

ワークショップ 項目反応理論を用いたテスト運用 の実際

加藤 健太郎(株式会社ベネッセコーポレーション)

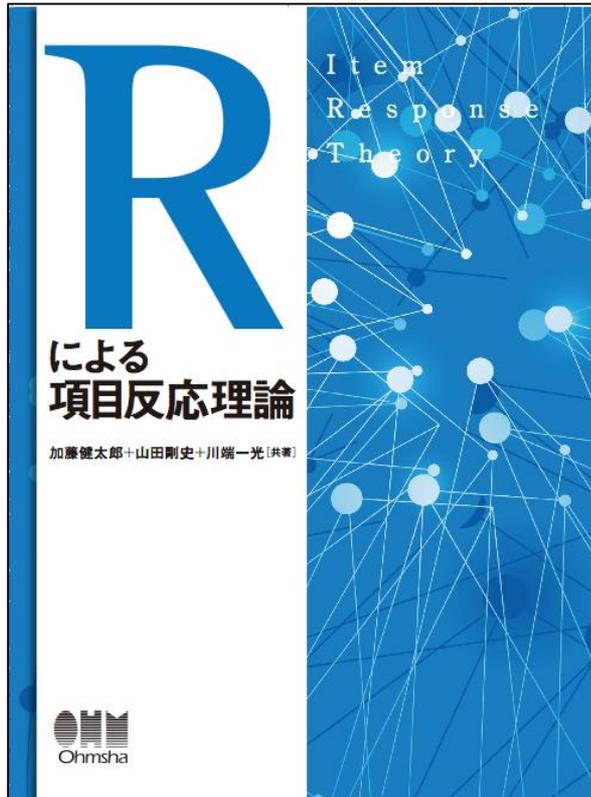
山田 剛史(岡山大学大学院教育学研究科)

川端 一光(明治学院大学心理学部)

本ワークショップの概要

- 近年，我が国の教育評価・測定領域で，項目反応理論(IRT)への注目が高まっている
- IRTは初学者にはなかなか取っつきにくく，理解しにくいという声も
- 加藤他(2014)は，Rを利用しながら，IRTの理論と実践を分かりやすく学ぶためのテキストを執筆
- 加藤他(2014)の執筆の背景から，項目プールを核にしたテスト運用とその留意点，IRT分析と等化の実際。テスト構成の実際，について解説する

Rによる項目反応理論



加藤健太郎・山田剛史・川端一光
Rによる項目反応理論
オーム社
2014年7月26日発売

本ワークショップのタイムテーブル

- 導入: 「Rによる項目反応理論」執筆の背景, IRTによるテスト運用のプロセス(山田, 15分)
- 等化: 項目プールへのinputとして。IRTに基づく等化の実際。IRT分析から等化へ(川端, 30分)
- テスト構成: 項目プールからのoutputとして。IRTに基づくテスト構成の実際(複数テストの同時構成)(加藤, 30分)
- 質疑応答(15分)

執筆の背景1:学習ツールとしてのR

- R は, オープンソースの統計ソフトウェア
- 誰でも無料で利用することができ, さまざまな分析ができる
- 加藤他(2014)では, Rを用いてIRT や関連する分析を実行する実践に役立つ方法を紹介
- さらに, RをIRTに関するさまざまな知識を習得するための学習ツールとして活用しているところがポイント

執筆の背景2: 数式をRで読む

- イメージしにくい数式の意味や中身について, Rを使って実際の数値を入れて計算したり, グラフ化したりすることで, より具体的に理解できる
- 例:「尤度関数を用いて, ある特定の項目反応パターンを取った受験者の能力を推定する」
- Rスクリプトを用いることで, 「尤度が最大となる θ を能力の推定値とする」という最尤推定の基本的な考え方を視覚的に理解しやすく

執筆の背景3:理論の学習から実践へのシームレスな移行

- ソフトウェアを用いる→数式や概念についての理解が促進
- これ自体は, Rでなくても, 他のソフト, Excel等でも可能
- Rを用いることで, 基本的な概念の理解から, 更に応用の段階へと無理なくステップアップができる→大きなメリット
- 例: Newton-Raphson 法により, 能力値 θ の最尤推定値を求める
- まず, 概念の理解のため, スクリプトを手作業で反復する→対数尤度の値の変動が徐々に小さくなり, 特定の値に収束していく様子を確認できる
- 基礎的な概念が理解できたら→同一の処理(計算)を, スクリプトで記述することで作業(反復計算)をRに実行させる→プログラミングの意義の理解と基礎技術の習得へと学びを深める

執筆の背景4:理論と実践の橋渡し

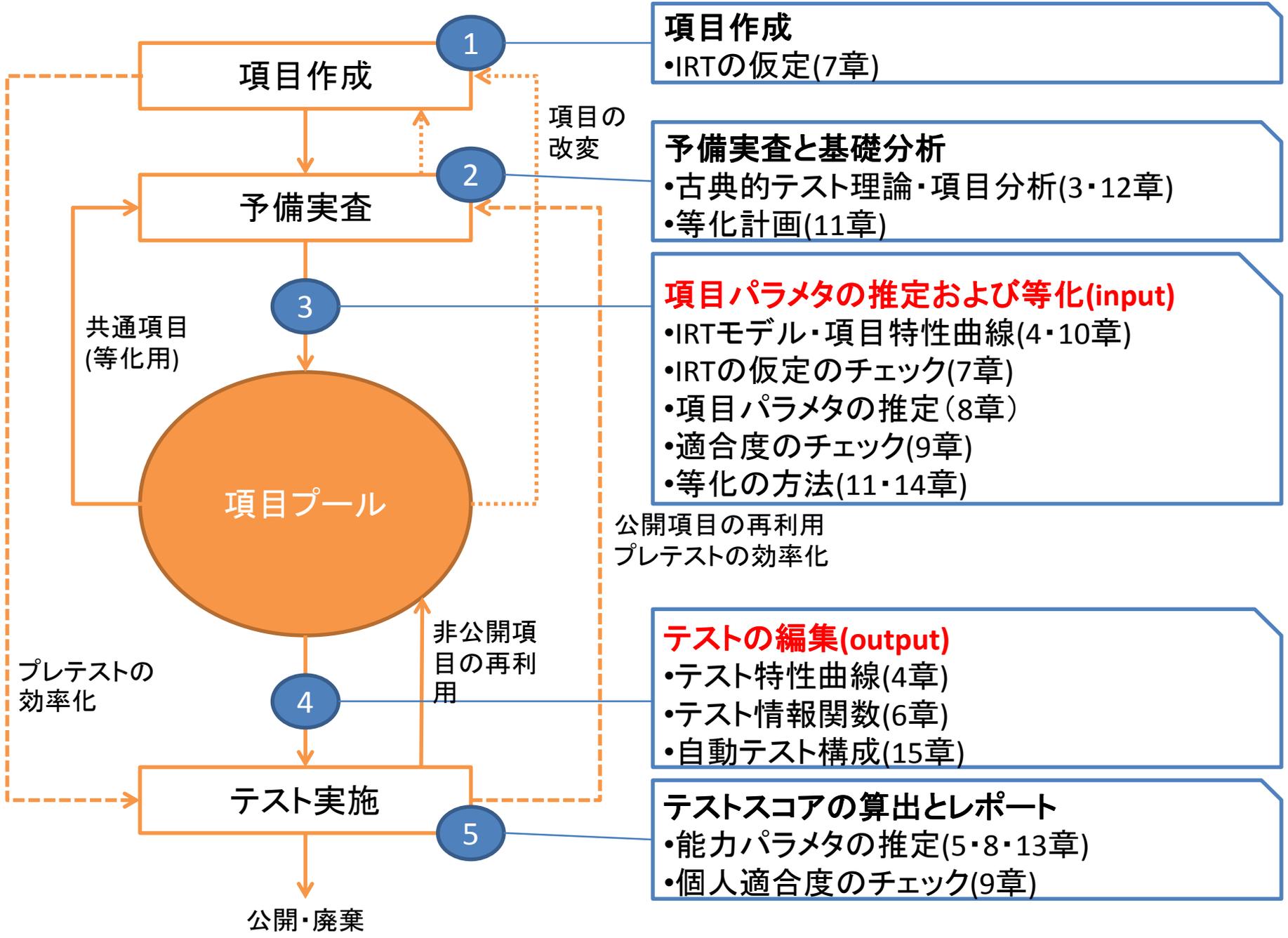
- 現実のテストデータ分析場面においては、「定型的な処理を自動化することで人為的なミスを減らす」ことが基本的な方略に
- 加藤他(2014)では、以下の情報を提供している
 - 定型処理の自動化について
 - irtoysパッケージを用いた項目分析・IRT分析の具体的な方法
 - 実際のテストデータの分析(実務)ですぐ利用できる実践的な内容
- 基本的な理論の理解と実践的な内容の両者を含むことを、Rの活用により実現

項目プールの活用を中心に据えた テスト運用

- IRT を用いたテストの開発・運営の中心→項目プールをいかに構築・維持するか
- 同じ尺度(共通尺度)上に等化された項目が項目プールに揃っていることが大前提
- 長期にわたる安定したテスト運営→項目プールに恒常的に新項目を追加していくことが必要
- 項目作成→IRT の仮定を満たすことを目指して項目を作成
- 予備実査と基礎分析→作成された項目の性能を吟味

事前の等化計画の重要性

- 等化計画(equating design): 等化を実現するためのテスト実施に関する計画。詳細は, 加藤他(2014)の11章を参照のこと
 1. 共通受験者計画
 2. 共通項目計画
 3. 係留テスト計画
- 等化の条件を満たしつつ, 実施・作問・測定における様々なコスト・リスクを勘案し, 実現可能な等化計画を設定する
 - IRTを利用したテスト運営では, 項目プールのメンテナンスに莫大なコストがかかる
 - テスト運営開始時に詳細な等化計画を立てることで, 結果的にコストを最小化できる



項目作成
 •IRTの仮定(7章)

予備実査と基礎分析
 •古典的テスト理論・項目分析(3・12章)
 •等化計画(11章)

項目パラメタの推定および等化(input)
 •IRTモデル・項目特性曲線(4・10章)
 •IRTの仮定のチェック(7章)
 •項目パラメタの推定(8章)
 •適合度のチェック(9章)
 •等化の方法(11・14章)

テストの編集(output)
 •テスト特性曲線(4章)
 •テスト情報関数(6章)
 •自動テスト構成(15章)

テストスコアの算出とレポート
 •能力パラメタの推定(5・8・13章)
 •個人適合度のチェック(9章)

項目の
 改変

共通項目
 (等化用)

公開項目の再利用
 プレテストの効率化

非公開項
 目の再利
 用

プレテストの
 効率化

公開・廃棄

テスト運用計画の重要性

- 等化計画だけではなく、それも含む「全体的なテスト運用計画」をしっかりと建てておくことが肝要
- 項目プールへのinput(パラメタ推定, 等化)と項目プールからのoutput(テスト編集)のバランスを考慮する
 - input < outputだと, いずれ項目プールは枯渇してしまう
 - 実際の運用では, 大規模な(あるいは継続的な)フィールドテストの実施が難しかったり, コストの制約があったりする
 - 現実の制約の中でどうやってinputを増やしていくか, あるいは, outputを無駄遣いしないか

テスト運用計画の重要性

- 大量のテスト項目を扱うための工夫
 - 重複テスト分冊法: PISAやNAEPといった国際的な大規模学力調査で用いられる
 - 多くの項目をいくつかのブロックに分け, 異なるブロックを組み合わせて複数のテストフォーム(分冊)を作る
- 等化, テスト構成についての文献
 - Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. 2nd ed. New York, NY: Springer.
 - 柴山直・佐藤喜一・熊谷龍一(2012). 全国規模の学力調査における重複テスト分冊法の展開可能性について(平成23年度文部科学省委託研究「学力調査を活用した専門的課題分析に関する調査研究」研究成果報告書) 国立大学法人東北大学.
 - van der Linden, W. J.(2005). *Linear models for optimal test design*. New York, NY: Springer.

柴山直先生の推薦のことば



- Rに関する高度なテクニックを学びながら、次世代の心理・教育テストを支える項目反応理論/IRTの最新の技術体系を深く理解し、実践的な知識を的確に修得できる、実に行き届いたテキストです。さらには、テストの開発から運用まで、実際に直面するさまざまな課題を解決するための数多くのヒントもこのテキストから得られることでしょう。
- [fpr 3669] 書評：Rによる項目反応理論：加藤・山田・川端：オーム社