

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



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.



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.



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.



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



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE
Robotics and AI for socio-economic empowerment
Ecosistema dell'Innovazione della Liguria
PNRR - M4C2 - 11.5

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 analysis to evaluate accuracy and efficiency



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.

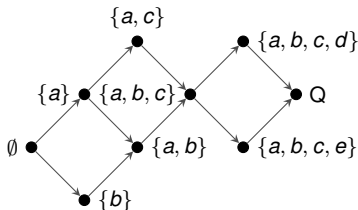
Formal Psychological Assessment (FPA; Spoto, Stefanutti & Vidotto, 2010): The objective from the clinical perspective is to give an in-depth evaluation of the construct investigated by the questionnaire.



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



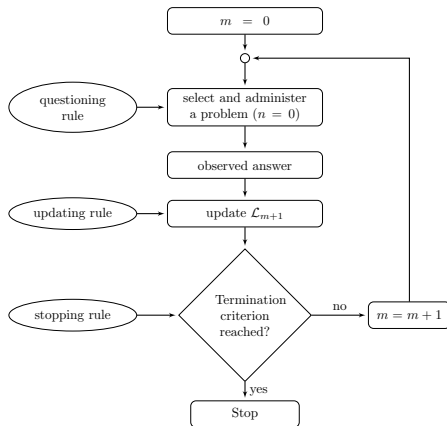
Flowchart of the Adaptive Assessment

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



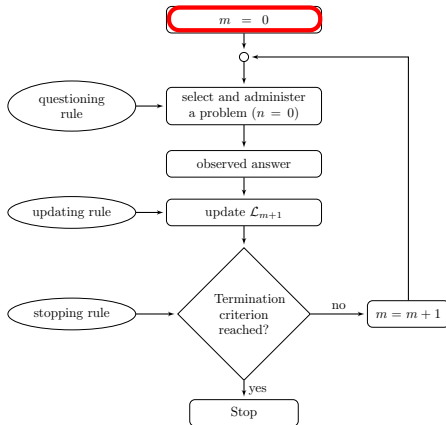


Probability distribution on the states

A probability distribution

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

Without prior knowledge is a uniform distribution





Questioning Rule

Half Split Rule

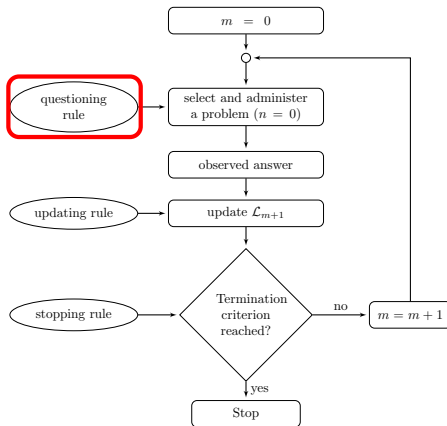
Select the “best” question to ask

Doignon & Falmagne, 1988

half_split: Let $\mathcal{L}_m(K)$ the likelihood of K at the step m , and the subset $\mathcal{K}_q \subset \mathcal{K}$ such that $q \in K$ for each $K \in \mathcal{K}_q$.

It selected problem $q \in Q$ that minimize

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





Updating Rule

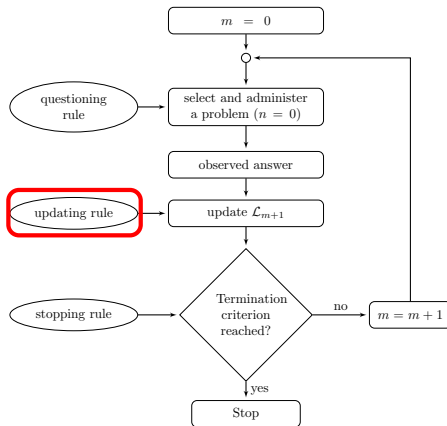
Bayesian Updating Rule

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

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





Termination rule

Heller & Repitsch, 2012

Likelihood Maximization:
The assessment terminate at step m if

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

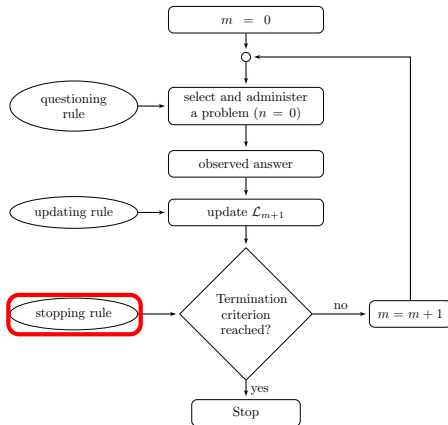
with $SC \in (0.5, 1]$.

Donadello, et al., 2017

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

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

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



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE
Robotics and AI for socio-economic empowerment
Ecosistema dell'innovazione della Liguria
PNRR – MAC2 – I1.5

Create and Try the Test ...

... without coding knowledge `run_Practice()`

Practice Assessment
Introduction
Upload your Test
Adaptive Assessment
Results

Welcome to the Adaptive Assessment Tools App

This application allows you to conduct adaptive assessments using the continuous Markov procedure by Falmagne and Delgion (CMP, 1988) from Knowledge Space Theory (KST) and for Adaptive Testing System for Psychological Disorders (ATS-PD, Donatello et al., 2017) from Formal Psychological Assessment (FPA). Whether you are assessing knowledge, evaluating clinical conditions, or any other application, this app provides a structured, efficient, and accurate way to collect your data.

What is Adaptive Testing in KST and FPA?

Adaptive testing is a dynamic process that tailors the assessment to each individual based on their responses. In both Knowledge Space Theory and Formal Psychological Assessment, this means recovering the state that represents the learner's or patient's abilities, among the collection of possible states on the structure.

The Underlying Structure

Both KST and FPA rely on an underlying structure—be it a knowledge structure in KST or a clinical structure in FPA. This structure guides the adaptive process, ensuring that each question or evaluation step is meaningful and relevant. The app can also accommodate a random presentation of questions, providing flexibility in how the assessment is conducted.

How the App Works

- 1. Upload Your Assessment:** On the second page, you can upload your assessment data. This includes questions and a skill map or clinical context, which should be in a CSV file format. Alternatively, if you have pre-created an assessment in R, you can upload an RDA file containing a variable of the 'assessment' type.
- 2. Adaptive Assessment:** The third page dynamically displays questions based on the adaptive algorithm. You can visualize the progress through a graph diagram, which highlights the likelihood of different states with color coding. This diagram helps to understand the current state. At the top of the diagram are the states that contain most of the items (with the whole collection of items on top), and at the bottom of the diagram are the states with fewer items (with the empty set on the bottom).
- 3. Results:** The final page provides a detailed outcome of the assessment. You'll see the cardinality of the recovered state (the score), the identified state (knowledge or clinical), and insights on potential changes in status. This comprehensive result allows for informed decision-making or further evaluation. Moreover, the result of the test can be downloaded and analyzed using R.

This app is designed to be a powerful tool for educators, psychologists, and clinicians, providing deep insights through adaptive assessments tailored to each individual's responses. Explore the pages and see how adaptive testing can enhance your understanding of knowledge or clinical states!

References

Donatello, I., Spoto, A., Simbo, F., Badaloni, S., Granziol, U., & Vidotto, G. (2017). ATS-PD: An Adaptive Testing System for Psychological Disorders. *Educational and Psychological Measurement*, 77(5), 792-815. <https://doi.org/10.1177/00131644166652180>

Falmagne, J.C., & Delgion, J.P. (1988). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, 41, 1-23. <https://doi.org/10.1111/j.2044.8317.1988.tb0084.x>



Finanziato
dall'Unione europea
NextGenerationEU



Italiadomani

PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE

Robotics and AI for socio-economic empowerment
Ecosistema dell'innovazione della Liguria
PNRR - M4C2 - I1.5

Create and Try the Test ...

... without coding knowledge `run_Practice()`

Practice Assessment

Introduction

Upload your Test

Adaptive Assessment

Results

Which dataset to use?

☒ Kfata

☐ Upload own test (CSV file)

Choose CSV file

Browse... No file selected

☒ Random

☒ Regression

☒ Header

Separator

☒ Comma

☐ Semicolon

☐ Tab

Quote

☒ None

☐ Double Quote

☐ Single Quote

Decimal

☒ Comma

☐ Period

Download Test .RData

Download

[X] "Data are not ok"

Your data

Test Parameters



Finanziato
dall'Unione europea
NextGenerationEU



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE
Robotics and AI for socio-economic empowerment
Ecosistema dell'Innovazione della Liguria
PNRR - M4C2 - 11.5

Create and Try the Test ...

... without coding knowledge `run_Practice()`

Practice Assessment
Introduction
Upload your Test
Adaptive Assessment
Results

Which dataset to use?

- ☒ RData
- ☐ Upload own test (CSV file)

Choose CSV File

Browse... Example.csv

Upload content

☒ Rankins
☒ Expedition
☒ Header

Separator

- ☐ Comma
- ☒ Tab
- ☐ Other

Quote

- ☐ None
- ☐ Double Quote
- ☐ Single Quote

Decimal

- ☐ Comma
- ☐ Period

Download Test.RData

Download

[1] "Data are ok"

Your data

Item	Maggiore	Misure	Uguale	Frazione	Decimale	addizione.1.dita	addizione.2.dita	ripeto	correct	incorrect	Incorrect1	Incorrect2
Inserisci maggiore, minore o uguale 7 ____ 12	0	1	0	0	0	0	0	0	0	+	+	NA
Inserisci maggiore, minore o uguale 190 ____ 164	1	0	0	0	0	0	0	0	0	+	+	NA
Inserisci maggiore, minore o uguale 4.5 ____ 4.30	0	0	1	0	1	0	0	0	+	+	+	NA
Inserisci maggiore, minore o uguale 0.06 ____ 0.60	0	1	0	0	1	0	0	0	+	+	+	NA
Inserisci maggiore, minore o uguale 1/5 ____ 0.2	0	0	1	1	1	0	0	0	+	+	+	NA
Inserisci maggiore, minore o uguale 1/2 ____ 0.2	1	0	0	1	1	0	0	0	+	+	+	NA
20 + 40 =	0	0	0	0	0	1	0	0	60	24	42	6
32 + 15 =	0	0	0	0	0	1	1	0	47	43	37	30
35 + 22 =	0	0	0	0	0	1	1	1	40	50	30	NA
9 + 5 =	0	0	0	0	0	1	0	1	10	9	9	NA

Test Parameters

Knowledge Structure

	1	2	3	4	5	6	7	8	9	10
0000000000	0	0	0	0	0	0	0	0	0	0
0100000000	0	1	0	0	0	0	0	0	0	0
1000000000	1	0	0	0	0	0	0	0	0	0
0010000000	0	0	1	0	0	0	0	0	0	0
0000001000	0	0	0	0	0	0	1	0	0	0
1100000000	1	1	0	0	0	0	0	0	0	0

Parameters

	beta	eta
1	0.15	0.15
2	0.15	0.15
3	0.15	0.15
4	0.15	0.15
5	0.15	0.15
6	0.15	0.15
7	0.15	0.15
8	0.15	0.15
9	0.15	0.15
10	0.15	0.15

Termination criterion: likelihood_maximization
with values of 0.5

Items

	question
1	Inserisci maggiore, minore o uguale 7 ____ 12
2	Inserisci maggiore, minore o uguale 190 ____ 164
3	Inserisci maggiore, minore o uguale 4.5 ____ 4.30
4	Inserisci maggiore, minore o uguale 0.06 ____ 0.60
5	Inserisci maggiore, minore o uguale 1/5 ____ 0.2
6	Inserisci maggiore, minore o uguale 1/2 ____ 0.2
7	20 + 40 =



Finanziato
dall'Unione europea
NextGenerationEU



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE
Robotics and AI for socio-economic empowerment
Ecosistema dell'innovazione della Liguria
PNRR - M4C2 - 11.5

Create and Try the Test ...

... without coding knowledge `run_Practice()`

Practice AssessmentIntroductionUpload your TestAdaptive AssessmentResults

Welcome on the adaptive assessment page, you will engage in an adaptive assessment where questions are presented based on your previous answers.

At any point, you may choose to restart the assessment by clicking the "Restart Assessment" button. Please note that restarting the assessment will delete all previous responses, resetting the evaluation process entirely. This allows you to begin the assessment anew, free from any prior input or results.

Restart Assessment

Visualizing Your Progress

A key feature of this page is the interactive graph that visualizes the structure underlying the assessment. This graph represents each possible state, with color coding to indicate the probability of each state being your current one based on the responses observed so far.

- Lighter colors represent states with a lower probability, suggesting they are less likely to match your current knowledge or clinical status.
- Darker colors represent states with a higher probability, indicating a closer match to your current knowledge or clinical status.

As you continue to answer questions, the colors of these states will dynamically update, providing you with a real-time visualization of the system's evolving understanding of your current state.

Adaptive Assessment

20/11/2024

60
42
24
6

Confirm

Dynamic plot



Finanziato
dall'Unione europea
NextGenerationEU



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RAISE
Robotics and AI for socio-economic empowerment
Ecosistema dell'Innovazione della Liguria
PNRR - M4C2 - 11.5

Create and Try the Test ...

... without coding knowledge `run_Practice()`

[Practice Assessment](#) [Introduction](#) [Upload your Test](#) [Adaptive Assessment](#) **[Results](#)**

Welcome to the Results page. Here, you will find a comprehensive summary of the assessment you have completed.

The outcome of the assessment is displayed below, providing detailed insights into your performance or condition. You can review your results, analyze your scores, and download them.

Show Me the Output

The score

The assessment ended with 4 questions.
The score of the individual is 4 out of 10.
The score has been estimate on 4 observed correct responses and 2 incorrect ones.

Your current status

The individual current status have a probability of 0.0286 and it contains the following problems:

Inserisci maggiore, minore o uguale 7 ____ 12

Inserisci maggiore, minore o uguale 190 ____ 164

Inserisci maggiore, minore o uguale 0.06 ____ 0.60

Inserisci maggiore, minore o uguale 1/2 ____ 0.2

20 + 40 =

32 + 15 =

What can change in your status

The problems that are not in the current status:

Inserisci maggiore, minore o uguale 4.3 ____ 4.30

Inserisci maggiore, minore o uguale 1/5 ____ 0.2

18 + 22 =

9 + 1 =

Outline

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



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 $\beta_q = .15$



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 $\beta_q = .15$



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 likelihood** $\bar{\mathcal{L}}_m(\hat{K}_m)$.
- 3 The **average Hamming distance** $\bar{D}_m(K, \hat{K}_m)$ computed by

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

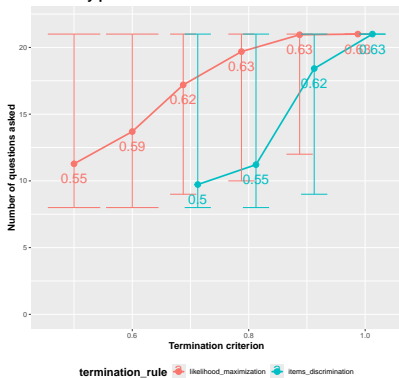
where Δ represents the symmetric set difference.



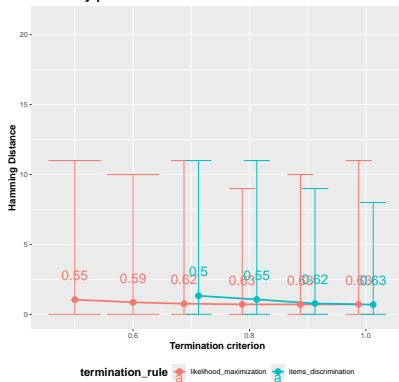
Performance Analysis

`performance_simulation(...)`

Efficiency plot



Accuracy plot





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.



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.



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.



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()`



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



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Robotics and AI for socio-economic empowerment
Ecosistema dell'Innovazione della Liguria
PNRR - M4C2 - 11.5

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.