# Computerized Adaptive Testing: a unified approach under Markov Decision Process

Patricia Gilavert[1][0000−0001−8833−9209] and Valdinei Freire[1][0000−0003−0330−3931]

School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, Brazil
{patifernan, valdinei.freire}@usp.br

**Abstract.** Markov Decision Process (MDP) is the most common planning framework in the literature for sequential decisions under probabilistic outcomes; MDPs also underlies the Reinforcement Learning (RL) theory. Computerized Adaptive Testing (CAT) is an assessment approach that selects questions one after another while conditioning each selection on the previous questions and answers. While MDP defines a well-posed optimization planning-problem, shortsighted score functions have solved the planning problem in CATs. Here, we show how MDP can model different formalisms for CAT, and, therefore, why the CAT community may benefit from MDP algorithms and theory. We also apply an MDP algorithm to solve a CAT, and we compare it against traditional score functions from CAT literature.

**Keywords:** Computerized adaptive testing · Markov Decision Process.

## 1 Introduction

Computerized Adaptive Testing (CAT) is an approach to assessment that tailors the administration of test items to the trait level of the examinee. Instead of applying the same question to every examinee, as in a traditional paper and pencil test, CATs apply questions one after another and each question selection is conditioned on the previous questions and answers [17]. The number of applied questions to each examinee can also vary to reach a better trade-off between precision, a correct trait estimation, and efficiency, a small number of questions. CATs reduce the burden of examinees in two ways; first, examinees do not need to complete a lengthy test; second, examinees answer questions tailored to their trait level avoiding too difficult or too easy questions [18].

Because examinees do not solve the same set of questions; an appropriate estimation of the latent trait level of the examinee must be considered. In the case of dichotomic questions, the item response theory (IRT) can be used to find the probability of an examinee to score one item as a function of his/her trait and therefore provide a coherent estimator. CAT in combination with IRT makes it possible to calculate comparable proficiencies between examinees who responded to different sets of items and at different times [6, 8]. This probability is influenced by item parameters, as difficulty and discrimination.

In every CAT we identify at least six components [22, 23]: (*i*) an item bank, (*ii*) an entry rule, (*iii*) a response model, (*iv*) an estimation mechanism, (*v*) an item selection criterion, and (*vi*) a stop criterion. The item bank determines questions that are available for the test; usually, items are selected without replacement. The entry rule specifies *a priori* knowledge from the examinee; in a Bayesian framework, it represents an *a priori* distribution over latent traits, and, in a Likelihood framework, it represents an initial estimation. The response model describes the chance of scoring for each examinee on each question in the item bank; the response model supports the estimation mechanism to estimate the latent trait of the current examinee. The item selection criterion chooses the question to be applied to the current examinee, while the stop criterion chooses when to stop the test; usually, both criteria may be supported by the current estimation, the item bank, and the response model.

Markov Decision Process (MDP) models problems where an agent makes sequential decisions and the results of these actions are probabilistic [12]. The agent's objective is to choose a sequence of actions so that the system acts optimally according to previously defined criteria. Such a sequence of actions follows a policy, which maps each observed state into an action. While MDP algorithms consider the model of the optimization problem to be known; Reinforcement Learning (RL) considers the same optimization problem when the model is unknown and must be learnt through trial-and-error interaction with the process [20].

Although CAT may be seen as a planning problem, where a long horizon must be taken into account to act optimally, mostly CAT methods consider shortsighted score-functions which select questions through an immediate analysis. Just a few works in literature has applied MDP framework to CAT, mostly by framing the sequential decision problem as Reinforcement Learning. [5] has applied RL to Sequential Mastery Testing, when a CAT is designed to classify each examinee as master or non-master; in this work, only two examinee is considered and questions follow the same response model; therefore, state can be simply defined as the difference between wrong and right answer and the only decision is when to stop the test. [15] make use of Recurrent Neural Networks to model the state space, mainly to account for an embedded representation of questions already applied; overexposure of items are avoided by penalizing overexposed items.

Although a few previous works frame the CAT problem as an MDP, they do so under a specific scenario. In this paper we show how MDP can model different formalisms for CAT, in particular, we frame CAT problems under the POMDP formalism, when agents does not observe the state fully [7]. We also gives an empirical example of how MDPs can be used for solving CAT problems, and we compare it against traditional shortsighted solutions from CAT literature.

## 2 Computerized Adaptive Testing

CATs are applied in an adaptive way to each examinee by computer. Based on predefined rules of the algorithm, the items are selected sequentially during the test after each answer to an item [18]. A classic CAT can be described by the following steps [9]:

1. The first item is selected;
2. The latent trait is estimated based on the first item answer;
3. The next item to be answered is selected;
4. The latent trait is recalculated based on previous answers; and
5. Repeat steps 3 and 4 until an answer is no longer necessary according to a pre-established criterion, called stop criterion.

### 2.1 Response Model and Latent Trait Estimator

Usually, a CAT is used to estimate some latent trait of an examinee. A university may use CAT to rank student by estimating a grade for each candidate. A doctor may diagnose some patient condition under a multidimensional spectrum. The Government may give a driver license to a teenager.

Generically, we can consider a latent trait $\theta_j \in \Theta$ for an examinee $j$, where $\Theta$ is the support set for latent trait. This trait can be unidimensional or multidimensional. When a question $i$ is submitted to the examinee $\theta_j$, an answer is given and a result $x_{ij}$ is observed.

Results may be dichotomic, polithomic, or continuum. A response model considers parameterized questions, i.e., a question $i$ is represent by a parameter vector $\gamma_i$ and a support set $\mathcal{X}_i$ for possible results. After a question $i$ is submitted to an examinee $j$, a random variable $X_{ij} \in \mathcal{X}_i$ representing the observed result is generated and the response model defines a probability distribution for the possible results[1], i.e., for all $x \in \mathcal{X}_i$:

$$\Pr(X_{ij} = x|\gamma_i, \theta_j).$$

Given an examinee $\theta$ and a sequence of $n$ answers $\mathbf{x}_n = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$, the latent trait $\theta$ can be estimated by Bayesian procedure or Maximum Likelihood (ML) [1].

We consider here a Bayesian estimator based on expected *a posteriori* (EAP), i.e.,

$$\hat{\theta} = \mathrm{E}\left[\theta|\mathbf{x}_n\right] = \int \theta \frac{f(\theta)\prod_{k=1}^{n}\Pr(X_{i_k} = x_{i_k}|\theta)}{\Pr(\mathbf{X}_n = \mathbf{x}_n)} d(\theta), \tag{1}$$

where $f(\theta)$ is the *a priori* distribution on the latent trait $\theta$, usually considered the standard normal distribution.

[1] We consider the case when $\mathcal{X}_i$ is enumerate for a briefer exposition.

The ML estimator estimates the latent trait by $\hat{\theta} = \max_\theta L(\theta|\mathbf{x}_n)$ where the likelihood is given by:

$$L(\theta|\mathbf{x}_n) = \prod_{k=1}^{n} \Pr(X_{i_k} = x_{i_k}|\theta). \tag{2}$$

If the objective of the CAT is to classify an examinee among options category $\mathcal{C}$, a classification function $C : \Theta \to \mathcal{C}$ may be defined after latent trait estimation $\hat{\theta}$ or a category estimator $\hat{C}$ may be directly defined.

### 2.2 Efficiency, Precision and Constraints

A CAT may be evaluated under two main criteria: precision and efficiency. Precision is related to how good is the estimation $\hat{\theta}$, while efficiency is related to the effort made by the examinee or examiner.

If the objective of the CAT is to ranking examinees, a common evaluation for precision is the mean square error (MSE):

$$MSE = \mathrm{E}\left[(\theta - \hat{\theta})^2\right] = \int (\theta - \hat{\theta})^2 f(\theta) d(\theta).$$

If the objective of the CAT is to classify an examinee, a common evaluation for precision may be accuracy (ACC):

$$ACC = \mathrm{E}[\mathbb{1}_{C_\theta = \hat{C}}],$$

where $\mathbb{1}_A$ is the indicator function for the condition $A$, and $C_\theta$ is the correct category.

Efficiency is usually evaluated by the length of the CAT, i.e., how many questions the examinee answered. Efficiency may also be evaluated by the time spent by the examinee because some questions may require more time than others, or some other effort measure.

Finally, CAT may pursue precision and efficiency under some restriction. The commonest restriction is regarding question repetition; in a test a question can only be applied once. A examiner may also be concerned with test content balancing and item exposure control. Content balancing considers that the item bank is clustered into groups of question and the test should choose question from all of the groups with a minimum rate. Item exposure control considers that a question should not be submitted to many examinee; if a question is overexposed, examinees will know it beforehand and the CAT will lost precision.

A particular restriction is considered in the multistage CAT [11]. In the multistage CAT, the test is subdivide into stages, and, in any stage, a set of question must be revealed at once to the examinee. After the examinee answers all the questions in a stage, he/she is routed individually to a new stage.

### 2.3 CAT as an optimization problem

All the elements described in the previous section allow us to define an optimization problem. Given a bank of items $Q$, we first model the CAT problem as a process:

1. A examinee $\theta$ is drawn from a distribution $f(\theta)$. Although the distribution $f(\theta)$ may be known to the examiner, the process does not reveals the examinee $\theta$.
2. While examiner does not decide to stop the test, at any stage $t = 1, 2, \ldots$
   (a) the examiner chooses a question $q_t \in Q$ and submits to the examinee[2].
   (b) By answering the question, the examinee generates a result $x_t$.
3. The examiner estimates the latent trait $\theta$ through an estimation $\hat{\theta}$[3].

The result of the CAT process is a random history $h = (\theta, q_1, x_1, q_2, x_2, \ldots, q_N, x_N, \hat{\theta})$. An optimal examiner is obtained from the following optimization problem:

$$\begin{array}{ll} \min & \mathrm{E}_\theta[U(h)] \\ \text{s.t.} & Z(f(\theta), h) \end{array}.$$

$U$ is an objective function. For example, if the examiner wants to minimize MSE, then, $U(h) = (\theta - \hat{\theta})^2$. If the examiner wants to minimize the test length, then, $U(h) = N$. The objective function can also combine both objectives.

$Z$ is a constraint function. If the examiner wants to constraint item to maximum $\alpha$ exposition, then, $Z(\cdot) = \{h|\mathrm{E}[\sum_{t=1}^{N} \mathbb{1}_{q_t=q}] \leq \alpha \ \forall q \in Q\}$. If the examiner wants to avoid question repetition, then, $Z(\cdot) = \{h|\sum_{t=1}^{N}\sum_{t'=t+1}^{N} \mathbb{1}_{q_t=q_{t'}} = 0 \ \forall q \in Q\}$. Note that in the first example, constraints are on the expectation over the random variable $h$, while in the second example constraints are on the samples of the random variable.

## 3 Markov Decision Process

We consider a special case of Markov Decision Process, the Stochastic Shortest Path (SSP) problem [2]. A SSP is defined by a tuple $\langle \mathcal{S}, \mathcal{A}, P, C, s_0, \mathcal{G} \rangle$ where: $s \in \mathcal{S}$ are the possible states; $a \in \mathcal{A}$ are possible actions; $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ is the transition function; $\mathcal{C} : \mathcal{S} \times \mathcal{A} \to R^+$ is the cost function; $s_0 \in \mathcal{S}$ is the initial state; and $\mathcal{G}$ is the set of goal states.

A SSP defines a process of an agent interacting with an environment and at all time step $t$: $(i)$ the agent observes a state $s_t$, $(ii)$ the agent chooses an action $a_t$, $(iii)$ the agent pays a cost $c_t$; and $(iv)$ the process moves to a new state $s_{t+1}$. The process ends when a goal state $s \in \mathcal{G}$ are reached. Transitions and costs present Markov's property, i.e., they both depend only on the current state $s_t$ and chosen action $a_t$.

[2] In the multistage CAT, the examiner must choose a set of questions.

[3] In a classification CAT, the examiner can estimate directly a category instead of a the latent trait $\theta$.

The solution for a SSP consists of policies that describe which actions should be taken in each situation. Here we consider probabilistic stationary policies $\pi : \mathcal{S} \to (\mathcal{A} \to [0,1])$ that maps each state for a probability distribution over actions.

The objective of a SSP is to find an optimal policy $\pi^*$ that minimizes the expected cumulative cost, i.e., we define a value function of a policy $\pi$ by $V^\pi = \mathrm{E}[\sum_{t=0}^{\infty} c_t|\pi]$ and defines the optimal policy as $\pi^* = \arg\max_\pi V^\pi$. There are many algorithms in the literature to find optimal policies for MDPs [12]. Here, we focus on Linear Programming algorithms, in particular the dual criterion formulation [21].

### 3.1 Linear Programming Solution

The Linear Programming Dual Formulation considers variables $x_{s,a}$ for all $s \in \mathcal{S}, a \in \mathcal{A}$ that indicates an expected accumulated occurrence frequency for every pair state-action. The SSP dynamics restricts the solutions by specifying an $in(s)$ and $out(s)$ flow model for every state $s$. Every state but initial state and goal state must equalize $in(s)$ and $out(s)$. Initial state $s_0$ presents a unity in-out difference, while goal states in $\mathcal{G}$ has no output. For every state $s \in \mathcal{S}$, we have:

$$in(s) = \sum_{s' \in \mathcal{S}, a \in \mathcal{A}} x_{s',a} \mathcal{P}(s|s',a) \quad \text{and} \quad out(s) = \sum_{a \in \mathcal{A}_{(s)}} x_{s,a}. \tag{3}$$

Define the linear programming LP1 as follows:

$$\begin{array}{lrll} \text{LP1} & \min_{x_{s,a}} & \sum_{s \in \mathcal{S}, a \in \mathcal{A}} x_{s,a} \mathcal{C}(s,a) & \\ & \text{s.t.} & x_{s,a} \geq 0 & \forall s \in \mathcal{S}, a \in \mathcal{A}(s) \\ & & out(s) - in(s) \leq 0 & \forall s \in \mathcal{S} \setminus (\mathcal{G} \cup s_0)\,. \\ & & out(s_0) - in(s_0) \leq 1 & \\ & & \sum_{s_g \in \mathcal{G}} in(s_g) = 1 & \end{array} \tag{4}$$

The optimal policy $\pi^*$ to the SSP can be obtained from expected accumulated occurrence frequency $x_{s,a}$ in LP1 by:

$$\pi^*(a|s) = \frac{x_{s,a}}{\sum_{a' \in \mathcal{A}} x_{s,a'}} \quad \forall s \in \mathcal{S},$$

and $\pi^*$ incurs in expected cost $c_{min} = \sum_{s \in \mathcal{S}, a \in \mathcal{A}} x_{s,a} \mathcal{C}(s,a)$.

### 3.2 Partial-Observable MDP

A formulation to sequential decision problem more generic than MDPs or SSPs is the Partial-Observable MDP (POMDP) [7]. In an MDP, the agent is considered to observe immediately the state of the process, this guarantee that an optimal policy can be find in the state space, i.e., the optimal decision is based only on the current state observation.

In a POMDP, the agent observes the process state mediate by a probabilistic observation function $O : \mathcal{S} \to (\mathcal{O} \to [0, 1])$. A POMDP defines a process of an agent interacting with an environment and at all time step $t$: $(i)$ the agent makes an observation $o_t \sim O(s_t)$, $(ii)$ the agent chooses an action $a_t$, $(iii)$ the agent pays a cost $c_t$; and $(iv)$ the process moves to a new state $s_{t+1}$.

In a POMDP, the optimal policy must consider the history for the current time step t, i.e., $h_t = (o_0, a_0, c_0, \ldots, o_t)$ or a belief state $b_t$ which is a probability distribution over state space given the history of observations. Note that the history grows exponentially with the time step $t$ and finding an optimal policy to a POMDP is only computational practical when the state space is finite and small. Usually, algorithms try to find a quasi-optimal policy [7].

### 3.3 Augmented States

MDPs and SSPs are very strict regards Markov property. First, the agent must observes the full state of the process. Second, the cost function must depend only on the current state. Third, the objective function considers a random variable that is simply the sum of immediate costs. If any of this properties is not presented, most of time, the optimal policy is not restricted to the MDP state space.

The POMDP solution is an example. In this case, the policy must be defined in an augmented state space, the history space. In general, Markov property can be recovered under an appropriated augmented state space. For example, if the process ends after some fixed times step $T$, the state space must be augmented with the current time $t$, i.e., $(s, t)$.

Although augmented state space can be a general technique to find out optimal policies, we have already seem with POMDP, that the augmented state space may grows exponentially.

## 4 CAT as a POMDP

Remember that in the process of a CAT, initially an examinee is drawn from a probability distribution $f(\theta)$. Then, given any question, the examiner observes the answer for that question, which depends on the question itself and the latent trait $\theta$ of the examinee. However, the examiner never observes directly the latent trait $\theta$. A CAT can be seen as a POMDP by:

1. Considering the state space as the tuple $(\theta, q_t)$, i.e., the latent trait and the last question.
2. Considering the action space as the set of question in the item bank, plus a termination action.
3. Considering deterministic transitions. First, $\theta$ never changes and the question $q_t$ changes following the examiner choices.
4. Considering the observation of the process as the result of the examinee answering the last question, i.e., the CAT response model.

Figure 4 shows a scheme connecting CAT and POMDP. There, the belief state is obtained from the *a posteriori* distribution based on Bayesian method. Next, we show how POMDPs can represent different formulations of CAT. $I$ is the belief state, while $\tau$ is the model to generate the *a posteriori* belief state.

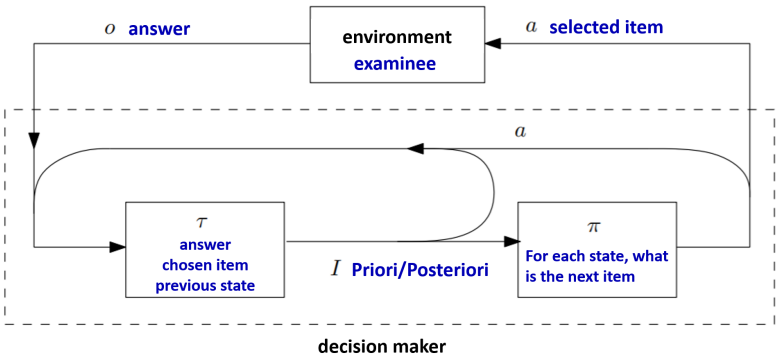


**Fig. 1.** CAT framed in a POMDP with belief state.

**Minimize MSE with fixed-length $N$ without repeated questions.** Note that we do not formulate yet the cost function. Essentially it depends on the trade-off between precision and efficiency. In this scenario, it is desired to obtain the smallest MSE (utility function $U$) with at maximum $N$ questions (restriction function $Z$).

To do so, we must augment the observation space with all the question submitted to the examinee and respective results. Then, a terminal cost is payed based on the variance of the *a posteriori* distribution. The action space depends on the observation, i.e., $a_t \in \mathcal{A}(o_t)$ and if a question was already submitted, such a question is not included in $\mathcal{A}(o_t)$. After $N$ questions, the only question available in $\mathcal{A}(o_t)$ is the termination action.

**Item Exposure Control: every action is exposed equally.** Consider the LP1 formulation in Section 3.1, if every question must be exposed equally, the following constraint must be considered for every pair $a, a' \in \mathcal{A}$:

$$\sum_{s \in \mathcal{S}} x_{s,a} = \sum_{s \in \mathcal{S}} x_{s,a'}.$$

Remember that $x_{s,a}$ indicates the expected amount of occurrence of the pair $(s, a)$. The sum $\sum_{s \in \mathcal{S}} x_{s,a}$ is the expected total occurrence for action $a$, therefore, a question $a$.

**Multi-stage CAT: $K$ questions per stage.** In this case, an action is a subset of the item bank $Q$ with $K$ questions, i.e., $\mathcal{A} = \{A \in 2^Q \text{ and } |A| = K\}$, where $2^Q$ is the power set of $Q$.

Note that although the formulations here presented may be not practical to be solved optimally, they are described under the same framework, MDPs and POMDPs. In the next section we give a practical result of our theoretical formulation.

## 5 Approximating a CAT by an MDP

In this section we show that the theoretical result showed in the previous section can elucidate results in the CAT literature. We consider a traditional CAT problem: fixed-length with a dichotomous item bank. We approximate this CAT by an MDP, which can be solved optimally, and we show that traditional selection criterion such as Fisher Information may be close to optimal solution.

### 5.1 Response Model: Item Response Theory

It is possible to build a CAT based on the item response theory (IRT), a mathematical model that describes the probability of an individual to score an item as a function of the latent trait level $\Pr(X_i = 1 \mid \theta)$. We consider the logistic model with three parameters [3]. A bank of 45 items calibrated from a national mathematics exam was used in our experiment [19].

### 5.2 Normalized and Discretized Bayesian Method

As said before, one way of recovering Markov property in a process is to consider augmented state or belief state. The *a posteriori* distribution by following Bayes method is exactly the belief state of our CAT POMDP. However, the state space of such distribution is the continuum $\aleph_1$. We construct a finite MDP from such belief states.

First, we normalized every *a posteriori* distribution by a Gaussian distribution with equivalent mean and variance. Second, we discretized such pair of values; we consider 100 mean values (between -4 and 4) and 1,000 variance values (between 0.001 and 1). Finally transitions are defined among normalized discretized belief states.

Usually, a CAT does not repeat an item in the same exam. Therefore, the state space must be augmented with the applied questions. In this case, the number of state is exponential in the number of questions. To reduce the number of states, we allow the CAT to repeat questions; in this case the state space does not need to be augmented. Remember that answers to questions are probabilistic; therefore, the user may answer differently for the same question.

We consider the CAT optimization problem of minimize MSE with fixed-length $N$ with repeated questions. In this case, an optimal policy can easily be founded. In the next section we compare it to traditional approach from CAT literature.

### 5.3 Experimental Result

We consider two score functions from CAT literature: Fisher Information (FI) [10, 16] and Minimizing the Expected Posterior Variance (MEPV) [14]. The first one, FI, is well-known in the literature and is also the cheapest method. The second one, MEPV, is a costly score function, which consider the next potential belief states given every possible question in the item bank; it is optimal in our normalized discretized MDP when only one question is allowed. We define policies based on each score function and evaluated them under the MDP framework; because of the MDP framework the value function $V^{\pi}$ for each score function can be calculated exactly.

Score functions are shortsighted, in fact, they do not even take into account the fixed-length horizon $N$. They are clearly suboptimal, however, it is far from clear how far from the optimal policy they are. We make use of an MDP algorithm to obtained an optimal policy for every length between 1 and 45 questions. Every other policies, we compare against this gold standard.

Figure 5.3 shows our results. We plot the difference between the root MSE (RMSE) for every policy and the gold standard. Besides FI and MEPV policies, we also plot optimal policies behaviour for fixed-length 15, 30, and 45. Note that optimal policies may be far from gold standard during the first questions, but in the end (15, 30, or 45 questions) the gold standard is reached. Because policies based on score functions choose questions that gives return immediately, they are always close to the gold standard, but never is optimal. Despite score function not being optimal, we can see that the largest difference is not more than 0.01.

## 6 Conclusion

We formulated CAT formalisms as MDPs and showed in an experiment that despite being optimal, the gain with MDPs may not compensate since suboptimal solution must be considered in real scenarios. However, the formulation as MDPs allows to formulate CAT as optimization problem and describe many CAT formulations under the same framework.

We believe that such a framework and experiments as the one here showed, may elucidate the limits of a myriad of methods in the CAT literature, mainly regarding CAT under constraints. Recently we showed that Fixed-length stop criterion has great advantages against other stop criteria, with this framework we can investigate how it compares against optimality under different evaluations [4]. For example, the MDP framework allows to define risk-sensitive optimality [13]; Risk-sensitive MDPs would allow a CAT to weight worst scenarios regarding the length of the test or the MSE so that CAT is fair for every examinee.

## References

1. de Andrade, D.F., Tavares, H.R., da Cunha Valle, R.: Teoria da resposta ao item: conceitos e aplicações. ABE, São Paulo (2000)

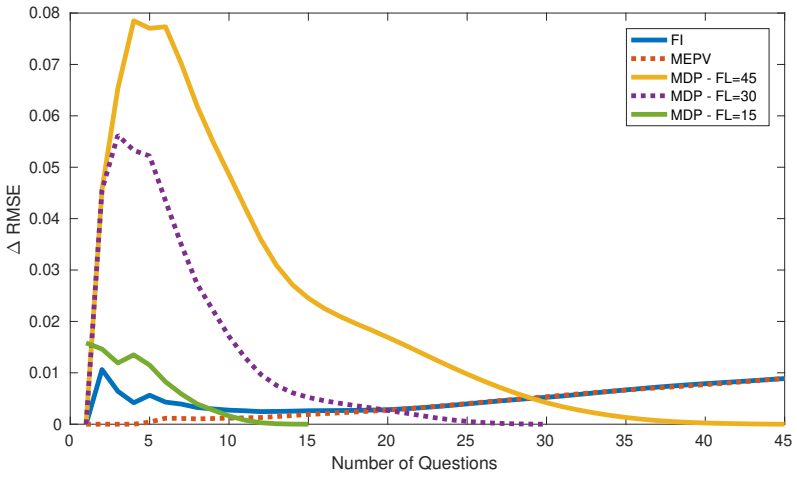


**Fig. 2.** Comparison of policies based on shortsighted score-functions and MDP optimal policies.


2. Bertsekas, D.P., Tsitsiklis, J.N.: An analysis of stochastic shortest path problems. Mathematics of Operations Research **16**(3), 580–595 (Aug 1991)
3. Birnbaum, A.L.: Some latent trait models and their use in inferring an examinee's ability. Statistical theories of mental test scores (1968)
4. Blind: Comprehensive empirical analysis of stop criteria in computerized adaptive testing. In: Submitted to International Conference on Computer Supported Education (CSEDU) (2021)
5. El-Alfy, E.S.M.: A reinforcement learning approach for sequential mastery testing pp. 295–301 (2011)
6. Hambleton, R.K., Swaminathan, H.: Item response theory: Principles and applications. Springer Science & Business Media (2013)
7. Hoerger, M., Kurniawati, H.: An on-line pomdp solver for continuous observation spaces (2020)
8. Kreitzberg, C.B., Stocking, M.L., Swanson, L.: Computerized adaptive testing: Principles and directions. Computers & Education **2**(4), 319–329 (1978)
9. van der Linden, W.J., Glas, C.A.: Computerized Adaptive Testing: Theory and Practice. Springer Science & Business Media, Boston, MA (2000)
10. Lord, F.M.: Applications of item response theory to practical testing problems. Routledge (1980)
11. Magis, D., Yan, D., von Davier, A.A.: Computerized Adaptive and Multistage Testing with R: Using Packages CatR and MstR. Springer Publishing Company, Incorporated, 1st edn. (2017)
12. Mausam, Kolobov, A.: Planning with markov decision processes: An ai perspective. Synthesis Lectures on Artificial Intelligence and Machine Learning **6**(1) (2012)
13. Minami, R., Silva, V.F.d.: Shortest stochastic path with risk sensitive evaluation. In: 11th Mexican International Conference on Artificial Intelligence (MICAI 2012). Lecture Notes in Artificial Intelligence, vol. 7629, pp. 370–381 (2012)

14. Morris, S.B., Bass, M., Howard, E., Neapolitan, R.E.: Stopping rules for computer adaptive testing when item banks have nonuniform information. International journal of testing **20**(2), 146–168 (2020)
15. Nurakhmetov, D.: Reinforcement learning applied to adaptive classification testing. In: Theoretical and Practical Advances in Computer-based Educational Measurement, pp. 325–336. Springer, Cham (2019)
16. Sari, H.I., Raborn, A.: What information works best?: A comparison of routing methods. Applied psychological measurement **42**(6), 499–515 (2018)
17. Segall, D.O.: Computerized adaptive testing. Encyclopedia of social measurement **1**, 429–438 (2005)
18. Spenassato, D., Bornia, A., Tezza, R.: Computerized adaptive testing: A review of research and technical characteristics. IEEE Latin America Transactions **13**(12), 3890–3898 (2015)
19. Spenassato, D., Trierweiller, A.C., de Andrade, D.F., Bornia, A.C.: Testes adaptativos computadorizados aplicados em avaliaçoes educacionais. Revista Brasileira de Informática na Educação **24**(02), 1 (2016)
20. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, USA (1998)
21. Trevizan, F., Teichteil-Königsbuch, F., Thiébaux, S.: Efficient solutions for stochastic shortest path problems with dead ends. In: Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence (UAI) (2017)
22. Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J.: Computerized adaptive testing: A primer. Routledge (2000)
23. Wang, C., Chang, H.H., Huebner, A.: Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. Journal of Educational Measurement **48**(3), 255–273 (2011)