を目指す。従来の ID 管理（Google や Facebook など）は中央集権的で情報漏洩リスクがあるが、DID では**ユーザー自身が情報をコントロールし、必要な情報だけを提示できる**点が特徴である（自己主権型アイデンティティ：SSI）。
- 5) 卒業生向け科目等履修生受け入れ制度
https://www.cyber-u.ac.jp/support/grad_nondegree.html（2026/2/1 参照）

生成 AI 時代における大学教育の質向上

— 上位認知レベル育成と教育システム —

除 村 健 俊¹

1. 概要

近年、生成 AI (Generative AI) の急速な発展は、大学教育の在り方に大きな変化をもたらしている。ChatGPT や Gemini などに代表される生成 AI は、知識の収集だけでなく、文章作成、プログラミング支援、画像・動画・音楽生成、知的創作など、従来人間が担ってきた知的活動の一部を代替可能にしつつある。従来、大学教育で教員は、知識を教え、知識の定着を評価し、研究において知識の応用を指導し、論文作成を支援する、という教育を行ってきた。しかし、生成 AI の登場により、学生は生成 AI を使い、自ら知識を修得し、自分で考えることなく問題を解くことができる。英作文、レポート、論文などの創作物は、学生が自分で作成したものか、生成 AI による創作物であるかの判断は困難になっている。このため、生成 AI は学生に対し個別化された学習体験や即座のフィードバックを提供し、学生指導の効率化を図る強力なツールとなる可能性がある一方、教育的誠実性 (Academic Integrity)、評価の公平性や倫理的枠組みへの懸念などについて、多数の論文で論じられている^{1) 2) 3)}。さらに、教育の在り方が激変する中で、より根本的な問いである、教育とは何か、学びとは何か、という哲学的な問いに立ち返ることが必要になっている。

本稿は、生成 AI 時代における大学教育の在り方を問い直し、改めて大学教育の質的向上を目的とした授業設計と教育評価の実践を報告するものである。授業設計と教育評価の実践として具体的には、サイバー大学 (以下、「本学」) において実施した①生成 AI 概論、②ソフトウェア工学の 2 つの授業を対象とし、それぞれの授業における生成 AI 活用による教育の質向上の工夫と学習成果を報告する。さらに、大学教育の質向上を支える教育システムに求められる要件と、生成 AI を活用した教育の姿について論じる。

¹ サイバー大学 IT 総合学部・教授

2. 生成 AI 時代の学びと大学教育に対する影響

従来、学びは知識の獲得や理解に重きが置かれていたが、生成 AI はこの領域を容易に自動化してしまう。このため、生成 AI の登場は、学びそのものを根底から問い直す契機となった。

学びを問い直す契機となった3つの事例を紹介しよう。

一つ目は、筆者が教えている生成 AI 概論という講義での出来事である。生成 AI が写真の中の手書き文字を認識できる事例として、ChatGPT に図1を入力した。図1は、自宅のBD-R ディスクに録画した映画のタイトルを手書きしたものである。

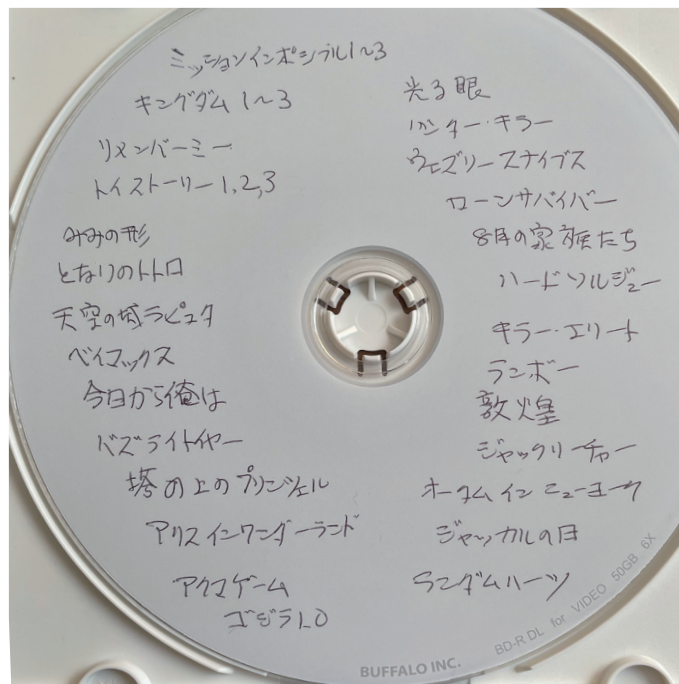


図1 ディスクに録画した映画の手書きタイトル

そして、以下のようなプロンプトを入力した。

「画像の文字は映画のタイトルです。映画のタイトルとして認識してください。」

その結果、以下のような回答が返ってきた。

ミッションインポッシブル 1~3

通常は『ミッションインポッシブル』1作目~3作目。

キングダム 1~3

『キングダム』実写映画は現在（2023年時点）3作あるため、その1~3作を指していると推測。

リメンバー・ミー

ディズニー/ピクサー映画『リメンバー・ミー』（原題“Coco”）。

トイ・ストーリー 1,2,3

ディズニー / ピクサー映画『トイ・ストーリー』シリーズ 1~3 作。

みみの形

おそらく『聲の形』（こうのかたち / A Silent Voice）を誤記している可能性があります。あるいは『耳をすませば』と混同しているのかもしれませんが、明確ではありません。

以下省略

ChatGPT は映画のタイトルという知識も使って手書き文字を認識している。ここで、「みみの形」と手書きで書いたのは、映画名の「聲の形」を読むことができず「みみ」と読んでしまい、また、ひらがなで書いたのは漢字が書けなかったためである。ChatGPT は、これを「こうのかたち」と読んだ。筆者は、これを信じてしまい、講義において学生に、「こうのかたち」と話してしまった。これを聞いた学生から、これは「こえのかたち」と読むという指摘を受けた。この出来事で筆者は自分の無知を恥じるとともに、知識がないと生成 AI の出力を容易に信用してしまうことを体験し、**基礎知識の重要性**を改めて認識した。

二つ目の事例は、「同じ 5 行のコードが全く違って見える 12 の瞬間」⁴⁾ という記事である。記事で紹介しているのは、userID から user 名を取り出す JavaScript の 5 行のコードである。

```
async function getUsername(userId) {  
  const response = await fetch(`https://api.example.com/users/${userId}`);  
  const user = await response.json();  
  return user.name;  
}
```

生成 AI を使うとコードは簡単に生成できてしまう。このコードは動作するので、初心者は、動けば OK で完成したと思ってしまう。一方、熟練者は、エラーハンドリング、パフォーマンス、セキュリティなど、記事では 12 の視点からコードの問題点を検証している。その結果、改善が反映されたコードは 98 行になっている。これは、同じ 5 行のコードを見ても、初心者と経験を積んだ熟練者では見えている世界が違うということである。生成 AI の出力に対しても、生成 AI との多面的な対話を通して、その品質を高めていくために、基礎知識と経験、さらにそれらに基づく**批判的思考力**が重要であることを示している。

三つ目の事例は問いを立てる力の重要性である。PBL などで「プロダクトの改善案を出す」という課題に取り組む場合があるが、単に提示された機能の良し悪しを評価するだけでは、表面的な分析にとどまってしまう。しかし、「本当に解くべき課題は何か」「ユー

ザーが不便だと感じているのはどこか」「そもそもこの機能は必要なのか」といった問いを自ら立てることで、検討対象が明確になり、新たな視点や改善の方向性が立ち上がってくる。問いを立てる力は思考の起点をつくり、解決すべき問題領域を適切に定義する役割を果たし、結果として批判的思考をより高いレベルへ導くことができる。このように、検討する対象そのものを創り出し、対話を通して的確な問いを立て、思考を深めていくためには、問いを立てる力が不可欠である。

すなわち、生成 AI の時代の学びとは、基礎知識の重要性が改めて認識され、さらに、批判的思考力や問いを立てる力を養い、対象物の「見えなかったものを見えるようにする力」を養うプロセスである、と考えることができる。

学習者の思考のレベルを整理する枠組みとして、表1に示すブルームの認知分類⁵⁾が知られている。

ブルームの認知分類において従来の教育はレベル1から3に対応すると考えることができる。これに対し、批判的思考力や問いを立てる力は、表2のようにレベル4から6に対応すると考えられる。

これらの上位領域こそが、批判的思考力や創造的問題解決力を支える中核である。たとえば、ソフトウェアの設計や AI 活用の学習において、初心者は「正しく動くコード」を求める段階に留まるが、熟練者は「なぜこの設計が適切なのか」「他により良い方法はあるか」を問い直す。この「問い直しの思考」が、まさに上位認知レベルへの到達を意味する。AI が生成した出力を鵜呑みにせず、設計意図やリスク、倫理的影響を分析・評価し、より望ましい解決を創り出すことができるかどうか、学びの成熟度を決定する。したがって、生成 AI 時代における大学教育は、知識伝達型の講義から、「問いを立て、比較し、創造する学び」へと再設計することが求められる。教員は、AI を禁止するのではな

表1 ブルームの認知分類

レベル	英語表現	日本語訳	代表的な行為例
①	Remembering	記憶・再生	用語・事実を覚える
②	Understanding	理解	説明する・要約する・概念を整理する
③	Applying	活用／応用	手法を使って解決する
④	Analyzing	分析	構造・関係を分解し意味づける
⑤	Evaluating	評価	判断基準に基づいて比較・評価する
⑥	Creating	創造	新しい構成・統合・設計を行う

表2 上位認知レベルと批判的思考力／問いを立てる力との関係

能力	ブルーム分類との関係	説明
批判的思考力	④分析 + ⑤評価	情報の構造を理解し、根拠や妥当性を見抜く思考。AI の出力を吟味する力。
問いを立てる力	④分析 + ⑥創造	問題を分解し、新しい問いや視点を構築する力。学びの出発点となる。

表 3 上位認知レベルを育成する教育アプローチ

教育方法	ねらい	実践例（ソフトウェア工学や AI 教育に応用）
① 問い直し型学習	学生に「なぜそう考えるのか」「他の方法は？」を問いかけ、思考の根拠を明示化させる。	コードレビューで「この設計選択の根拠は？」「他に可能なパターンは？」と問う。
② 比較・対照学習	複数のモデルや理論を比較する中で、評価軸を自分で構築させる。	2つのアーキテクチャ（例：MVC vs MVVM）を比較し、利点・欠点を分析。
③ ケースベース学習	現実事例を分析し、判断の妥当性を議論させる。	プロジェクト失敗事例を分析し、要因を特定して改善策を提案。
④ メタ認知リフレクション	「自分の考え方を振り返る」ことで、思考過程を内省化。	議論後に「自分の判断基準はどこにあったか」「どのデータを重視したか」を記述。
⑤ PBL + AI 活用	問題解決・意思決定・協働を通じて批判的思考を実践的に鍛える。	生成 AI を用いた要件分析・設計比較・意思決定プロセスを学生がチームで評価。

く、学生の思考を促す「学びの伴走者」として活用する立場に立たなければならない。AI を通じて思考のプロセスを可視化し、学生が自らの判断や価値基準を形成する過程を支援することこそ、教育の質向上の核心である。上位認知レベルを育成する教育アプローチとしては、表 3 のような教育方法がある。これらの教育方法に「ブレンディッド・ラーニング (Blended Learning)」、「反転授業 (Flipped Teaching)」を組み合わせ、批判的思考力を育成する教育も報告されている⁶⁾。

3. 実践事例

「2. 生成 AI 時代の学びと大学教育に対する影響」で述べた事柄は、生成 AI 時代の学びの構造モデルとして、図 2 のように示すことができる。

本項ではこのモデルに沿って実施した 2 つの講義事例を紹介する。事例の中には、①教材の提示や課題・レポート・演習などによる思考レベルの向上と②生成 AI 活用により可能になった教材の質向上も含まれる。

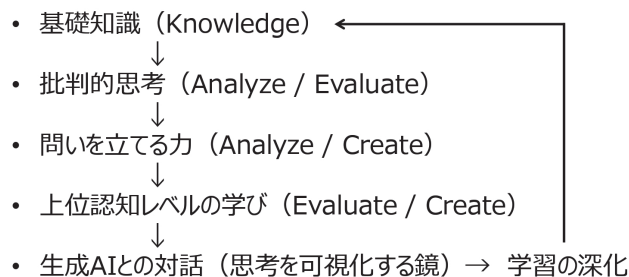


図 2 生成 AI 時代の学びの構造モデル

なお、Wagner は、生成 AI を使って楽をすることや効率を最優先し、学習のプロセス（苦労や試行錯誤）を省略する Vibe-Teaching を批判し、学生が AI を「思考のパートナー」として使い、適切なレベルの課題に挑み続けられる Flow-Teaching（没入型教育）の重要性を述べている⁷⁾。本項で述べる学習内容は、この没入型の教育につながる試みであるとも考えられる。

3.1. 生成 AI 概論

生成 AI は、学生の個別対応を可能にし、飛躍的に学習意欲や能力を高めることができる可能性があり、うまく使うと非常に有用なツールである。一方で、学生の剽窃意識が薄れる可能性や学習プロセスを省略してしまうリスクも懸念される。このため、生成 AI の基本的な活用方法を教えるリテラシー教育と生成 AI を活用する上で注意すべき倫理教育を行うことが必要であり、このために本学では生成 AI 概論という講義を提供している。

生成 AI 概論は大きく、基礎編、事例編、実践編の3部で構成され、各編の概要は表4の通りである。

表4 生成 AI 概論の構成

編	授業回	主要な学習テーマ
基礎編	第1回～ 第5回	・ AI の基本原理と技術 ・ 応用分野と活用事例 ・ 社会的影響、倫理、法規制の基礎
実践編	第6回～ 第10回	・ プロンプト・エンジニアリングの実践 ・ データ分析とビジネス企画への応用 ・ 業務効率化とエージェント AI への進化
事例編	第11回～ 第15回	・ デジタル遺産とアイデンティティの AI による再定義 ・ AI の自律化（エージェント）と分散 AI ・ 物理的世界の基幹産業における AI 応用フロンティア

事例編は、米国在住の講師が作成した講義であり、テーマ音楽、映像、講師まですべて AI で作成されている。内容は、大学の講義とは思えない、まるでドキュメンタリー番組を見るような高品質の講義となっている。

生成 AI 概論では、倫理教育とリテラシー教育が適切に行われている。筆者が担当している第6回、第7回の講義では、プロンプト・エンジニアリングや生成 AI を活用した各種事例を紹介し、生成 AI 活用の基礎となるリテラシー教育を行っている。各回の講義の終わりに下記のような演習を行っている。

生成 AI 活用事例をヒントに、日常生活や仕事での活用についてのレポートを提出してください。

問1. 作成したプロンプトをコピーしてください。（入力ファイルがある場合は添付してください。）

問2. このプロンプトを作成した意図を説明してください。

- 問 3.** プロンプト作成上の工夫を記述してください。(プロンプト・エンジニアリングのある手法を使った、など)
- 問 4.** 生成 AI の出力をコピーしてください。(出力ファイルがある場合は添付してください。)
- 問 5.** この生成 AI の演習で学んだことは何ですか。

この演習では、体験を通して生成 AI の活用方法を学ぶことに加え、2つの上位思考レベルの育成につながる問いを設定している。

問 2 : プロンプト作成により問いを立てる力の育成に加え、メタ認知リフレクションにより、なぜその問いを考えたかの振り返りを行い、認知レベル④(分析)の育成を試みている。

問 5 : 学んだことを振り返ることで、認知レベル④(分析)とレベル⑤(評価)の育成を試みている。

この演習で、学生は自らプロンプトの書き方による生成 AI 出力の精度変化を試してみたり、講義で教えた生成 AI 活用事例の応用を試したりするなど、自律的に学習する意欲も引き出すことができた。

本講義は 2025 年の秋学期に 460 人の学生が受講した。それらの学生が回答した、問 5 : 学んだことの振り返り、に書かれた主要コメントを紹介する。

- 生成 AI に対して「どういう形式で」「何を含めて」出力してほしいのかを明確にプロンプト化することで、専門的で実務に使えるドキュメントを短時間で作成できることを学びました。特に、手順書のように階層構造が重要なドキュメントは、Markdown 形式を指定することで整理された内容を出力でき、人間がゼロから書くよりも効率的で正確であると実感しました。
- プロンプトを生成 AI に提案してもらいながら修正を加える(生成 AI で壁打ちをする)と、自分の期待をより明確にしたプロンプトを作成でき、出力の精度をあげることができると感じました。
- 生成 AI が pdf ファイルを扱えることを利用して、内容の分析や要約を行えることを学びました。また、実用性を意識してプロンプトを作成した今回の経験は、今後、より実用的なプロンプトを作るための良いきっかけになりました。

学生コメントと認知レベルの対応は以下のように考えられる。

(a) 構造の理解と効率性の比較(分析・評価)

学生コメント:「…特に、手順書のように階層構造が重要なドキュメントは、Markdown 形式を指定することで整理された内容を出力でき、人間がゼロから書くよりも効率的で正確であると実感しました。」

・ **レベル④ 分析 (Analyzing)** : 学生は「手順書」というドキュメントが持つ「階層構造」という特性を見抜き、それに対して「Markdown 形式」という具体的な手法が適合するという構造的な関係性を理解している。これは情報の構造を分解し、意味づけを行う「分析」の段階にあたると思われる。

・ **レベル⑤ 評価 (Evaluating)** : 「人間がゼロから書くよりも効率的で正確」という記述は、AI による出力結果と人間の作業を比較し、効率性と正確性という「判断基準に基づいて評価」を行っている。これは単なる感想ではなく、比較対象を持った批判的な吟味 (Critical Thinking) ができていると考えられる。

(b) 対話による問いの精緻化 (分析・創造)

学生コメント : 「プロンプトを生成 AI に提案してもらいながら修正を加える (生成 AI で壁打ちをする) と、自分の期待をより明確にしたプロンプトを作成でき、出力の精度をあげることができると感じました。」

・ **レベル④ 分析 (Analyzing) + ⑥ 創造 (Creating) = 「問いを立てる力**」 : 「生成 AI で壁打ちをする」というプロセスは、AI からのフィードバックを元に、自身の要求 (問い) を再構築する行為である。これは本稿で「問いを立てる力」として定義されている能力であり、問題を分解 (分析) し、新しい問いや視点を構築 (創造) するプロセスに対応する。

・ **レベル④ 分析 (Analyzing) + メタ認知リフレクション (Metacognitive Reflection)** : 「自分の期待をより明確にした」という気づきは、思考のプロセス自体を振り返るメタ認知的な分析であり、思考の根拠を明示化させる高度な学習活動が行われたことを示している。

(c) 応用可能性の拡張と学習の統合 (応用・創造)

学生コメント : 「生成 AI が pdf ファイルを扱えることを利用して、内容の分析や要約を行えることを学びました。また、実用性を意識してプロンプトを作成した今回の経験は、今後、より実用的なプロンプトを作るための良いきっかけになりました。」

・ **レベル③ 活用 / 応用 (Applying)** : pdf ファイルを扱うという具体的な手法を学び、それをどのようなタスク (要約など) に使えるかを理解している点は、知識を具体的な場面で使う「活用」の段階である。

・ **レベル⑥ 創造 (Creating)** : 「今後、より実用的なプロンプトを作るための良いきっかけ」という記述は、今回の学習体験を将来の活動に統合しようとする姿勢を示している。これは単なる知識の記憶にとどまらず、将来の新しい問題解決に向けて自らの学習を設計・再構成しようとする「創造」的な態度の芽生えと解釈できる。

これらのコメントから本授業における学習活動と認知レベルは表5のように整理できる。

表 5 生成 AI 概論における学習活動と認知レベル

活動内容	主な認知レベル	学習で生じた思考
手順書構造の理解	④分析	構造の分解と手法適合性
効率性の比較	⑤評価	判断基準による評価
生成 AI との壁打ち (修正)	④分析 + ⑥創造	問いの再構築
思考プロセスの振り返り	④分析 + メタ認知リフレクション	思考の根拠の明示化
知識の学び	③応用	知識の活用
将来学習の創造	⑥創造	学習の設計・再構成

これらのコメントは、学生が単に AI に答えを出させる (Vibe-teaching) レベルに留まらず、AI を「思考のパートナー」や「壁打ち相手」として活用し、出力結果を批判的に吟味・評価しており、図 2 の生成 AI 時代の学びの構造モデル、に対応づけることができる。これは、AI が学習者の思考を可視化し、知識の伝達から「思考と創造の支援」へと教育の質が転換されたことを裏付ける証拠となっていると考えられる。

3.2. ソフトウェア工学

筆者が担当するソフトウェア工学の授業に生成 AI をいち早く取り入れた⁸⁾。その意図は以下の通りである。

- 生成 AI により高速にコードの自動生成が可能になっており、人間の付加価値は、ソフトウェアの要求仕様、アーキテクチャ設計など、より上流にシフトし、システム開発におけるソフトウェア工学の重要性が増すと考えたこと。
- 生成 AI を活用することで、従来、学生の理解が難しかったソフトウェア構造などの基礎概念に対し、悪い例や良い例の具体的な例を生成 AI で作成し、理解を促すことができると思ったこと。
- ソフトウェア開発における生成 AI の活用事例を示すことで、学生は自分でそれらを活用し応用できるようになると考えたこと。この事例の中には、従来、手作業では実現が難しかった作業などを含む。(例：Mermaid 記法によるソフトウェア構造の可視化など)
- 生成 AI を使いシステム開発の事例を示すことで、産業界で役立つ実践的な内容を扱うことができると思ったこと。

以上の考えを基に、ソフトウェア工学では、従来の基礎知識の教育に加え、生成 AI 活用により、概念理解の促進、ソフトウェア開発における生成 AI の活用法教育、システム設計方法の教育、などによる上位認知レベルの教育を意識し、表 6 のような講義を追加した。

表6 ソフトウェア工学で扱った生成 AI を活用した学習

学習内容	主な対応レベル	理由
Mermaid 記法	②理解～③応用	・構文や図式表現を理解し、システム構造を図示する。 → 例：「クラス図やシーケンス図を正確に描ける」。
オブジェクト指向設計の原則 (SOLID 原則)	④分析～⑤評価	・責務や依存関係を分析し、設計の妥当性を評価する。 → 例：「単一責任原則を守るようにコードを改善できる」。
デザイン・パターン	③応用～⑥創造	・既存の設計パターンを適用し、必要に応じて新しい構造を創出。→ 例：「適切なパターンを選び、複数組み合わせで設計できる」。
ドメイン駆動開発 (DDD)	⑤評価～⑥創造	・ドメインモデルを抽象化し、概念を統合してシステムを構築。→ 例：「ユビキタス言語を用いて、ドメインモデルを自ら設計できる」。
ソフトウェア開発における生成 AI の活用事例	②理解～③応用	・コード自動生成・変換 ・デバッグとバグ修正 ・学習と改善のサポート ・ドキュメントの生成とメンテナンス ・テスト自動化 ・デザイン・パターンの適用支援 ・要件定義と仕様生成の支援 ・リファクタリング ・開発プロセス全体の最適化、など

本講義は 2025 年度の春学期に 153 人の学生が受講した。それらの学生が授業後のアンケートで回答した主要コメントを紹介する。

- ソフトウェア工学は、現役のシステムエンジニアにとっても有意義であり、重要な分野だと思います。しかし、なかなか勉強する機会がないため、生成 AI と組み合わせて普及していただけるとエンジニアの能力の底上げにつながると思いました。また実践編や AI 活用編の講義を作っていただけると嬉しいです。生成 AI の普及に伴って、レビュー力としてもソフトウェア工学の基本を理解していることが必須になると思います。
- ソフトウェア工学は、上の立場に立つ方には必須の知識だと思います。私に足りないものだったのでとても興味深く学習できました。

学生コメントと認知レベルの対応は以下のように考えられる。

(a) 教育システムへの建設的提案と業界分析 (分析・創造)

学生コメント：「ソフトウェア工学は、現役のシステムエンジニアにとっても有意義であり…生成 AI と組み合わせて普及していただけるとエンジニアの能力の底上げにつながると思いました。また実践編や AI 活用編の講義を作っていただけると嬉しいです。」

・レベル④ 分析 (Analyzing) + ⑥ 創造 (Creating)：「エンジニアの能力の底上げにつながる」という発言は、現状のエンジニア教育の課題と、本講義のコンテンツ (生成 AI

×ソフトウェア工学)の構造的な関係を分析した結果と考えられる。さらに、「実践編や AI 活用編を作ってほしい」という要望は、単なる受動的な感想にとどまらず、分析結果に基づいて新しい教育プログラムの必要性を提言するものであり、広義の創造や、「問いを立てる力」につながる姿勢が見られる。

(b) AI と人間の役割分担に対する洞察 (評価・分析)

学生コメント:「生成 AI の普及に伴って、レビュー力としてもソフトウェア工学の基本を理解していることが必須になると思います。」

・レベル⑤ 評価 (Evaluating) + 批判的思考力: この学生は、AI がコードを書く時代だからこそ、人間にはそれを判定する「レビュー力」が必要であり、その判断基準として「ソフトウェア工学の基本 (知識)」が不可欠であると結論付けている。これは、AI の有用性と限界を比較検討し、人間の役割を「判断基準に基づく評価」と定義づける評価 (Evaluating) レベルの思考と考えられる。

(c) 知識の構造化とメタ認知 (分析・理解)

学生コメント:「ソフトウェア工学は、上の立場に立つ方には必須の知識だと思います。私に足りないものだったのでとても興味深く学習できました。」

・レベル④ 分析 (Analyzing) + メタ認知リフレクション: 「上の立場に立つ方には必須」という記述は、ソフトウェア開発における役割 (プログラマ、設計者、マネージャ等) と知識の対応関係を構造的に理解しており、分析 (Analyzing) レベルに該当する。また、「私に足りないものだったので」という記述は、表 3 にある「メタ認知リフレクション (自分の考え方や知識状態を振り返る)」が実践できており、自分の学習プロセスを客観視できていることを示している。

これらのコメントは、本稿の「3.2 ソフトウェア工学」で意図された「人間の付加価値は、ソフトウェアの要求仕様、アーキテクチャ設計など、より上流にシフトする」という教育目標が、学生に深く浸透していることを裏付けている。特に、AI を単なる「時短ツール」としてではなく、「レビュー対象」あるいは「能力底上げのパートナー」として認識している点は、学生が知識の受容 (レベル①-②) を超え、知識の価値を評価 (レベル⑤) する段階で思考できることを示唆している。

3.3. 2つの講義の比較

2つの講義を、教育が影響を与えた認知レベルや、教育における AI との関係性など観点から比較した表 7 に示す。

このように、比較の観点を定めることで、授業の位置づけを明確に理解することができる。一方で、本項で取り上げた 2 科目を対象とした教育実践事例は、以下のような課題があり、今後、対応が必要であると考えている。

表7 実践2科目の比較 (生成AI概論 vs ソフトウェア工学)

観点	生成AI概論	ソフトウェア工学
主な教育目標	批判的思考・倫理判断	設計力・レビュー力
生成AIの役割	出力の吟味・壁打ち相手	設計比較・コード評価
上位認知レベル	Analyze/Evaluate/Create	Analyze/Evaluate/Create
学習効果 (質的)	思考基準の明確化・内省	役割理解・設計妥当性の判断力
AIとの関係	思考の可視化	レビュー対象・共創相手

- 質的データ (学生コメント) が中心であり、定量的な学習成果を測定していない。
- ICT系科目を中心に検討しており、文系や技能系科目への適用可能性が検証されていない。
- 生成AI活用の効果はモデル・ツールに依存しており、継続的な技術進化に伴う再検証が必要である。

4. まとめと今後の課題

生成AIは、教育現場においてこれまでにない革新と課題を同時にもたらしている。AIが知識を自動生成できる時代に、大学教育の価値は「知識の伝達」から「思考と創造の支援」へと転換しつつある。本稿では、本学における複数の授業実践を通して、生成AIがどのように教育の質向上に貢献し得るかを述べた。

まず、生成AI概論においては、AI出力の信頼性や社会的リスクを批判的に吟味する活動を通じて、学生の「判断する力」と「情報の責任ある扱い方」を養うことができた。

次に、ソフトウェア工学の授業では、AIが生成したコードや設計案を比較・評価する学習を通して、学生が上位認知レベルである「分析」「評価」「創造」の段階へと発達した。

これら2事例を通じて得られた重要な示唆は、生成AIが単に知識の補助者ではなく、学習者の思考を可視化し、批判的・創造的学びを支援する「対話的存在」になり得るという点である。AIは学びのプロセスを鏡のように映し出し、学生が自分の理解の浅さや偏りに気づくきっかけを提供する。この「AIとの協働的省察」は、上位認知能力の発達に寄与するだけでなく、学びの主体性を高める教育的効果をもたらしている。

一方で、生成AI活用にはいくつかの課題も残されている。

1. LMS (学習管理システム) との連携の不十分さである。ブルームの認知分類③～⑥の教育では、学生が生成AIを活用して、独自の提出物を作成するレポートや演習が増える。これら提出物を、生成AIを用いて作成したルーブリックを用いて、自動評価できることが強く求められる。もちろん、教員による評価の確認は必須である。提出物をダウンロードするなどの手間がなく、LMSの中で自動評価できるシステムの

実現が望まれる。

2. LMS の中で AI が学修履歴やレポートデータを参照し、学生の理解度に応じた個別最適化やフィードバック自動化の実現が望まれる。
3. 個人情報漏洩などのリスクがない、学生が使える閉域生成 AI が利用できる環境も必要であろう。
4. 教育的誠実性 (Academic Integrity) の再定義が求められている。AI 生成物をどこまで自己成果として扱うか、引用や倫理の新しい基準づくりも必要である。
5. 教員側の AI リテラシー向上も不可欠であり、AI を拒絶するのではなく、教育目的に応じて適切に設計・制御できるスキルの育成が必要である。

今後の大学教育は、AI を単なる効率化の道具としてではなく、学びを深める共創のパートナーとして位置づける方向へ進むだろう。AI と人間の役割を峻別するのではなく、両者が補完し合う「協働的学習 (Collaborative Learning)」の設計が鍵となる。さらに、教員個人の工夫 (授業設計) だけでなく、それを支える LMS や制度設計 (教育システム) が連動して初めて、生成 AI 時代の教育の質保証が可能になると考えられる。

本稿で示した取り組みは、生成 AI 時代の大学教育における質保証の新しい試みを提示するものであり、今後は実践の継続と定量的評価を通じて、その教育的有効性をさらに検証していく予定である。本研究は JSPS 科研費 JP24K06314 の助成を受けて実施した。

参考文献

- 1) Junaid Qadir, Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education, 2023 IEEE Global Engineering Education Conference), May 2023
- 2) Miriam Sullivan, Andrew Kelly, Paul McLaughlan, ChatGPT in higher education: Considerations for academic integrity and student learning, Journal of Applied Learning & Teaching, Vol.6 No.1 (2023)
- 3) Frank Angelo Pacala, ChatGPT in education: Its value, accuracy, and regulation, HCMCOUJS-Social Sciences, 15 (2), 116-127, 2024
- 4) 同じ 5 行のコードが全く違って見える 12 の瞬間
<https://zenn.dev/coconala/articles/reasons-for-continuing-to-learn/>
(参照: 2025/11/29)
- 5) Anderson, L. W., & Krathwohl, A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, New York: Longman, 2001
- 6) J.S. Artal-Sevil, ARTIFICIAL INTELLIGENCE AND CRITICAL THINKING TRAINING IN HIGHER EDUCATION: SKYNET AND TERMINATOR ARE NOT HERE YET, Proceedings of 18th annual International Conference of Education, Research and Innovation, November 2025
- 7) M. Wagner, FROM VIBE-TEACHING TO FLOW-TEACHING: RE-GROUNDING AI-ENHANCED PEDAGOGY IN GOAL-DIRECTED ENGAGEMENT, Proceedings of 18th annual International Conference of Education, Research and Innovation, November 2025
- 8) 除村健俊, ソフトウェア工学の講義資料作成における生成 AI 活用、日本工学教育協会 第 73 回 年次大会・工学教育研究講演会 予稿集、2025/08/27-29

Operational Efficiency and Feasibility of Generative AI in English Writing Instruction: A Study on Teacher-led Assessment Model

Kensaku Ishimaki¹

Abstract

The increasing integration of Large Language Models (LLMs) offers potential approaches to addressing the persistent challenge of teacher workload in English writing instruction. This study explores a “Teacher-led AI Utilization Model,” operating as a semi-autonomous workflow wherein the instructor assumes the role of a “System Architect.” Utilizing a custom-built prototype system developed with Google Spreadsheet and Apps Script (GAS) alongside parallel prompt architectures (Claude-4-Sonnet and Gemini-Pro), the model was implemented in an advanced university English course (n=57). The research examines how structured system design and task re-engineering can align AI-generated outputs closely with pedagogical intent.

Operational observations indicated a degree of efficiency, enabling the finalization of comprehensive feedback for all submissions within three working days. However, comparative analysis highlighted a persistent need for human mediation. While the system achieved a 77.7% concordance rate at the individual evaluation item level, the rate of complete agreement per submission was limited to 10.5%. This discrepancy was primarily due to the AI’s tendency toward “excessive rigor,” necessitating instructor modifications for 89.5% of the raw scores.

These findings imply that while LLMs has potential to expand instructional capacity, the educator’s role remains indispensable for pedagogical calibration and the final scrutiny of feedback. Ultimately, within the context of the proposed model, this study presents an example of how the role of the human instructor can evolve and deepen into a more advanced pedagogical dimension.

Keywords: Generative AI, English writing instruction, Automated writing assessment, Teacher-led model, System Architect.

¹ サイバー大学 IT 総合学部・准教授

1. Introduction

1.1 Background and Rationale

Since the emergence of ChatGPT at the end of 2022, the rapid spread of Generative AI, including Large Language Models (LLMs), has brought about a major paradigm shift in the field of education. English writing instruction is one of the areas experiencing the most significant impact.

In the field of automated writing assessment, Automated Essay Scoring (AES) existed as a conventional system. However, this was limited to evaluations focused on surface-level features based on the analysis of linguistic characteristics. The advent of LLMs, however, has demonstrated the potential to realize more flexible feedback, such as understanding context, interpreting logic, and even inferring and reflecting the learner's intent for individualized optimization.

In traditional writing instruction, the limit of the teacher's workload has always been a bottleneck in aiming to achieve both the immediacy of correction/grading and the qualitative comprehensiveness of feedback in large classes. When aiming to provide instruction optimized for individual students with varying basic academic abilities, comprehension, and proficiency levels, it inevitably resulted in immense operational burden and "Brain Fatigue" for teachers.

Within our institution, initiatives have been undertaken to address the workload bottleneck in large-scale, mandatory foundational courses, including the introduction of a feedback system that outsources primary corrections and evaluations (Fujisawa et al., 2024). While this outsourcing model has effectively supplemented human resources for massive cohorts, elective advanced courses, characterized by limited enrolment, do not necessarily require such external outsourcing. Furthermore, assignments in advanced courses tend to involve relatively more sophisticated integrated tasks and evaluation criteria. Consequently, given these distinct educational settings, a different approach from the aforementioned system was implemented.

Specifically, the integration of Generative AI was deemed a viable means to facilitate complex evaluations in advanced courses while reflecting the instructor's pedagogical intentions more directly. However, when corrections are made using general-purpose AI tools such as ChatGPT or Google Gemini with abstract prompts, risks such as "Over-correction"—which may not match the learner's proficiency level—or evaluations that diverge from the instructor's intent have been pointed out. Consequently, there is a concern that the intended learning outcomes may not be achieved.

Therefore, to effectively bridge the gap between educational objectives and AI-generated outputs, the present study introduces a prototype of a “teacher-led AI system”. This experimental attempt aims to enable flexible operation tailored to the class size while mitigating the risks of generic AI use. In this approach, teachers take on the role of a mediator, applying their existing instructional insights to govern the behaviour of the technology. The ‘Teacher-led AI Utilisation Model’ is thus positioned as a viable methodology designed to ensure that AI serves as a high-fidelity extension of the teacher’s intent.

1.2 Research Objectives

This study aims to position Generative AI as a core component of a system designed to augment instructional expertise. This approach seeks to enable higher-order instruction by simultaneously enhancing the quality and quantity of pedagogical feedback. Furthermore, the paper investigates the operational feasibility and efficiency of this model, focusing on the following three objectives:

- **First: Literature Review and Theoretical Framework**

This study examines the transition from traditional Automated Essay Scoring (AES) to Generative AI-based evaluation, organizing the challenges inherent in this shift. This includes confirming the precision of AI in addressing both surface-level elements and deep context understanding, while establishing the validity of the “Hybrid Assessment Model” as a foundational best practice.

- **Second: Development of the System Architecture and the “Architect” Role**

The research explores the practicality of an original system that utilizes dedicated prompts on a custom platform. This section focuses on the methodology of encoding pedagogical rubrics into system logic, illustrating how the instructor functions as a “System Architect” to ensure the integrity of the evaluation process.

- **Third: Empirical Analysis of Implementation and Teacher Mediation**

Through a case study of an Advanced English course, the study evaluates the actual outcomes of the proposed workflow. By analysing the concordance between AI-generated assessments and instructor modifications, the analysis clarifies the necessity of human mediation and the model’s impact on both instructional quality and operational efficiency.

1.3 Structure of the Paper

Based on the above objectives, this paper unfolds with the following structure.

- Chapter 2 reviews the literature on English writing instruction, automated scoring, and Generative AI. It outlines the transition from traditional AES to LLMs and

examines prior research regarding evaluation reliability.

- Chapter 3 reports on the implementation within an elective Advanced English course, detailing the development of a custom-built prototype scoring and correction system. It explains the architecture of a “semi-autonomous workflow” utilizing both Claude-4-Sonnet and Gemini-Pro, and assesses the operational efficiency achieved through AI-driven batch processing. Furthermore, it examines the specific evaluative characteristics of the system and the necessity of teacher mediation, based on an analysis of instructor modifications to AI-generated feedback drafts.
- Chapter 4 addresses the limitations of the current study and future prospects for the proposed model. By synthesizing the empirical results, it discusses how the model presents an example of instructors building upon their existing expertise to fulfil a more advanced pedagogical role in an AI-integrated environment.

2. Literature Review: The Evolution of Automated Writing Assessment

2.1 The Shift from Automated Essay Scoring (AES) to Large Language Models (LLMs)

Prior to the emergence of Generative AI, exemplified by ChatGPT in 2022, the field of automated English writing assessment was characterised by Automated Essay Scoring (AES) systems, such as ETS’s “e-rater®.” AES systems scored essays by statistically analysing and extracting pre-defined “linguistic features,” such as lexical diversity, sentence length, and mechanical errors including spelling and grammar. While these attempts achieved a moderate level of accuracy, balancing cost reduction in large-scale exams with evaluation consistency, scholars have noted that they fell short of interpreting higher-order abilities such as the overall logical structure, rhetorical intent, or content creativity (Hannah et al., 2023).

The advent of Large Language Models (LLMs) in the 2020s has been a gamechanger for this field. Indeed, English writing correction and evaluation are arguably one of the domains most profoundly affected by Generative AI. A defining characteristic of LLMs is their capacity to “interpret” the text holistically and generate context-aware feedback. This has enabled “logical dialogue regarding the learner’s arguments” and “proposals for vocabulary and expressions aligned with context and intent”—tasks deemed difficult for traditional AES. Consequently, the focus of assessment is shifting from “formal correctness” to “communicative validity” (Hannah et al., 2023).

2.2 Evaluative Precision and Tendencies in Generative AI

Research indicates that the accuracy of Generative AI in correcting English text varies significantly by error type. First, regarding “surface-level errors” such as subject-verb agreement, number agreement, spelling, and basic word order, AI models demonstrate high identification accuracy, exceeding 90%. Conversely, accuracy declines notably for items requiring “deep context understanding,” such as unnatural sentence structures, logical cohesion via conjunctions, and interpreting the author’s intent (Alsaweed & Aljebreen, 2024). Comparative studies between GPT-4 and experienced teachers confirm that, at present, AI reliability and accuracy remain inferior to human performance (Pack et al., 2025).

Second, the appropriateness of “correction proposals” poses a challenge. Due to the inherent nature of their generation mechanisms, Generative AI models produce highly fluent text. Consequently, there is a marked tendency towards “over-correction,” wherein the learner’s original expressions are substantially altered, often deviating from the original intent (Fang et al., 2024). In writing pedagogy, the principle of “Minimal Correction”—respecting the learner’s voice as much as possible—is considered educationally desirable. However, studies note that unless constrained by precise prompts or few-shot examples, AI tends to rewrite text into overly fluent prose that overrides the learner’s intent (Loem et al., 2023).

Third, concerns persist regarding the quality of AI-generated “explanations and commentary.” Even where correction proposals are immediate and linguistically sound, they may not necessarily be pedagogically appropriate for the learner. Research highlights challenges regarding psychological acceptability, noting that explanations are frequently perceived as “ambiguous” or “lacking empathy” (Chen et al., 2024).

2.3 Assessment Reliability and the Rationale for the ‘Hybrid Model’

Concerns regarding the reliability of Generative AI assessment centre on the issue of “model and prompt dependency”. Given that evaluation outcomes are susceptible to fluctuations driven by frequent AI model updates and the specific phrasing of user prompts, relying solely on a single AI judgment requires cautious consideration. (Mizumoto et al., 2024; Guo & Wang, 2023).

Consequently, leading English testing organisations maintain that a “Hybrid Assessment Model”—combining primary AI evaluation with final judgment by skilled human experts—remains best practice. For instance, ETS has ensured reliability through concurrent use with humans since the early development of AES (Monaghan, 2005). Similarly, Cambridge English’s “Linguaskill” employs a mechanism wherein

human experts intervene when the reliability of the AI judgment is low (Cambridge, 2022). Although the specific mechanisms for human intervention differ, both represent hybrid assessment models that perform judgments utilising both AI and humans in separate stages.

In keeping with this framework, this study adopts a hybrid approach wherein AI undertakes surface-level checks and the initial drafting of feedback. This allows the educator to concentrate on higher-order instruction—such as moderating assessment severity and ensuring constructive feedback tone—and final evaluation judgment. This represents a pragmatic strategy to maximise educational validity while mitigating the inherent limitations of current AI assessment models (Automated Writing Evaluation Tools, 2023; Examining AI-Based Accuracy Assessment, 2024).

3. Practice of Generative AI Utilization in Advanced Proficiency Classes: Integration of System Design and Task Development

3.1 Overview of Practice: Semi-Autonomous Evaluation and Feedback Generation

3.1.1 Concept and Implementation of the ‘Semi-Autonomous Workflow’

This chapter reports on the implementation of Generative AI in an Advanced English class (1A) during the Spring 2025 semester. A defining feature of this practice is the establishment of a ‘semi-autonomous workflow’ in which AI is positioned as an assistant that undertakes the primary operational workload. Within this framework, the system generates the initial drafts of the feedback content, while the instructor finalizes the output through a process of review and editing.

In this advanced course, students are required to demonstrate the ability to execute complex ‘Integrated Tasks’. These tasks involve, for instance, composing an English email by applying structural conventions learned from instructional content to information synthesised from an English source video (e.g., tips on airport logistics and cost-saving). Attempting to provide comprehensive and detailed feedback (FB) on such assignments independently would impose a significant cognitive and physical burden (‘Brain Fatigue’) on the instructor, while inevitably resulting in prolonged turnaround times.

In this practice, the ‘semi-autonomous workflow’ refers to a system wherein the instructor functions as an ‘Architect’, designing a process that enables the AI to replicate the instructor’s pedagogical judgments as faithfully as possible. Consequently, the system drives the operational workflow—from generating evaluations and corrections to shaping feedback comments—under the architect’s design. By rigorously

encoding the instructor's assessment logic into the system, a systematic automated workflow was established. As a result, this system enabled the generation and finalisation of comprehensive feedback sheets for all 57 submissions within three actual working days—a feat unfeasible for a solitary instructor using traditional methods. This initiative demonstrates a novel approach in which the instructor's conventional expertise in correction and evaluation is extended and deepened into the realm of AI supervision and pedagogical calibration. In this context, “architecting” the system represents a higher-order extension of traditional teaching skills, allowing for a concentrated focus on complex evaluative decisions and the final refinement of AI-generated outputs.

3.1.2 Ethical Considerations

In this study, strictly anonymized data was used for analysis. The processing of student submissions via Generative AI tools (Claude and Gemini) was conducted in accordance with institutional data handling guidelines, ensuring that no personally identifiable information was exposed to model training data.

Regarding transparency in the instructional process, the utilization of Generative AI within the assessment workflow was not explicitly stated to the students. This approach was adopted to establish the AI strictly as a backend administrative tool to support the instructor. Therefore, it was deemed ethically imperative that the instructor assume full and unmitigated authorship responsibility for all feedback content delivered.

While Generative AI was integrated into the assessment workflow, its role was strictly limited to the generation of primary drafts for correction and evaluation. Every final feedback comment and score underwent a comprehensive review and modification process by the instructor. This “Human-in-the-Loop” framework ensures that all outputs represent the instructor's professional judgment and pedagogical intent. Furthermore, the feedback examples presented in the Appendix are synthesized samples created to protect student privacy.

3.2 Design and Development of an Original Scoring and Correction System

3.2.1 System Architecture

For this implementation, a proprietary scoring and correction system was constructed using Google Spreadsheet (GSS) and Google Apps Script (GAS) as the platform. The design attempted to synchronise the pedagogical intent of the task with the behaviour of the AI models. The operational workflow executed during the Spring

2025 semester is outlined below (see Figure 1).

1. Primary Processing: Automated Batch Execution via GSS and GAS

Following the conclusion of the submission period, all student answer sheets were retrieved from the school LMS. The response sections from these individual submissions were consolidated into a central GSS to facilitate the subsequent single batch operation. The system then executed independent prompts concurrently for each of the ten assessment criteria (e.g., Subject, Address, Greetings, Content Points) using Claude-4-Sonnet. This process generated evaluations, correction proposals, and explanatory commentary collectively for the entire cohort.

2. Secondary Processing and Final Confirmation: Teacher-Mediated Synthesis and Refinement

The discrete evaluation components generated during primary processing were manually transferred by the teacher into the chat interface of a 'Custom Gem' (a dedicated AI assistant) on Google Gemini for each individual submission. Using this interface, the teacher synthesised the raw data into cohesive feedback comments, refining the tone and content before finalising and recording them in the feedback sheet (see figure 1).

Secondary processing was initially intended to be purely manual. However, trials in a colleague's course revealed that most of the tasks were formal and repetitive, indicating they were well-suited for automation via Generative AI. After initially using general AI, a specialized 'Custom Gem' was developed for Advanced 1A to enhance

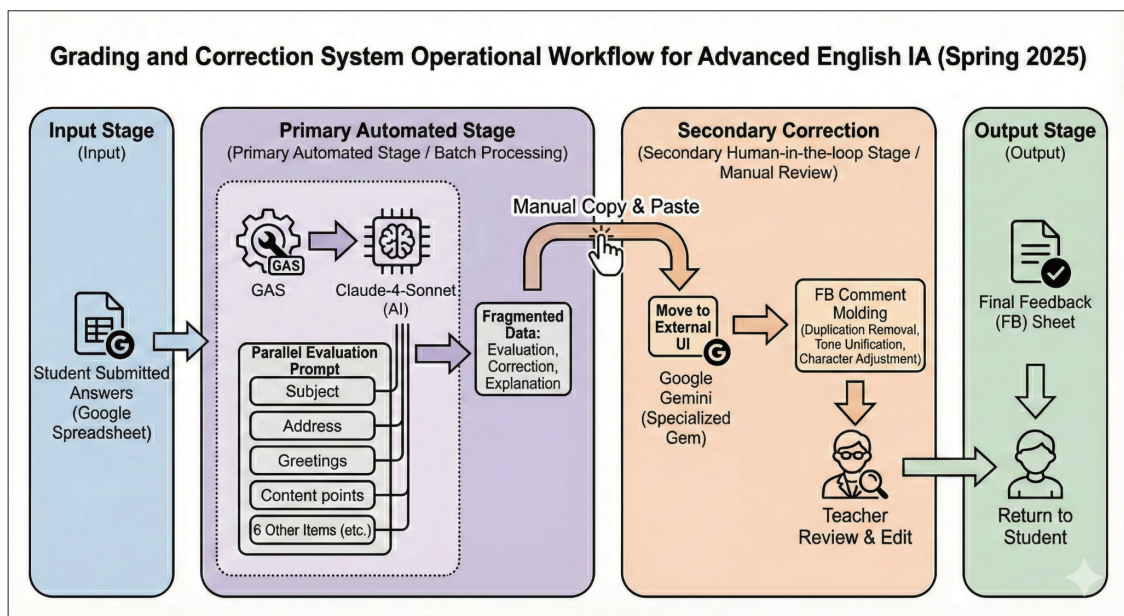


Figure 1: Operational Workflow of Scoring and Correction System in Advanced English IA (Spring 2025)

efficiency. These insights led to the decision to integrate the Gem-based processing into the batch workflow for the Autumn 2025 semester, leaving only the final review and editing as a manual stage for the instructor. While this refinement streamlines data handling, the Spring 2025 implementation serves as a proof-of-concept where manual review remains vital to pedagogical validity.

3.2.2 Parallel Prompt Architecture for 'Watertight' Evaluation

A central tenet of the system design was the assurance of logical robustness in evaluation, embodied by a 'Watertight' (leak-proof) design philosophy.

During the initial prototyping phase, attempts were made to process all evaluation criteria simultaneously via a single comprehensive prompt. However, this approach resulted in frequent 'leakage' due to 'attention dispersion', wherein the model failed to address multiple items consistently. To resolve this structural vulnerability, the implementation adopted a 'Parallel Evaluation Structure', in which the ten evaluation criteria are assessed independently. By executing dedicated prompts for each specific item, the architecture directs the LLM to concentrate its attention resources on discrete elements, thereby maximising both the comprehensiveness and accuracy of the evaluation (see Figure 2).

Specifically, evaluation criteria were deconstructed and explicitly articulated to eliminate interpretative ambiguity, aiming for a state where the AI could render consistent judgments. For instance, distinct judgment logic was embedded into the

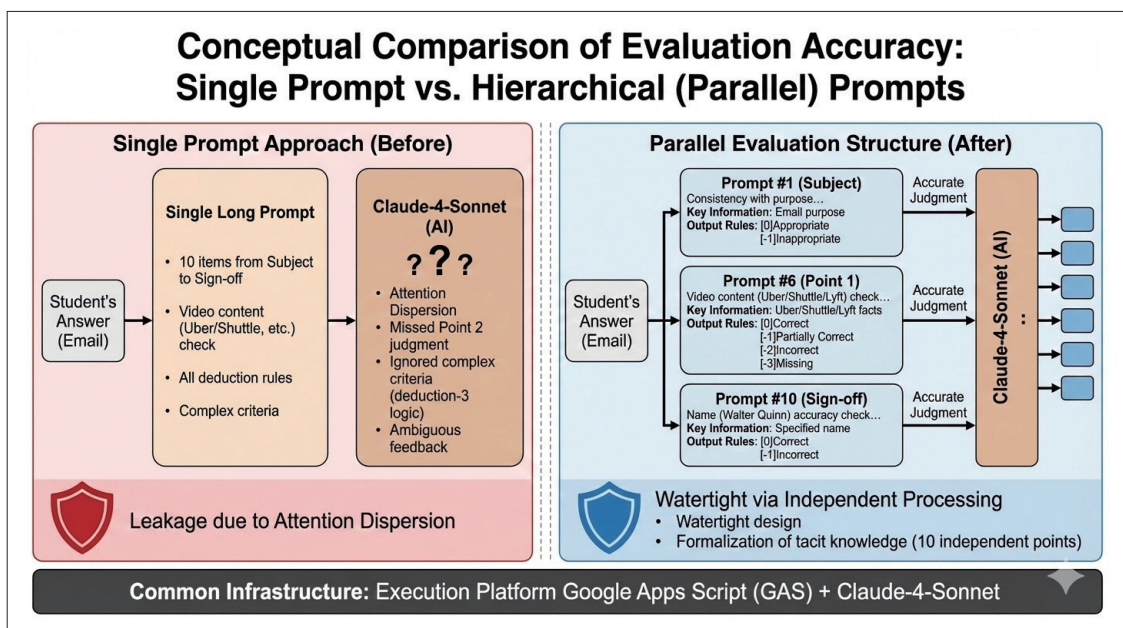


Figure 2: Conceptual Comparison of Evaluation Accuracy between Single Prompt and Hierarchical (Parallel) Prompts

prompts for each category, spanning from formal conventions such as ‘Subject’ and ‘Address’ to substantive ‘Content Points’ —specifically verifying the inclusion of accurate references to the video lesson. This configuration, akin to a ‘checklist method’, facilitated practical operation by significantly mitigating evaluation variability.

This process, wherein the instructor functions as a ‘System Architect’ to systematically encode pedagogical expertise into a suite of prompts, constitutes the core of the ‘Watertight’ design—a central theme of this study. This design philosophy aims to empower the AI to operate as a sophisticated evaluation engine, replicating the instructor’s pedagogical intent to the best extent possible.

3.3 Task Redesign for AI Compatibility

3.3.1 Task Design as ‘Requirements Definition for AI’

In this implementation, assignments were re-engineered as ‘requirements definitions for AI’. This approach ensured a logical structure that facilitated AI processing while accurately reflecting educational intent.

- **Stipulating Persona and Scenario:** For the writing task, a highly specific context was established to reduce ambiguity. The student assumes the role of ‘Walter Quinn, a new staff member at the travel agency Landhome’, replying to a customer, Paige O’Brien, on behalf of a colleague, Cabe Dodd.
- **Codification of Structural Constraints:** Answers were strictly limited to ‘100–160 words’, and specific formatting constraints, such as the ‘prohibition of bullet points’, were imposed to facilitate accurate automated scoring. Additionally, the instructions explicitly required the inclusion of specific references to the video content within the text.
- **Alignment with Prompt Architecture:** The study aimed to synchronise the pedagogical intent of the task with the behaviour of the AI model. This alignment was pursued by mirroring these task constraints within the scoring prompts, incorporating directives such as ‘Respect the intent of the original text (Minimal Corrections)’ and ‘Provide explanations suitable for CEFR B1–B2 level students’.

3.3.2 Articulation of Assessment Rubrics and Integration of Materials

To ensure that AI judgments accurately reflect the instructor’s pedagogical intent, efforts were made to explicitly articulate the underlying evaluation rubrics. Furthermore, the prompts, source materials, and instructional texts were configured to function as a cohesive ‘tripartite’ unit.

Specifically, ‘Key words/information’ were rigidly defined for each evaluation category (see Table 1). For instance, criteria verified the presence of formulaic

expressions in the introduction (e.g., “Thank you for contacting Landhome”) and the inclusion of specific solutions within the content (e.g., “Uber, Lyft, or airport shuttles”). The central objective was to transfigure evaluation criteria that teachers traditionally assess ‘intuitively’ into concrete, AI-reproducible indicators—effectively converting ‘tacit knowledge’ into ‘explicit knowledge’.

3.3.3 Establishing ‘AI Resistance’ through Integrated Task Design

To mitigate the risk of students bypassing learning objectives via general-purpose Generative AI, the structure of the ‘Integrated Task’ was fortified to necessitate the comprehension of external materials.

- **Mandatory Synthesis of External Sources:** Students were required to view specific video content on the university’s e-learning platform, CU ENGLISH, and integrate distinct information retrieved from it (e.g., alternatives to public transport and cost-saving tips) into their responses.
- **Requirement for Personalised Input:** Beyond synthesising video content, instructions mandated the inclusion of ‘original advice’ tailored to a specific scenario—specifically, a parent travelling with a five-year-old child.
- **Implementation of Verification Logic:** Within the scoring prompts (Points 1 and 2), logic was implemented to strictly cross-reference student responses against the source material. This ensured the inclusion of specific solutions mentioned in the video (e.g., Uber, airport shuttles) while penalising the presence of inaccurate or ‘hallucinated’ information.

Through this strategic design, the task incorporated a level of ‘AI resistance’ — rendering it impossible to achieve high scores via simple, generic prompting—while

Table 1: Excerpt of Independent Evaluation Criteria and Design Rationale

Evaluation Item (Category)	Required Elements / Judgment Logic	Assessment Objective
Subject (Form)	Logic: Keywords: Travel Tips, NYC, Family Trip	Intent: Verification of conciseness and relevance regarding the email’s purpose.
Point 1 (Content)	Logic: Required Lexis: Uber, Lyft, airport shuttles	Intent: Validation of accurate synthesis and comprehension of the source video.
Point 3 (Application)	Logic: Input Requirement: Original advice for traveling with a child	Intent: Assessment of original contribution beyond the source material.
Sign-off (Form)	Logic: Fixed Phrase: Best regards, Walter Quinn	Intent: Confirmation of adherence to the designated persona.

establishing an environment where the assessment AI could execute precise judgments as the instructor's effective 'alter ego'.

3.4 Analysis of Implementation Results

3.4.1 Comparative Analysis of AI-Generated Assessment and Instructor Evaluation

An analysis of the concordance between the AI's automated scoring and the instructor's final judgment revealed distinct tendencies in the system's evaluative performance, particularly when contrasting item-level accuracy with submission-level agreement.

First, when analysed at the "item-level," the system executed a total of 570 discrete evaluation instances (57 submissions \times 10 independent criteria). Among these, the AI's judgments completely aligned with the instructor's final evaluations in 443 instances, resulting in an overall item-level concordance rate of 77.7% (see Table 2). This level of concordance suggests that the "watertight," parallel prompt architecture enables the AI to execute specific, compartmentalized assessment tasks with a certain degree of precision.

However, a stark contrast emerges when the data is examined at the "submission-level" (or essay-level). The rate of complete agreement—where all 10 criteria within a single submission matched the instructor's judgment perfectly—was observed in only 10.5% of cases (6 out of 57 submissions). Conversely, in the remaining 89.5% (51 submissions), the raw score (out of a maximum of 12) necessitated at least one modification by the instructor. Furthermore, even regarding the converted grade scale (out of 5), modifications were required for 75.4% (43 submissions) of the scripts.

The observed difference between the item-level concordance (77.7%) and the submission-level complete agreement (10.5%) highlights a characteristic of the current AI assessment model: while the AI demonstrated a certain degree of accuracy in individual verification tasks, cumulative discrepancies tend to arise when synthesizing multiple criteria into a holistic evaluation. This reaffirms the necessity of the instructor's final scrutiny.

Regarding the nature of these modifications, a clear tendency towards "excessive rigour" was confirmed. The AI applied assessment criteria with significantly greater strictness than the instructor, consistently calculating lower scores. Consequently, the overwhelming majority of instructor interventions involved upward adjustments to the grades. Where the AI mechanically applied deductions based on rigid criteria, the instructor intervened to interpret context and learner intent, thereby recalibrating the score.

Table 2: Concordance Rates between AI Assessment and Instructor Judgment by Criterion (Advanced IA, N=57)

Evaluation Criterion	Concordance Rate (Unmodified)	Upward	Downward	Operational Observations
1. Subject	98.25%	1	0	Binary verification of format yielded near-perfect precision.
2. Address	71.93%	15	1	Ambiguity in name entity recognition. Resolvable via prompt calibration.
3. Greetings	85.96%	7	1	—
4. Self Introduction	92.98%	4	0	Verification of mandatory information presence demonstrated high consistency.
5. Purpose	91.23%	5	0	—
6. Point 1 (Video Content 1)	56.14%	23	2	Increased divergence observed due to complexity of multi-level scoring (4-point scale).
7. Point 2 (Video Content 2)	63.16%	17	4	(Same as above)
8. Point 3 (Own Advice)	84.21%	5	4	—
9. Closing	77.19%	13	0	—
10. Sign-off & Name	56.14%	25	0	Discrepancies regarding persona adherence; necessitates alignment of prompt constraints with source material.
Total (N=570)	77.72% (442 items)	115	12	Overall performance confirms high individual item accuracy, contrasting with the lower complete agreement rate per essay.

However, a granular analysis reveals that this need for modification was not uniform across all criteria. While complex, multi-stage evaluation items often triggered the AI's excessive rigour, binary assessment items requiring simple verification—specifically 'Self Introduction' (Item 4) and 'Purpose' (Item 5)—demonstrated exceptionally high concordance rates.

3.4.2 Operational Efficiency and Mitigation of Instructor Burden

While a quantitative validation of the linguistic accuracy of AI-generated feedback falls outside the scope of this study, the operational reality—specifically the substantial reduction in working hours—warrants attention as a noteworthy outcome. Despite the necessity for score adjustments noted in Section 3.4.1, the deployment of this semi-autonomous workflow significantly mitigated the instructor's administrative and cognitive load, enabling the rapid completion of the feedback generation process—a

feat difficult to achieve through traditional solitary correction.

In terms of actual time allocation, the entire process was completed within approximately three working days. This duration comprised a Batch Processing Phase (approx. 4 hours) for the retrieval and pre-processing of 80 submissions across three courses, followed by an Individual Processing Phase (2 days) dedicated to the review and refinement of the 57 'Advanced English IA' submissions.

The primary driver of this efficiency was the AI's capacity to instantaneously generate detailed draft comments, which relieved the instructor from the burden of composing text *de novo*. This allowed for a redirection of attentional resources towards higher-order pedagogical tasks—specifically, moderating the AI's scoring severity and calibrating the tone of guidance to align with the course's target proficiency and the specific quality of the individual submission. The fact that the instructor could ratify the majority of these explanatory drafts with only minor refinements suggests the potential of the feedback to offer practical utility sufficient for higher education contexts.

3.4.3 Impact on Learner Perception and Perceived Utility

Although quantitative measurement of learning outcomes is outside the scope of this study, the course evaluation survey elicited favourable responses regarding the depth and precision of the feedback provided through this system. Although the number of specific qualitative comments was limited, students explicitly noted the quality of the guidance, with remarks such as “The feedback on the assignment was corrected very carefully” and “Mistakes were pointed out in detail.”

For a representative example of the actual feedback architecture delivered to students, please refer to the Appendix (Note: To ensure privacy preservation, the data presented constitutes a synthesised example based on actual submissions and feedback content). It is inferred that the feedback of a volume and granularity previously unattainable for a solitary instructor directly correlated with the observed learner satisfaction.

3.5 Summary of Findings from the Advanced Course Implementation

The implementation within the Advanced English course demonstrated the operational feasibility of a semi-autonomous workflow. The key findings from the practice reported in sections 3.1 through 3.4 are summarized as follows:

- **Instructional Capacity and Efficiency:** By delegating the generation of primary feedback drafts to AI, the instructor finalized comprehensive feedback sheets for 57 submissions within three actual working days. This efficiency enabled the

delivery of detailed guidance at a volume and speed that would be difficult to sustain through traditional manual correction.

- **System Design and Task Integration:** The efficacy of the workflow was found to rely on the precision of prompt engineering to encode pedagogical expertise into “watertight” parallel prompt architectures. This structural measure suggested potential utility for compartmentalized assessment, with the system indicating a 77.7% concordance rate at the individual item level, demonstrating a certain degree of accuracy in executing discrete verification tasks.
- **The Necessity of Instructor Mediation:** While the system demonstrated the potential to significantly reduce instructor burden and shorten feedback creation time, the contrast between the item-level concordance and the 89.5% raw score modification rate per submission reaffirmed the vital importance of the human rater. This result indicates that cumulative discrepancies tend to arise when synthesizing multiple criteria into a holistic evaluation, underscoring the instructor’s essential role in final scrutiny and the pedagogical calibration of the feedback tone.

In summary, the practice indicates that the instructor’s role involves transmuting tacit pedagogical knowledge into explicit system logic. The system demonstrates the potential for drastically amplifying the instructional capacity, while it was also reaffirmed that the role of the human rater remains vital to this model.

4. Limitations and Future Prospects

4.1 Limitations of the Study

While the proposed model indicated a degree of operational efficiency, several limitations must be acknowledged:

- **Scope of Implementation:** The practical verification was limited to a specific course and a single instructor, which may constrain the generalisability of the findings to different instructional styles or settings.
- **Technical Dependency:** The evaluative performance remains highly dependent on the specific characteristics of the LLMs used (Claude-4-Sonnet and Gemini-Pro) and the precision of the current prompt engineering. Future model updates will likely require the recalibration of “watertight” architectures and prompts.
- **Absence of Long-term Learning Outcomes:** This study focused primarily on operational efficiency and assessment consistency. Further research is required to measure the long-term impact of AI-generated feedback on learner motivation and the actual improvement of writing proficiency.

4.2 Future Prospects and Concluding Remarks

The findings of this research suggest the possibilities of the following trajectories for English writing instruction in the AI era:

- **Expansion of Instructional Capacity:** While teacher workload has traditionally been a bottleneck in providing intensive writing instruction to large cohorts, the model presented in this study suggests that AI integration may have the potential to overcome these constraints. The model is expected to lead to enabling a single instructor to deliver immediate and individualized feedback at a volume that previously would have required a significantly greater investment of educational resources.
- **Course Design Premised on AI Feedback:** As an extension of this study, it is conceivable that individual instructors will be able to design their courses based on the premise that AI-generated feedback is available. For instance, within such a framework, an instructor could implement more frequent, iterative writing cycles, as learners would be supported by immediate and detailed guidance that encourages continuous refinement.
- **Elevation of the Instructor's Role:** By delegating a substantial portion of routine, repetitive, and administrative workloads to AI, instructors are able to concentrate their resources on higher-order responsibilities. These include sophisticated task design and high-level evaluative judgments that require an understanding of nuanced contexts. In this context, the role of the human educator is poised to evolve and deepen into a more advanced pedagogical dimension.
- **Challenges in Sharing Expertise Among Instructors:** However, if this system is to be used collaboratively by multiple instructors, a potential challenge arises from the isolation of specialized system design and operational know-how. Additionally, the possibility of discrepancies in the application and interpretation of assessment criteria, which is inevitable with multiple human raters, still persists. Therefore, the accumulation of shared knowledge among colleagues becomes essential.

The proliferation of Generative AI provides a significant opportunity to reflect on the indispensable role of the human instructor. The author concludes this paper with the hope that the “Teacher-led AI Integration Model” proposed herein will serve as a modest contribution to a new perspective on English education in an era of human-AI symbiosis.

Acknowledgements

The author extends sincere appreciation to Lecturer Jared Baierschmidt for his significant contributions to the development and practical operation of the system. Gratitude is also owed to Lecturer Fumi Wakui for her valuable advice during the writing of this paper. Finally, this study is presented as one of the academic outcomes based on the organizational management of courses within our institution. Any remaining errors or shortcomings are the sole responsibility of the author.

References

- Aldosemani, T., Assalahi, H., Lhothali, A., & Albsisi, M. (2023). Automated writing evaluation in EFL contexts: A review of effectiveness, impact, and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, 13(1), 1-19.
<https://doi.org/10.4018/IJCALLT.329962>
- Alsaweed, W., & Aljebreen, S. (2024). Investigating the accuracy of ChatGPT as a writing error correction tool. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1-18.
<https://doi.org/10.4018/IJCALLT.364847>
- Leading, L. L., Monaghan, W., & Bridgeman, B. (2005). E-rater as a Quality Control on Human Scores. *Connections*.
https://www.kr.ets.org/Media/Research/pdf/RD_Connections2.pdf
(accessed 8 January 2026)
- Cambridge, C. E. M., & Dictionary, C. Technology-Enhanced Language Assessment.
<https://www.cambridgeassessment.org.uk/insights/technology-enhanced-language-assessment-innovative-approaches-for-better-learning/>
(accessed 8 January 2026)
- Chen, Q. (2024). Students' perceptions of AI-powered feedback in English writing: Benefits and challenges in higher education. *International Journal of Changes in Education*. Advance online publication.
<https://doi.org/10.47852/bonviewIJCE52025580>
- Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435-8463.
<https://doi.org/10.1007/s10639-023-12146-0>
- Fang, T., Yang, S., Lan, Y., Wong, D. F., Hu, J., Chao, L. S., & Zhang, Y. (2023). Is ChatGPT a highly fluent grammatical error correction system? A comprehensive evaluation. arXiv.
<https://doi.org/10.48550/arXiv.2304.01746>
- Fujisawa, K., Shirasu, Y., Sato, K., Kawachi, K., & Inoue, M. (2024). e-learning による英文ライティング指導のための結果フィードバックシステム導入 [Implementation of a result feedback system for English writing instruction via e-learning]. *サイバー大学研究紀要 [Cyber University Research Bulletin]*, 1, 9-16.
https://www.cyber-u.ac.jp/about/pdf/research_bulletin/001/CU_RB001_02.pdf
- Hannah, L., Jang, E. E., Shah, M., & Gupta, V. (2023). Validity arguments for automated essay scoring of young students' writing traits. *Language Assessment Quarterly*, 20(4-5), 399-420.
<https://doi.org/10.1080/15434303.2023.2288253>
- Lee, O. (2024). Examining AI-based accuracy assessment in L2 learners' writing. *Journal of Pan-Pacific Association of Applied Linguistics*, 28(2), 39-55.
<https://eric.ed.gov/?q=source%3A%22Journal+of+Pan-Pacific+Association+of+Applied+Linguistics%22&id=EJ1457668>
- Loem, M., Kaneko, M., Takase, S., & Okazaki, N. (2023). Exploring effectiveness of GPT-3 in

grammatical error correction: A study on performance and controllability in prompt-based methods. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023) (pp. 205-219). Association for Computational Linguistics.
<https://doi.org/10.48550/arXiv.2305.18156>

Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3(2), 100116.

<https://doi.org/10.1016/j.rmal.2024.100116>

Pack, A., Hartshorn, K. J., Escalante, J., & Gillette, N. (2025). How well can GenAI (GPT-4) provide written corrective feedback on English-language learners' writing? *International Journal of English for Academic Purposes: Research and Practice*, 5(1).

<https://doi.org/10.3828/ijeap.2025.2>

Appendix

- Note: The following feedback sheet is a synthesized sample for illustrative purposes and does not contain actual student data.

2025SS Advanced English 1-A Writing Task		
FB_code	25SS_AE1A_Sample	Raw Score
Student ID	Sample	成績評価点数
		9 3
<p>全体コメント 課題の基準を概ね満たしています。減点となったところをよく見直して復習しましょう。</p>		
<p>Your Answer</p> <p>Subject About travel inquiry</p> <p>Dear Ms. Paige,</p> <p>I hope this message find you well. My name is Walter Quinn, and I work as support desk at Landhome. I am writing to you behalf on my colleague Cabe Dodd who received your call yesterday. Regarding your family trip to New York City, I have some advices for you.</p> <p>First, if you want to avoid public transport, I recommend you should choice Uber or a taxi to the airport. To save money during your travel, you should bring a refillable water bottle. Also, because you are traveling with a child, I recommend to bring some toys or books to keep him seated.</p> <p>I hope these suggestions are helpful. Please feel free to contact us if you have any questions.</p> <p>Best regards,</p> <p>Walter Quinn</p>		
<p>添削後の解答例(修正部分は「J」で表示されます)</p> <p>Dear Ms. Paige,</p> <p>I hope this message [finds] you well. My name is Walter Quinn, and I work as support desk at Landhome. I am writing to you [on behalf of] my colleague Cabe Dodd who received your call yesterday. Regarding your family trip to New York City, I have some [advice] for you.</p> <p>First, if you want to avoid public transport, I recommend you [choose] Uber or a taxi to the airport. To save money during your travel, you should bring a refillable water bottle. Also, because you are traveling with a child, I recommend [bringing] some toys or books to keep him seated.</p> <p>I hope these suggestions are helpful. Please feel free to contact us if you have any questions.</p> <p>Best regards,</p> <p>Walter Quinn</p>		
<p>解説</p> <p>#1 原文: I hope this message find you well. 添削後: I hope this message finds you well. 解説: 主語が「this message」(三人称単数)なので、動詞に三単現のsを付けると、文法的に正しくなります。</p> <p>#2 原文: I am writing to you behalf on my colleague Cabe Dodd who received your call yesterday. 添削後: I am writing to you on behalf of my colleague Cabe Dodd who received your call yesterday. 解説: 「〜を代表して」という熟語は「on behalf of」です。正しい前置詞と語順をセットで覚えること、よりスムーズに書けるようになります。</p> <p>#3 原文: I have some advices for you. 添削後: I have some advice for you. 解説: adviceは不可算名詞(数えられない名詞)なので、常に単数形で使うように意識しましょう。</p> <p>#4 原文: I recommend you should choice Uber or a taxi to the airport. 添削後: I recommend you choose Uber or a taxi to the airport. 解説: recommendの後には「主語 + 動詞の原形」とするのが一般的です。名詞のchoiceではなく、動詞のchooseを使うと正確に伝わります。</p> <p>#5 原文: I recommend to bring some toys or books to keep him seated. 添削後: I recommend bringing some toys or books to keep him seated. 解説: recommendの直後に動詞を置く場合は、動名詞(-ing)の形にしましょう。</p> <p>評価の解説 [-1] 解説: 宛名は、相手の姓を用いた「Dear Ms. O'Brien,」としましょう。ビジネスメールでは姓(last name)を使用するのが標準的なマナーです。 [-1] 解説: 課題の要件に正確に沿うため、本文では指定された選択肢(Uber、Lyft、またはairport shuttle)のみを用いるようにしましょう。 [-1] 解説: 節約方法について、空港での飲食費の高さや荷物の預け入れ手数料を避ける点も具体的に説明しましょう。必要な情報を網羅することで、より説得力のある回答になります。</p>		

英語ライティング指導における 生成 AI の運用効率と実現可能性： 教員主導型評価モデルに関する研究

石 卷 賢 作

概 要

大規模言語モデル (LLM) の普及は英語ライティング指導に多大な影響を与えつつあり、添削評価のための教員の業務負担という長年のボトルネックが解消される可能性を示している。本研究では、教員が「システム・アーキテクト (設計者)」として機能する「教員主導型 AI 活用モデル」を提案し、これに基づく「半自動型ワークフロー」の運用可能性を検証した。Google スプレッドシートと Apps Script (GAS)、および並列プロンプト・アーキテクチャ (Claude-4-Sonnet および Gemini-Pro) を用いた独自のシステムを構築し、大学の上級英語コース (n=57) にて実践を行った。本稿では、特にシステム設計と課題の再構築を統合的に行ったことで、AI の出力をいかに教育的意図に合致させうるかを検証している。

実証結果として、全答案に対する包括的なフィードバックを実働3日以内に完了するなど、一定の運用効率を示された。一方で、AI による自動評価と教員の最終判断を比較した結果、人間による介入の継続的な必要性も明らかになった。評価項目単位での評価一致率は 77.7% であったのに対し、答案単位での完全一致率は 10.5% にとどまった。この差異は主に AI の「過剰な厳密さ (excessive rigor)」に起因しており、結果として全体の 89.5% の答案において教員による素点の修正が発生した。

これらの知見は、LLM が教員の指導キャパシティを拡張しうる一方で、教育的な配慮による調整やフィードバックの最終精査において、人間の教員の役割が依然として不可欠であることを示している。結論として本稿は、提案する AI 活用モデルの環境下において、人間の教員の役割がより高度な教育的次元へと発展・深化する具体的な一例を提示するものである。

キーワード：生成 AI、英語ライティング指導、自動添削評価、教員主導型モデル、システム・アーキテクト

サイバー大学における 実践的データサイエンス教育の統合設計

— イベント・自由参加コース・ハイブリッド型
ハッカソン／アイデアソンの役割と効果 —

藤澤 弘美子¹、浅田 麻菜²、安間 文彦³

1. はじめに

データサイエンス教育においては、理論や分析手法の習得にとどまらず、実社会の課題と結びついたデータを扱う学習経験が重要である。特に、実世界で収集された実データ (authentic data) を用いることは、学習者の学習動機を高めるのみならず、データの背景にある文脈や制約を理解し、分析結果を解釈する力、さらには探究的・批判的思考力の育成に資するものとなる¹⁾。

一方で、教育機関において、こうした条件を満たす実データを継続的かつ安定的に教育に取り込むことは容易ではない。教育用に加工されたデータは扱いやすい反面、実社会における不完全性や多義性が捨象されがちであり、学習者が自らの課題意識を持ち、この解決に向けた計画をたてる経験が限定される可能性がある。

今回、サイバー大学 (本学) では宇宙航空研究開発機構 (JAXA) と協力し、JAXA が公開する地球観測衛星データを用いたデータサイエンス教育の取り組みを設計した。この JAXA 公開の地球観測衛星データは特定の教育目的のために設計されたものではなく、研究・実務・社会的課題解決など広く多様な用途に供される汎用的な実データである。また、近年では Earth Dashboard²⁾、Earth API³⁾、G-Portal⁴⁾ 等のツール整備により、専門的な設備や契約を伴わずともデータにアクセスできる環境が整いつつあり、データサイエンス教育において authentic data を扱うための条件を備えていると考えられる。

データサイエンス教育における実践的学習方法としては、課題解決型学習 (Project-Based Learning : PBL) の重要性も指摘されている。数理・データサイエンス・AI 教育強化拠点校コンソーシアム⁵⁾ では、PBL を「実社会の課題や実データを用いて学習者が主体的に学ぶ学習形態」と整理し、数理・データサイエンス・AI 教育における有効な教

¹ サイバー大学 IT 総合学部・准教授

² サイバー大学 IT 総合学部・講師

³ サイバー大学 IT 総合学部・教授

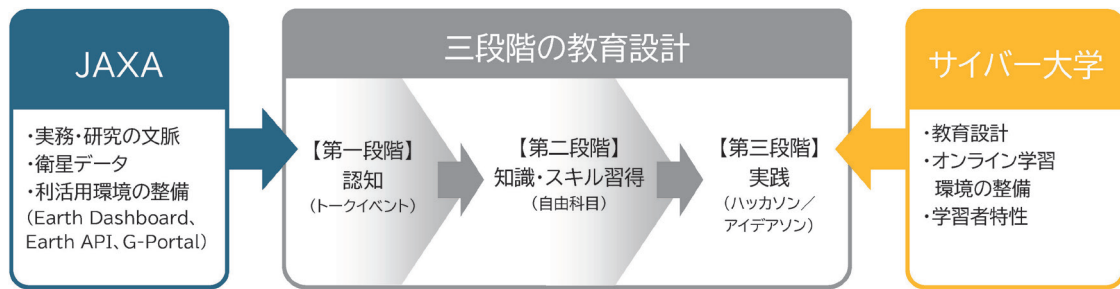


図1 産学連携による三段階の教育設計

育方法としてその活用を推奨している。しかしその一方で、PBLを機能させるためには、学習者間の協働や実務に近い課題設定、一定期間にわたる活動が必要であり、その設計と実施は容易ではない。数理・データサイエンス・AI教育強化拠点校コンソーシアムにおいても、企業等との連絡調整やデータの秘密保持などに膨大な時間と注意を要する事や、指導できる教員の確保などの困難さから、多くの大学がPBLの導入を希望しているにもかかわらず、そのうち実現できている割合は半数に満たないとの報告がある。

本学は、情報技術を基盤とした完全オンライン大学として、社会人学生や卒業生を含む多様な学習者が学び続ける環境を特徴としている。本学のディプロマ・ポリシーにおいても、課題解決力や他者との協働を通じた実践的能力の育成が重視されており、実社会のデータを題材とした学習設計は重要な位置づけを占める。

今回、地球観測衛星データの提供主体であるJAXAの専門家と、本学においてデータサイエンスのオンライン教育を担う教員が協力し、地球観測衛星データに関するトークイベント、使用法を解説する自由参加コース、そして活用を実践するハッカソン/アイデアソンの開催を通じて、教育と実社会を接続する学習機会の構築を目指した。そして本稿では、JAXAとの連携によるこの一連の取り組みを、データサイエンス教育における教育設計の観点から整理し、「認知」「知識・スキル習得」「実践」という三段階で構成した学習設計の意図と妥当性について論じる(図1)。

2. 教育プログラムの設計理念

データサイエンス教育において学習者が実データを活用し、自ら課題を設定して分析に取り組むためには、段階的な学習設計が不可欠である。本教育プログラムの設計にあたってJAXA側は、地球観測衛星データがどのような目的やプロセスで取得・提供されているかという実務的・研究的文脈を提供し、本学教員は、完全オンライン環境における学習者の特性や段階的な学習設計の観点から教育プログラムの構成と実際の展開を担った。この体制により、実世界データ(authentic data)に加え、データを扱う専門家の知見(authentic expertise)を教育設計に組み込むことを意図している。

なお、本学では正課科目として、「データサイエンス入門」「データサイエンス応用」等を通じて、統計学、データ分析の考え方、機械学習の基礎など、データサイエンスに関す

体系的な教育を提供している。一方で、本稿で扱う連携プログラムは、これら正課科目自体の再教授を目的とするのではなく、実データを用いた実践的学習へと展開することを目的とした、正課外の教育的取り組みとして位置づけられる。

本稿で扱う教育プログラムは、学習者の学習プロセスを「認知」「知識・スキル習得」「実践」の三段階で構成することを基本理念として設計されている。これは、データサイエンス教育における PBL の一般的困難性を前提としつつ、完全オンライン環境において顕在化しやすい課題に対応するための設計である。

このように学習過程を段階的に捉える理論として、Fitts と Posner⁶⁾ は、技能習得を「認知段階」「連合段階」「自動化段階」の三段階で整理している。「認知段階」では学習者は課題や操作の全体像を理解し、言語的・概念的な説明に依存して技能を獲得する。次の「連合段階」では試行錯誤とフィードバックを通じて知識や技能が統合され、操作の正確性や安定性が高まっていく。そして最後の「自動化段階」においては、基本的な操作が安定し、学習者は技能そのものに過度な注意を払うことなく、より高次の判断や他者との協働、課題全体の構造理解に認知資源を割くことが可能となる。

本稿で提案する「認知」「知識・スキル習得」「実践」の三段階からなる学習設計は、この技能習得モデルの考え方を教育設計の観点から実践的に応用したものである。「連合段階」に相当する過程を実データやツールを用いて知識・スキルを習得するフェーズとして位置づけ、基本的な操作や分析手法を前提とした上で、学習者が自ら課題を設定し、他者と協働しながら取り組む PBL 的ハッカソン／アイデアソンを、自動化段階に対応する「実践」フェーズとして設計している。ここでは、技能の自動化を単なる反復熟達ではなく、実社会的課題に向き合うための準備段階として捉え、段階的な学習プロセスの一部として構成している。各段階の内容は、具体的に以下の通りである。

第 1 段階の「認知」は、学習者が衛星データやその社会的活用可能性に触れ、関心や問題意識を形成する段階である。本プログラムにおいては、JAXA の研究者・実務担当者が登壇するトークイベントをこの段階に位置づけ、Earth Dashboard や Earth API、G-Portal といった具体的なツールを題材に、データの取得背景や特性、そして活用事例を解説した。講演に加えて質疑応答の時間を設けることで、学習者がデータの背後にある文脈や制約を理解する機会を提供している。

第 2 段階の「知識・スキル習得」は、認知段階で喚起された関心を、具体的な知識や技能の獲得へとつなげる段階である。本学における自由参加コースとは、教員裁量により設置する単位を伴わない正課外科目であり、特定の題材やツールを通じて、学習者が主体的に学ぶ機会を提供することを目的としている。本プログラムでは、この自由参加コースを、Earth Dashboard を用いた可視化から始め、QGIS によるデータ操作、Python を用いたデータ分析、さらに API を通じたデータ取得へと段階的に学習内容を構成することで、認知段階と実践段階を接続する「知識・スキル習得」フェーズとして位置づけている。

第 3 段階の「実践」は、学習者が自ら課題を設定し、他者と協働しながらデータ分析に

取り組む段階である。本プログラムでは、ハッカソン／アイデアソンを、地球観測衛星データを活用した課題解決をテーマに、課題定義、データ選択、分析、成果発表までを一連のプロセスとして経験する場として構想している。本学は完全オンライン教育を行う大学であるが、本プログラムは正課外の科目であるため、オンラインと対面を組み合わせたハイブリッド形式での展開とした。これにより、在学生、卒業生、社会人学習者など、多様な背景を持つ学習者が協働する実務に近い環境を実現することを意図している。

以上のように、本稿で扱う JAXA との連携プログラムは、正課外におけるトークイベント、自由参加コース、ハッカソン／アイデアソンを段階的に接続する教育設計として構想されたものである。イベントと正課外教育を補完的に組み合わせることで、完全オンライン環境においても、データサイエンス教育における PBL 的学習体験を実現することを目指している。

3. 予備的調査による学習者の現状分析

本章では、第1段階である「認知」フェーズに位置づけられたトークイベント終了後に実施したアンケート調査の結果をもとに、学習者の現状を整理する。本調査は、本稿で提示する三段階型教育設計の妥当性を検討するための予備的調査として位置づけられる。

アンケートは、2025年12月に実施された JAXA との連携トークイベント視聴者を対象にオンラインで実施し、有効回答数は82件であった。設問は、イベント内容に対する理解や関心、衛星データや関連ツールに対する印象、今後の学習意欲等に関する選択式設問および自由記述設問から構成した。

調査結果からは、衛星データやその社会的活用可能性に対する関心や理解が喚起され、イベントが「認知」段階として一定の役割を果たしていることが示唆された(図2)。

一方で、課題解決に向けた具体的なイメージや実際のアイデア創出については、肯定的回答が限定的であった(図3)。自由記述においても、データの可能性への関心が示される

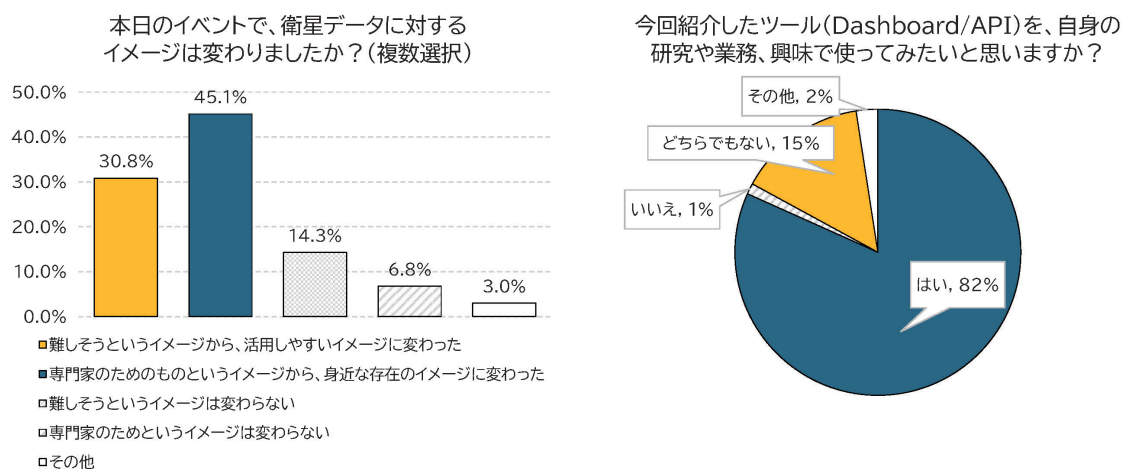


図2 トークイベント後の学習者の認知の変化

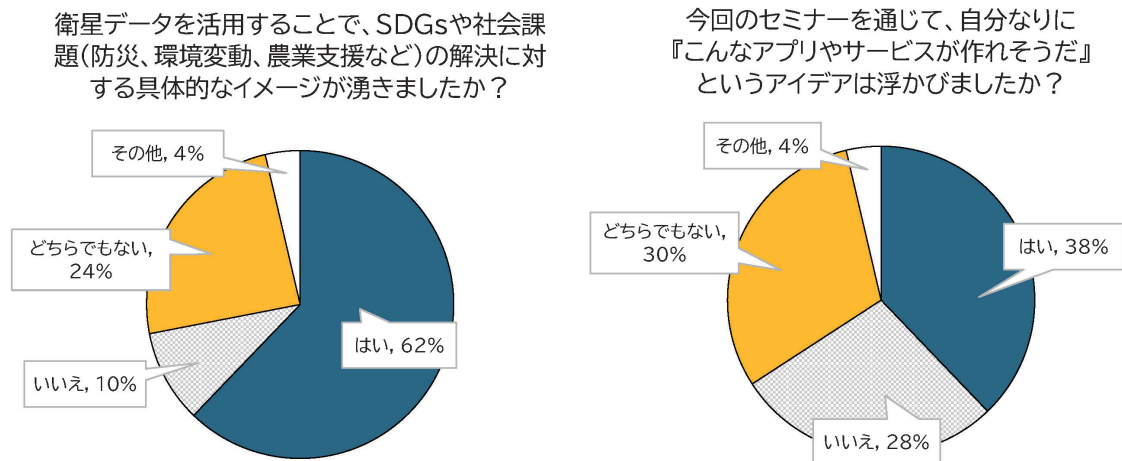


図3 トークイベント後における課題解決やアイデア創出に関する状態

一方で、具体的な活用方法や自身のスキルとの結びつきについては不明確であるとする記述が多く見られた。これらの結果は、認知段階にとどまることの限界と、次段階への橋渡しの必要性を示唆している。

4. 教育設計の妥当性の検討

予備的調査の結果から、トークイベントは学習者の関心を喚起する「認知」段階として機能している一方で、単発のイベントのみでは、学習者が実データを用いた分析や課題設定へと移行することが難しいことが示された。

この点から、認知段階と実践段階の間に、知識や技能を段階的に習得する「知識・スキル習得」段階を意図的に設計することは、データサイエンス教育において不可欠であると考えられる。また、イベントと科目を横断した学習設計により学習機会を連続的に構成することは、完全オンライン環境におけるPBL的学習体験の構築においても合理的である。

さらに、本取り組みでは、データ提供主体と教育実践主体が協働することで、教育内容と実社会との乖離を抑制する設計を試みている点も特徴的である。これは特定の組織に固有の取り組みというよりも、authentic dataを教育に取り入れる際に共通して求められる設計上の視点であり、他の高等教育機関においても参照可能な枠組みであると考えられる。また本設計は、特定のデータや組織に依存するものではなく、authentic dataを教育に取り入れる際の一般的な学習設計として位置づけることができる。

5. 結論と今後の展望

本稿では、JAXAとの連携による教育的取り組みを、データサイエンス教育における教育設計の観点から整理し、三段階型学習設計の意図と妥当性について論じた。本設計は、単発イベントにとどまりがちな学習機会を、正課外科目と横断して段階的に接続する

点に特徴がある。

一方で、本稿は教育設計および初期段階の状況分析を目的としたものであり、実践段階における教育効果の検証を行うものではない。今後は、自由参加コースおよびハッカソン／アイデアソンの実施を通じて、学習者の変化やアウトプットについての検証を行う予定である。

謝辞

本研究における教育プログラムの企画および実施にあたり、国立研究開発法人宇宙航空研究開発機構第一宇宙技術部門地球観測プログラム戦略室／地球観測研究センター（EORC）参事杉田尚子氏、同地球観測研究センター（EORC）主任研究開発員河村耕平氏、同衛星利用運用センター（SAOC）主任研究開発員池畑陽介氏には、地球観測衛星データおよび関連ツールに関する専門的知見の提供ならびに、トークイベント実施や教育プログラム設計に関する多大なるご協力を賜った。ここに記して、深く感謝の意を表する。

注および参考文献

- 1) De Veaux, R. D. et al. (2017). Curriculum guidelines for undergraduate programs in data science. *The Annual Review of Statistics and Its Application*, 4, 15-30.
<https://doi.org/10.1146/annurev-statistics-060116-053930>（最終閲覧日：2026年1月8日）
- 2) JAXA. Earth Dashboard. <https://earth.jaxa.jp/dashboard/>（最終閲覧日：2026年1月8日）
- 3) JAXA. Earth API. <https://data.earth.jaxa.jp/ja/>（最終閲覧日：2026年1月8日）
- 4) JAXA. G-Portal. <https://gportal.jaxa.jp/gpr/>（最終閲覧日：2026年1月8日）
- 5) 数理・データサイエンス・AI教育強化拠点校コンソーシアム教育用データベース分科会. (2025). 数理・データサイエンス・AI (MDA) 教育における課題解決型学習 (PBL) のつくり方 (ベータ版). https://www.mi.u-tokyo.ac.jp/consortium/pdf/PBLmanual_db_beta.pdf（最終閲覧日：2026年1月7日）
- 6) Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Model Context Protocol による 学習管理システム拡張の試み

— LTI との併用による生成 AI への接続 —

田 中 頼 人¹

1. はじめに

生成 AI、特に大規模言語モデル (LLM) の教育応用に対する関心が高まっている¹⁾。その一方で、多くの学習管理システム (LMS) は標準機能としての AI 教材配信をサポートしていない。そのため、従来は LMS の独自拡張による統合か、AI 教材を外部ホストに置く方法で対応してきた。しかし前者は特定 LMS への依存を招き、後者は学習体験の分断を引き起こしていた。

e ラーニング技術標準である Learning Tools Interoperability (LTI)^{2) 3)} は LMS 内外の学習体験を統合できるが、教材の開発では LTI の新しいバージョンに追従する対応コストを要していた。

本研究では、これらの制約をクリアするため、Model Context Protocol (MCP)⁴⁾ を導入し LTI と併用したアーキテクチャを提案する。提案システム上で「単一選択 + 理由入力」から「自動採点 + LLM フィードバック」の流れを持つ教材を動作させ、学習支援の質を向上させることを目指す。

2. 提案システムの構成

提案システムは

- I. LTI Tool が、LTI Platform からの認証と認可を受ける
 - II. LTI Tool が内包する MCP Client が学習者向けの画面インタフェースを提供し、MCP Server へ学習状況データを送る
 - III. MCP Server がプロンプトを生成し、LLM の呼び出しを行う
- の三層構成を採用する。

LTI Platform と LTI Tool 間は LTI 1.3 仕様に準拠する OAuth 2.0 フローで接続し、学

¹ サイバー大学 IT 総合学部・准教授

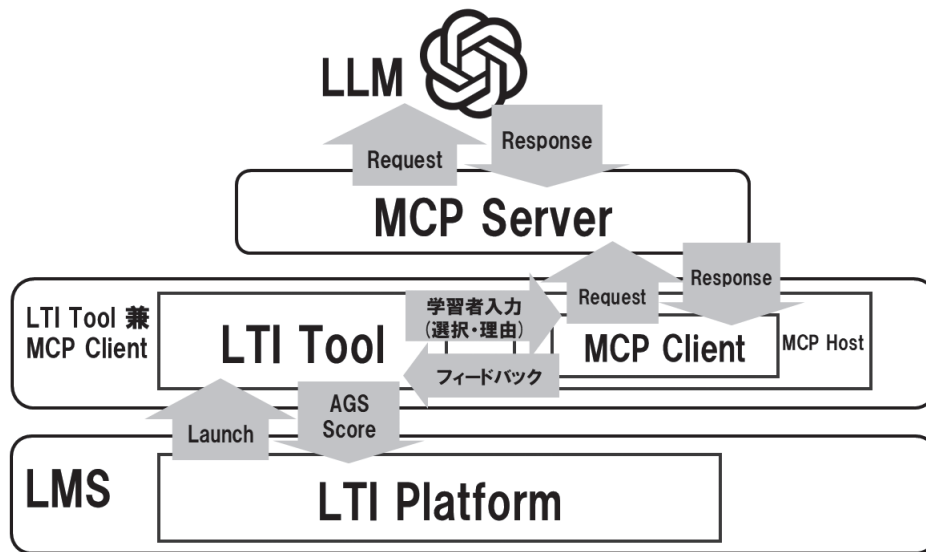


図1 システム全体の構成

習者には単一選択問題と「選択理由」入力欄を提示する。MCP Client と MCP Server 間は REST API (FastAPI の /mcp エンドポイント) で通信し、学習者の解答の正誤判定を MCP Client で、学習者への「選択理由」テキストへのフィードバックは LLM が行う。採点結果は MCP Client が LTI の Assignment and Grade Services (AGS) を介して LTI Platform に返却する。提案システム全体の構成を図1に示す。

3. 実行時の処理の流れ

提案システムにおける実行時の処理の流れを以下に示す。学習者が LMS 上で教材にアクセスし、クイズへの解答の結果が LTI Platform の成績表に記録されるまでを扱う。

- (1) 学習者が LMS の画面上で教材を選択する。
- (2) LTI Platform は OAuth 2.0 トークンを付与し、Launch リクエストを LTI Tool エンドポイントへ転送する。
- (3) LTI Tool は受信したリクエストを検証し、署名とパラメータの整合性をチェックする。
- (4) LTI Tool 内のクイズ出題モジュールが、クイズの設問セットを構築する。
- (5) 構築されたクイズの設問が HTML 形式で組み立てられ、学習者のブラウザ画面に表示される。
- (6) 学習者が単一選択問題への解答と「選択理由」のテキストを入力すると、LTI Tool 内の MCP Client は選択問題の正誤を判定し、MCP Server へ「選択理由」テキストを含むリクエストを送信する。図2は学習者が解答と選択理由を入力する様子である。

- (7) MCP Server は API を介して LLM を呼び出し、学習者の「選択理由」へのフィードバックテキストを生成して MCP Client へ返却する。
- (8) MCP Client を内包する LTI Tool は「単一選択問題への採点結果」と「選択理由へのフィードバックテキスト」を学習者の画面に表示する。図 3 は採点結果とフィードバックテキストの画面例である。

学習ツール相互運用性 (LTI) の主な目的は何ですか？

教育コンテンツの作成

外部ツールとLMSの連携

学習者の進捗状況追跡

学習管理システムの構築

理由を入力してください (オプション)

コンテンツが何よりも重要だから

Question 1 of 10 次へ進む →

図 2 解答と選択理由の入力

学習ツール相互運用性 (LTI) の主な目的は何ですか？

あなたの回答:

教育コンテンツの作成

正解:

外部ツールとLMSの連携

解説:

LTIはIMS Global Learning Consortiumによって開発された標準規格で、外部ツールと学習管理システム (LMS) をシームレスに統合することを可能にします。

🗉 AIフィードバック:

フィードバック：あなたが選んだ「教育コンテンツの作成」という選択肢について、理由として「コンテンツが何よりも重要だから」と述べています。確かに、教育コンテンツは学習において非常に重要な要素です。しかし、学習ツール相互運用性 (LTI) の主な目的は、外部ツールと学習管理システム (LMS) との連携を実現することです。LTIは、異なる教育プラットフォームやツールが簡単に統合できるように設計されています。これにより、教育者や学習者はさまざまなリソースにスムーズにアクセスできるようになります。つまり、LTIを通じて外部ツールとの連携が強化され、より豊かな学習環境が提供されるのです。あなたの考え方はコンテンツの重要性を強調していますが、LTIの本質的な目的を理解することも同様に重要です。この理解を深めることで、より効果的に...

図 3 フィードバックの例

- (9) LTI Tool は採点結果を LTI AGS 仕様に準拠する形で LTI Platform へ送信する。
- (10) LTI Platform は成績データを検証し、成績表データベースへ書き込む。

上記の流れにより、学習者は LMS 内外の区別を感知させないシームレスな受験体験を享受でき、成績データは安全かつ確実に LMS へ反映される。

4. 多肢選択式を超える学習体験の提供

多肢選択式の自動採点は即時性と客観性に優れる一方、学習者の思考過程を捉えるには限界がある。そこで提案システムでは選択後に理由を記述させ、その内容に応じて AI が個別フィードバックを返す構成を採用した。選択と理由を入力させる出題方式は、Clark らの実践⁵⁾を参考にしている。

成績には選択肢の正誤のみを反映し、理由記述は評価の対象外とすることで、公平性を担保しながら学習支援の幅を広げる仕様とした。AI による応答は正誤にかかわらず学習者が自らの考え方を見直すきっかけとなるよう設計されており、形式的な「当てずっぽう」を避け、根拠のある選択を促す効果がある。

Butler and Winne⁶⁾は、フィードバックを学習者の自己調整学習を支援する情報として位置づけ、成績に繋がる評価そのものよりも、学習者が自身の理解や方略を修正するための手がかりを与えることの重要性を指摘した。提案システムにおいて AI を理由記述へのフィードバックに限定し成績評価から切り離れた構成は、このような形成的フィードバックの理論的枠組みに整合するものである。

また、教員の負担を増やさずに個別対応を実現する手段としても、部分的に AI の処理を取り入れる提案システムの方式は有効である。このように、評価と支援を適切に分離した構成は、思考の質を高めるための設計的工夫の一つといえる。

5. 今後の拡張の可能性

LMS から LLM までを接続しフィードバックを得る提案システムの機能は、既存の Web API や LMS 固有機能を用いて実現することも可能であった。しかし、提案システムでは MCP と LTI を併用することで、LMS と LLM 間の情報伝達を汎用的かつ中立的な方法で統一できる点を重視した。特に、複数の LTI Platform との互換性を保ちながら、文脈情報の注入や AI フィードバックの処理を共通化できる利点は大きい。今後、他の学習支援ツールとの連携や異なるプロンプト設計手法の実装も視野に入れることで、より高度な教育支援システムへの発展が期待される。

6. まとめと今後の課題

本研究では、LTI Toolとして動作する MCP Client と MCP Server を連携させ、複数の LMS 上で統一的に「単一選択+理由入力」から LLM によるフィードバックを提供できることを示した。提案システムは教材アクセスから AGS での成績送信までをシームレスに運用し、学習者の振り返りの促進につながるフィードバックの生成を行うことができる。

今後の課題として (1) 教員用画面でのクイズ作成・履歴閲覧機能追加、(2) 閉域ネットワーク環境および学内設置サーバでセキュリティを担保した LLM 運用のサポート、(3) AI フィードバック効果の定量的分析とプロンプト最適化サイクルの仕組みの構築、が挙げられる。

注および参考文献

- 1) Wang, S., et al., 'Large Language Models for Education: A Survey and Outlook', arXiv:2403.18105, 2024, <https://arxiv.org/abs/2403.18105> (最終アクセス日: 2026年1月8日).
- 2) IEdTech Consortium, 'Learning Tools Interoperability (LTI)[®] Core Specification v1.3', <https://www.imsglobal.org/spec/lti/v1p3> (最終アクセス日: 2026年1月8日).
- 3) 田中頼人, 「学習基盤を拡張する国際技術標準 IMS LTI 1.3 — 第2回 LTI 1.3 開発のための資料とサービス —」, 『情報処理』, 63 巻 7 号, 2022 年, pp. 347-350.
- 4) Anthropic, LLC, 'Introducing the Model Context Protocol', <https://www.anthropic.com/news/model-context-protocol> (最終アクセス日: 2026年1月8日).
- 5) Clark, S.A., et al., 'The Role of Answer Justification in Multiple-Choice Testing: Effects on Performance and Metacognitive Accuracy', "Behavioral Sciences", Vol. 15, No. 4, Article 477, 2025, <https://www.mdpi.com/xxxx> (最終アクセス日: 2026年1月8日).
- 6) Butler, D. L., & Winne, P. H., 'Feedback and self-regulated learning: A theoretical synthesis', "Review of Educational Research", 65 (3), 1995, pp. 245-281.

Evaluating MEXT's Plan to Replace Japanese University English Entrance Exams with Commercial Assessments

Jared R. Baierschmidt ¹

Abstract

This paper analyzes Japan's proposed reforms to The Common Test for University Admissions, focusing on the English section's transition from a government-created multiple-choice format to commercially available four-skills English proficiency tests. Drawing on the history of English language education and assessment in Japan, the paper examines the long-standing emphasis on grammar and reading that stems from the "examination hell" culture and the resulting neglect of productive language skills. It discusses how previous Ministry of Education, Culture, Sports, Science and Technology (MEXT) reforms failed to change classroom practices, the potential mismatches between commercial English tests and Japanese curricular goals, and the expected social and educational impacts of these reforms. The analysis concludes that without coordinated curriculum redesign and teacher retraining, the reform may reinforce educational inequality while failing to substantially improve students' communicative English ability.

Keywords: English language learning, Japanese university entrance exams, English assessment reform, MEXT

1. Introduction

Every year, nearly 500,000 Japanese high school seniors will spend two grueling days in January participating in a Japanese rite of passage known as the National Center Test for University Admissions¹. The results of this test will determine which university the students will be able to attend and, by proxy, which careers they will be able to pursue². In 2017, Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) announced drastic changes to the portion of the test which assesses test-takers' knowledge of English³. Whereas previously the government

¹ サイバー大学 IT 総合学部・講師

subsidized National Center for University Entrance Examinations (NCUEE) developed the English portion of the test, MEXT is proposing to utilize commercially available language assessments. In this paper, I will analyze the reasons for this change by situating them in terms of the history of English language teaching and assessment in Japan. I will then describe how the current version of the National Center Test for University Admissions is conducted and the problems with the English portion of the test that have led to the announced revisions. Next, I will discuss potential problems with the proposed changes. Finally, I will analyze the potential impact of these changes on Japan's education system at the high school, university, and national levels.

2. A Brief History English Assessment in the Japanese University Entrance Exam System

The Japanese term *shiken jigoku*, roughly translated as “examination hell,” can be traced back nearly 150 years to Japan's Meiji era²⁾. Having just ended hundreds of years of isolation from the rest of the world, the Japanese government was dismayed to find itself at an extreme technological disadvantage compared with Western powers⁴⁾. Japan mobilized its education system as one way to correct this imbalance. Learning the English language was seen as the key to quickly acquiring the technology Japan needed in order to modernize, and thus English classes found a prominent place in the educational system and began being taught in middle school, high school, and university⁵⁾.

It was during this period of great political, social, and economic upheaval that the entrance exam system was introduced in Japan. Japan's only national university (currently The University of Tokyo) established English as the primary language of instruction in 1863, and as more universities were created, instruction in English became the norm⁵⁾. Therefore, English first appeared on university entrance exams at this time to ensure that applicants possessed adequate English language proficiency in order to be able to comprehend their university classes. Because of the low number of universities, prospective students needed to spend a large amount of their high school education preparing for the competitive entrance exams to secure a university spot. Hence, the term “examination hell” was coined to describe the “disproportionate amount of resources allocated towards preparing for [the exams]”²⁾ (p. 173).

Although English proficiency was initially required to enter university, as Japan established itself as a modern nation by defeating China in the Sino-Japanese War of 1895 and Russia in the Russo-Japanese War of 1904-1905 there came a nationalistic shift in both government and public attitude that saw Japanese re-established as the

language of instruction at Japanese universities⁴⁾. Yet English was not abandoned as a requirement to enter Japanese universities and indeed even became a requirement on some high school entrance exams. Sasaki notes that despite the nationalistic fervor of the time, people still viewed learning English as useful for “cultivating the learners’ minds”⁵⁾ (p. 66).

While the entrance exams may not have changed much at the start of the 20th century, English instruction at the high school and university level did undergo a drastic shift. Whereas previously instruction in the 1860s had utilized forerunners to the audiolingual method⁴⁾, the early 1900s saw grammar-translation become the primary method of instruction, with learners mostly focused on translating literary works from English into Japanese or vice versa⁵⁾.

This period also saw the rise of the *juken-eigo* industry⁴⁾. *Juken-eigo* roughly translates to “English for examinations.” A number of book publishers and private teaching schools rushed to fill the need of students who could not speak English but needed to use English in order to pass university or high school entrance examinations. This robust industry exists still today, with entire sections of bookstores devoted to *juken-eigo* materials and a plethora of examination cram schools crowded near major train stations.

English did not lose its place as an entrance requirement for universities and elite high schools even after the outbreak of World War 2. While the use of what was perceived as the language of the enemy for communicative purposes was discouraged, English still appeared on entrance exams and was still “taught” inasmuch as diagramming the grammatical structure of English sentences or translating randomly selected lines from English novels can be considered teaching English^{5) 6)}.

After World War 2, under American occupation, Japanese education underwent many drastic reforms⁶⁾. In the new system, although English was initially designated as an elective subject, due to its continued use on high school and university entrance exams it became a *de facto* compulsory course⁵⁾, with teachers continuing to utilize grammar-translation methods in preparing their students for the written exams which “tend[ed] to heavily emphasize reading and the grammatical aspects of English and give little attention to oral/aural skills”⁶⁾ (p. 29).

3. English Assessment in the Modern Entrance Exam System

During Japan’s economic boom of the 1960s, the government became interested in standardizing the university entrance exam process, as students who were applying to multiple schools were suffering inordinate stress from having to prepare for each

school's unique entrance exam. In 1979, the Joint First-Stage Achievement Test (JFSAT) was established, with the test being updated and renamed the National Center Test (NCT) in 1990⁷⁾. This entrance exam has been used as the primary admissions criterion for all public universities. Private universities have been given the choice to use the standardized exam as the sole criterion for admission, as a weighted part of the admissions process, or instead administer their own admissions test. In practice, most private universities choose to use the standardized test as part of the admissions process. For example, as of 2017, of the 1100 public and private universities in Japan, 850 required applicants to take the NCT⁸⁾. This equates to roughly 86% of admitted university students ending up taking the exam⁹⁾.

English has always been one of the subjects tested by the NCT. Originally, the test featured only a written component, which measured several areas of English knowledge including: knowledge of pronunciation, knowledge of grammar and vocabulary, knowledge of discourse, knowledge of reading strategies, and comprehension of expository passages¹⁰⁾. All questions are multiple-choice, with each section's instructions written in Japanese.

In 2003, MEXT recommended a series of revisions to English education in Japan in an attempt to produce students more capable of utilizing functional English¹¹⁾. As part of these changes, a listening section was added to the NCT¹²⁾. The listening section assessed four listening skills: the ability to understand short conversations about everyday topics, the ability to discern the purpose of a conversation, the ability to comprehend information during a long exchange, and the ability to understand the main idea of a lengthy monologue¹⁰⁾. Like the written section, all questions are multiple choice. It was hoped that the washback effect from adding listening to the NCT would result in English high school teachers adjusting their courses to focus more on communicative English skills and away from grammar-translation as a primary means of instruction¹²⁾.

4. Issues with the English Assessment in the Current Entrance Exam System

One of the most glaring problems with English assessment on the National Center Test (NCT) is that, currently, only the English receptive skills of reading and listening are being tested. Despite MEXT issuing policy guidelines in both 2003 and 2013 requesting that teachers focus on four-skills training, going so far as to calling for classes to be conducted in English¹³⁾, the NCT focus on only receptive skills has inevitably led to a situation in which speaking and writing skills are mostly neglected in

English classrooms¹²⁾.

Another problem with the current NCT is how students prepare for the English portion of the exam. High school materials and textbook developers have been slow to embrace a balanced four-skills approach and focus more heavily on reading and grammar¹²⁾. While this might at first seem to be beneficial for students preparing to take the NCT, Underwood examined MEXT-approved senior year high school English textbooks and discovered a number of problems⁹⁾. One of the major problems revealed was that the NCT reading passages were significantly more difficult than the ones appearing in the textbooks. Additionally, the textbooks did a poor job of preparing students to answer the types of questions that frequently appear on the NCT. For example, while the NCT questions often required students to synthesize information from a reading, textbooks merely required students to scan for a particular word in a text.

Because the English textbooks do not fully prepare students for the NCT, any student that wishes to score highly on the NCT is forced to spend large sums of money attending one of the many cram schools in Japan in order to prepare for the test. These schools use publicly released previous versions of the NCT to develop and teach effective test-taking strategies⁴⁾. Unfortunately, as a result of this focus on test-taking strategies, most students learn how to answer questions on the NCT without actually having to use English²⁾. Underwood also notes that many of the cram schools have learners memorize a large number of low-frequency vocabulary, 40% of which in his analysis of publicly released exams never actually appeared on the test⁹⁾.

Yet another problem with the current university entrance exam system is that it is optional for private universities. While most universities in Japan do in fact use the NCT as part of their admissions process, many also administer a second English exam in February²⁾. This second exam differs greatly from the NCT. Brown and Yamashita, in a comparison of private school exams to the NCT, found that whereas the NCT required about a 7th-grade level of English reading comprehension and vocabulary, the private school exams required university levels of each¹⁴⁾. Furthermore, the private school exams focused much more intensively on grammatical questions than the NCT. Brennan points out that the discrepancy between the two tests lies in the fact that there is no national standard for university entrance examinations²⁾. For private schools that use both the NCT and their own secondary examination, applicants must spend even more time preparing for their exams and focus on "examination English" as opposed to developing productive skills in the language.

Brennan summarizes the problems with the current English exam system as follows²⁾. First, most students' entire motivation for passing the exams is to get into a

good university, not to use English productively. Second, in order to pass the exams, students must learn advanced vocabulary and grammar that is unnecessary for basic proficiency in English, while at the same time failing in the process to acquire skills that are essential to using English productively. Finally, most students end up attending private after-school classes that focus on test-taking strategies, which will enable them to answer questions without actually using the intended amount of English to answer the exam questions. Given the problems listed above, it is perhaps unsurprising that MEXT has announced drastic reforms for the Japanese university entrance exam system.

5. Proposed Changes to the English Section of the NCT

In 2013 and 2014, MEXT conducted two large-scale surveys of the English abilities of Japanese high school seniors and were dismayed to find that, despite educational reform plans initiated in 2003 and 2013, after six years of English education, the high school seniors only achieved CEFR levels of high A1 and low A2 range in reading and listening. Even worse, more than 80% of students only achieved the lowest level of A1 in speaking and writing¹⁵⁾. Partly in reaction to this, in 2016 MEXT proposed an overhaul of the entire NCT to begin in 2020. The reforms called for replacing the formerly multiple-choice question format with short-answer questions, and initially to eliminate the English section entirely from the NCT and replace it with commercially available four-skills tests of English, such as TOEFL iBT. However, due to protests from many universities that they would be unable to modify their admissions process sufficiently by 2020 to accommodate the proposed changes, MEXT agreed to continue to develop an English section for the NCT (now renamed to the Common Test) for the foreseeable future. Additionally, all universities could decide whether to use the Common Test English section or commercially developed English proficiency tests—or some combination of both—during this transition period. Confusing matters further, MEXT has yet to announce which commercially available tests will be considered valid for the purposes of entrance examinations, leading to a number of companies actively lobbying the government for approval¹⁶⁾.

6. Issues with the Proposed Change to Commercially Available Tests

For years, many researchers have blamed Japanese students' poor English abilities on the long shadow of university entrance exams, which have focused almost

exclusively on reading, listening, and grammar to exclusion of speaking and writing^{4) 12) 17)}. MEXT's decision to revise the entrance exam to use commercially available tests of English, which assess all four skills, is therefore a welcome change. However, will this change actually lead to Japanese students finally improving their English proficiency?

It is hard to be optimistic about the proposed changes having a significant impact on the English abilities of Japanese students when one considers the nearly 15 years of English educational reforms that MEXT has already attempted. In both 2003 and 2013, MEXT announced policy changes to English education in Japanese schools that, for the most part, went unsupported and ignored. Despite MEXT's calls for a more four-skills approach to teaching English, materials and textbook developers only marginally changed their designs¹²⁾. Furthermore, the bulk of public-school teachers were never retrained to accommodate the proposed teaching methodologies, and those that did receive training reported it to be poorly organized and executed¹¹⁾. To date, MEXT's proposed changes to the university entrance exam system have not been accompanied by either directives to materials and textbook developers to update their materials or with plans for the retraining of teachers. Even if the university entrance exam system is reformed, if teachers are not provided with the materials and training to help learners prepare for the exams, it is doubtful that a noticeable improvement in students' English abilities will manifest.

Another potential problem with the proposed changes lies within the commercially available tests themselves. The tests may not be appropriate measures of English ability for Japanese learners of English. For example, Kaneko analyzed the suitability of the reading section of the TOEFL iBT as a replacement for the current NCT exam¹⁸⁾. Kaneko found that although the size of vocabulary MEXT requires high school students to learn includes the 3,000 most common word families, the TOEFL iBT reading section required up to the 6,000-word level in order to achieve 95% coverage of the texts, which is generally accepted as the amount of coverage necessary for clear comprehension of a text. Assuming students had mastered the 3,000 word families they were exposed to in high school, Kaneko estimated that students would only be able to comprehend about 50% of the texts that appear on the TOEFL iBT. This demonstrates a clear mismatch between the English knowledge that commercially available tests assess and the English knowledge that MEXT promotes as the goal of English education in Japan, making the use of commercial tests for university entrance exam purposes questionable.

7. Potential Effects of the Proposed Changes to the NCT

Without a doubt, the proposed use of commercial English proficiency tests for university entrance exam purposes is going to have wide-ranging effects on Japanese society. At the middle school and high school level, teachers who have been preparing students for an exam that has primarily assessed grammar and reading skills must completely revise their curriculum if they wish to help their students prepare for the new tests. However, as has been mentioned previously, no plans have been announced to retrain teachers in preparation for the exam reforms, nor have materials developers announced any plans to modify their primarily grammar and reading-focused textbooks. Without adequate support, it is possible that only the most earnest of middle and high school teachers will change their class design. Overwhelmed by the new changes, the remaining teachers may choose to ignore this new policy direction, as has happened with past attempted MEXT reforms¹²⁾.

One area that will likely change drastically will be the *juken* (test preparation) private sector. Already booming, it is likely to be expanded as there will now be multiple entrance exams to prepare for: the Common Test, commercial English assessments, and private university exams. Unfortunately, *juken* education is only available to those students who can afford it. This is a problem for Japanese society, as it is a commonly held belief that education should be egalitarian and therefore available equally to all children¹⁷⁾. If, however, the majority of teachers do not adjust their English teaching curriculum to accommodate the new testing format, students will be required to attend the expensive test preparation cram schools to adequately prepare for the entrance exams and achieve a competitive score.

Japanese universities are also likely to see a shift in policy, particularly in the admissions process. First, they must decide which English assessment to use for admissions purposes. As part of this process, universities will also need to decide whether English is to be assessed purely to determine the extent a student studied in high school or whether they are using it as a measure of how competent the student will be in actually using English in their studies at the university. Additionally, they will need to re-evaluate whether they are training students to play international roles in their chosen field of study, which will require additional training in the four primary skills, or whether English is merely used as a tool to acquire knowledge, in which case only reading skills would primarily matter¹⁵⁾.

To examine the current use of English at Japanese universities, Sawaki conducted an exploratory investigation of how English was used by six university

teachers in the sciences and math fields¹⁵⁾. Through structured interviews, he uncovered that while all of them used English textbooks in their classes, productive skills and listening skills were almost wholly neglected. All of the teachers explained their choice of using English textbooks in their classes by pointing out that Japanese textbooks are often of poor quality compared to their English-language counterparts and that future researchers would need to use English to stay up to date on the latest trends by reading published English research. The professors were generally positive about the proposed university entrance exam reforms, but also expressed some worry that the reforms would lead to incoming students unable to comprehend the English textbooks used in university classes, since students would now split their high school study time across all four skills rather than focus mostly on reading and grammar as had been previously required with the NCT. Given the results of Sawaki's study, one is left to wonder what the point of reforming the university entrance exam system to focus on four skills is supposed to be when the majority of students do not appear to need any English skills beyond reading for their university subjects.

8. Conclusion

For the first time since English language assessment began in Japan over 150 years ago, Japan's university entrance exams are poised to assess all four English skills as opposed to just the receptive skills of reading and listening. The reasons for this drastic change are many, but primarily result from MEXT acknowledging the failure of decades of educational policy reforms that have attempted to raise the English proficiency level of Japanese students. It remains to be seen whether these proposed test reforms will have the desired washback effect on middle school and high school education that results in higher English proficiency for Japanese students. What is clear is that with the English portion of the entrance exams being split across multiple commercial tests and the nationally prepared Common Test, the already robust test preparation industry in Japan is about to experience a major boom. Meanwhile, universities will need to reflect on the place English has in their educational training. Is it merely an indicator of prior academic success? Is it simply a tool to gain knowledge about research topics? Or is it preparation for students to play a role on the global stage? How the majority of universities answer these questions is likely to influence whether or not the Japanese population as a whole will increase in overall English proficiency over the coming decade.

References

- 1) Watanabe, Y., 'The National Center Test for University Admissions', "Language Testing", 2014, Vol. 30, Issue 4, pp. 565-573. <https://doi.org/10.1177/0265532213483095>
- 2) Brennan, J. T., 'University English Entrance Examinations in Japan', "CELE Journal", 2016, Vol. 22, pp. 176-191. Accessed October 20, 2025. <https://asia-u.repo.nii.ac.jp/records/25356>
- 3) MEXT, 「高大接続改革の実施方針等について（平成29年7月13日）」 (On the Direction of Implementing the Upper Secondary School-University Articulation Reform, July 13, 2017), August 22, 2017. Accessed October 20, 2025. https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/siryo/_icsFiles/afieldfile/2017/10/17/1396986_05.pdf
- 4) Fujimoto-Adamson, N., 'Globalization and History of English Education in Japan', "The Asian EFL Journal Quarterly", 2006, Vol. 8, Issue 3, pp. 259-282.
- 5) Sasaki, M., 'The 150-Year History of English Language Assessment in Japanese Education', "Language Testing", 2008, Vol. 25, Issue 1, pp. 63-83. <https://doi.org/10.1177/0265532207083745>
- 6) Butler, Y.G., & Iino, M., 'Current Japanese Reforms in English Language Education: The 2003 Action Plan', "Language Policy", 2005, Vol. 4, pp. 25-45. <https://doi.org/10.1007/s10993-004-6563-5>
- 7) Kuramoto, N., & Koizumi, R., 'Current Issues in Large-Scale Educational Assessment in Japan: Focus on National Assessment of Academic Ability and University Entrance Examinations', "Assessment in Education: Principles, Policy & Practice", 2018, Vol. 25, Issue 4, pp. 415-433. <https://doi.org/10.1080/0969594X.2016.1225667>
- 8) Kamiya, N., 'Can the National Center Test in Japan be Replaced by Commercially Available Private English Tests of Four Skills? In the Case of TOEFL Junior Comprehensive', "Language Testing in Asia", 2017, Vol. 7, Issue 15, pp. 1-22. <https://doi.org/10.1186/s40468-017-0046-z>
- 9) Underwood, P., 'A Comparative Analysis of MEXT English Reading Textbooks and Japan's National Center Test', "RELC Journal", 2010, Vol. 41, Issue 2, pp. 165-182. <https://doi.org/10.1177/0033688210373128>
- 10) Watanabe, Y., 'The National Center Test for University Admissions', "Language Testing", 2014, Vol. 30, Issue 4, pp. 565-573. <https://doi.org/10.1177/0265532213483095>
- 11) Gorshkov, V., & Lange, E., 'Motivating Factors in Bridging English Communication and International Understanding', "Journal of Nihonbashi Gakkan University", 2015, Vol. 14, pp. 107-119. https://doi.org/10.24581/nihonbashi.14.0_107
- 12) Glasgow, P. G., & Paller, D. L. (2015). 'English language education policy in Japan: At a crossroads', "English Language Policy in Asia", Vol. 11, pp. 153-180. https://doi.org/10.1007/978-3-319-22464-0_7
- 13) MEXT, "English Education Reform Plan Corresponding to Globalization", January 23, 2014. Accessed October 21, 2025. https://www.mext.go.jp/en/news/topics/detail/_icsFiles/afieldfile/2014/01/23/1343591_1.pdf
- 14) Brown, J., & Yamashita, S., 'English entrance examinations at Japanese universities: What do we know about them?', "JALT Journal", 1995, Vol. 17, Issue 1, pp. 7-30. Accessed October 21, 2025. <https://jalt-publications.org/jj/articles/2734-english-language-entrance-examinations-japanese-universities-what-do-we-know-about->
- 15) Sawaki, Y., 'University Faculty Members' Perspectives on English Language Demands in Content Courses and a Reform of University Entrance Examinations in Japan: A Needs Analysis', "Language Testing in Asia", 2017, Vol. 7, Issue 13, pp. 1-16.

<https://doi.org/10.1186/s40468-017-0043-2>

- 16) McCrostie, J., 'Spoken English tests among entrance exam reforms Japan's students will face in 2020', "The Japan Times", July 5, 2017. Accessed October 21, 2025.
<https://www.japantimes.co.jp/community/2017/07/05/issues/spoken-english-tests-among-entrance-exam-reforms-japans-students-will-face-2020/#.WjxZJ3mYPX5>
- 17) Aspinall, R. W., "International Education Policy in Japan in an Age of Globalisation", Brill, 2013.
- 18) Kaneko, M., 'Is the Vocabulary Level of the Reading Section of the TOEFL Internet-Based Test Beyond the Lexical Level of Japanese Senior High School Students?', "Vocabulary and Learning Instruction", 2014, Vol. 3, Issue 1, pp. 44-50. <http://dx.doi.org/10.7820/vli.v03.1.kaneko>

日本のオンライン大学における 英語学習 SNS 活動の実施

バイアシュミット・ジャレド・リチャード

概要

本稿では、日本の大学入学共通テストにおける英語試験改革を分析する。従来の政府作成による選択式試験から、市販の4技能型英語検定試験への移行を中心に、英語教育および評価の歴史的背景を踏まえて検討する。日本の「受験地獄」文化に起因する文法・読解偏重の傾向と、それに伴う発信型技能の軽視を指摘し、過去の文部科学省（MEXT）による英語教育改革がなぜ実質的な変化をもたらさなかったのかを考察する。また、市販試験と日本の教育課程との不一致、さらに今回の改革が社会的不平等や教育現場に与える影響についても論じる。結論として、教育課程の再設計や教師の再研修が伴わない限り、今回の改革は英語力の向上にはつながらず、むしろ格差を拡大させる可能性があるを指摘する。

キーワード：英語学習、大学入学共通テスト、英語教育改革、文部科学省（MEXT）

外部資金による研究プロジェクト紹介

相関関係に基づくラーニングパス最適化による 定量化個人学習支援

外部資金名：科学研究費助成事業 学術研究助成基金助成金（基盤研究（C））

課題番号：22K12323

研究期間：令和4（2022）年度～令和8（2026）年度

研究代表者：陳 健（サイバー大学, IT 総合学部, 教授）

研究分担者：池田 大樹（サイバー大学, IT 総合学部, 講師）

K A K E N：https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-22K12323/

研究内容の紹介：

正規教育の講義が導入された MOOC が 2006 年から米国等の国で公開されて以降、MOOC を利用して自律的に学ぶ学習者が急速に増えたが、一方で修了率に課題がある。主な原因として、1) 学習者が適した学習コンテンツを選択できない；2) 自ら学習目標を策定、計画を管理できない等があると考えられる。

オンデマンド教育における学習の個性化を支援するため、マイクロラーニング（Micro-learning、以下 ML）及びラーニングパス（Learning path、以下 LP）最適化の研究結果を用いて、以下の解決策を提案する。a) 講義をラーニングユニット（Learning unit、以下 LU）に細分化し、履修時間を定量化する。b) 領域知識と分離可能性基準を利用し、LU の相関性によって、領域、難易度別で階層的にクラスタリングを行う。また、c) 対象学習者と類似する参照学習者グループの成功した事例の LP を抽出し、学習ステータスと LU の相関関係に基づいて、学習目標ごとに LP を最適化、推薦を目指している。

一般的なオンデマンド教育の特徴として、基本的にクラス担任教員のような学習進度を管理する役割はおらず、学習者は、①自らの力で適した学習コンテンツの探索や学習方法を探索する必要があり、学習そのものに対する負担感が大きい。加えて、②適切な LP を発見することが困難である、という 2 つの問題があると考えられる。

問題①を下記の課題 1) と 2)、問題②を下記の課題 3) に分類し解決を目指している。

- 1) 領域知識と分離可能性基準を利用し、ラーニングユニットにラベル（Label）を付ける；
- 2) 1) を元に、相関ルール（Correlation rule）とクラスタリング方法を融合し、領域・難易度によって LU を階層的にクラスタリングする；
- 3) 成功例の LP を元に、学習ステータスと LU の相関関係によって、LP を最適化する；

現在の一般的なクラスタリングの手法では、テキスト・音声・ビデオ・画像のようなマ

マルチソース異種データに対しては十分な分析が難しいことから、上記の1)と2)の課題へ取り組むことは、マルチソースによる異種データのクラスタリング手法の改善に資する研究課題でもある。

現在、次の2つの取組みを進めている。

- ①学習コンテンツを定量化するための基準を定めた検証用学習コンテンツの作成を行っており、学習効率を測定する基準を定義し、学習効果の検証を目指している。実際に学習コンテンツを定量化するためにコンテンツの属性に応じてラベル付けを検討している。
- ②学習者の類似度を基に、公開されている実際の学習履歴からLPを抽出及び解析し、成功例となるLPの構築を目指している。

研究の進捗：

同一の学習コンテンツを用いて、LPに刺激を与えた検証を行ったところ学習者ごとに異なる学習の結果が得られることを確認した。(研究報告：“オンデマンド教育におけるプロセスフィードバックと学習効率の関係に関する考察”日本高等教育学会第25回全国大会(2022))

コンテンツのラベルに関するアンケートを実施し、学習者に対して学習コンテンツの属性が影響することを確認した。(研究報告：“オンデマンド教育における学習内容の適応性と個性化学習支援に関する考察”日本高等教育学会大会第27回全国大会(2024))

ロジスティック回帰要因分析とランダムフォレスト特徴量重要度分析を用いて、学生のプロフィールと学習履歴情報に基づき、学習効果に影響を及ぼす要因を分析した。(研究報告：“オンラインオンデマンド教育における離脱率に影響を与える要因の発見”教育システム情報学会第3回研究会(2025))

二段階一般化回帰モデルを提案し、学生のプロフィールと学習履歴による学習傾向を分析した。(学術会議：“Learning Tendency Discovery with a Two-Stage Generalized Regression Model,” The 10th IEEE Cyber Science and Technology Congress (2025))

今後、検証用の学習コンテンツを利用した実証検証を行い、さらなる分析と考察を進めていく予定である。

AI 社会実装人材育成のための 3 段階プロトタイプ構築法の開発と分野対応教育法

外部資金名：科学研究費助成事業 学術研究助成基金助成金（基盤研究（C））

課題番号：24K06314

研究期間：令和6（2024）年度～令和8（2026）年度

研究代表者：大江 信宏（サイバー大学, IT 総合学部, 教授）

研究分担者：井上 雅裕（大正大学, 情報科学研究所, 教授）

中島 毅（大正大学, 情報科学研究所, 教授）

海津 裕（東京大学, 大学院農学生命科学研究科（農学部）, 准教授）

神戸 英利（東京電機大学, 理工学部, 特定教授）

秋山 康智（東京電機大学, 理工学部, 教授）

除村 健俊（サイバー大学, IT 総合学部, 教授）

研究協力：M2M・IoT 研究会（理事長 小泉寿男 東京電機大学名誉教授）

K A K E N：https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-24K06314/

研究内容の紹介：

1. 概要（研究の背景と目的）

産業分野、とりわけ中小事業体において、業務課題の解決に AI を適用し、運用までつなげられる「AI 社会実装人材」の不足が指摘されている。大学等での教育カリキュラムは整備が進む一方で、社会人が自組織の課題に AI を適用するまでには、(a) 学習時間の制約、(b) AI 応用分野の多様性、(c) 実装環境・ツール群の理解負荷、(d) 成果物（プロトタイプ）を業務要件へ接続するための支援不足、といった障壁が重なり、学習から実装への移行が長期化しやすい。

本研究は、これらの障壁を短縮し、社会人が「理解」にとどまらず「最小限のプロトタイプを構築・検証できる」レベルまで到達する教育手法として、次の3要素からなる教育法を提案・整備し、その教育効果を評価することを目的とする。

- ・3段階プロトタイプ構築法（学習から実装・応用へ至る段階的ステップの設計）
- ・オンライン・ブレンド型学習法（オンライン環境での同期×非同期の統合設計）
- ・分野対応教育法（各産業分野の代表要件に基づき、教材・課題・評価を展開する手順）

さらに近年の生成 AI の発展を踏まえ、生成 AI を「学習支援」だけでなく「実装支援 (プロトタイピング支援)」として位置づけ、識別・予測系 AI を含むプロトタイプ構築演習へ組み込み、教育効果と実装上の有効性を検討する。

2. 提案する教育手法の構成

(1) 3段階プロトタイプ構築法

本構築法は、知識背景や専門分野の異なる社会人が、段階的に「理解→実装→応用」へ進むための設計である。各段階は次の到達目標を持つ。

- 第1段階：AI 応用事例から学ぶ AI 基礎
 - ・到達目標：業務課題と AI タスク (分類・回帰・クラスタリング・異常検知等) の対応関係を説明できる。
 - ・ねらい：抽象的知識ではなく「課題—手法—成果物」の対応を事例で獲得する。
- 第2段階：AI 実装の仕組みとツール理解を目的とする教材プロトタイプ演習
 - ・到達目標：データ前処理、モデル学習、評価、推論の一連の流れを、テンプレートを用いて再現できる。
 - ・ねらい：学習内容を“動く形”で確かめ、再利用可能な最小プロトタイプを構築する。
- 第3段階：実用的課題を対象とする応用プロトタイプ構築
 - ・到達目標：自分の業務課題 (または擬似要件) に基づき、要件・データ・評価指標の策定と改善サイクルの実施。
 - ・ねらい：現場要件に即した「小規模・迅速な検証」を自律的に遂行する能力を養う。

本研究の新規性は、社会人教育で生じやすい「理解の獲得」でとどまる問題に対して、プロトタイピングによる段階を踏んで、それを分野要件へ接続する枠組みとして整備する点にある。

(2) オンライン・ブレンド型学習法 (同期×非同期の統合)

オンライン環境における同期 (ライブ) と非同期 (自己学習・演習・フィードバック) を統合して学習を成立させる設計である。社会人は学習時間が断片化しやすいため、① 同期で概念理解とつまづき解消を行い、② 非同期で演習と成果物作成を進め、③ 発表・解説によるフィードバックで改善サイクルを回す、という役割分担を明確にする。

(3) 分野対応教育法 (分野の代表要件から教材展開)

分野対応教育法は、各産業分野で頻出する要件 (例：目的、制約、データ種別、評価観

点、運用条件)を整理し、各段階の教材・課題・評価ルーブリックへ展開する手順である。これにより、受講者が自分の分野に近い前提で学習を開始でき、また第3段階での応用課題設定が容易になる。

3. 2025年度の研究実施内容

2025年度は、生成AI活用を含め、現場での実践活用につながる内容へ拡張した社会人向け研修講座を表に示すように実施し、教育効果のデータを収集した。

項目	内容
対象	社会人 15名 (11社)
期間	2025年6月～9月 (8日間、計52時間)
実施形態	オンライン・ブレンド型 (同期 (講義・質疑・実装デモ) + 非同期 (教材学習 + 講義ビデオ) + フィードバック)
テーマ	生成AIの実践的応用 (RAG等)、機械学習・深層学習の基礎と応用 各テーマについて、基礎知識、事例、基本プログラムの実行確認、応用テーマでの構築演習を含む。

4. 教育効果の評価設計

本研究は、提案する教育手法の有効性を、学習到達 (理解・実装・応用) の到達度と課題提出による成果の両面から評価する。2025年度は以下のデータを収集した (分析は進行中)。

(1) 評価指標

学習到達：講義後の到達度、課題実施後解説を聞いた後の到達度

満足度：自分の課題にAIを適用できる見通し、実装への自信

成果物評価：講師による課題成果の評価

(2) データ収集方法

講義後および課題実施後解説を聞いた後のアンケート (Likert尺度 + 自由記述)

成果物に対するルーブリック評価 (講師による)。

5. 研究成果 (2025年度)

(1) 研修講座実施による成果

2025年度研修では、3段階構築法を基盤としつつ、生成AIにおけるプロンプトエンジニアリングを解説することで、プロトタイプ構築プロセスでの支援として導入した。収集

データの分析を進めており、教育効果（到達度・満足度・成果物品質）との関係を整理している。

(2) これまでの講座実施による成果の論文化

以下の要素を中核とする教育手法の実践と評価結果を論文として取りまとめて投稿し、採録された。

- ・3段階プロトタイプ構築法（学習から実装・応用までの段階設計）
- ・オンライン・ブレンド型学習法（オンライン環境での同期×非同期の統合設計）
「複数企業の技術者向け研修型オンライン AI 実習教育法と実践評価」, 山崎 貞彦 他, 工学教育, Vol. 74, No. 1, 2026, pp. 1-63 https://doi.org/10.4307/jsee.74.1_63

(3) 質保証に関する成果

本研究と関連して、社会人に対するリカレント教育の新しい仕組みであるマイクロクレデンシャルの質保証のフレームワークとアーキテクチャに関する研究成果を発表した。

(4) 分野対応の成果

複数分野の応用事例・要件を抽出し教材化を進め、以下の分野で「分野の代表要件→教材展開→プロトタイプ例」の形で整理し、教育資源として蓄積した。

- ・農学分野：カラー画像および深度画像を用いた対象の姿勢推定に関する研究開発を通じ、画像系 AI の要件・評価観点を整理。
- ・XR × AI 分野：事例調査とアイデア創出に基づくプロトタイプ的设计・実装・評価を進め、XR 環境特有の制約（計測・UI・体験品質）を教育課題へ反映。
- ・UX 分野：UX 実習における AI 応用の知見を整理し、ユーザ価値／評価指標／プロトタイプングを教育設計に組み込む枠組みを蓄積。
- ・ソフトウェア工学分野：状態遷移図からのテスト手順生成等、生成 AI の活用を「品質・検証」へ接続する教材例を作成し、実用上の留意点を整理。

6. 最終年度（2026年度）に向けた計画

2026年度は、2025年度までに整備した教材体系と評価設計を基盤に、2026年度講座を設計し、次を重点的に進める。

- ・プロトタイプ構築過程における生成 AI の活用の設計と評価。
生成 AI を第2段階では実装テンプレート生成とデバッグ支援、第3段階では要件整理と検証観点抽出に用いる。
- ・分野対応の拡張と標準化：分野要件の抽出手順、教材展開、評価ルーブリックをテンプレート化し、再利用可能な形で整理。
- ・成果の発表と論文化。

サイバー大学研究紀要 第3号

2026年（令和8年）3月31日 発行

発行者 サイバー大学

〒813-0017

福岡県福岡市東区香椎照葉 3-2-1

編集 サイバー大学

印刷所 株式会社 外為印刷

■サイバー大学

URL <https://www.cyber-u.ac.jp/>

■お問い合わせ

e-mail research@cyber-u.ac.jp

