A Reinforcement Learning Approach for Sequential Mastery Testing

El-Sayed M. El-Alfy
College of Computer Sciences and Engineering
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
alfy@kfupm.edu.sa

***Abstract*—This paper explores a novel application for reinforcement learning (RL) techniques to sequential mastery testing. In such systems, the goal is to classify each examined person, using the minimal number of test items, as master or non-master. Using RL, an intelligent agent autonomously learns from interactions to administer more informative and effective variable-length tests. Empirical results are also provided to evaluate the performance of the proposed approach as compared to two common approaches for variable-length testing (Bayesian decision and sequential probability ratio test) as well as to the fixed-length testing.**

Keywords-sequential mastery testing; reinforcement learning; Bayesian decision theory; sequential probability ratio test; intelligent tutoring.

I. INTRODUCTION

The role of computers in administering more effective sequential tests is increasingly growing in several domains; for example, psychometric measurements, screening and medical diagnosis, professional licensure, machine troubleshooting, software testing, and quality control in manufacturing systems. The general type of this problem belongs to inductive logic or plausible reasoning where the value of an underlying hidden variable can be inferred by observing its consequences. In this paper, we focus on educational testing where the hidden variable is the examined person's mastering ability and the consequences are his/her response pattern to the presented test items. Educational tests are still the mainly adopted method for measuring students' ability levels, pinpointing their strengths and weaknesses, and classifying them into categories based on their knowledge states. Various subsequent decisions are also frequently made such as college admittance, assigning letter grades, recommending pre-requisite courses and curriculum changes, monitoring progress, and granting awards and scholarships.

One of the latest advances is computer-based testing, that attracted considerable interest of several researchers and enables greater precision and efficiency, is sequential mastery testing (SMT) [1, 2, 13, 14, 15]. Rather than using a fixed set of test items, which has been pre-determined for all examinees by a domain expert, a sequential testing procedure automatically tailors the test length and/or items to the ability level of each examinee. A sequential test starts with some prior belief about the examinee's ability and sequentially updates this belief after each item response is collected. Testing continues until enough confidence is achieved. Thus, the test length can be tailored for each examinee. An appealing advantage of this approach is that higher classification precision can be accomplished even with fewer test items. It has been shown that substantial reduction of the test length can be attained without severely sacrificing the measurement accuracy [2]. Moreover, if items in the item pool have different characteristics, then the administered items can be appropriately selected to maximize the information revealed about the examinee. This persuasively makes sense because fewer test items are needed to evaluate examinees with very high or very low ability than those on the boundary where a decision is not a clear-cut. The score takes into account not just the number of questions answered correctly, but also the statistical characteristics of questions and the performance of each examinee on the question set. Most of the earlier studies on this problem have been concerned with the use of item response theory (IRT) for probabilistic modeling and reasoning in these systems [3].

The objective of this study is to investigate the application of reinforcement learning techniques in sequential mastery testing where the goal is to find an optimal testing strategy. After a brief review of the state-of-the-art of the problem, we describe how techniques based on reinforcement learning (RL) [4] can be alternatively deployed in building such variable-length sequential tests. Using RL, an intelligent agent autonomously learns from interaction with the examinees to administer more informative and effective tests. Empirical results are also provided to evaluate the performance of the proposed approach as compared to fixed-length testing as well as to two existing techniques for variable-length testing.

The rest of this paper is organized as follows. Section 2 reviews the state-of-the-art of sequential mastery testing. Section 3 presents the proposed methodology. Section 4 describes the experimental work. Section 5 presents and discusses the results. Finally Section 6 concludes the paper.

II. RELATED WORK

Most of the earlier work on computer-based testing is concerned with the efficient estimation of an unknown latent trait by observing the examinee's responses to a pre-determined set of test items. The latent trait is represented by a continuous variable, θ, and statistical methods based on item response theory (IRT) are used for modeling the examinee behavior on the test items [3].

978-1-4244-9888-8/11/$26.00 ©2011 IEEE

Another direction, which is the main focus of this paper, aims at classifying examinees into two or more mutually exclusive categories, such as {master, non-master}, {pass, fail}, {excellent, good, average, poor}, or {*A*, *B*, *C*, *D*, *F*}, rather than the accurate estimation of θ. One way to tackle this problem is to treat it as a special case of the first category and after estimating the latent trait, it is compared with pre-defined cutoff points or threshold values (chosen by the test maker). Alternatively, there has been a growing interest in applying sequential decision theory to this mastery testing problem [5, 6, 7, 8]. Two statistical methods that have been intuitively used and intensively studied by many researchers for deriving optimal sequential decision rules are the *sequential probability ratio test* (SPRT) [2, 6, 9, 10, 15] and *sequential Bayesian decision theory* [1, 11, 12].

SPRT is relevant in situations where data is made available sequentially and its first application to designing variable-length mastery tests dates back to [16]. Test items are randomly selected from a calibrated item pool, the examinee's responses to test items are treated as a sequence of independent and identically distributed (*i.i.d.*) Bernoulli trials, and SPRT is used as a test termination criterion. SPRT has been extended and investigated under varying assumptions. For instance, Reckase [9] allowed items of varying difficulty and discrimination characteristics to be considered by using IRT models. SPRT was found to be a very effective tool in modern educational and psychological testing procedures. It leads to reducing the average test length; and for specified error rates, it outperforms the traditional fixed-length test. Typically, one-half to one-third the number of items can be only required [18]. Chang [2] concentrated on the properties of the stopping time (*i.e.* the expected test length) of multiple-category SPRT without the *i.i.d.* assumption of item responses. Weissman [17] provided a general approach for item selection in adaptive multiple-category classification tests using mutual information (MI) along with SPRT. It was shown, through simulation, that this approach classified the highest proportion of examinees correctly and yielded the shortest test lengths as compared with two other item selection methods, namely Fisher information (FI) and posterior-weighted FI (FIP).

On the other hand, Bayesian decision theory with IRT has been used for classifying examinees into two categories [1]. Glas and Vos [19] have used a multidimensional IRT model and Bayesian sequential decision theory for classifying a student as master/non-master or continuing testing and administering another item or testlet (*i.e.* a group of highly related items). Spray and Reckase [20] empirically compared the performance of SPRT with sequential Bayes' methodology. They found that under the considered conditions, the SPRT procedure required fewer test items to achieve the same level of classification accuracy. However, the Bayesian sequential decision strategy for variable-length mastery testing allows costs associated with administering additional items to be taken into account. These costs can be assigned based on psychological, social, or economical consequences of all possible decision outcomes.

III. The Proposed RL-Based Approach

A. RL Background

Reinforcement learning (RL) is a class of machine learning problems and associated solution methods [4]. Under RL, a decision-making agent (controller) interacts with an environment (controlled system) and learns by trial-and-error what to do in each situation (state) in order to optimize a pre-specified objective. In other words, over time the agent learns an optimal policy or strategy (also called control law or decision rule) that maps each state of the environment to an appropriate action. Unlike other machine learning paradigms (both supervised and unsupervised), the agent is not told which action is the best in each state, but instead it autonomously discovers it by trying and evaluating all actions (*i.e.* explore and exploit). The agent-environment interaction is split into stages over finite or infinite time horizon. At each stage, the agent observes the environment state, and selects and performs an action according to the current policy and the exploration criterion. Consequently, the environment state changes either deterministically or stochastically and the agent receives an immediate evaluative feedback (reinforcement signal or payoff) which rewards (positive payoff) or penalizes (negative payoff) the action taken. Based on the expected improvement in the objective function, the tendency to reproduce actions is strengthened or weakened over time accordingly. Thus, the agent progressively adjusts its behavior (policy) based on the received payoffs. This process is repeated until certain stopping criteria are satisfied. The agent's objective is to learn a policy for which some function of the received payoffs is optimized.

RL provides a flexible framework that is widely applicable to several cost-sensitive sequential decision making problems in artificial intelligence, control theory and operations research. It has been applied to numerous problems such as adaptive games (*e.g.* backgammon), elevator scheduling, robot path planning, manufacturing processes and resource allocation in telecommunication networks [4]. RL can be applied in two different modes: online in which the agent interacts directly with the real system and offline in which the agent interacts with a simulation model. The online mode is preferred when it is difficult to model the system or part of it. Moreover, it allows dynamic strategies to be learned when the system operating conditions are changing. The main drawback of this mode is that the agent takes time to learn a good approximation of an optimal policy because at the beginning of interactions no experience is available to the agent. Thus, it may jeopardize the system before learning a good policy.

B. RL-Based SMT Framework

A sequential mastery testing system under RL methodology can be visualized as shown in Fig. 1. A decision-making agent interacts with examinees. The decision maker mainly consists of two modules. The first module (SE) estimates a belief state based on the examinee's responses to previously administered items and prior knowledge. The state for each examinee is estimated sequentially from the previous estimate and the response for the lately administered item. Hence, at each time step n, the agent observes the examinee's current response z_n for the administered item q_n then it will estimate a new state θ_n as function of z_n, q_n, and θ_{n-1}. The second module (π) decides

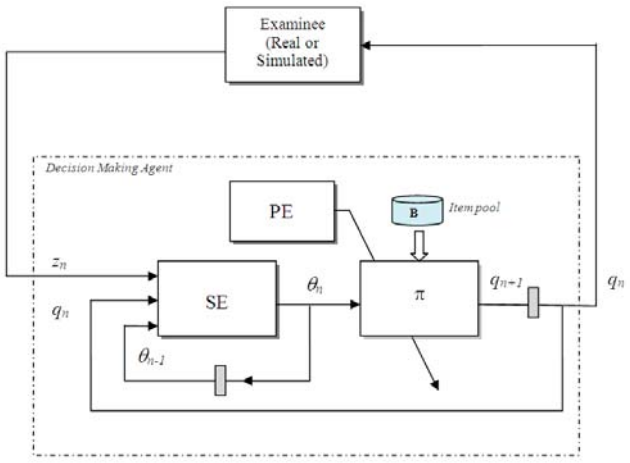

Figure 1. Structure and agent-environment interaction under the RL methodology for sequential mastery testing (SE: State estimator, PE: Policy evaluation, π: Decision Strategy).

whether to stop testing and declare the examinee category or to continue testing. Whenever the decision is to continue testing, this module will select an appropriate item, q_{n+1}, from the set of items in the item pool (bank) that were not previously presented *i.e.* $q_{n+1} \in \mathbf{B} \backslash \mathbf{T}_n$ where $\mathbf{B}$ is the set of items in the item pool, and $\mathbf{T}_n = \{q_1, q_2, \ldots, q_n\}$ is the set of items that were previously presented to the examinee at stage n. The set of all decision rules are known as a decision strategy (*a.k.a.* testing strategy) π, which can be pre-determined and pre-stored, previously approximated and updated online, or completely learned online.

The sequence of interaction with the examinees breaks naturally into sub-sequences (or repeated interactions). Each subsequence represents the interaction with one examinee where system starts at some initial state and ends at a terminal state. This can be illustrated by Fig. 2. The agent starts interacting with examinee j at time t where the state is s^j_t. It takes an action a^j_t which causes the state to change to s^j_{t+1} and receiving a payoff r^j_t. Then, the agent takes an action a^j_{t+1} and the interaction continues until the terminal state s^j_{t+T}. The goal of reinforcement learning is to determine a testing strategy that optimizes some function of received payoffs over the long run. Typical forms of this function are: expected sum of discounted payoffs, expected sum of undiscounted payoff, and average payoff.

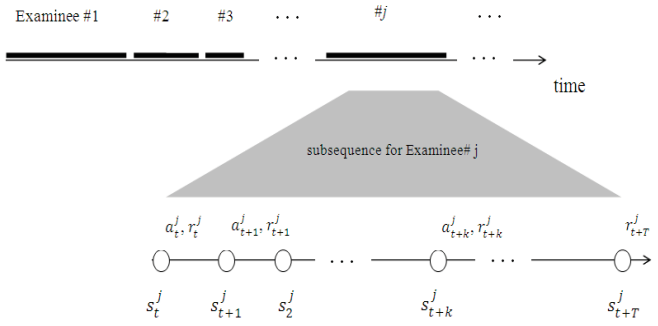

Figure 2. Decision-maker interaction with examinees over time.

C. *Deriving Optimal Testing Strategy*

Here, we consider a case similar to the sequential mastery testing problem for SPRT and Bayesian approaches in [1, 11, 12, 14]. We assume that there are two categories of examinees: w_1 means master and w_0 means non-master. The items follow the Rasch IRT model [3] where the conditional probabilities of item response is $p(z_i = 1|w_0) = \mu$ and $p(z_i = 1|w_1) = 1 - \mu \; \forall i$ where $0 < \mu < 0.5$. Let p_1 denote the prior belief that the examinee is w_1 and p_0 denote the prior belief that the examinee is w_0. Since the examiner does not have enough confidence to decide the examinee category, several items are presented and after each response the examiner updates his/her belief about the examinee. After k items are administered and applying Bayes' rule, the ratio of the posterior probabilities is given by,

$$\frac{p(w_1 \mid z_1, z_2, \ldots, z_k)}{p(w_0 \mid z_1, z_2, \ldots, z_k)} = \frac{p_1}{p_0} \cdot \frac{p(z_1, z_2, \ldots, z_k \mid w_1)}{p(z_1, z_2, \ldots, z_k \mid w_0)} \quad (1)$$

The state of the system is define by the number of correct responses after k items and is denoted by s_k where,

$$s_k = \sum_{i=1}^{k} z_i \quad (2)$$

Thus, it can be shown that (1) becomes,

$$\frac{p(w_1 \mid z_1, z_2, \ldots, z_k)}{p(w_0 \mid z_1, z_2, \ldots, z_k)} = \frac{p_1}{1 - p_1} \cdot \left(\frac{\mu}{1 - \mu} \right)^d \quad (3)$$

where $d = 2k - s_k$ is the difference between the number of correct responses and the number of wrong responses. It is clear that the posterior probability depends only on d but not on k. The systems state is completely defined by d. If the system is in a non-terminal state d and one more item is presented to the examinee, then based on his response the next state will be d+1 or d-1 depending on whether the response is correct or incorrect, respectively. The state transition diagram is as shown in Fig. 3. In each state there are three actions: continue, declare master, declare non-master. If the decision maker chooses to continue, a payoff equal to the cost of administering one item is incurred. If the decision maker chooses to stop and declare the mastery level, a classification cost is incurred based on whether the classification is correct or incorrect. This system satisfies the Markov property as the next state is determined by the current state and the action taken.

A testing strategy can be defined by a single parameter Δ which defines the terminal states such that if $|d| < \Delta$, the decision maker continues testing; otherwise it stops and declares the examinee's category as w_1 if $d = \Delta$ and as w_0 if $d = -\Delta$. Thus, given Δ, the decision rule at state d is defined as follows:

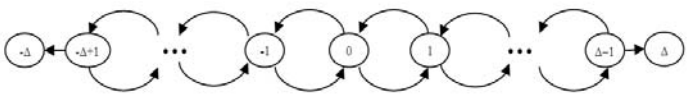

Figure 3. State transition diagram for the testing strategy.

$$\pi^{\Delta}(d)=\begin{cases} w_1 & d=\Delta \\ w_0 & d=-\Delta \\ continue & |d|<\Delta \end{cases} \quad (4)$$

Given the system state, the predictive distribution of next item response is defined by the transition probabilities. If the system state is i and the continue testing decision is applied then the probability that the next state is j is given by,

$$p_{ij}=\begin{cases} p(w_1\,|\,i)(1-\mu)+(1-p(w_1\,|\,i))\mu & i=-\Delta,...,\Delta-1,\ j=i+1 \\ p(w_1\,|\,i)\mu+(1-p(w_1\,|\,i))(1-\mu) & i=\Delta,...,-\Delta+1,\ j=i-1 \\ 0 & otherwise \end{cases} \quad (5)$$

The expected cost is given by,

$$J^{\Delta}=p_1[c_{11}p(w_1\,|\,\Delta)+c_{01}(1-p(w_1\,|\,\Delta))]+ p_0[c_{00}p(w_0\,|\,-\Delta)+c_{10}(1-p(w_0\,|\,-\Delta))] \quad (6)$$

The expected number of items to get this expected cost given that you are currently in state d and using strategy Δ is calculated as follows,

$$N^{\Delta}(d)=\begin{cases} 0 & d=\Delta \\ 0 & d=-\Delta \\ 1+p_{d,d+1}N^{\Delta}(d+1)+p_{d,d-1}N^{\Delta}(d-1) & otherwise \end{cases} \quad (7)$$

Thus, the average cost per item is,

$$\rho^{\Delta}=\frac{J^{\Delta}}{N^{\Delta}(0)} \quad (8)$$

The optimal strategy is defined by Δ^* that minimizes the average cost per item, where,

$$\Delta^*=\arg\min_{\Delta}\{\rho^{\Delta}\} \quad (9)$$

In order to determine the optimal testing strategy using RL, the decision maker defines a utility function $Q^{\pi}(s, a)$ which determines the expected value of taking action a in state s under policy π. The Q-values are determined using the expected value as follows:

$$Q^{\pi}(s,a)=E_{\pi}\{R_t\,|\,s_t=s,a_t=a\} \quad (10)$$

If $Q^*(s,a)$ is the value of taking action a in state s under the optimal policy, which we denote as π^*, then the best action to be taken in state s is the one that has the minimum action value and is given by,

$$a^*=\arg\min_{a}\{Q^*(s,a)\} \quad (11)$$

The determination of Q^* values depends on solving the Bellman optimality equations. Alternatively, we estimate them for all state-action pairs using a similar approach to R-learning [4],

$$Q(d,a)=(1-\delta_1)Q(d,a)+ \delta_1[c-\rho+\min_{a'}\{Q(d',a')\}] \quad (12)$$

where ρ is an estimated average, given by

$$\rho=(1-\delta_2)\rho+ \delta_2[c+\min_{a'}\{Q(d',a')\}-\min_{a}\{Q(d,a)\}] \quad (13)$$

and δ_1, d δ_2 are learning step parameters such that $0 < \delta_1, \delta_2 < 1$. The outline for the algorithm is as shown below:

```
Given the maximum test length N
Set δ1, δ2 to small values
Initialize Q(d, a) arbitrarily for all |d| < N, a ∈ {continue, declare master, declare non-master}
Initialize Q(d, a) arbitrarily for all |d| = N, a ∈ {declare master, declare non-master}
Initialize ρ
For (each examinee)
  Set k = 0;
  Set current state d = 0;
  While (k < N) do
     Select an action a using an ε-greedy
     Apply action a and observe next state d' and cost c
     Update Q values
     If Q(d,a) = min_a{Q(d,a)} then update ρ
     d = d'  // change the current state
  End While
End For
```

IV. Experimental Work

A. Data Preparation

As in [1, 11, 12, 14], we use *simulated data* to evaluate and compare the performance of different testing methodologies. We consider an item pool of 100 items; each of which is described by a two-parameter logistic model (2PL) with discrimination parameter equal to one and difficulty parameter of zero. Item responses are generated for 1000-simulated examinees drawn from a normal distribution N(0, 1). For each item and examinee, a random number from a uniform distribution U(0, 1) is generated and compared with the probability of correct response. If the probability of correct response is greater than the random number, the examinee's response to this item is considered correct; otherwise it is incorrect. An examinee level is assumed to be mastery if his/her ability is higher than a pre-specified threshold value (taken to be zero in our case) otherwise it is assumed to be non-mastery. If items have different characteristics, then we can use the 3-parameter logistic (3PL) IRT model instead to describe the probability of getting a correct response.

B. Performance Measures

The performance of a mastery testing methodology can be visualized using a confusion matrix or contingency table. It shows the actual and predicted classification of each category. Let A denotes to the number of non-master examinees that are failed (*i.e.* true negative), B denotes to the number of master examinees that are falsely failed (*i.e.* false negative), C denotes to the number of non-master examinees that are falsely passed (*i.e.* false positive), and D denotes to the number of master examinees that are correctly passed (*i.e.* true positive). In our experimental work, we have used the following statistical measures.

- *Overall Classification Accuracy* (*Acc*): It measures the effectiveness of the test in terms of the proportion of correctly classified examinees. It is defined as,

$$Acc = \frac{A+D}{A+B+C+D} \tag{14}$$

- *False positive rate (type-I error)*: The proportion of non-master examinees that passed the test. It is defined as,

$$FPR = \frac{C}{A+C} \tag{15}$$

Its complementary measure is known as *test specificity* and is given by,

$$SPC = 1 - FPR = \frac{A}{A+C} \tag{16}$$

- *False negative rate (type-II error)*: The proportion of master examinees that failed the test. It is defined as,

$$FNR = \frac{B}{B+D} \tag{17}$$

Its complementary measure is known as *test sensitivity*, *recall* or *true positive rate*; it is given by,

$$SNS = 1 - FNR = \frac{D}{B+D} \tag{18}$$

- *Test precision (or positive predictive value of the test)*: It measures the fraction of passed examinees that certainly are master. It is defined by,

$$\text{Precision} = \frac{D}{C+D} \tag{19}$$

- *Phi-correlation*: The correlation between the true and predicted categories is computed using phi-correlation coefficient as given by,

$$\phi = \frac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \tag{20}$$

These metrics are often expressed as percentages. Other important metrics, that we have explored, are the average test length and the proportion of examinees classified at the end of the test. These metrics are defined as follows:

- *Average test length*: For variable-length tests, the average number of items that must be administered is an important performance measure. It is given by,

$$T_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} T_i \tag{21}$$

where N is the number of examinees and T_i is the test length for the i-th examinee.

- *Proportion classified*: the proportion of examinees for whom a classification decision is made (regardless of its correctness).

We also studied the impact of the maximum test length on all these metrics. It is important to note that a test is required to have high accuracy, specificity, sensitivity, precision, correlation, and proportion classified. On the other hand, it should have low false positive rate, false negative rate, and average number of items. Also, the compromise of the trade-off between false positive rate and false negative rate (or equivalently between specificity and sensitivity) depends on the purpose of the test and the preference of the domain expert.

C. Experiments

In this set of experiments, the above-mentioned performance measures are evaluated for fixed-length conventional tests (CT), SPRT-based sequential tests, Bayesian-based sequential tests, and RL-based sequential tests. Various experiments are conducted and performance measures are computed as a function of the maximum test length in the range of 2 items up to 100 items. For each testing method and maximum test length, we repeat the experiment 100 times (each time a test is randomly generated and administered). The details of the settings for each testing method is as described below.

Fixed-Length Conventional Tests (CT). For a given test-length, a fixed set of items is randomly drawn from the 100-item pool and is used to test all examinees in a conventional manner. The examinee's ability is estimated after the test and compared with a cutoff point to declare master or non-master. Since observations are assumed to be independent and identically distributed (*iid*) random variables, the proportion of correct responses is used instead. For a test of length n, the response pattern is $(z_1, z_2, \ldots, z_n)$ where $z_i \in \{1, 0\}$. The number of correct responses is s_n. The cutoff threshold is set to 0.6, *i.e.* an examinee is declared master if he/she correctly answers 60% or more of the administrated items and declared non-master otherwise. Thus the decision rule is given by,

$$\pi(s_n) = \begin{cases} s_n \geq 0.6n \Rightarrow master \\ s_n < 0.6n \Rightarrow non-master \end{cases} \tag{22}$$

Then the predicted category is compared with the true category for each examinee.

SPRT-Based SMT. Here, the system sequentially observes the examinee's responses to administered items until enough information is gained to stop testing and make a classification decision or until a maximum number of items N are administered. Recall that we run experiments for different values of N from 2 up to 100. In each experiment, we apply Wald's SPRT after each item response to determine whether to terminate the test and declare the examinee's category or to continue testing. The error rates α and β are set to 0.05. Since SPRT can terminate after N without identifying the category of the examinee, we use an additional performance measure, the proportion of examinees classified. The accuracy is computed as the proportion of the classified examinees that are correctly classified (*i.e.* excluding unclassified examinees).

Bayesian SMT. We conducted a number of experiments using Bayesian sequential strategy for mastery state testing. As in [11], we assumed a binomial distribution for modeling the response behaviour and a beta distribution for the prior knowledge about the true state. The policy is derived using backward induction. Similar to SPRT, the system sequentially observes the examinee's responses to administered items until enough information is gained to stop testing and make a classification decision or until a maximum number of items are administered. The decision is now made based on the expected cost. The cost incurred for each additional item, e, is set to 1, the costs of correct classification are to zeros (*i.e.* $c_{11} = c_{00} = 0$), the costs of misclassifications are $c_{10} = c_{01} = 100$.

RL-Based SMT. Using RL, we conducted a number of experiments for mastery state testing similar to Bayesian SMT. However, instead of determining the policy a priori and then deploys it during the test, the decision maker sequentially learns the policy during the interaction with the examinees. We used *Q*-learning approach with similar costs as Bayesian SMT, *i.e.*, $e = 1$, $c_{11} = c_{00} = 0$, and $c_{10} = c_{01}$=100. We repeated the experiments for $N = 2$ up to N =100.

V. Results and Discussions

The average performance measures are computed and compared as shown in Figs. 4 to 10 for each of the following testing strategies: Fixed-length conventional tests (CT), SPRT-based SMT, Bayesian-based SMT, and RL-based SMT. Fig. 4 compares the percentage of classified examinees as a function of the maximum test length. As depicted in this figure, all strategies except SPRT can classify all examinees. The reason is that SPRT can terminate early before reaching to a decision because of the constraint on the maximum length. However, the other techniques always make a classification based on the information available to the decision maker. As shown in Fig. 5, the classification accuracy initially increases as the number of administered items increases except for SPRT. However, as the maximum test length reaches certain value, the rate of increase starts to decrease because the costs of items compromise the gains from correct classifications. The reason that the curve for SPRT is decreasing is that the unclassified examinees are considered as incorrectly classified. The precision is compared in Fig. 6 versus the maximum test length. Again we can notice that as the maximum length increases the precision improves but beyond certain number of items the rate of improvement starts to decrease. Fig. 7 and 8 show the false positive and false negative performance metrics. We can see that although the fixed-length tests have less false positive rates, they have very high false negative rates. On the other hand, SPRT has less false negative but higher false positive. The Bayesian approach has almost the same rate for false positive and false negative. Finally, we can see that RL approach has better false positive and false negative than the Bayesian approach. The average test length decreases drastically for variable length approaches as compared to the fixed-length traditional tests, see Fig. 9. Although RL has a bit higher average test length than the SPRT and Bayesian approaches, the improvement in other metrics, *e.g.* the accuracy in Fig. 5 and the correlation in Fig. 10, is justified.

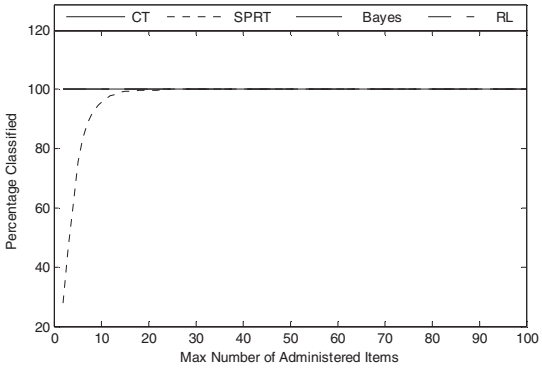

Figure 4. Percentage of classified examinees vs. the maximum test length.

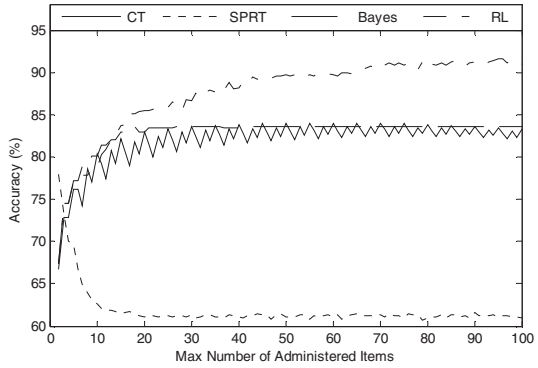

Figure 5. Classification accuracy as function of the maximum test length.

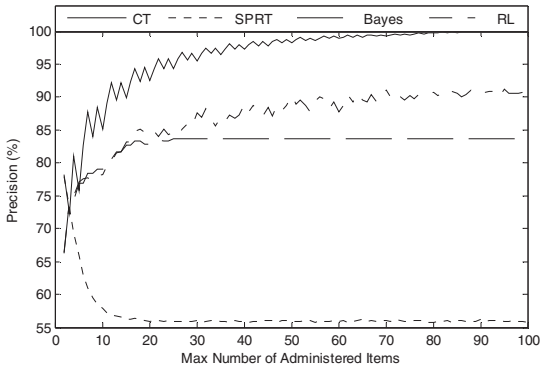

Figure 6. Precision as function of the maximum test length.

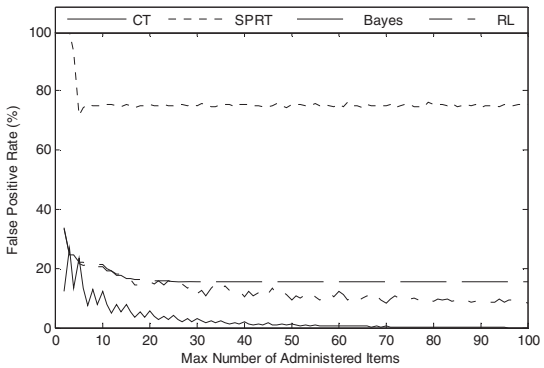

Figure 7. False positive rate as function of the maximum test length.

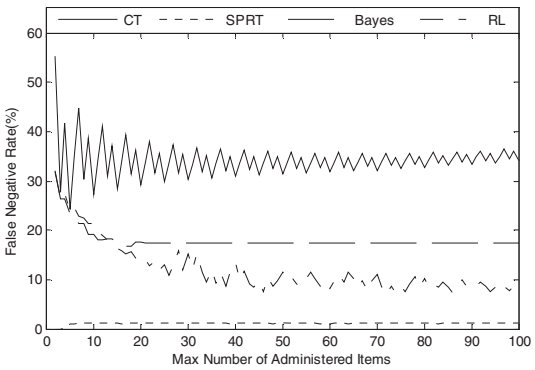

Figure 8. False negative rate as function of the maximum test length.

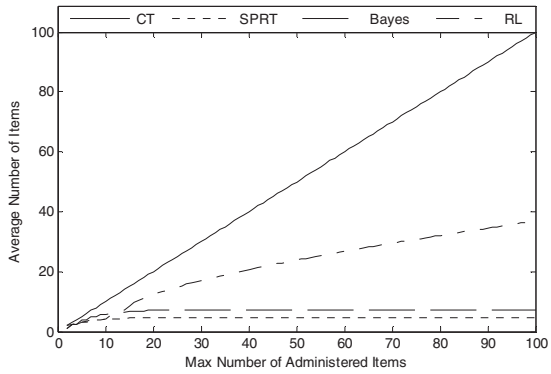

Figure 9. Average test length as function of the maximum test length.

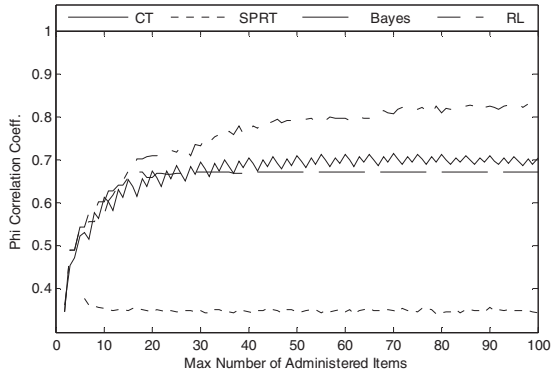

Figure 10. Phi-correlation as function of the maximum test length.

VI. CONCLUSIONS

In this paper, we formulated the sequential mastery testing problem as a reinforcement learning problem. Then, after reviewing the state-of-the-art methodologies for solving it, we provided a general framework based on the application of reinforcement learning as an alternative. The goal of sequential mastery testing is to tailor the number of test items to each examined person's ability level. This allows the system to build more informative and effective tests using reduced test lengths, on average, and consequently reducing the costs incurred in administered test items and the costs of making a classification. Different performance metrics have been computed and compared. The results attained for the reinforcement learning approach are promising as compared to other techniques. Moreover, the existing techniques depend on the derivation of analytical models which is possible under simplifying conditions. However, as the problem gets into its general settings, the derivation of analytical models become too complicated and infeasible. This motivates further investigation of the problem under various conditions using reinforcement learning techniques.

ACKNOWLEDGMENT

The author would like to acknowledge King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia for the support and sponsorship during this work under Grant Number JF050001.

REFERENCES

[1] H. J. Vos, "Applications of Bayesian decision theory to sequential mastery testing," *Journal of Educational and Behavioral Statistics*, vol. 24, no. 3, 1999, pp. 271-292.

[2] Y. Chang, "Application of sequential probability ratio test to computerized criterion-referenced testing," *Sequential Analysis*, vol. 23, no. 1, 2004, pp. 45-61.

[3] F. Baker, and S.-H. Kim, Item response theory: Parameter estimation techniques, 2/e. CRC, 2004.

[4] R. S. Sutton, and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[5] B. K. Ghosh and P. K. Sen, *Handbook of Sequential Analysis*. Marcel Dekker, Inc., New York, 1991.

[6] J. Bartroff, M. Finkelman and T. L. Lai, "Modern sequential analysis and its applications to computerized adaptive testing," *Psychometrika*, vol. 73, no. 3, 2008, pp. 473-486.

[7] J. Bartroff, M. Finkelman, and T. Lai, "Modern sequential analysis and its applications to computerized adaptive testing," *Psychometrika*, 2007.

[8] E.-S. M. El-Alfy, Computerized adaptive testing methodologies. Technical Report, KFUPM-CCSE-2006-001/ICS, King Fahd University of Petroleum and Minerals, KSA, 2006.

[9] M. Reckase, "A procedure for decision making using tailored testing," In D. J. Weiss, Ed. *New Horizons in Testing – Latent Trait Test Theory and Computerized Adaptive Testing*, Academic Press, New York, 1983, p. 238-257.

[10] T. Eggen, "Item selection in adaptive testing with the sequential probability ratio test," *Applied Psychological Measurement*, vol. 23, 1999, pp. 249–261.

[11] H. J. Vos, "A Bayesian sequential procedure for determining the optimal number of interrogatory examples for concept learning," *Computers in Human Behavior*, vol. 23, 2007, pp. 609-627.

[12] H. J. Vos, "A Monto Carlo simulation to sequential mastery testing in education using a backward induction computational procedure," in Proc. of the 14th European Simulation Multiconference on Simulation and Modelling, 2000.

[13] Y. Chang and Z. Ying, "Sequential estimation in variable length computerized adaptive testing," *Journal of Statistical Planning and Inference*, vol. 121, 2003, pp. 249-264.

[14] L. M. Rudner, "An examination of decision-theory adaptive testing procedures," *Paper presented at the annual meeting of the American Educational Research Association*, New Orleans, LA, 2002.

[15] T. Eggen, "Item selection in adaptive testing with the sequential probability ratio test," *Applied Psychological Measurement*, vol. 23, no. 3, 1999, pp. 249-261.

[16] R. L. Ferguson, The development, implementation, and evaluation of a computer assisted branched test for a program of individually prescribed instruction. PhD Thesis, University of Pittsburgh, Pittsburgh PA, 1969.

[17] A. Weissman, "Mutual information item selection in adaptive classification testing," *Educational and Psychological Measurement*, vol. 67, no. 1, 2007, pp. 41-58.

[18] C. W. Baum, and V. V. Veeravalli, "A Sequential Procedure for Multihypothesis Testing," *IEEE Transactions on Information Theory*, vol. 40, no. 6, 1994, pp. 1997-2007.

[19] C. A. W. Glas, and H.J. Vos, "Testlet-based adaptive mastery testing," Computerized Testing Report 99-11. LSAC Research Report Series, 2006.

[20] J. Spray and M. Reckase, "The selection of test items for decision making with a computer adaptive test," *Paper presented at the Annual Meeting of the National Council on Measurement in Education*, New Orleans, LA, 1994.