

G2

多肢選択問題によるIRTに基づく 評価手法の構築

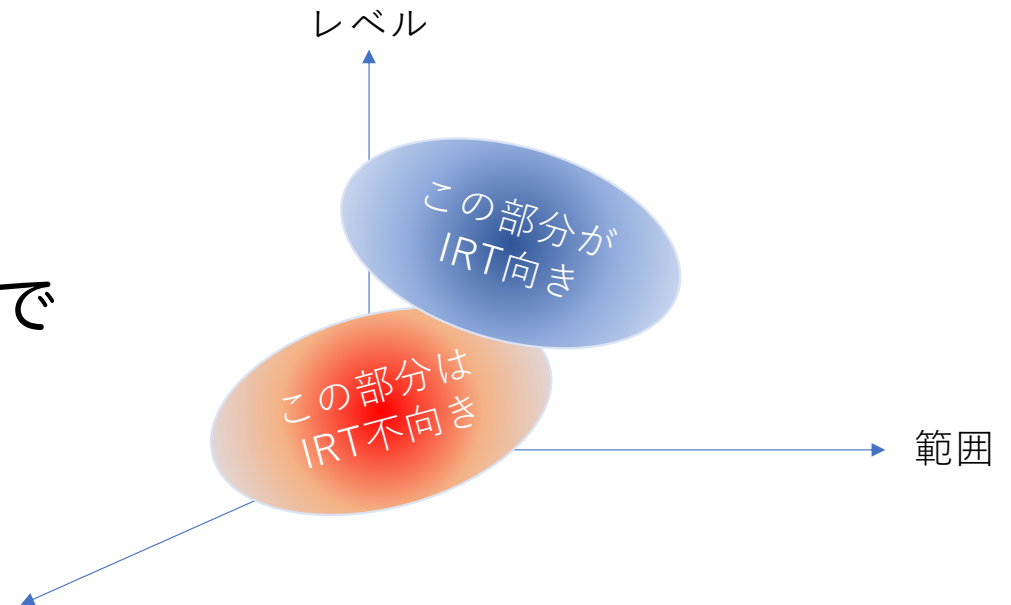
慶應義塾大学 環境情報学部
植原啓介 (kei@sfc.keio.ac.jp)

概要

- 目的

- 多肢選択問題等の自動採点可能な問題による評価手法を開発する
- 作問のBest Current Practiceとしての手順書を作成する

- 多肢選択問題でどれくらいの
範囲（分野）、レベル（深さ）、
思考力・判断力・表現力・応用力まで
確認できるかを明らかにする。



思考力・判断力・表現力・応用力

大学入試を中心とした情報分野の学力評価手法の検討シンポジウム2025

多肢選択問題の作問方法についての検討状況

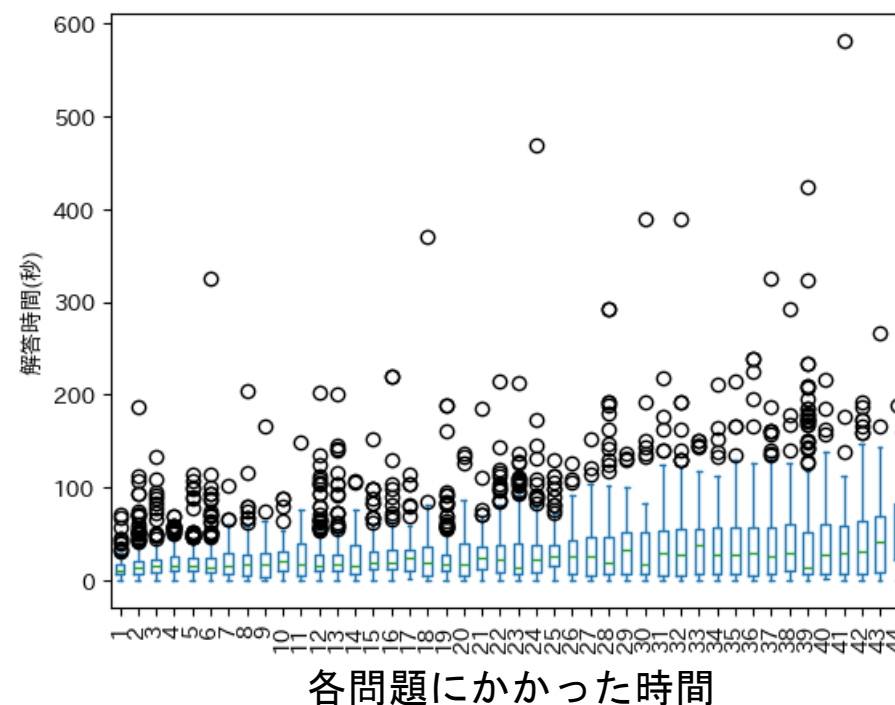
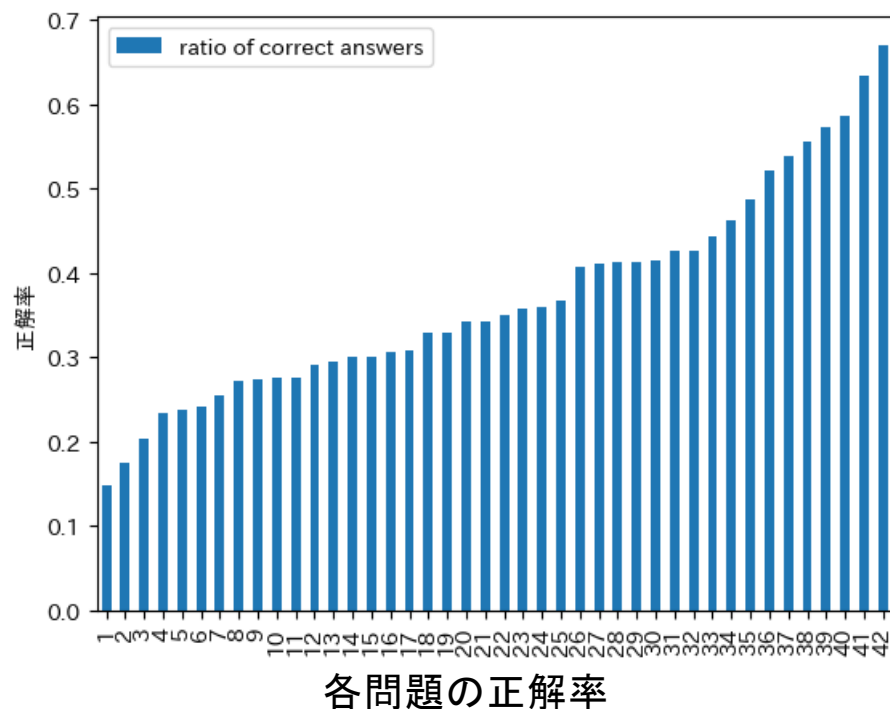
情報に関連する語句のリストがあるとした場合、下記のような作問の方法がある。

1. 説明文を読んで適切な句を選ばせる。（語句を知っている）
同義語（あるいは反対語や非同義語）を選ばせる。
2. 語句の説明として正しいものを選ばせる。（語句を説明できる）
語句があるクラスを表している場合はそのインスタンス（あるいは間違っ
たインスタンス）を答えさせる。
3. 語句の説明文の空欄を埋めさせる。（語句を説明できる）
4. 基数変換などの操作をさせる。（語句の内容を使える）
5. 知識を使って美しく解答を導く。（語句の内容を応用できる）

現在、上記の方法に基づいて作問を進めている状況。

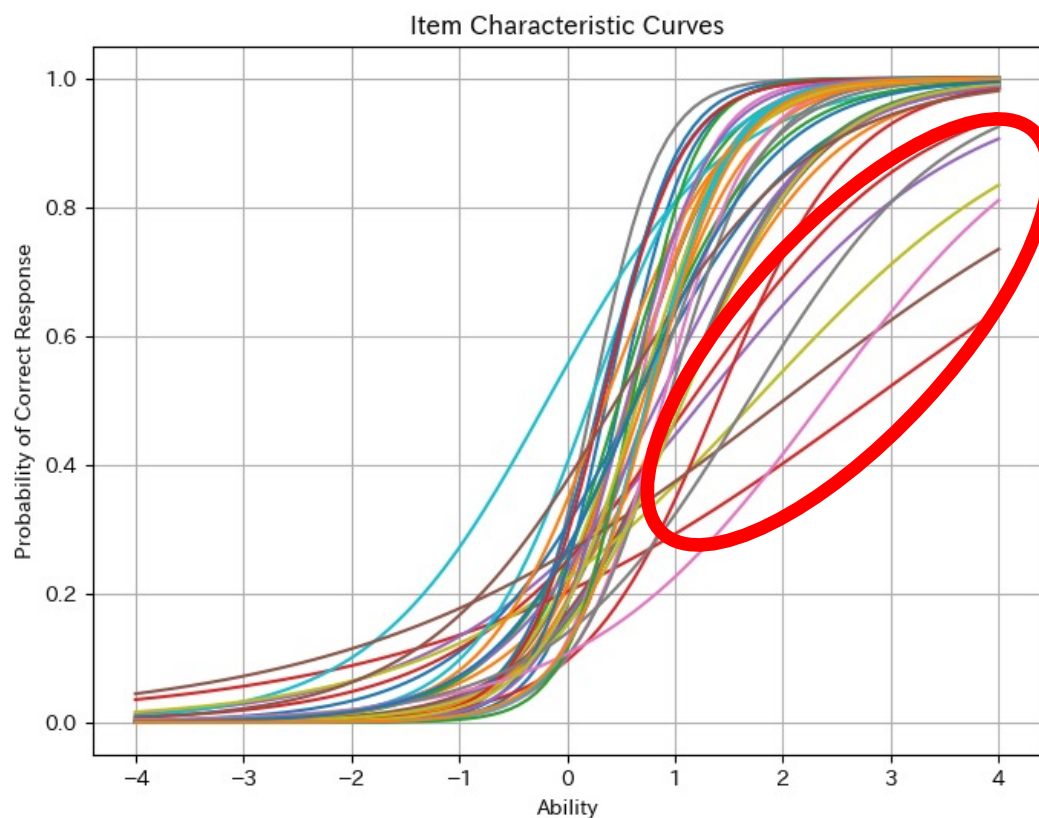
IRTを想定した多肢選択問題の結果(2025春)

- 簡単な問題から難しい問題まで適切に出題できていたと考えられる
- 1問1分想定であったが、十分な時間が与えられたと考えられる



IRTを想定した多肢選択問題の結果(2025春)

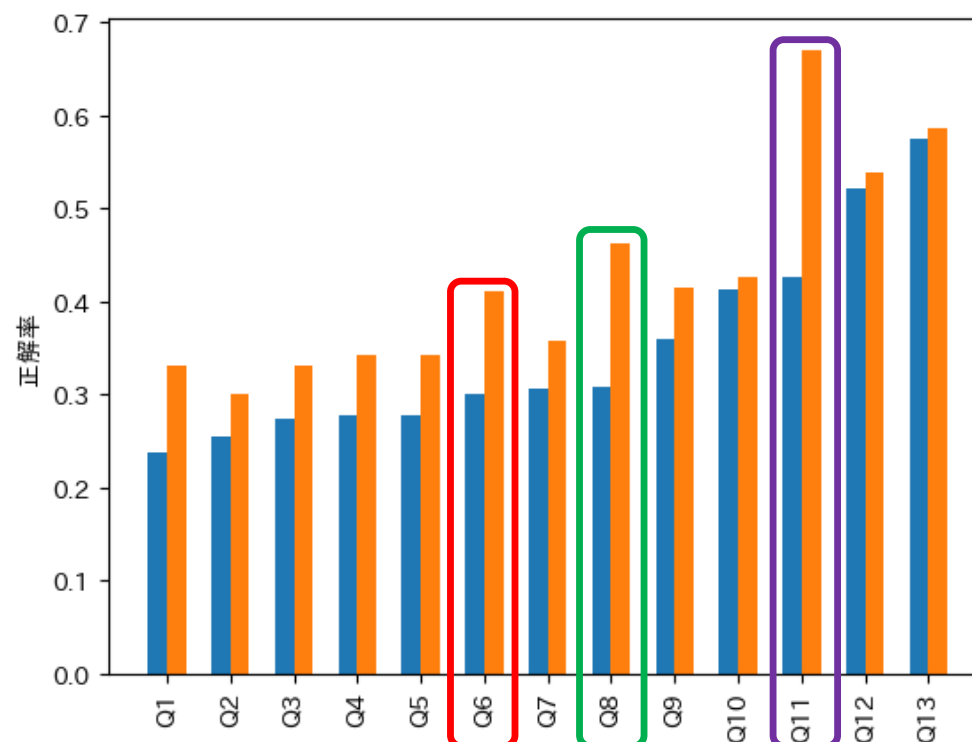
- 一部難しすぎる問題○も見られるが、概ね弁別性のある適切な問題が出題できている



※2PLで分析

IRTを想定した多肢選択問題の結果(2025春) (枝番問題)

- **Q6**は、繰り返しで入力された値の平均を求める問題と、合計を求める問題であり、前者の方が正解率が悪い。前者は繰り返しの中でいわゆる足し込み操作を行って合計を得た後、繰り返しの後の合計値を個数で割る操作が空欄となっている。後者はこの足し込み操作が空欄となっている。このことから、繰り返しで足し込みをおこなうことについてはプログラミングで学んでいるが、その後の変数の扱いが理解できていないことが想定される。
- **Q8**は、2つの変数に入っている値のうち、より大きな値をセットする問題と、より小さな値をセットする問題であり、前者の方が正答率が悪い。この問題は変数に入っている値が等しい場合はその値をセットすることとなっているが、その部分で間違った受験生が多いと考えられる。特に枝番間でアルゴリズムの複雑さや問題構造に大きな違いはないと考えられ、なぜ大きな差が生じたかについては、より詳細な分析が必要である。
- **Q11**は、単語に対して説明文を選ぶ問題と、説明文に対して用語を選ぶ問題であり、前者の方が正答率が悪い。この問題は、枝番間で特に大きな差が見られている。もともと、文献(植原、2024)で筆者らが述べているように、作問にあたっては単語に対して説明文を選ぶ問題より説明文に対して用語を選ぶ問題の方が難易度が低いという仮説を立てていた。Q11の結果は、この仮説をサポートするものとなる。



IRT問題の例（2024年秋）

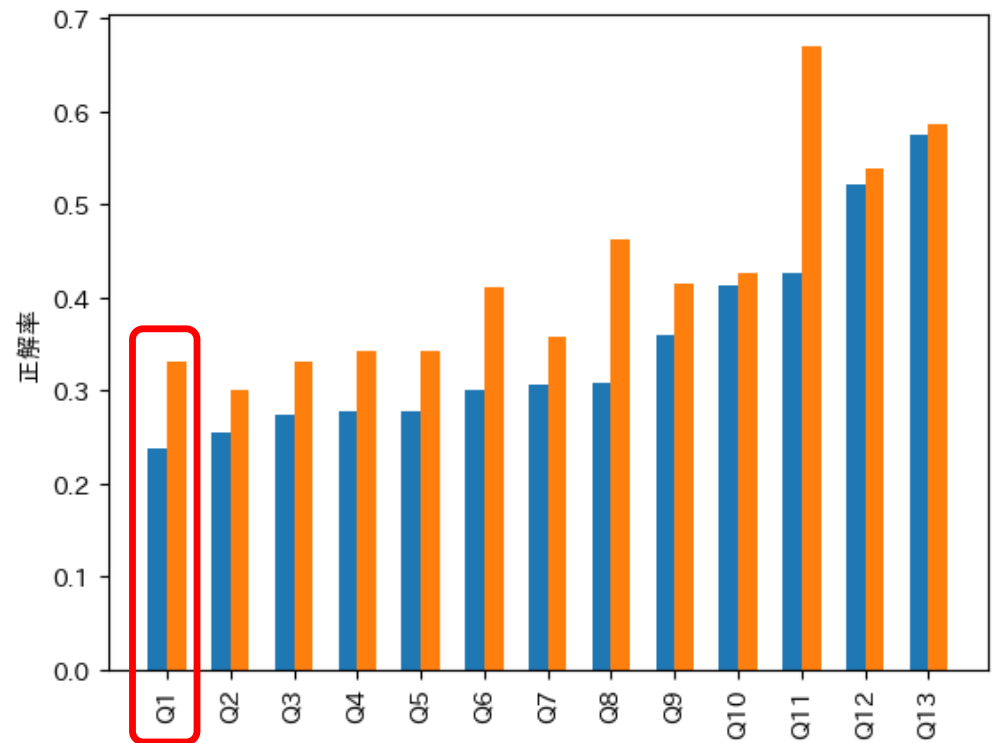
変数 a, b, c の値のなかで、最も小さな値を d にセットする処理（記述）はどれか。

- | | | | |
|--------------|-------------------|-----------------|-------------------|
| 1. $d = a$ | 2. $d = a$ | 3. もし $a < b$: | 4. もし $a < b$: |
| もし $d > b$: | もし $d > b$: | └ $d = a$ | └ $d = a$ |
| └ $d = b$ | └ $d = b$ | もし $b < c$: | そうでなくもし $b < c$: |
| もし $d > c$: | そうでなくもし $d > c$: | └ $d = b$ | └ $d = b$ |
| └ $d = c$ | └ $d = c$ | $d = c$ | そうでなくもし $c < a$: |
| | | | └ $d = c$ |

- 最も正答率が低かった問題（正答率20.3%）
- それぞれの選択率は1. 20.7%、2. 26.8%、3. 13.8%、4. 38.7%であった。選択肢4のプログラムが最も強く誤答に誘導したが、その理由は明らかでない。
- 変数の上書きが行われる問題の正答率は低くなる傾向にある。（手元に紙が無い?）
- 選択肢1において比較演算子が「大なり」を使っているが、「最も小さな値」という問題文との不整合が影響を及ぼした可能性もある。

IRTを想定した多肢選択問題の結果(2025春) (枝番問題)

- **Q1** 正解については、不等号の左側に大きな数があるか小さな数があるかの違い。大きな数を求めよ、という問題については左側に大きな数があったほうが自然で、正解率も10ポイント程度高くなる。



今後について

- 今回得られた知見をもとに、例えば正答率が低かった問題はなぜ正答率が低かったのか、といった研究を進める。
- 他の分野の作問を継続する。