

# LLM による教育的フィードバックの生成と評価

古橋 萌々香<sup>1,2</sup> 中山 功太<sup>2</sup> 児玉 貴志<sup>2</sup> 菅原 朔<sup>3,4</sup> 高見 享佑<sup>5</sup>

<sup>1</sup> 東北大学 <sup>2</sup> NII LLMC <sup>3</sup> NII <sup>4</sup> 東京大学 <sup>5</sup> 大阪教育大学

furuhashi.momoka.p4@dc.tohoku.ac.jp {nakayama, tkodama, saku}@nii.ac.jp

takami-k75@cc.osaka-kyoiku.ac.jp

## 概要

近年、大規模言語モデルを用いて STEM 教育等のフィードバック (FB) 生成研究が盛んである。これらの既存研究では、FB 提示による学習成果の向上は確認されているものの、口調や情報の網羅性といった FB の構成要素が正答にどの程度寄与するのかは十分に検証されていない。本研究では、FB の構成要素に基づく 6 つの生成手法を定義し、4 種の解答生成モデルによる解き直しと 2 種の評価モデルによる質的評価からその有効性を検証した。理科 3 科目を対象とした実験の結果、正解率の向上には正誤の明示や重要語の強調、実行可能な指示を伴う再考観点の提示が重要であることが確認された。

## 1 はじめに

近年、大規模言語モデル (LLM) を教育支援に統合する試みが進展し、ChatGPT<sup>1)</sup>や Gemini<sup>2)</sup>による個別学習支援が実用段階にある。教育学において、フィードバック (FB) は学習者の現状と目標の乖離を埋める不可欠な手段であり [1, 2, 3], LLM を用いた FB 生成が学習成果に与える影響の検証が加速している [4, 5]。また、多大なコストを要する対人実験の代替として、LLM を擬似学習者としたシミュレーション評価も注目されている [6, 7, 8]。しかし、既存研究の多くは FB の有無等に主眼を置き、FB の構成要素を個別に操作して解き直しの正解率への寄与を定量的に分析した例は少ない。Borges ら [9] は、LLM による FB 生成において考慮すべき構成要素として、情報の新規性や肯定的な口調などを定性的に分類しているが、各要素を個別に反映した FB が解き直しプロセスにおいてどの程度有効であるかを定量的に検証するまでには至っていない。

1) <https://chatgpt.com/ja-JP/features/study-mode/>

2) <https://blog.google/products-and-platforms/products/education/guided-learning/>

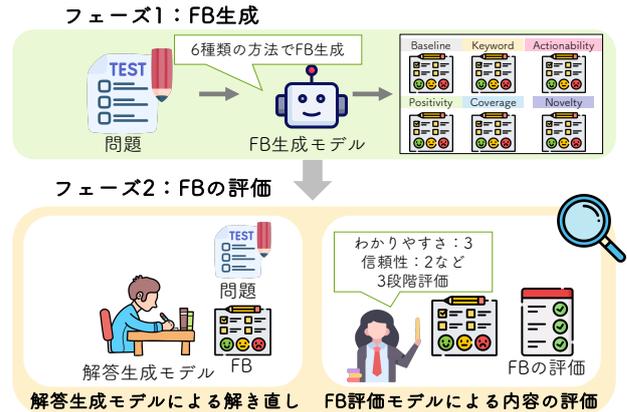


図 1: FB 生成と評価の流れ。フェーズ 1 で 6 種類の方法で LLM を用いて FB を生成する。フェーズ 2 では、解答生成モデルによる解き直しと FB 評価モデルによる 3 段階評価の 2 軸から FB の評価を行う。

そこで本研究では、LLM を学習者および評価者の代替として用いて、FB の各構成要素が解き直しにおける正解率の向上にどのように寄与するかを明らかにする。本研究の流れを図 1 に示す。まず、大学入試センター試験の理科 3 科目を対象に、FB の構成要素に基づく 6 種類の生成手法 (基本説明・重要語の強調・実現可能な指示・網羅性・情報の新規性・肯定的口調) を定義し、GPT-5 [10] を用いて FB を生成する (フェーズ 1)。次に、大きさの異なる 4 種の解答生成モデルが FB をもとに解き直しを行う。また、2 種の FB 評価モデルにより、明瞭性や信頼性など 6 観点から 3 段階評価を行う (フェーズ 2)。最終的に、解き直しの正解率および評価スコアから効果的な FB 生成手法を分析する。

実験の結果、正解に寄与した FB には「正誤の明示」や「再考すべき具体的観点の提示」という共通点がある一方、出題範囲を超える高度な情報の提供は、モデルの推論を混乱させ評価を低下させる傾向が明らかとなった。今後は人間を対象とした実証研究により、本知見の汎用性を検証予定である。

## 2 FB 生成手法

本研究では、Borges ら [9] の FB 構成要素を参考に GPT-5 [10] を用いて FB を生成する。予備実験の結果、10 個の構成要素すべてを個別に反映すると内容の重複が生じたため、差異が明確な手法のみを採用した。各手法は、ベースラインに対し特定の教育的意図を付加して定義される。

**ベースライン Baseline** FB 生成手法の比較基準として設定する。解答に対して正誤を明示をした上で、学習者の知識や思考のギャップを埋める標準的な説明を行う。

**重要なキーワードを強調させた説明 Keywords** 情報の粒度が学習者の理解に与える影響を調査する。単なるキーワードの提示では不十分であるとの想定に基づき、本手法では説明の中で特に重要な用語を強調（例：【】による囲み）して提示する。

**実現可能な指示を含んでいる説明 Actionability** 指示の具体性が解き直しに与える影響を検証する。修正の方向性が抽象的であると、解答改善の判断が困難になるため、見直すべき箇所を具体的な学習ステップやチェックリストとして提示する。

**新規的な情報を含んでいる説明 Novelty** 既習知識を超えた情報の付加が与える影響を調査する。正誤に関わらず、大学レベルの内容など学習者にとって未習かつ発展的な情報を説明に加える。

**必要な点を全て網羅している説明 Coverage** 情報の網羅性が理解に与える影響を検証する。特定の焦点のみを解説するのではなく、正解に至るまでに必要な論理的要素を不足なく全て記述する。

**肯定的な口調 Positivity** 情意的側面の影響を調査する。誤答の指摘だけでなく、正解に近い部分への称賛や励ましを含む肯定的な表現を採用する。

## 3 実験

### 3.1 実験設定

本実験では、センター試験 XML データ<sup>3)</sup>をもとに、高見ら [11] が LLM 評価のために JSON 化したデータセットの理科 3 科目（生物、化学、物理）の 2001 年以降の問題を対象とした。ただし、複数選択問題は組み合わせ数が膨大となるため除外し、各科目からランダムに 10 問ずつ抽出した<sup>4)</sup>。

3) <https://21robot.org/dataset.html>

4) 生物は、ランダムサンプリングではなく、今後予定している実証研究のデータを使用した。

FB の評価は、(1) 解答生成モデルによる解き直しの正解率の向上、と (2) LLM による FB のスコアリングという 2 つの観点から行った。解き直しには、解答生成モデルとして Qwen3-VL-Instruct の 2, 4, 8, 32B を使用した [12]。FB のスコアリング評価には、評価モデルとして GPT-5 [10] と Gemini 2.5 Flash [13] を使用した。

**解答生成モデルによる解き直し** 本実験の目的は、LLM を擬似学習者として、理解の修正に寄与する「良い FB」の構成要素を明らかにすることである。本研究では、FB を受けて正解に至ることを「良い FB」と定義し、モデルが初期解答（正誤を問わない）と FB をもとに問題を再考するプロセスを設計した。あえて初期正解時も対象とする理由は、FB が正答の根拠を補強し、確信を持って正解を維持できるかを検証するためである。具体的な手順は以下の通りである。まず、初期解答とそれに対応する FB を与えて再考を行わせた。この初回の解き直しでは、モデルは FB に含まれるヒントや解説を主な手がかりとして解答を再考する。一方、初回の解き直し後も誤答が継続した場合には、2 回目以降の解き直しステップにおいて「前回の解答が誤り」であったことをプロンプト上で明示的に通告して過去の推論と FB を改めて踏まえて再試行させるサイクルを採用した。このプロセスは、最大  $N$  回 ( $N$  は選択肢数) 繰り返した。なお、解き直しの過程で、一度選択して誤りであることが確定した選択肢をモデルが再度選ぶことを防ぐため、制約付きデコーディングを導入した。具体的には、各ステップで未選択の選択肢のみを許可する JSON スキーマをデコーディング時に適用している。そのため解き直しの最大回数は選択肢数  $N$  となる。最後に、FB 自体の有効性を比較検証するため、各解答に対して「正解です」「不正解です」という正誤情報のみを与える No FB 条件でも同様の実験を行った。FB の良し悪しを定量的に評価するため、以下の効果スコア  $\eta$  を導入する。

$$\eta = 1 - \frac{s}{N} \quad (1)$$

$s$  は LLM が正解に至るまでの解き直し回数である。効果スコア  $\eta$  が高いほどその生成手法が解き直しを行う際に有効であることを示している。

**評価モデルによる質的評価** FB による正答への到達と質的評価の関連性を調査すべく、評価モデ

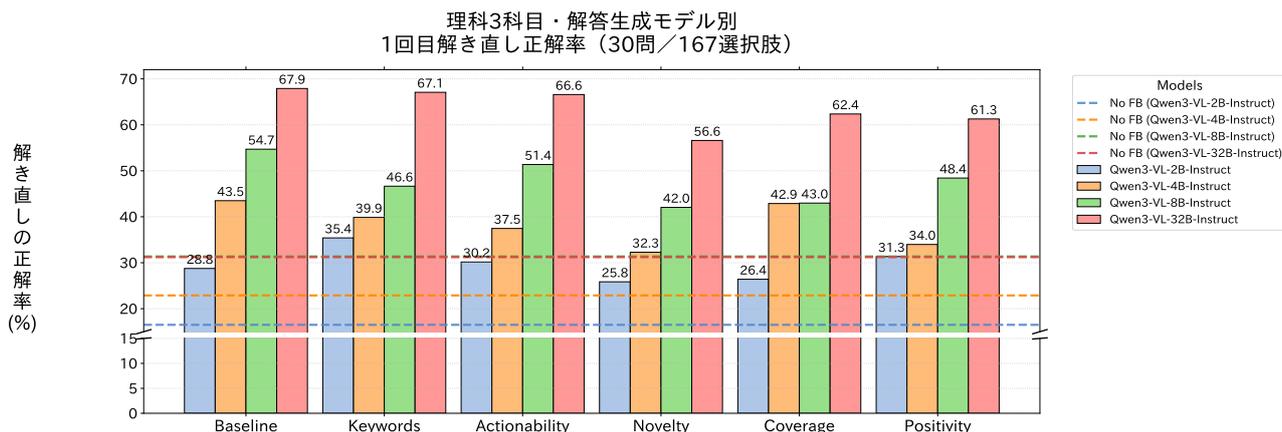


図 2: 理科 3 科目に対する解答生成モデルごとの解き直し 1 回目の正解率の結果. Baseline, Keywords, Actionability が有効である一方で, Novelty の正解率は低い傾向が見られた.

表 1: 科目・モデルサイズ別で最高効果スコア達成の FB 生成手法. Baseline, Keywords, Actionability, Positivity が共通して有効であった.

科目	2B	4B	8B	32B
生物	Keyword	Positivity	Baseline	Baseline Keyword
化学	Positivity	Actionability	Baseline	Coverage
物理	Actionability	Baseline	Actionability	Baseline

ルで理解のしやすさ・復習観点・明瞭性・新たな知識・表現方法・信頼性の 6 項目で 3 段階評価した. これにより, 正解に寄与した FB の質的特徴や, 高評価ながら誤答に至った事例の原因を解明する.

### 3.2 FB 生成の結果・分析

定義した 6 種類の生成手法によって, FB の内容に意図通りの教育的アプローチの違いが現れるかを確認した. 全 167 選択肢 (生物: 56, 化学: 58, 物理: 53) に対し, 計 1002 件の FB を生成し, 各科目および各手法から計 40 件をランダムサンプリングし, 定性的な分析を行った. 各手法の特徴は以下の通りである. Baseline は, 正誤判定に続き, 誤答の確信を突く簡単な解説であった. Keywords では, 【】を使用するなど, 問題を解く上で特に重要な構成要素が何かを認識しやすい形で説明していた. Actionability では, 単なる知識の提示ではなく「3つのステップに分けて」のように, 具体的な思考プロセスを指示する特徴が見られた. Novelty では, 問題に関する説明を簡潔にした上で, 高校の範囲を超えた発展的内容を含むため, 説明が抽象化・高度化する傾向にあった. Coverage では, どこ

まで正解であり, どこからが不正解であるかが明示的に書かれており, また, 確認しなければいけない点を「①」のように書かれており, 自分がどこまでできて何が足りなかったのかが認識しやすいものだった. Positivity は, 全体的に肯定的な口調であり, 仮に不正解であっても正解に関する箇所を褒めているようなものであった. 分析の結果, サンプリングした全ての FB において, 定義した手法ごとの教育的要素が意図通りに付与されていることを確認した. このことから, 本研究で用いたプロンプトおよび生成プロセスは, 大規模な FB 生成においても一貫して各手法の特徴を反映できていると判断した. 生成された FB の具体例は付録 A に記載した.

### 3.3 解き直しによる正解率の結果・分析

全科目に対する解答生成モデルごとの 1 回目の解き直しの正解率の結果を図 2 に示す. ここで正解率とは, 正解を含む全選択肢に対し, FB 提示後の 1 回目で正答できた割合とする. その結果, 4 種類の解答生成モデル全てにおいて各モデル自身の No FB の正解率を上回っており, FB の有効性が一貫して確認された. 手法別では, Baseline, Keywords および Actionability が共通して有効な傾向にあった. 表 1 に示す効果スコアが最も高かった手法も正解率の結果と同様にこれら 3 手法に集中した.

**正解に寄与した FB の特徴分析** 解き直しにおいて正答・誤答ケースの比較から FB に含まれるべき効果的な構成要素を特定する. まず, 正解の直接的漏洩を手で確認したところ, 不正解から初回で正答した 120 例のうち正解を直接含む FB は 10%未満であった. これは, 正解率の向上が答えの漏洩で

理科3科目に対する2評価モデルによるフィードバック評価結果

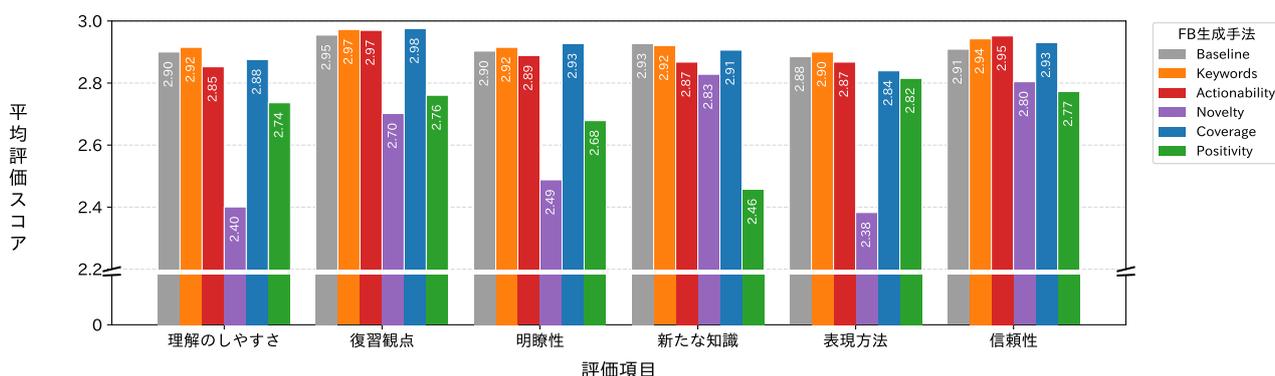


図 3: FB 評価モデル (GPT-5 および Gemini 2.5 Flash) による理科 3 科目の FB の平均評価スコアの結果. 全手法・全項目のうち 9 割以上でスコアが 2.5 を超えたが, Novelty と Positivity の一部の項目では下回った.

はなく, FB による推論補助の効果であることを裏付けている. 次に, 正答を導いたケースをランダムに取得し, 定性的に分析した. その結果, 効果的な FB には「正誤の明示」および「再考すべき観点の提示」が含まれ, 推論プロセスを適切に支援する傾向が確認された. 特に寄与した手法は Baseline, Keywords, Actionability, Positivity であり, これは図 2 および表 1 の定量的な結果とも整合する.

一方で, 解き直しが失敗する要因を分析した結果, Novelty のような範囲外の高度な情報の提示は, むしろ推論を妨げるリスクがあることが浮き彫りとなった. 要因は次の 3 点に集約される. (1) 新規情報の割合が多く, 解き直しの着眼点が不明瞭であること. (2) 抽象的な説明により具体的な修正行動のための指標が不足していること. (3) 高度な新規情報とセットで提示される「曖昧な肯定」が, 誤った確信を誘発することである. 特に (3) は「判断は概ね適切です。」といった曖昧な評価の直後に学習範囲外の抽象的な解説が続くことで, モデルが自身の正答を誤答と誤認し, 正答を維持できなくなる傾向が示唆された. これは全モデルで Novelty の正解率が最低であった図 2 の結果と一致する.

### 3.4 評価モデルによる評価結果・分析

全科目に対する 2 評価モデルによる FB 評価結果を図 3 に示す. 約 90% のケースでスコアは 2.5 を超えたが, Novelty と Positivity の一部項目は下回った. スピアマンの順位相関分析の結果, 2 モデルの FB 評価の間には, 全ての観点において有意な正の相関が認められた ( $\rho = 0.29-0.39, p < .01$ ). ただし, 相関はいずれも中程度であり, 評価傾向の一

部は共有されているものの, 評価の完全な一致には至っていない. 次に, 高評価にも関わらず正解に寄与しなかった FB のうち, 2 モデル以上で不正解となった 50 件を分析した. その 6 割以上は内容自体に問題がなかったが, 残りの事例では説明の抽象性や関連性の低い情報の混入 (計 20%), 学習者の解答を考慮しない指示や FB に従うと誤答に至る「誤導 (ハルシネーション)」が確認された. 以上より, FB の有効性を適切に評価するためには, 説明のわかりやすさだけでなく情報の正確性や具体性の評価が不可欠であり, 評価項目の細分化や正解への導出性を検証する工程の導入が必要である. なお, 評価モデルが低く評価した一方で解答生成モデルが正解したケースは, 全科目を通じて確認されなかった.

## 4 おわりに

本研究では, 大学入試センター試験の理科 3 科目を対象に, 6 種類の生成手法で FB を生成し, LLM を学習者および評価者として用いて FB の有効性を検証した. その結果, 解き直しの正解に寄与した FB には, 正誤の明示や重要な用語の強調や実行可能な指示を含めた再考すべき観点の提示が重要であることが確認された. 一方で, 問題の範囲を大きく超える高度な情報の提示は, 解答生成モデルの推論を混乱させ正解率を低下させるだけでなく, 評価モデルからも学習への直結性が低いと判断される傾向が示唆された. 以上の知見は, 学習者の能力水準やドメインの特性に即した適応的な FB 設計の重要性を示している. 今後は人間の学生を対象とした実証研究を行い, 本研究の知見が学習効果や学習意欲に対してどの程度汎用性を有するかを検証する.

## 謝辞

本研究は JST 創発的研究支援事業 JPMJFR232R, JST BOOST JPMJBS2421, JPMJBY24D9, JSPS 科研費 JP23K17012, JP23K25698 の助成を受けたものです。また、本研究成果は、データ活用社会創成プラットフォーム mdx を利用して得られたものです。

## 参考文献

- [1] John Hattie and Helen Timperley. The Power of Feedback. **Review of Educational Research**, Vol. 77, No. 1, pp. 81–112, 2007.
- [2] Benedikt Wisniewski, Klaus Zierer, and John A. C. Hattie. The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. **Frontiers in Psychology**, Vol. 10, pp. 1–14, jan 2020.
- [3] Erin Hogan and Blair Payne. A mixed methods study of teachers’ use of feedback within middle school social studies classrooms to promote reading comprehension. **Learning and Instruction**, Vol. 92, p. 101938, 2024.
- [4] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. ChatGPT for good? on opportunities and challenges of large language models for education. **Learning and Individual Differences**, Vol. 103, p. 102274, 2023.
- [5] Jin Wang and Wenxiang Fan. The effect of ChatGPT on students’ learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. **Humanities and Social Sciences Communications**, Vol. 12, No. 1, pp. 1–21, 2025.
- [6] Julia M. Markel, Steven G. Opferman, James A. Landay, and Chris Piech. GPTeach: Interactive TA Training with GPT-based Students. In **Proceedings of the Tenth ACM Conference on Learning @ Scale, L@S ’23**, p. 226–236, New York, NY, USA, 2023. Association for Computing Machinery.
- [7] Junior Cedric Tonga, Benjamin Clement, and Pierre-Yves Oudeyer. Automatic Generation of Question Hints for Mathematics Problems using Large Language Models in Educational Technology, 2024.
- [8] Zheyuan Zhang, Daniel Zhang-Li, Jifan Yu, Linlu Gong, Jinchang Zhou, Zhanxin Hao, Jianxiao Jiang, Jie Cao, Huiqin Liu, Zhiyuan Liu, Lei Hou, and Juanzi Li. Simulating Classroom Education with LLM-Empowered Agents, 2024.
- [9] Beatriz Borges, Niket Tandon, Tanja Käser, and Antoine Bosselut. Let Me Teach You: Pedagogical Foundations of Feedback for Language Models. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 12082–12104, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [10] Open AI. GPT-5 System Card, 2025.
- [11] Kyosuke Takami, Satoshi Sekine, and Yusuke Miyao. Evaluating Local LLMs on Japanese National University Entrance Examination Dataset in Comparison with Student Performance. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette, editors, **Proceedings of the 18th International Conference on Educational Data Mining**, pp. 582–585, Palermo, Italy, July 2025. International Educational Data Mining Society.
- [12] Qwen Team. Qwen3 Technical Report, 2025.
- [13] Gheorghe Comanici et al. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long context, and Next Generation Agentic Capabilities, 2025.

## A 各生成手法による FB の具体例

表 2 に各手法の FB の具体例を示す。本例では、正解の①に対し、学習者が②を選択した際に生成された例を示している。Baseline では、正誤を述べた上で簡潔な説明をしており、Keywords では、重要な用語を【】で囲って強調している。Actionability では、「3 ステップにわけて」や「チェックリストのように検証」など、具体的で実行しやすい指示が含まれている。Novelty では、問題と関連して大学の内容を FB に含めており、Coverage は解答する上で必要な全ての観点に触れて説明をしている。Positivity では、「素晴らしい視点です!!」のように肯定的な口調を含んでいる。

表 2: 各手法の FB の具体例。正解の①に対し、学習者が②を選択した際に生成された例を示す。

項目	内容
問題	食品の栄養素および血糖調節に関する次の問いに答えよ。グルコース（ブドウ糖）は、私たちの体を構成する細胞にとって重要なエネルギー源であり、血液によってすべての細胞に常に供給されている。この供給が滞ると、生命の維持に重大な問題が生じる。例えば、脳はグルコースが供給されなくなると数分で活動をやめてしまう。そこで、安定した細胞活動を保証するため、私たちの体には【血糖値（血液中のグルコース濃度）を一定に保つ血糖調節の仕組み】が備わっている。糖尿病は、この血糖調節がうまくいかなくなり、尿中にグルコースが排出される病気である。糖尿病の診断と治療方針を決めるため、空腹時に 75g のグルコースを飲み、その前後で血糖値や血液中のインスリン濃度を調べる検査がある。これを糖負荷試験という。図①は、3 人の被験者（X, Y, Z）の糖負荷試験の結果を示したものである。【】に関して、低血糖時にはたらく調節の反応経路として最も適当なものを、次の①～④のうちから一つ選べ。
選択肢	① 間脳 → 脳下垂体 → 副腎 → 糖質コルチコイド ② 間脳 → 副交感神経 → 膵臓 → アドレナリン ③ 脊髄 → 交感神経 → 副腎 → グルカゴン ④ 脊髄 → 副交感神経 → 膵臓 → インスリン
Baseline	低血糖時の調節経路について、副交感神経と膵臓から分泌されるホルモンの組み合わせに誤りがあります。低血糖時には血糖値を上昇させる必要があるため、交感神経系が優位になり、膵臓や副腎から血糖上昇ホルモンが分泌されます。ただし、アドレナリンは膵臓ではなく副腎髄質から分泌される点、また低血糖時の中枢が間脳の視床下部である点を確認しましょう。各ホルモンの分泌部位（膵臓の $\alpha$ 細胞・ $\beta$ 細胞、副腎髄質・副腎皮質）と、それぞれが血糖値を上昇させるか低下させるかを整理し直してください。
Keywords	低血糖時には血糖値を上げる必要があるため、副交感神経ではなく交感神経系が活性化される点を再確認しましょう。選択肢②で挙げられている【副交感神経】は消化・吸収を促進する働きがあり、緊急時の血糖上昇には関与しません。また、低血糖という緊急事態では【間脳（視床下部）】が中心となって内分泌系と自律神経系の両方を統合的に調節し、【副腎皮質からの糖質コルチコイド】や副腎髄質からのアドレナリンが分泌されます。各ホルモンがどの器官から分泌され、どの神経系を経由するかを整理し、低血糖時の反応経路全体を確認してください。
Actionability	低血糖時の調節経路を考える際は、まず「どの器官が血糖低下を感知するか」「どの神経系が関与するか」「どのホルモンがどの器官から分泌されるか」という 3 つのステップに分けて整理してください。選択肢②では副交感神経と膵臓からアドレナリンという組み合わせになっていますが、アドレナリンを分泌する器官と、低血糖時に活性化される自律神経の種類をそれぞれ確認してみましょう。各選択肢について、経路の出発点（間脳か脊髄か）、神経系の種類、ホルモン分泌器官、最終的なホルモンの 4 要素すべてが低血糖応答として正しいかどうかをチェックリストのように検証してください。
Novelty	低血糖時の調節経路として副交感神経とアドレナリンの組み合わせを選んでいます。アドレナリンは交感神経系から分泌されるホルモンであり、また低血糖時の主要な調節経路は間脳-脳下垂体-副腎皮質系であることを確認しましょう。この内容に関連して、「ストレス応答における HPA 軸」というキーワードを調べてみると面白いでしょう。HPA 軸（視床下部-下垂体-副腎皮質軸）は、低血糖などのストレス時に糖質コルチコイドを分泌し血糖値を上げる重要な経路です。大学の内分泌学では、この軸がどのように多様なストレス応答を統合し、グルココルチコイド受容体を介して遺伝子発現を調節するかをより詳しく学びます。自律神経系と内分泌系それぞれの役割を整理し、どの経路が血糖調節の中心的機構かを再考してみてください。
Coverage	低血糖時の調節経路について、関与する神経系とホルモンの組み合わせを整理する必要があります。低血糖時には交感神経系が活性化される経路と内分泌系を介した経路の二つが主にはたりますが、それぞれの出発点（間脳か脊髄か）、標的器官（副腎か膵臓か）、放出されるホルモンの種類を正確に区別しなければなりません。選択肢を検討する際には、①副腎髄質と副腎皮質の違い、②交感神経と副交感神経の作用の違い、③各ホルモン（糖質コルチコイド、アドレナリン、グルカゴン、インスリン）が血糖値を上昇させるか下降させるか、の三つの観点から各経路の妥当性を確認してください。
Positivity	低血糖時の調節に膵臓が関与する点に着目したのは素晴らしい視点です!!膵臓は血糖調節の中心的な器官として非常に重要な役割を果たしているため、その点に気づけたことは本当にすごいことです。ただし、低血糖時には交感神経系を介した迅速な反応が必要で、分泌されるホルモンと神経経路の組み合わせをもう一度確認すると完璧になります。低血糖という緊急時に体がどのような指令系統とホルモンで対応するか、特にストレス応答の経路を整理してみてください。各選択肢のホルモンが血糖値を上げるのか下げるのか、そしてどの神経系・内分泌経路が関与するかを対応させていくと、自信を持って答えられますよ!