

# 教育における生成AI活用推進リーダープログラム

## 生成AIの活用

# 学習評価



**吉田 壘**

東京大学 大学院工学系研究科 准教授

LLM 寄附講座 特任准教授

2026年4月

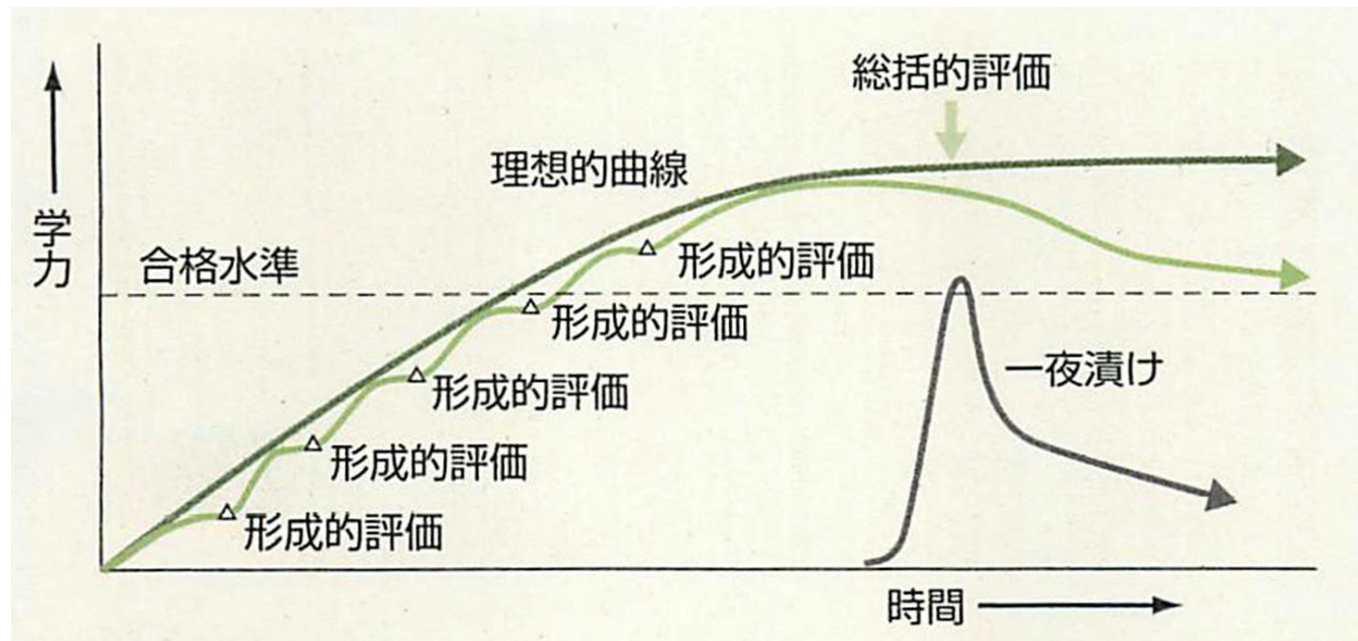
# 学習評価

- 学習者の学びの状態を捉え、学習改善や指導改善につなげるための情報収集と判断のプロセス
- 評価の役割
  - 学習の前進: 学習者が自分の状態を知り、次の一歩を決める手がかり
  - 指導の改善: 教員が授業の効果を捉え、指導を調整する手がかり
  - 到達度の確認: 学んだ成果を記録し、証明する（いわゆる評定）
  - 質の保証: 教育機関として社会に対して学習の質を示す

**生成AI時代、評価の再設計が鍵となる**

# 評価の3機能

機能	時期	目的
診断的評価	授業前	既有知識・誤概念の把握
形成的評価	授業中	学習状況の把握と指導改善
総括的評価	単元末・学期末	到達度の確認



# 形成的評価の重要性

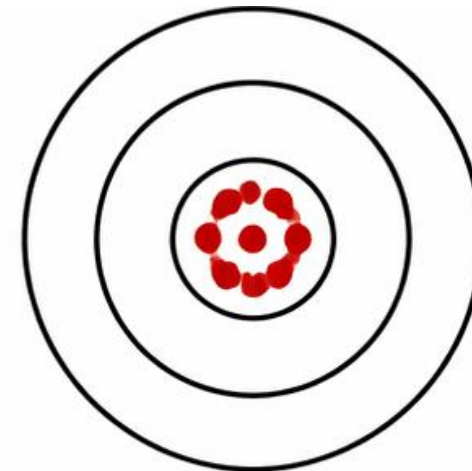
- **大規模レビュー** (Black & Wiliam 1998)
  - 形成的評価の学習効果は効果量  $d = 0.4 \sim 0.7$  (中程度)
  - 数ある教育的介入の中でも最大級の効果
  - 特に学力下位層に効果が大きい = 学力格差の縮小にも寄与
- **形成的評価を支える5つの戦略** (Wiliam & Thompson 2007)
  - 学習目標・成功基準を明確化し学習者と共有する
  - 効果的な発問や課題で「学習の証拠」を引き出す
  - 学びを前進させるフィードバックを与える
  - 学習者を互いの学びのリソースにする (相互評価・協働)
  - 学習者を自身の学びの当事者にする (自己評価)

# 評価の質

- **妥当性**: 測りたいものを本当に測れているか
- **信頼性**: 結果がどれだけ安定・一貫しているか

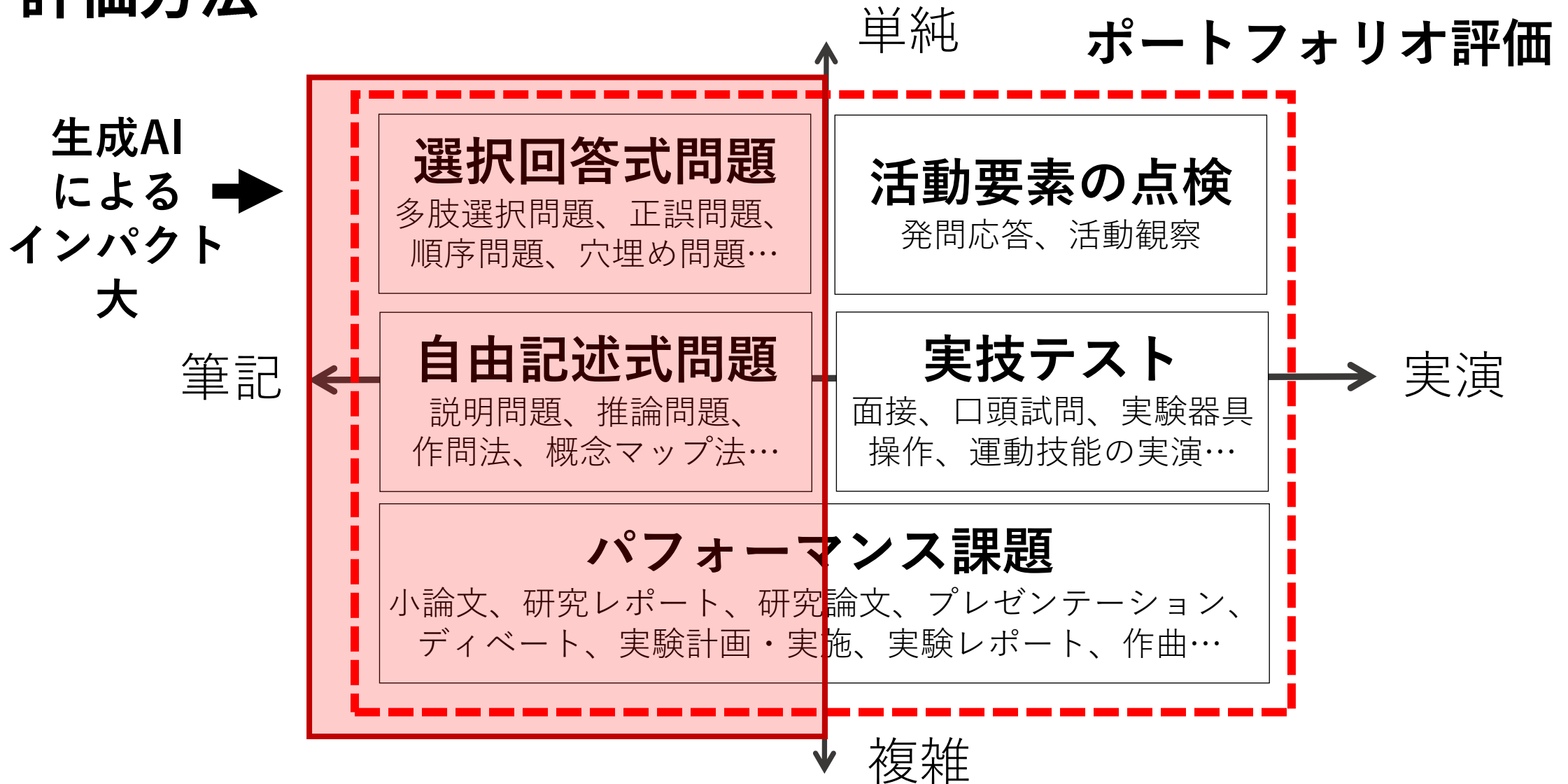


的の中心 (真の値) を  
狙えている状態が「**妥当性が高い**」



結果のばらつきが小さく、  
安定している状態が「**信頼性が高い**」

# 評価方法



# 学習と評価の一体化

- ピアインストラクションは1回目と2回目の投票が形成的評価に対応
- 学習を促しながら評価を実現可能

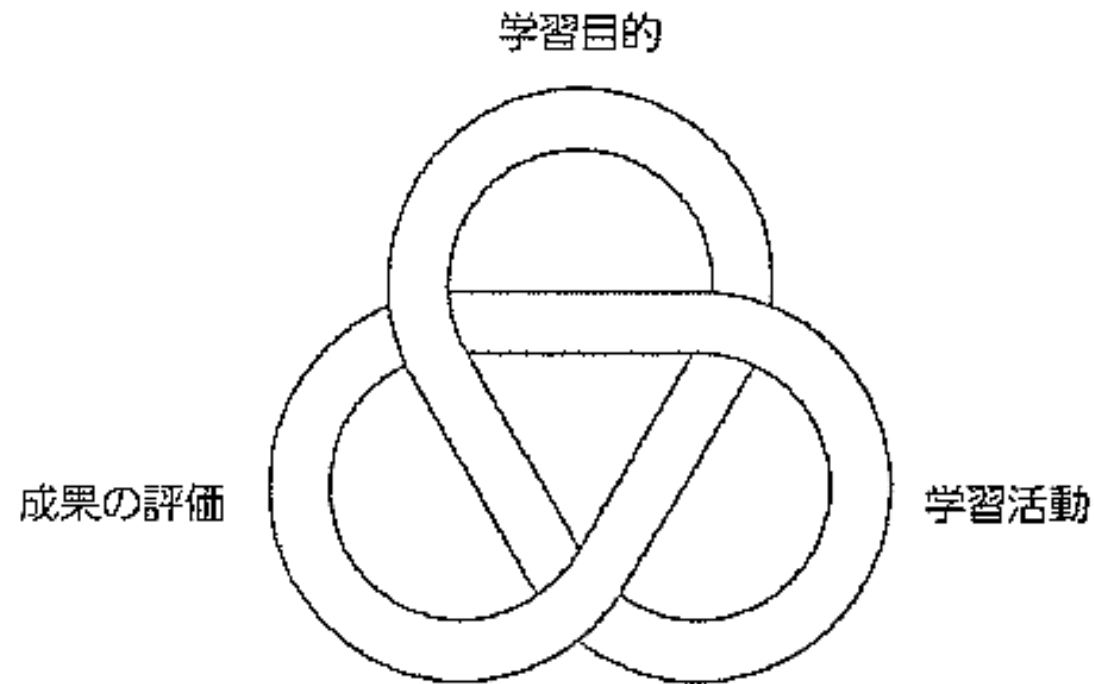


図 0.3 相互に関連しあう LAT の性質：教授、学習、評価

# 生成 AI と評価

- **生成AIは「評価の妥当性 (Validity)」を揺るがしている**
  - 学習者がAIを使ってレポートや課題を完成させた場合、提出物は学生本人の能力ではなく「学生+AI」の能力を反映している
- **「不正の問題」ではなく「評価設計の問題」**
  - 「どうやってAI利用を見破るか (検出)」に注力しても解決しない (AI検出は難しく、ツールを使っても誤検出のリスクがあり)
  - 課題そのものをAIを使われにくい/AIと協働する形に再設計することが肝要
- **シドニー大学の評価アプローチ Two-Lane Approach (Liu & Bridgeman 2023)**
  - Lane 1: Secure (確証のための評価) : 対面・監督下で実施。学生が独力で到達目標を満たしているかを確証する (例: 対面試験、口頭試問、実演、ライブデモ)
  - Lane 2: Open (学習のための評価) : AI利用を含め、あらゆるリソースの使用を認めるAIを使いこなす力・批判的評価・統合力を育てる場 (例: 長期プロジェクト、AI協働レポート、省察レポート)
  - どちらか一方ではなく、両方をプログラム全体にバランス良く配置する

# 生成 AI と評価 ～生成AIに対応した評価方法案～

- **まず利用に関するポリシーを明確に伝える（利用範囲を明確化する）**
  - AIAS（AI Assessment Scale）の考え方: Level 1: AI使用不可（対面試験など） Level 2: 構想段階のみ可、Level 3: 草稿作成に協働可（最終判断は学生）、Level 4: 全面活用可（活用の質で評価）、Level 5: AI活用法自体を探究
- **物理的に禁じる**
  - 対面でレポートを書かせる・テストする、画面ロックした状態でオンラインテストする、プレゼンテーションや口頭試問にする
- **成果物を自動生成しにくい課題にする**
  - 概念マップなど図の作成を取り入れる、協同学習を取り入れる、実験やフィールドワークなどを取り入れる、授業内容に特有の課題にする、自身の体験を具体的に交えて説明させる
- **学習過程そのものを評価対象にする**
  - 下書き・修正履歴の提出（Google Docsなどの編集履歴を活用）、AIとの対話ログの提出+どこをどう採用/修正したかの解説、中間発表・進捗チェックポイントを複数回設ける、学習ジャーナル/振り返りレポートで意思決定の過程を書かせる、ポートフォリオ評価（成果物+プロセスの記録+省察）
- **対話型生成 AI を積極的に使ってもらおう**
  - AI が生成したレポートを批評・修正させる（Mollick & Mollick 2022, [Yale 大学における実践例](#)）

**重要なのは、授業や学習目的に合った評価方法であるかどうか**

# まとめ

- **評価は「学びを前進させる対話」**
  - 評価は、学習と教授を「より良く」できる
- **形成的評価を設計の中心に置く**
  - 教育的介入の中でも最大級の効果
- **生成AIは評価の「妥当性」を問い直す契機**
  - 禁止でも全面許可でもなく、学習目的に応じた評価設計が重要
  - 例: Two-Lane Approach (Lane 1: 対面で確認、Lane 2: AI活用で育成)

# 参考文献

- バークレイ, E. F., & メジャー, C. H. (2020). 『学習評価ハンドブック：アクティブラーニングを促す 50 の技法』 (東京大学教養教育高度化機構アクティブラーニング部門・吉田壘 監訳) . 東京大学出版会. (原著: Barkley, E. F., & Major, C. H. (2016). *Learning Assessment Techniques: A Handbook for College Faculty*. Jossey-Bass.)
- Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Liu, D., & Bridgeman, A. (2023). *What to do about assessments if we can't out-design or out-run AI?* Teaching@Sydney, The University of Sydney.
- Mollick, E. R., & Mollick, L. (2022). *New modes of learning enabled by AI chatbots: Three methods and assignments* (Wharton/SSRN Working Paper).
- 日本医学教育学会 (編) (2008). 『医学教育白書 2008 年版』 . 篠原出版新社. (評価の 3 機能 = 診断的・形成的・総括的評価の分類典拠)
- 西岡加名恵 (編著) (2016). 『「資質・能力」を育てるパフォーマンス評価：アクティブ・ラーニングをどう充実させるか』 . 明治図書出版.
- William, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.