大学入試を中心とした情報分野の学力評価手法の検討

慶應義塾大学 環境情報学部

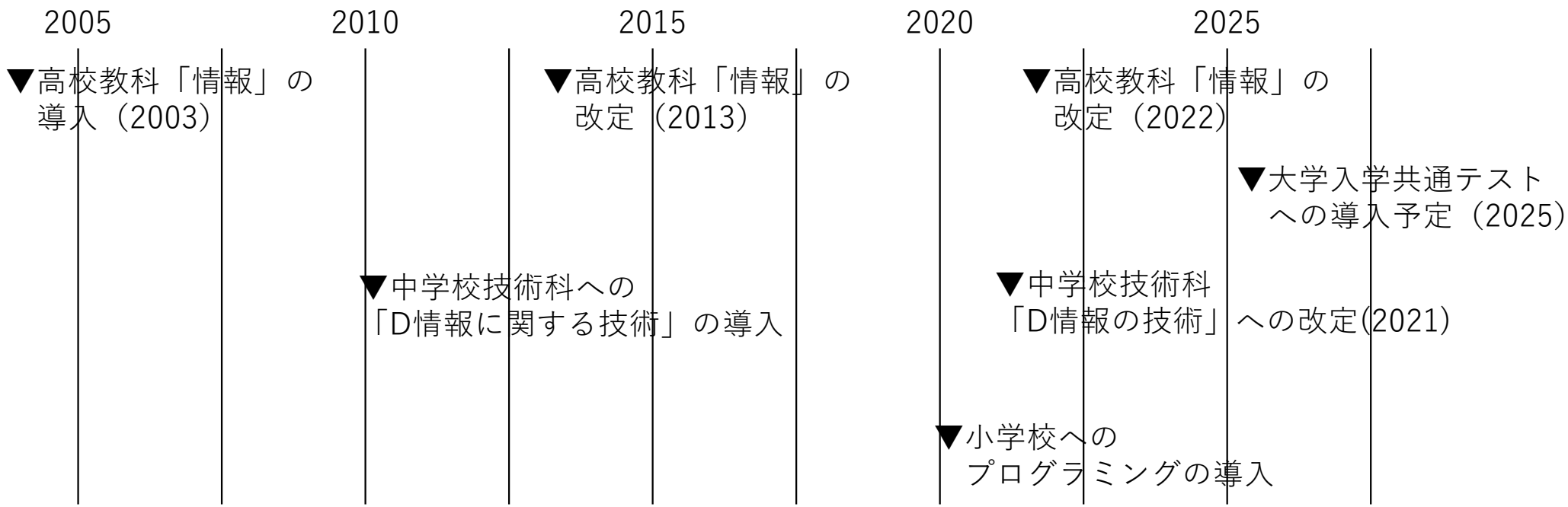
植原啓介 (kei@sfc.keio.ac.jp)

プロジェクト概要

- 背景: 情報分野においては、教育体系や学習効果の評価手法が十分に確立しているとは言い難い。円滑な高大接続のためには高校・大学双方の共通認識が必要である。
- 目的: 初等中等教育における情報分野の学習効果の評価手法の確立

- 実施枠組: 科学研究費助成事業（科研費） 基盤研究A
- 研究課題名:大学入試を中心とした情報分野の学力評価手法の検討
- 課題番号: 23H00068
- 実施期間: 2023年度～2027年度（5年間）
- 研究者: 植原啓介（慶應義塾大学 環境情報学部 教授）、他11名

背景



大学入学共通テストを始めとして、大学入試に教科「情報」の導入を検討している大学が増えているが、評価方法の共通認識に乏しい

想定される成果

- 作問マニュアル
  （典型的な問い、IRT（項目応答理論）を想定した多肢選択問題）

- CBTシステム（TAO用のPCIモジュールを想定）

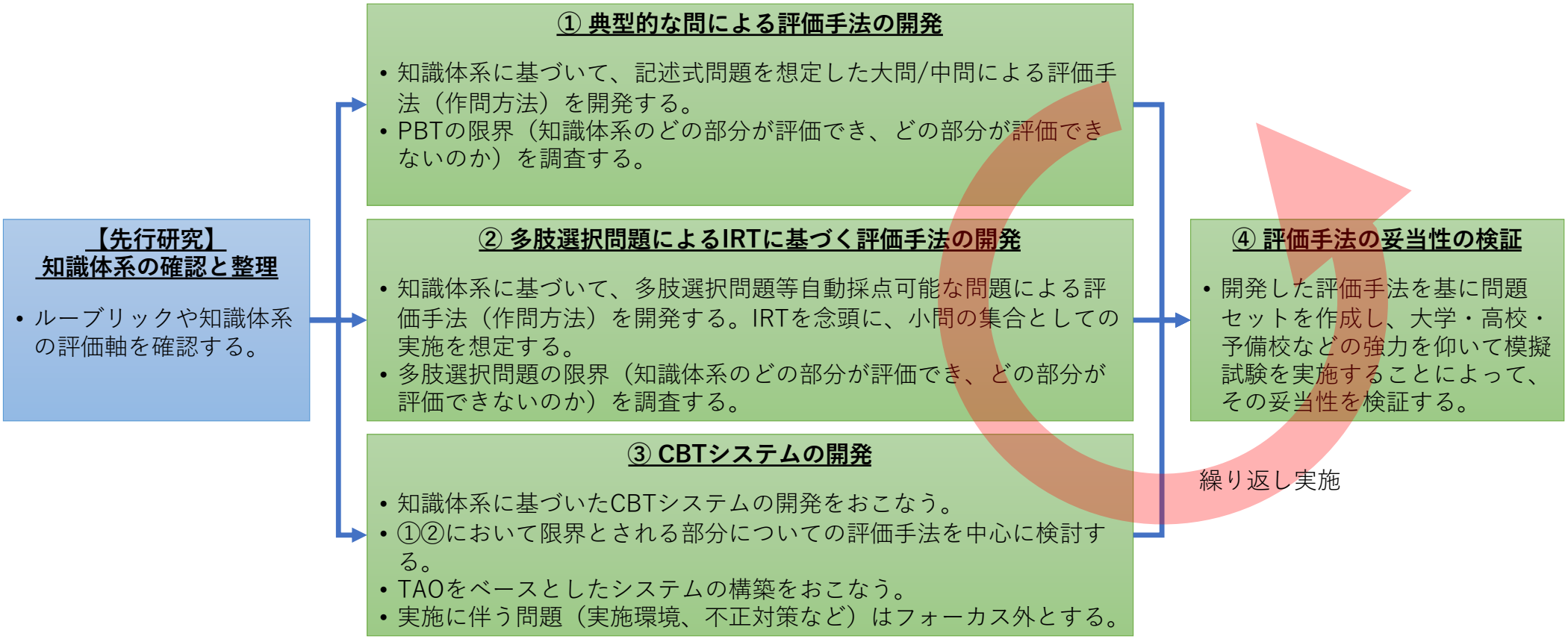
- 出題形態毎に、知識体系のどのような項目が評価可能で、どのような限界があるのかを示したもの

- 出題形態ごとのベストプラクティス（模試における知見の共有）

本研究において対象とする出題形態

- 従来の一般的な問題
  - 大問・中問による学力評価。長文読解、プログラミング、データ分析などを含む。
- IRTを想定した多肢選択問題
  - 4択問題のようなものを大量に出題することによる学力評価。IRT（項目応答理論）に基づき学力を評価することを想定する。
- CBTを前提とした問題
  - 実行環境付きのプログラミング問題、大量のデータを使ったデータ分析問題など、PBTでの実施が難しい出題方法による学力評価。

研究の流れ

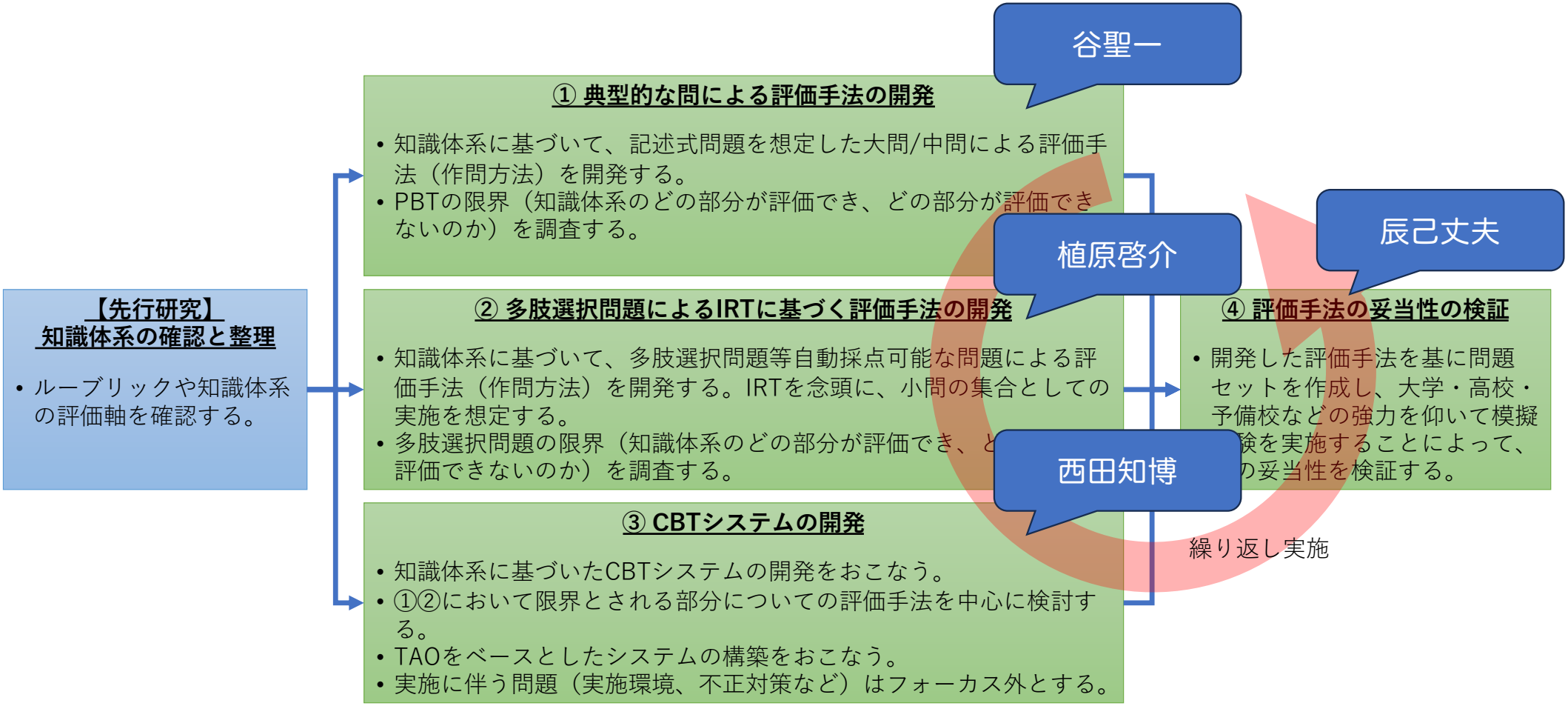

【先行研究】
知識体系の確認と整理
• ルーブリックや知識体系の評価軸を確認する。
① 典型的な問による評価手法の開発
• 知識体系に基づいて、記述式問題を想定した大問/中問による評価手法（作問方法）を開発する。
• PBTの限界（知識体系のどの部分が評価でき、どの部分が評価できないのか）を調査する。
② 多肢選択問題によるIRTに基づく評価手法の開発
• 知識体系に基づいて、多肢選択問題等自動採点可能な問題による評価手法（作問方法）を開発する。IRTを念頭に、小問の集合としての実施を想定する。
• 多肢選択問題の限界（知識体系のどの部分が評価でき、どの部分が評価できないのか）を調査する。
③ CBTシステムの開発
• 知識体系に基づいたCBTシステムの開発をおこなう。
• ①②において限界とされる部分についての評価手法を中心に検討する。
• TAOをベースとしたシステムの構築をおこなう。
• 実施に伴う問題（実施環境、不正対策など）はフォーカス外とする。
④ 評価手法の妥当性の検証
• 開発した評価手法を基に問題セットを作成し、大学・高校・予備校などの強力を仰いで模擬試験を実施することによって、その妥当性を検証する。
繰り返し実施

知識体系の確認と整理

- 学術会議: 5分野、11分野
- 大学入学者選抜改革推進委託事業（阪大他）: 12分野
- 学習指導要領: 4分野
  - それぞれの分野が3つに別れているので12分野とみることも可能
- ITSS: 13分野
- 「ITパスポート試験」シラバス: 大分類9分野、中分類23分野

- 高大連携のコンテクストでは、高校の指導要領にリンクすることは必須

研究の流れ


谷聖一
植原啓介
辰己丈夫
西田知博
【先行研究】
知識体系の確認と整理
• ルーブリックや知識体系の評価軸を確認する。
① 典型的な問による評価手法の開発
• 知識体系に基づいて、記述式問題を想定した大問/中問による評価手法（作問方法）を開発する。
• PBTの限界（知識体系のどの部分が評価でき、どの部分が評価できないのか）を調査する。
② 多肢選択問題によるIRTに基づく評価手法の開発
• 知識体系に基づいて、多肢選択問題等自動採点可能な問題による評価手法（作問方法）を開発する。IRTを念頭に、小問の集合としての実施を想定する。
• 多肢選択問題の限界（知識体系のどの部分が評価でき、と
評価できないのか）を調査する。
③ CBTシステムの開発
• 知識体系に基づいたCBTシステムの開発をおこなう。
• ①②において限界とされる部分についての評価手法を中心に検討する。
• TAOをベースとしたシステムの構築をおこなう。
• 実施に伴う問題（実施環境、不正対策など）はフォーカス外とする。
④ 評価手法の妥当性の検証
• 開発した評価手法を基に問題セットを作成し、大学・高校・予備校などの強力を仰いで模擬験を実施することによって、の妥当性を検証する。
繰り返し実施