An R Package for Adaptive Assessment Utilizing Knowledge Space Theory and Formal Psychological Assessment

Andrea Brancaccio & Umberto Granziol

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche

September, 19 2024

Outline

1. Introduction
2. mycaas Package
   - Algorithm
   - Shiny
   - Application on RAISE
3. Final Remarks

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Tests in Education and Clinical Psychology

- Time consuming
- Fatigue effect, social desirability, etc.

(Informal) Definition

A computerized adaptive assessment is an evaluation that adjusts the difficulty and nature of subsequent questions based on the test-taker's responses to previous ones.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Tests in Education and Clinical Psychology

- Time consuming
- Fatigue effect, social desirability, etc.

(Informal) Definition

A computerized adaptive assessment is an evaluation that adjusts the difficulty and nature of subsequent questions based on the test-taker's responses to previous ones.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Pros of Adaptive Assessment

- Increased Efficiency and Accuracy in Assessment:
  - Adaptive systems save time by focusing on the appropriate difficulty level.
- Personalized Learning or Therapy :
  - Feedback can be customized to individuals need.
- Immediate Feedback:
  - Results are available as soon as the assessment is finished.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Cons of Adaptive Assessment

- Dependent on the assumptions of the Model :
  - The validity of results depends on the correctness of the model used.
- Complexity of Implementation:
  - Requires sophisticated algorithms and data processing infrastructure.
- Difficulty in Tuning:
  - Fine-tuning the assessment to achieve accurate difficulty adjustments is challenging.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

My computerized adaptive assessment R package

`mycaas package`

```
devtools::install_github(''brancaccioandrea/mycaas")
```

- Based on the strong theoretical foundation
- User-friendly graphical interface made
- Performance analisys to evaluate accuracy and efficiency

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Theoretical Framework

Knowledge space theory (KST; Doignon & Falmagne, 1985): The objective is to precisely describe what the individual knows (their knowledge state) in a given domain of knowledge, rather than computing a numerical score.

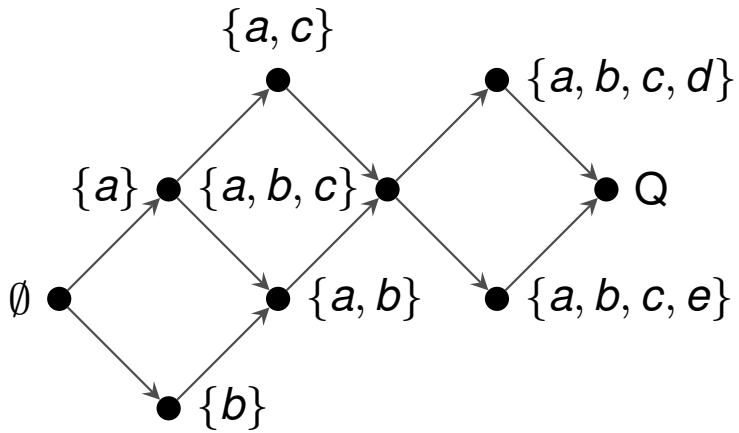
Formal Psychological Assessment (FPA; Spoto, Stefanutti & Vidotto, 2010): The objective from the clinical perspective is to give an in-depth evaluation of the construct investigated by the questionnaire.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Basic Definitions

Formal definitions of basic concepts encountered so far:

- Domain a either finite or infinite set Q of questions
- State the subset $K \subseteq Q$ of all questions that define the status of individual
- Structure a pair $(Q, \mathcal{K})$, where $\mathcal{K}$ is a collection of subset of Q, containing at least the empty set and Q



Outline

1. Introduction
2. mycaas Package
   - Algorithm
   - Shiny
   - Application on RAISE
3. Final Remarks

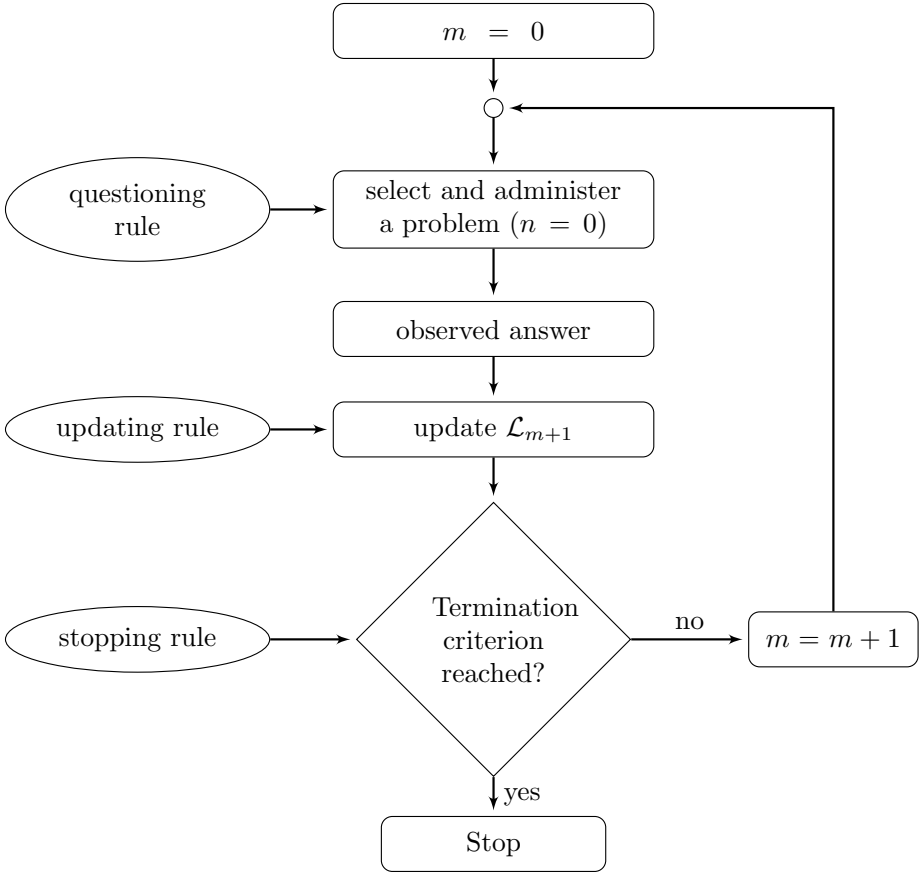
Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Flowchart of the Adaptive Assessment

Doignon & Falmagne, 1988; Donadello, et al., 2017

The goal of the assessment is to recover the true state of an individual asking the fewest possible questions

- Three rules guide the assessment:
  1. Questioning rule
  2. Updating rule
  3. Termination rule



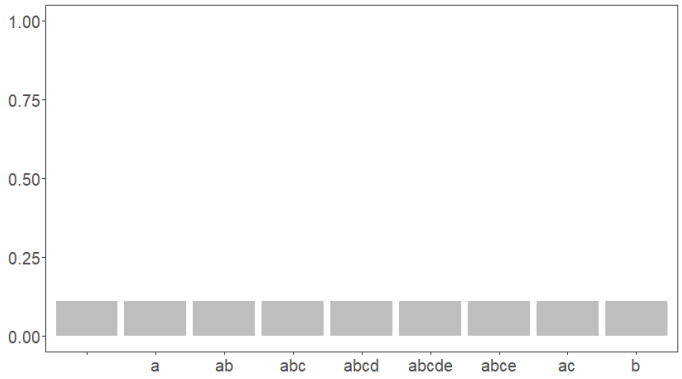
Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

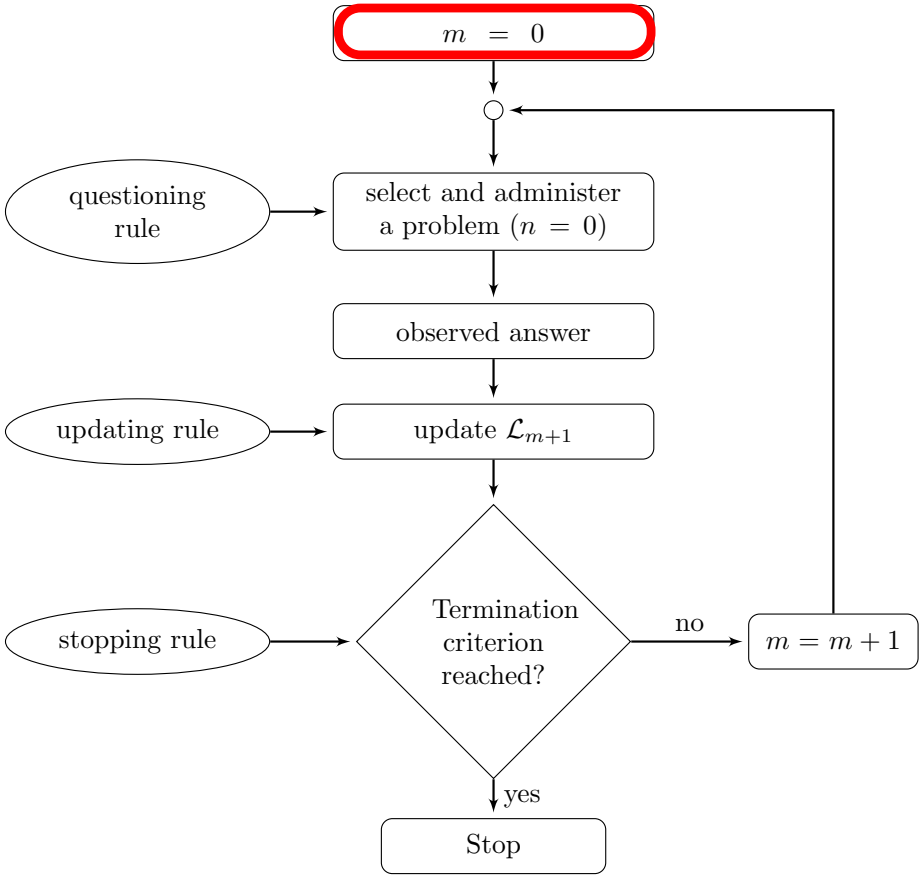
Probability distribution on the states

A probability distribution

$$\mathcal{L}_m : \mathcal{K} \to (0, 1)$$

Without prior knowledge is a uniform distribution



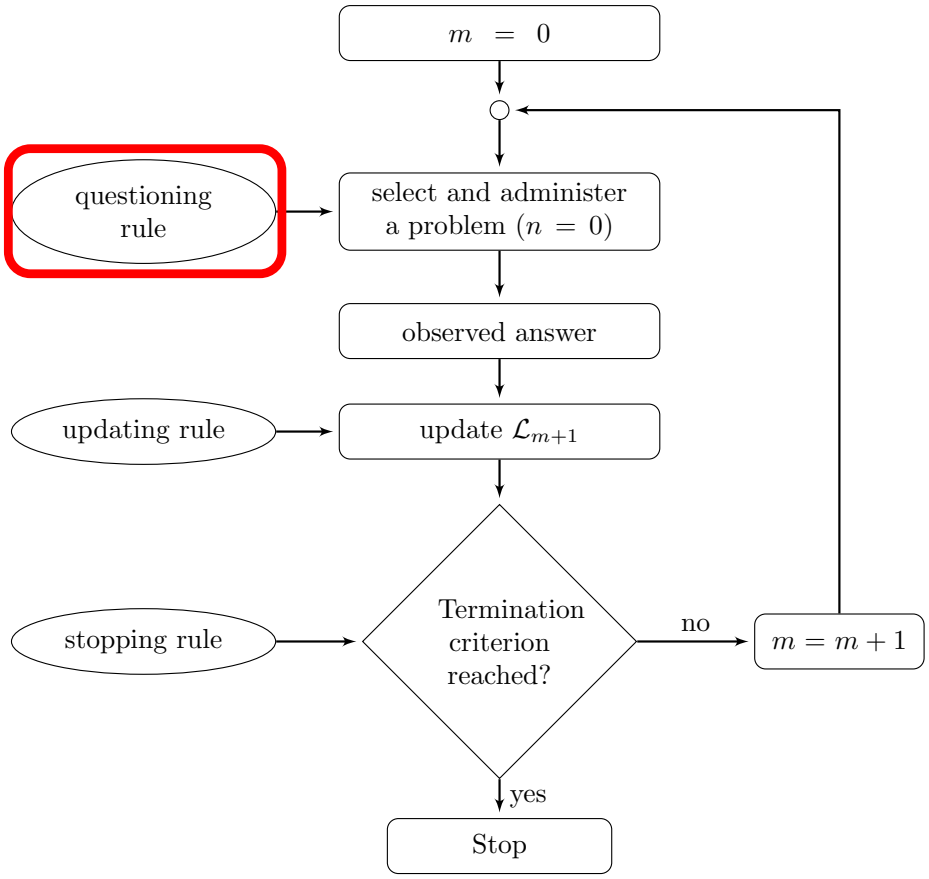


Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Questioning Rule

Half Split Rule

Select the "best" question to ask



Doignon & Falmagne, 1988

`half_split`: Let $\mathcal{L}_m(K)$ the probability of K at the step m, and the subset $\mathcal{K}_q \subset \mathcal{K}$ such that $q \in K$ for each $K \in K_q$.
It selected problem $q \in Q$ that minimize

$$|\mathcal{L}_m(\mathcal{K}_q) - 1/2|$$

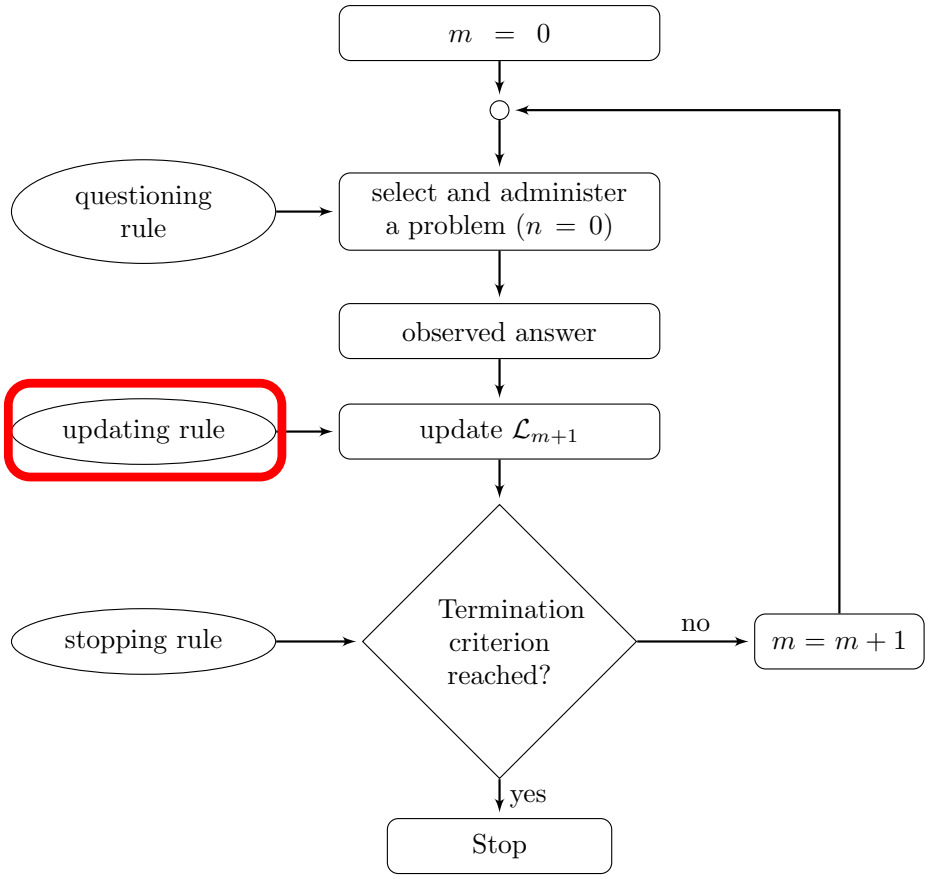
Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Updating Rule

Bayesian Updating Rule

The probability $\mathcal{L}_m(K)$, for each $K \in \mathcal{K}$ is updated in function of the the observed response r_q collected for problem q as follows.

$$\mathcal{L}_{m+1}(K) = \frac{P(r_q|K)\mathcal{L}_m(K)}{\sum_{K' \in \mathcal{K}} P(r_q|K)\mathcal{L}_m(K')}$$



Parameters

$$P(r_q|K) = \begin{cases} \beta_q & \text{if } r_q = 0 \ \& \ q \in K; \\ 1 - \eta_q & \text{if } r_q = 0 \ \& \ q \notin K; \\ 1 - \beta_q & \text{if } r_q = 1 \ \& \ q \in K; \\ \eta_q & \text{if } r_q = 1 \ \& \ q \notin K. \end{cases}$$

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Termination rule

Heller & Repitsch, 2012

`Likelihood Maximization:`
The assessment terminate at step m if

$$\max \mathcal{L}_m(K) > SC$$

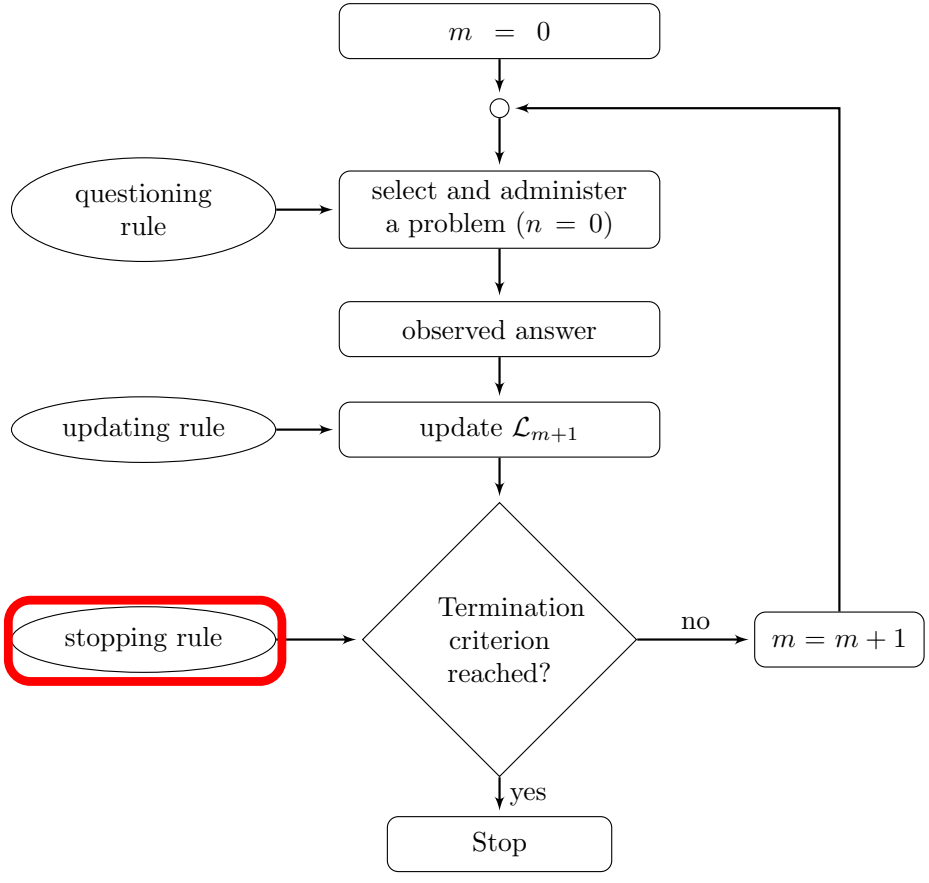
with $SC \in (0.5, 1]$.

Donadello, et al., 2017

`Item Discrimination:` The assessment terminate at step m if for each $q \in Q$

$$\mathcal{L}_m(\mathcal{K}_q) > SC \text{ or } \mathcal{L}_m(\mathcal{K}_q) < 1 - SC$$

with $SC \in [0.5, 1]$.



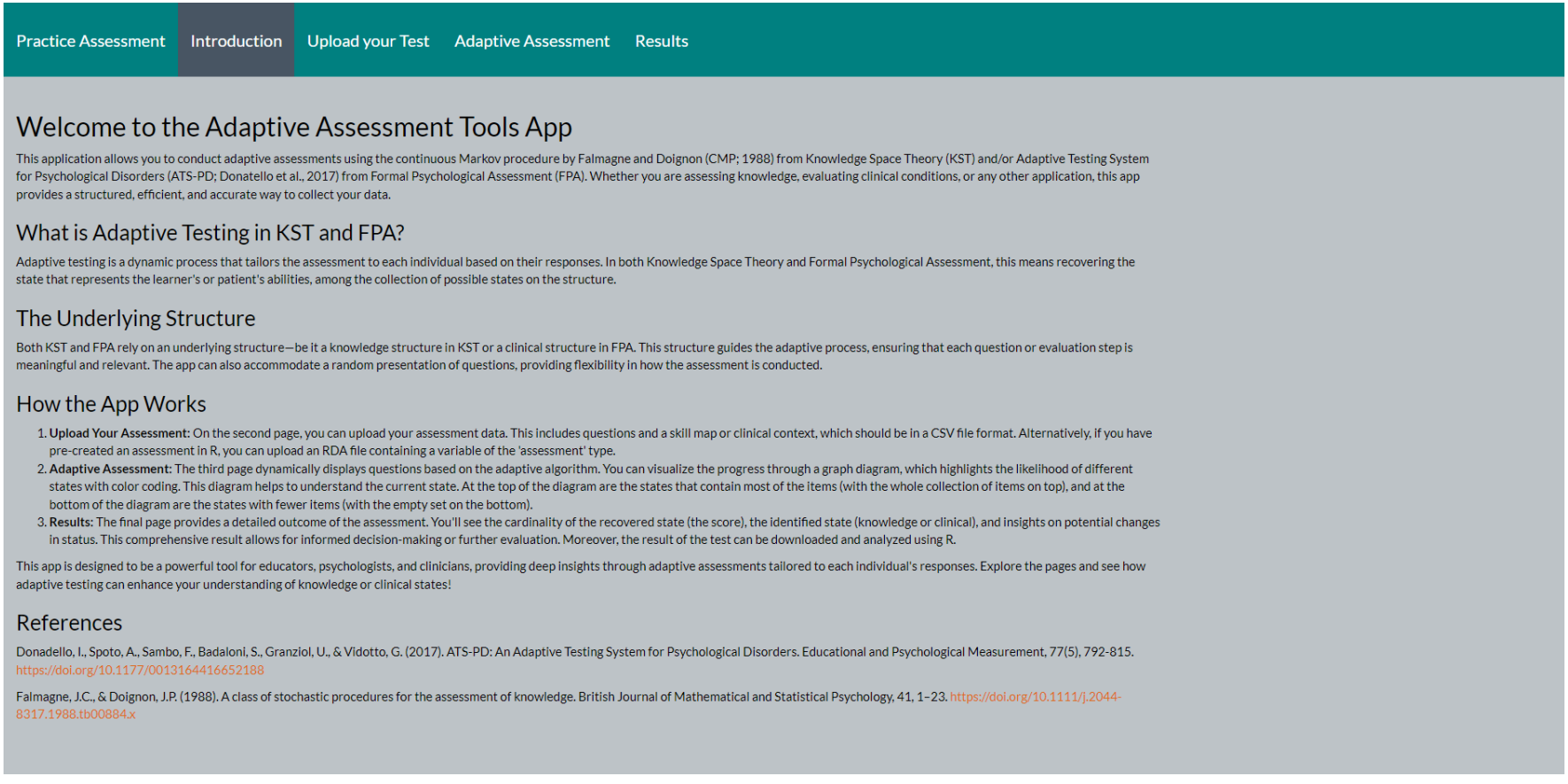
Outline

1. Introduction
2. mycaas Package
   - Algorithm
   - Shiny
   - Application on RAISE
3. Final Remarks

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE
Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Create and Try the Test ...

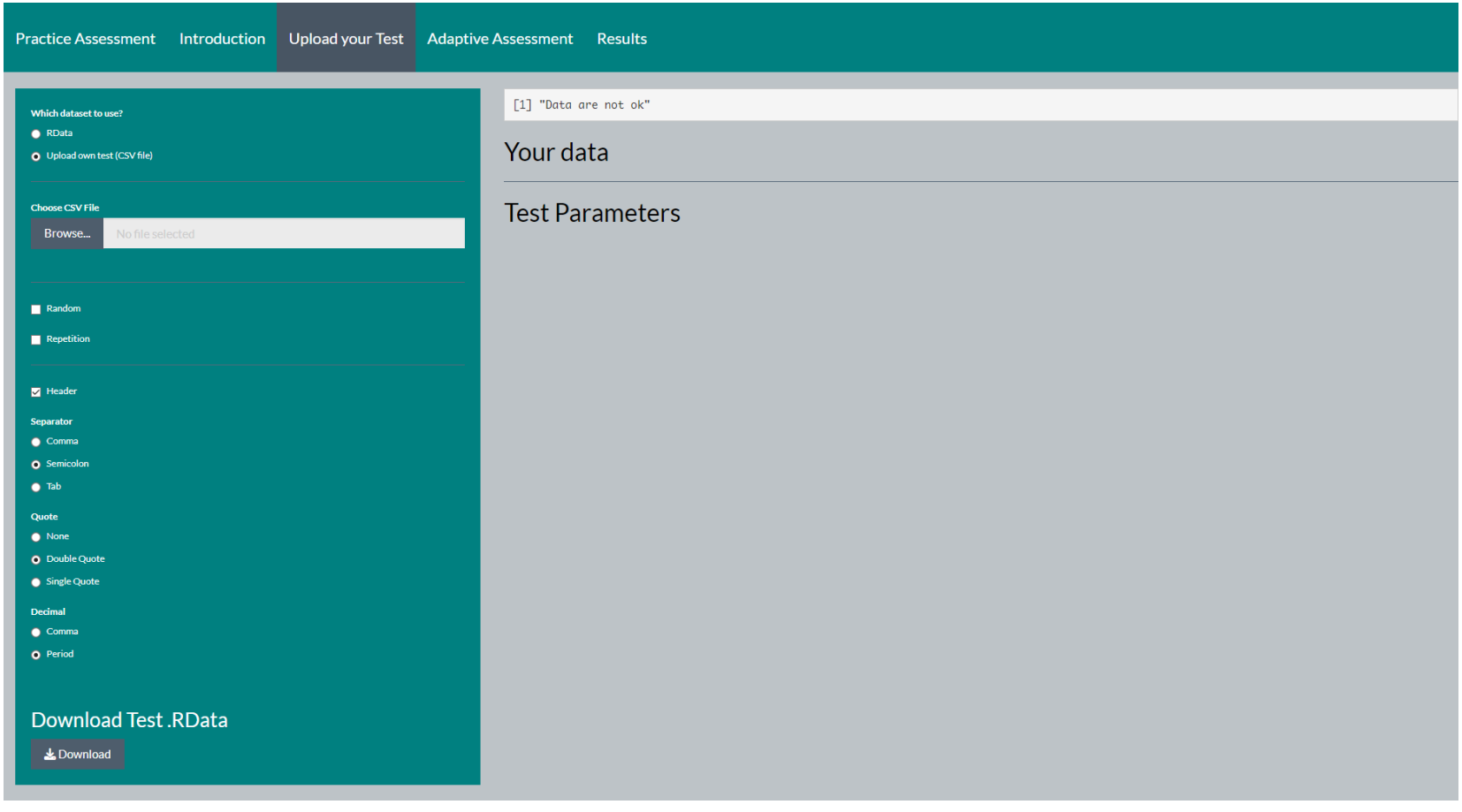
... without coding knowledge `run_Practice()`



Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE
Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Create and Try the Test ...

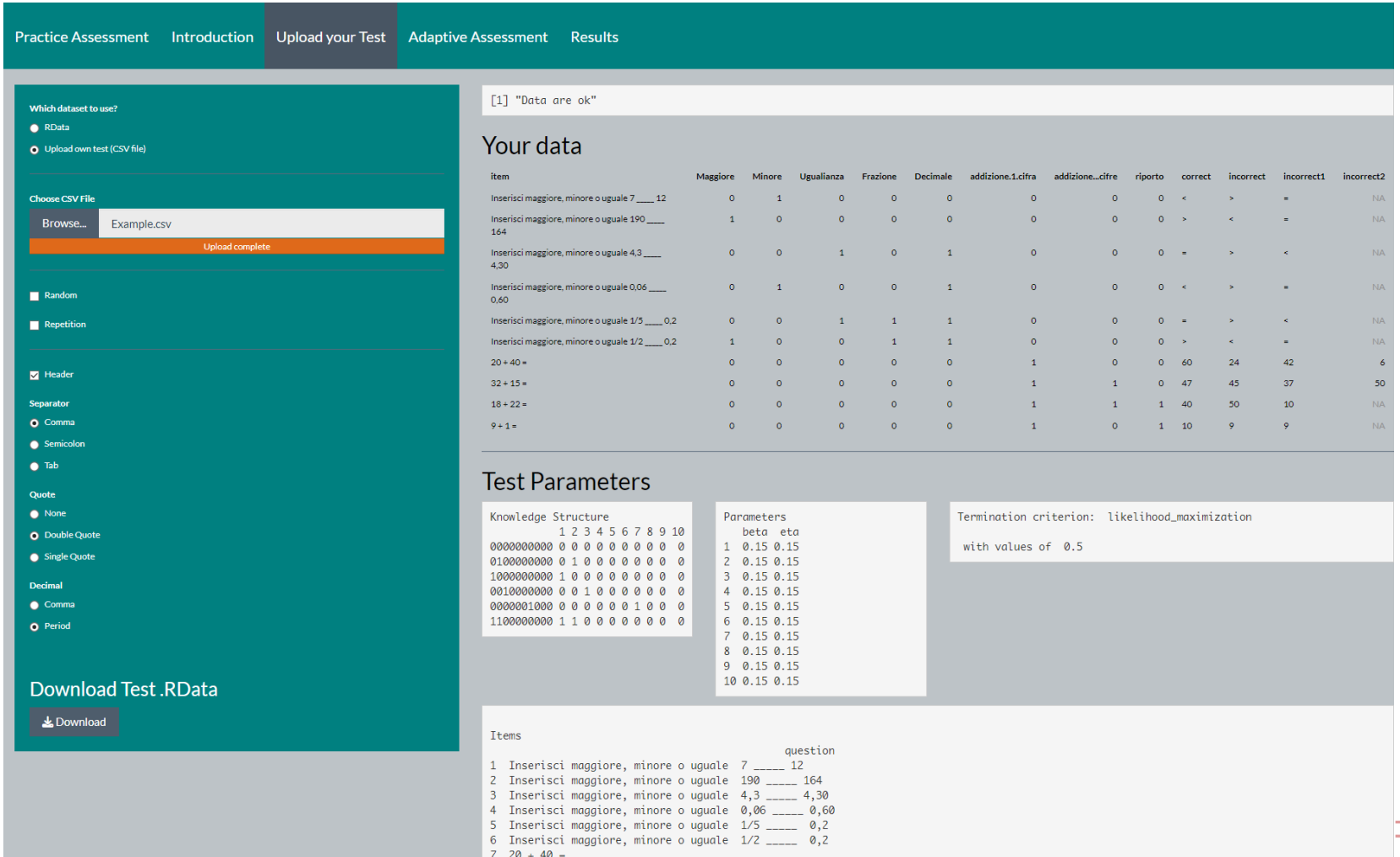
... without coding knowledge `run_Practice()`



Finanziato dall'Unione europea NextGenerationEU
MUR Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Create and Try the Test ...

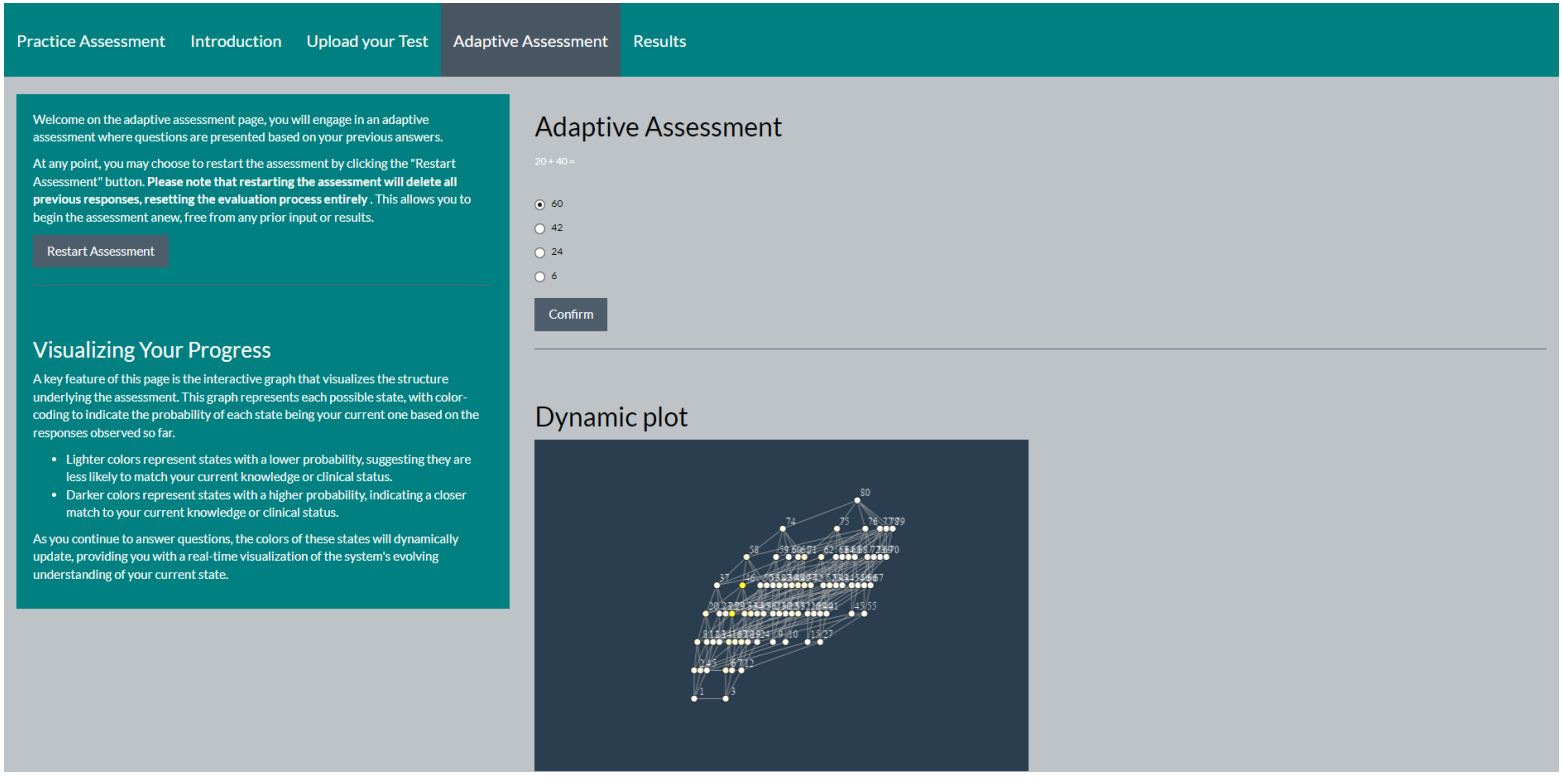
... without coding knowledge `run_Practice()`



Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE
Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Create and Try the Test ...

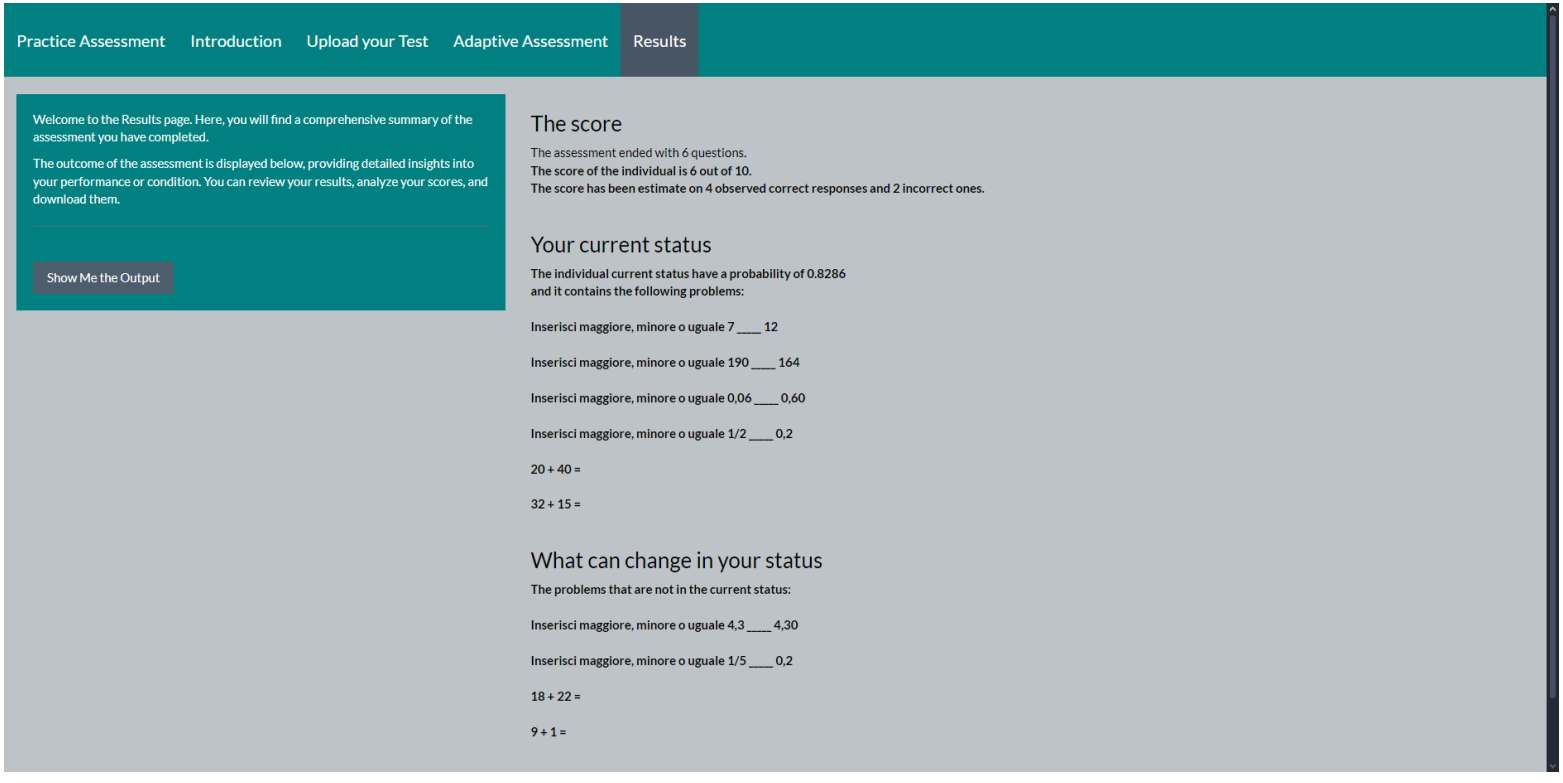
... without coding knowledge `run_Practice()`



Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE
Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Create and Try the Test ...

... without coding knowledge `run_Practice()`



Outline

1. Introduction
2. mycaas Package
   - Algorithm
   - Shiny
   - Application on RAISE
3. Final Remarks

Finanziato dall'Unione europea NextGenerationEU
MUR Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

RAISE - Robotics and AI for Socioeconomic Empowerment

- In collaboration with some middle schools in Lombardia and Liguria
- A pilot test covering the middle school program in in mathematics

Tuning a Test with mycaas

21 items with multiple choice and 595 states

Simulation parameters:

- two termination rules with six stopping criteria $SC = \{.5, .6, .7, .8, .9, 1\}$
- simulate ten response patterns for each $K \in \mathcal{K}$
- lucky guess $\eta_q = .2$ & careless error $\beta_q = .15$

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

RAISE - Robotics and AI for Socioeconomic Empowerment

- In collaboration with some middle schools in Lombardia and Liguria
- A pilot test covering the middle school program in in mathematics

Tuning a Test with mycaas

21 items with multiple choice and 595 states

Simulation parameters:

- two termination rules with six stopping criteria $SC = \{.5, .6, .7, .8, .9, 1\}$
- simulate ten response patterns for each $K \in \mathcal{K}$
- lucky guess $\eta_q = .2$ & careless error $\beta_q = .15$

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Performance Analysis

Accuracy and Efficiency Indexes

Given:

- true knowledge state K^w of a student w;
- the probability distribution $\mathcal{L}_m$ at the end of the assessment
- the recovered knowledge state $\hat{K}_m^w$

The following indexes were computed across simulated subjects:

1. The average number of questions asked.
2. The average maximum probability $\bar{\mathcal{L}}_m(\hat{K}_m)$.
3. The average Hamming distance $\bar{D}_m(K, \hat{K}_m)$ computed by

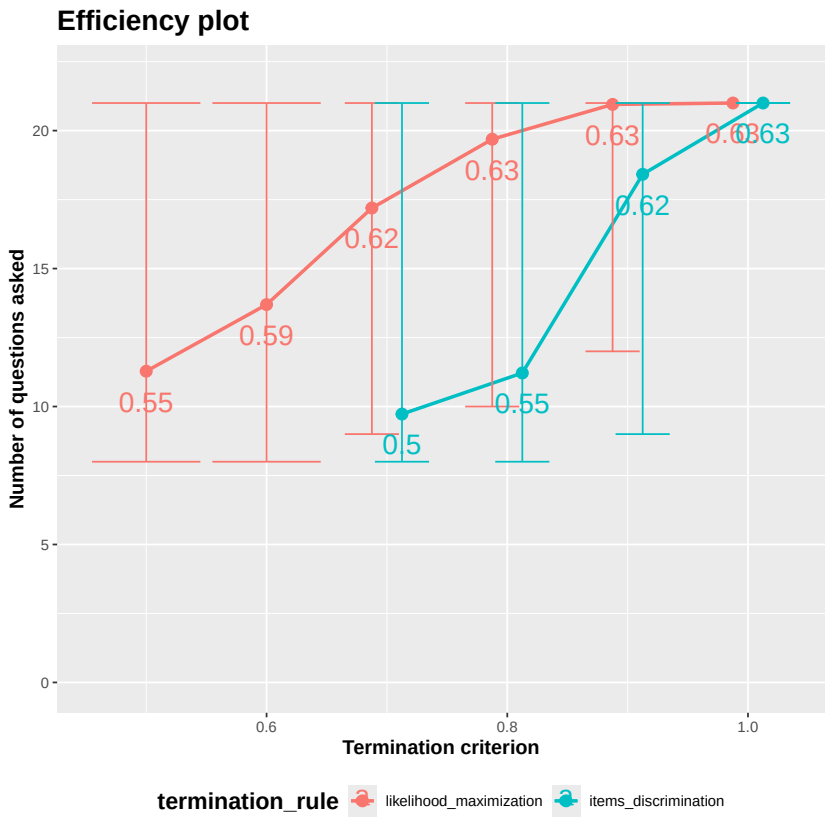
$$\bar{D}_m(K, \hat{K}_m) = \frac{1}{N} \sum_{w=1}^{N} |K^w \Delta \hat{K}_m^w|,$$

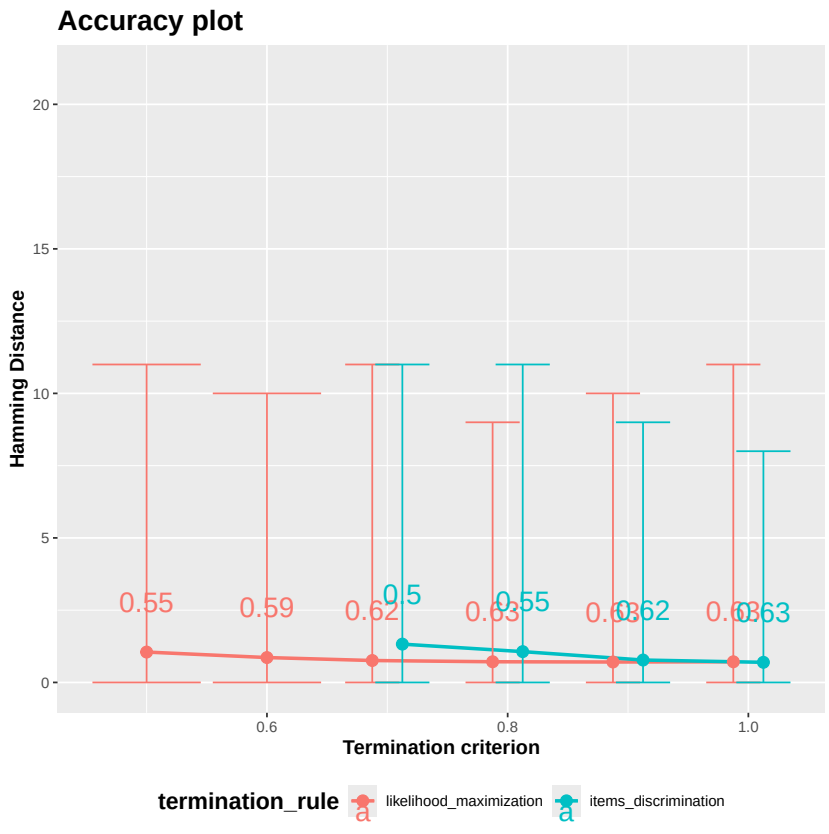
where Δ represents the symmetric set difference.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE
Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Performance Analysis

`performance_simulation(...)`





Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Cons of Adaptive Assessment

- Dependent on the assumptions of the Model :
  - The validity of results depends on the correctness of the model used.
- Complexity of Implementation:
  - Requires sophisticated algorithms and data processing infrastructure.
- Difficulty in Tuning:
  - Fine-tuning the assessment to achieve accurate difficulty adjustments is challenging.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Cons of Adaptive Assessment

- Dependent on the assumptions of the Model :
  - The validity of results depends on the correctness of the model used.
  - Response: Building upon KST and FPA
- Complexity of Implementation:
  - Requires sophisticated algorithms and data processing infrastructure.
- Difficulty in Tuning:
  - Fine-tuning the assessment to achieve accurate difficulty adjustments is challenging.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Cons of Adaptive Assessment

- Dependent on the assumptions of the Model :
  - The validity of results depends on the correctness of the model used.
  - Response: Building upon KST and FPA
- Complexity of Implementation:
  - Requires sophisticated algorithms and data processing infrastructure.
  - Response: GUI offline interface for adaptive assessment `run_Assessment()` and `run_Practice()`
- Difficulty in Tuning:
  - Fine-tuning the assessment to achieve accurate difficulty adjustments is challenging.

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Cons of Adaptive Assessment

- Dependent on the assumptions of the Model :
  - The validity of results depends on the correctness of the model used.
  - Response: Building upon KST and FPA
- Complexity of Implementation:
  - Requires sophisticated algorithms and data processing infrastructure.
  - Response: GUI offline interface for adaptive assessment `run_Assessment()` and `run_Practice()`
- Difficulty in Tuning:
  - Fine-tuning the assessment to achieve accurate difficulty adjustments is challenging.
  - Response: Performance analysis of the test `performance_simulation()`

Finanziato dall'Unione europea NextGenerationEU
MUR
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RAISE Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Final Remarks

- The Shiny environment is used to make adaptive assessments more accessible and user-friendly.
- Currently the usability of the package is tested within RAISE in Liguria school
- `https://github.com/brancaccioandrea/mycass`

Finanziato dall'Unione europea NextGenerationEU

MUR

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE
Robotics and AI for socio-economic empowerement
Ecosistema dell'Innovazione della Liguria
PNRR – M4C2 – I1.5

Acknowledgements:

This work has been supported by the European Commission, NextGenerationEU, Missione 4 Componente 2, "Dalla ricerca all'impresa", Innovation Ecosystem RAISE "Robotics and AI for Socio-economic Empowerment", ECS00000035.