AUTOMATED ASSESSMENT OF KNOWLEDGE AND SKILL ACQUIRED BY E-LEARNERS THROUGH ADAPTIVE TESTING

Nandakumar G S
Associate Professor, Department Of Computer Science And Engineering, Kumaraguru College Of Technology, Coimbatore, Tamilnadu, India
sunandakumarr@gmail.com

Thangasamy S
Professor, Department Of Computer Science And Engineering, Kumaraguru College Of Technology, Coimbatore, Tamilnadu, India
stsamy2001@yahoo.com

Geetha V
Professor, Department Of Computer Applications, Kumaraguru College Of Technology, Coimbatore, Tamilnadu, India
geethasuresh@rediffmail.com

Abstract: Automated assessment of very large numbers of e-learners registered for popular online courses is a critical issue in the process of e-learning. Due to the flexibility offered to e-learning through online courses, the tests taken by learners at their own chosen time schedules with varying levels of preparedness need to be designed to provide reliable and fairly accurate measurement of the learners' proficiency, the latter found to vary over a wide range. The experience with the adaptation strategy employed in the test scheme adopted in a particular course, the question bank creation and validation issues and the test score calculation options are all discussed with their relative merits clearly brought out.
Keywords: Adaptive Testing, E-Assessment, Test Strategy

INTRODUCTION

Universities aiming to choose meritorious students for admission into their elite programs or employers attempting selection of right candidates from a vast pool of applicants often resort to automated tests on basic knowledge and skill in specified essential subject areas. Objective type questions (often through selection from multiple choice answers), randomly picked from a large question bank on the specified subject have been used for this purpose for a long time. The increasingly popular e-learning or flexible learning in educational institutions also demand a shift from the conventional fixed timetable examinations with single question paper for a whole class or for a whole university. The new learners are more comfortable with flexible examinations to assess the knowledge and skills of the learners in their own convenient time schedules. Automated testing with focused efforts to generate sufficiently large question banks for diverse subjects is found essential to enable such learner assessment. A survey conducted among a student community about online assessment focusing on affective factors, validity, practical issues, reliability, security and teaching & learning shows positive expectations from e-assessment (John Dermo, 2009). Experience with E-assessments used across the world establishes the advantages of this approach over traditional assessments/examinations. (SIR, U. S. A. OSUJI, 2012)

Often candidates with widely varying levels of proficiency in a given subject area need to be tested within a finite time duration by a single test, still obtaining a realistic quantification of their attainments. This has been found feasible, if the candidates with high initial success rates are posed with questions of gradually increasing conceptual complexity. Such adaptive testing can utilize the test duration with maximum proportion of time spent in challenging the candidates at their own right level of proficiency. If a candidate answers correctly for a set of questions, then the next question will be of higher difficulty level. (Elena Papanastasiou, 2015). The experience gained in such testing of undergraduate engineering students' proficiency in computer program implementation in 'C' language is described here with multiple strategies for classification of learners. Rough Set Theory has been used earlier to model the students' performance in the E-Assessment data and to generate classification rules (Nandakumar G.S. et.al 2014).

QUESTION BANK CREATION

The success of any adaptive test is largely dependent on the quality and quantity of the question bank from which the administered items are drawn. The total number of questions in the bank must be sufficient to probe depth of knowledge on each of the topics in the subject area chosen. A good question bank should have enough

 Copyright © The Online Journal of Distance Education and e-Learning

questions to attain high measurement precision while assessing candidates with a wide range of variation in knowledge levels. This criterion essentially means that there should be sufficient number of questions at all levels of complexity. Calibrating the question bank to know the measurement characteristics of individual question is a necessary part of large E-Assessment. It is useful to assign a difficulty level to each question in the bank depending on the fraction of a large population of candidates correctly answering the question. A high-quality question bank will contain sufficient numbers of questions at each difficulty level, to evaluate students at varying levels of proficiency. If the bank is large enough, the chances of each student receiving a distinct subset of questions are enhanced. Students' motivation towards learning, subject areas, programs, characteristics of the online delivery format and other imposed constraints are the factors usually considered in the design of questions for the online tests.(Shijuan Liu, Retrieved March 12, 2016)

CALIBRATION AND VALIDATION OF QUESTION BANK

Every item in the question bank has to be calibrated before being used in the tests. The accuracy of calibration will have a direct impact on the reliability of the test as reflected in the scores of the candidates. A question bank consisting of several hundred multiple choice questions on C-Programming language features was created by collecting questions from multiple course experts. A conventional test was initially conducted where each of the questions in the question bank has to be answered by a large population of students. Calibration of questions was done with the proportion of the examinees who answered each question correctly to the total population, based on which the questions were initially classified into five levels of Degree of Toughness (DT) (Nandakumar G.S. et.al 2014). (1 – very simple and 5 – very tricky) as shown in Table-1.

DT	Classification
1	Very simple
2	Below average
3	Average
4	Tricky
5	Very tricky

Table 1. DT (Degree of Toughness) level and classification

DT	% Answered Correctly	No. of Questions
5	0 – 10	72
4	11 – 29	120
3	30 – 49	146
2	50 – 69	118
1	70 – 100	117

Table 2. Initial classification of questions into DT levels

The initial classification of each question into right DT levels is based on the proportion of correct answers from a large body of learners attempting the entire question bank. The outcome is shown in Table 2. The DT of an individual question has to be updated periodically, after a broad spectrum of students undergoes the tests and the question has been asked sufficiently large number of times.

ADAPTATION STRATEGY

The interesting aspect of this model is that it allows a student to choose initial DT level of the questions soon after logging into the system of examination. If he opts for the k^{th} DT (k=1, 2, 3, 4, 5) the system will start displaying the questions randomly from the chosen DT, for which the candidate selects answers from the given multiple choices.

The following is the algorithmic strategy followed for transitioning into adjacent DT levels during the test progress:

 Copyright © The Online Journal of Distance Education and e-Learning

Case 1: If the candidate answers the first three questions of the k^{th} DT level correctly, the system will shift to $k+1^{st}$ DT level provided $k \neq 5$. When $k = 5$, the system continues to ask questions from the same level.

Case 2: In case the candidate answers all the three questions of the k^{th} DT incorrectly, the system will shift to $k-1^{st}$ DT provided $k \neq 1$. For k=1, the system continues to display from the 1^{st} DT level irrespective of the number of wrong answers provided.

Case 3: If the examinee correctly answers only one or two out of the first three questions from the k^{th} DT level, the system provides one more question from the same DT.

Three correct answers out of a total of four questions leads to a shift to $k+1^{st}$ DT level, provided $k \neq 5$.

Three wrong answers out of four questions leads to a shift to $k-1^{st}$ DT level, provided $k \neq 1$.

In case examinee answers two out of these four questions correctly from the k^{th} DT level, one more question from the same DT level is given. A total of three correct answers out of five given questions, shifts to $k+1^{st}$ DT; otherwise shift is made to $k-1^{st}$ DT. However such shifting to next higher or lower DT does not take place when k=5 or k=1 respectively.

SCORE EVALUATION PROCEDURE

The final score is calculated using the relation

$$Score = \sum_{i=1}^{5} w_i n_i$$

Where

n_i – is the number of i^{th} DT level questions correctly answered
w_i - is the weightage associated with i^{th} DT level.

The weightage currently used for questions from each of the DT levels are given in Table 3. It is to be noted from the table that weightage in score calculation increases linearly with the DT level.

DT	1	2	3	4	5
Weighting Element w_i	0.2	0.4	0.6	0.8	1.0

Table 3. Weightage associated with each DT level

The above weightage parameters and the duration of the examination can be set according to the needs of the subject. The test will get terminated either on the expiry of the time frame or the examinee attempting questions for the prescribed maximum score whichever occurs first. The test score, the number of DT level-wise questions asked and correctly answered gets displayed at the end of the test.

SCORE CALCULATION OPTIONS

The above conventional method of score calculation assigns an incremental score for every correct answer, the increment size depending on the degree of toughness (DT) of the question answered. The final scores obtained by a sample of students is given as Total Marks (TM) in Table 4. To penalize answering questions without clear knowledge of the problem (i.e. through mere guess work), negative markings were assigned to wrongly answered questions with 25% and 50% negative marking of the incremental score for each DT level. The revised scores (TMs) with negative markings are also provided in Table 4. When the candidates are sorted in terms of descending TMs, we find multiple candidates sharing the same discrete values.

Stud-id	Total Marks (TM)	TM with 25% -ve	TM with 50%-ve	Success Rate Score
47	8.8	7.3	5.7	222.0
101	8.8	7.3	5.7	119.3
102	8.8	7.3	5.7	58.7
26	8.2	6.5	4.8	120.5
79	8.2	6.5	4.8	116.1
135	8.2	6.5	4.8	106.4
17	8.0	6.3	4.5	144.4
14	8.0	6.3	4.5	122.9
119	8.0	6.3	4.5	95.5
59	8.0	6.3	4.5	93.7
84	8.0	6.3	4.5	63.8
4	8.0	6.3	4.5	60.8
97	8.0	6.3	4.5	60.8
60	7.8	6	4.2	97.7
7	7.8	6	4.2	73.7
130	7.8	6	4.2	62.4

Table 4. Sample list of candidates with marks sorted in the descending order of 'Total marks'

The final scores could be alternatively calculated by following a strategy similar to the one adopted in all conventional (non-adaptive) tests, i.e. based on the proportion of right answers among the total questions posed. Such a score termed as success rate score is defined as

$$Success\ \ Rate\ Score\ = \frac{100}{3}\sum_{i=1}^{5} w_i nc_i \,/\, n_i$$

Where

n_i – is the total number of questions faced by the candidate in ith DT level
nc_i- is the number of correct answers in the ith DT level and
w_i- is the weight associated with ith DT level.

Since the maximum value of the summation is 3 for the assumed w_i values, to normalize the score in the range of (0-100), it is multiplied by 100/3.

The success rate scores included in the above table is found to provide superior discrimination over the entire range of scorers. It is particularly useful for finer discrimination among top scorers. The scatter diagram given in Figure 1 providing the correlation between conventional scores and the success rate scores throws further light on the relative merits of the different score calculation options.

 Copyright © The Online Journal of Distance Education and e-Learning

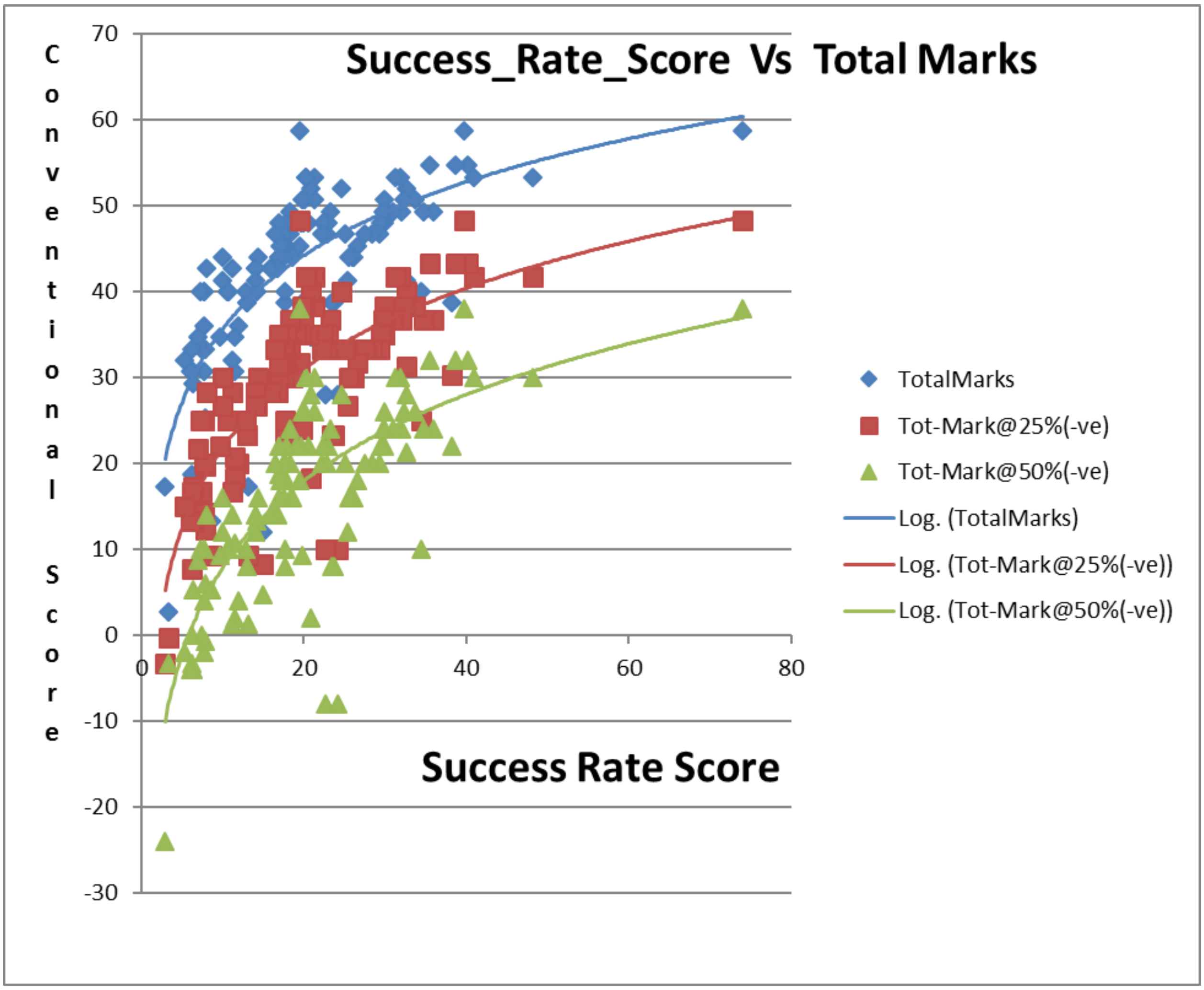


Figure 1: Scatter diagram showing the correlation between conventional scores and Success rate scores

The correlation between success rate score(X) and the total marks(Y) calculated as

$$\text{Correlation} = \text{covariance}(X,Y)/(\sigma_X * \sigma_Y)$$

is only around 0.62. This correlation with X was found to improve marginally with negative marking, 0.68 with 25% negative marking and 0.69 with 50% negative marking.

Since success rate score is similar to conventional proportion of correct answers at each difficulty level, they are to be preferred in most circumstances. However, to get a deeper insight into the profile of a particular candidate, the entire sequence of questions faced at different DT levels along with the consistency with which the candidate scores at different levels can be viewed graphically as shown in Figures 2, 3, 4 and 5. Student no.47 is found to fast climb to higher levels of DT, though he achieves right answers for only about 50% of the questions at DT levels of 3 and above. Student No.101 climbs to a higher level slowly and also makes as many mistakes as the correct answers. Student No. 102 starting at DT 2 is spending all the time in levels 2 and 3. He is able to attempt more number of questions at lower DT levels compared to the earlier two, all three getting the same total marks. But the success rate score for these three students found in the top 3 rows of Table 4 provides very good discrimination that is in conformity with the intuitive assessments of merit from Figures 2 to 5.

 Copyright © The Online Journal of Distance Education and e-Learning

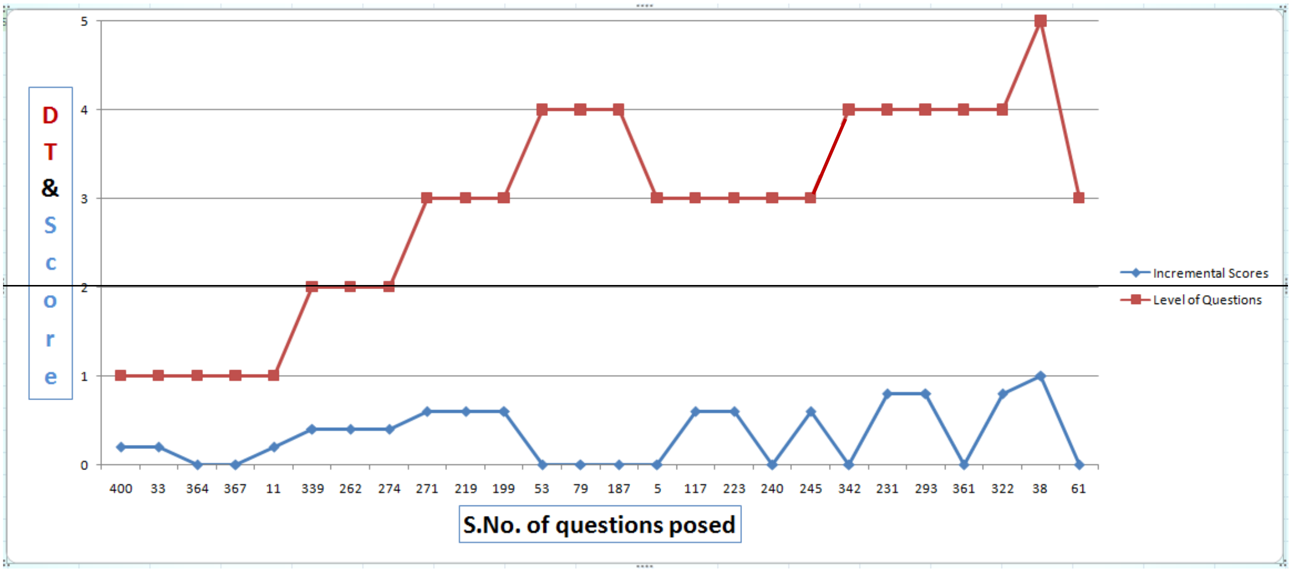


Figure 2: Sequence of questions posed and scores obtained by the candidate No.47

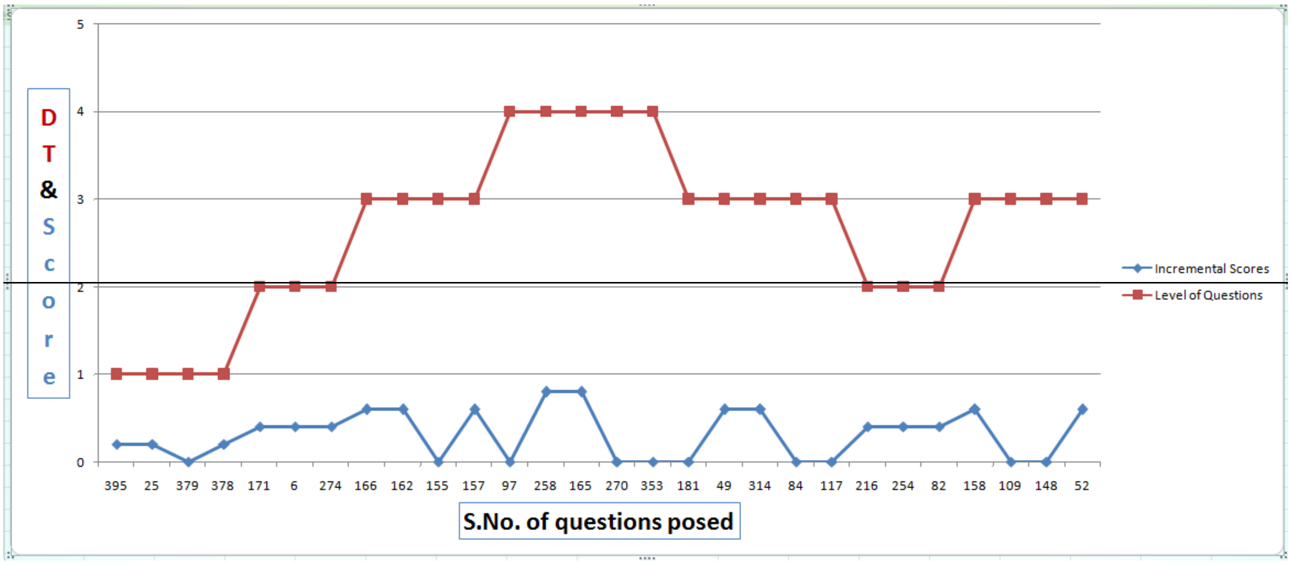


Figure 3: Sequence of questions posed and scores obtained by the candidate No.101

 Copyright © The Online Journal of Distance Education and e-Learning

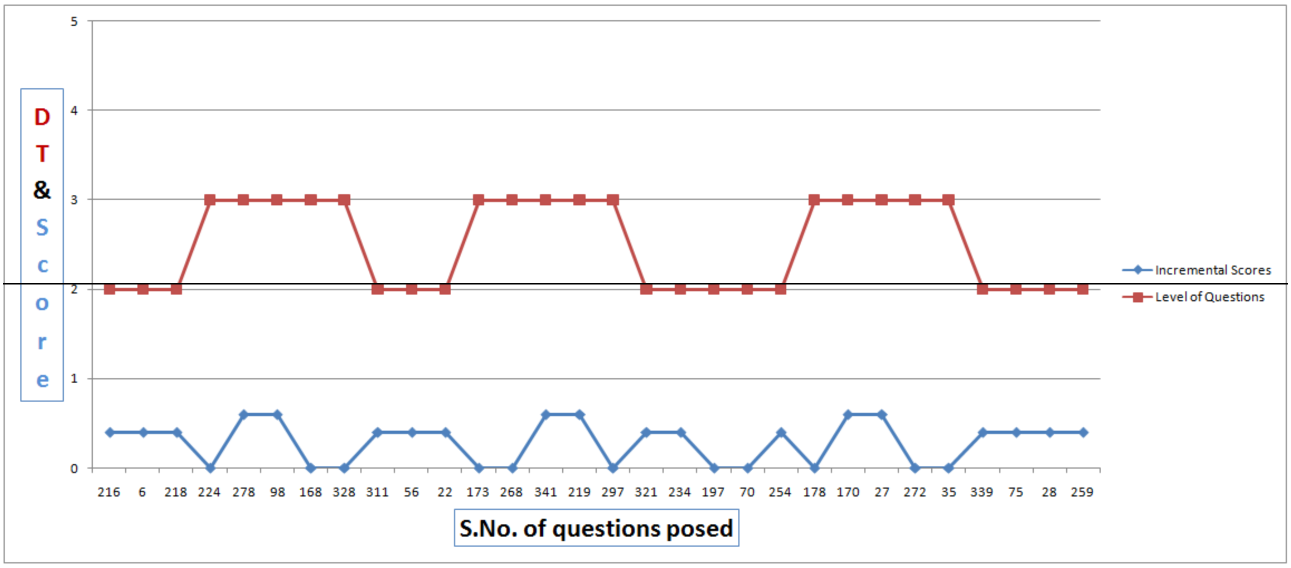
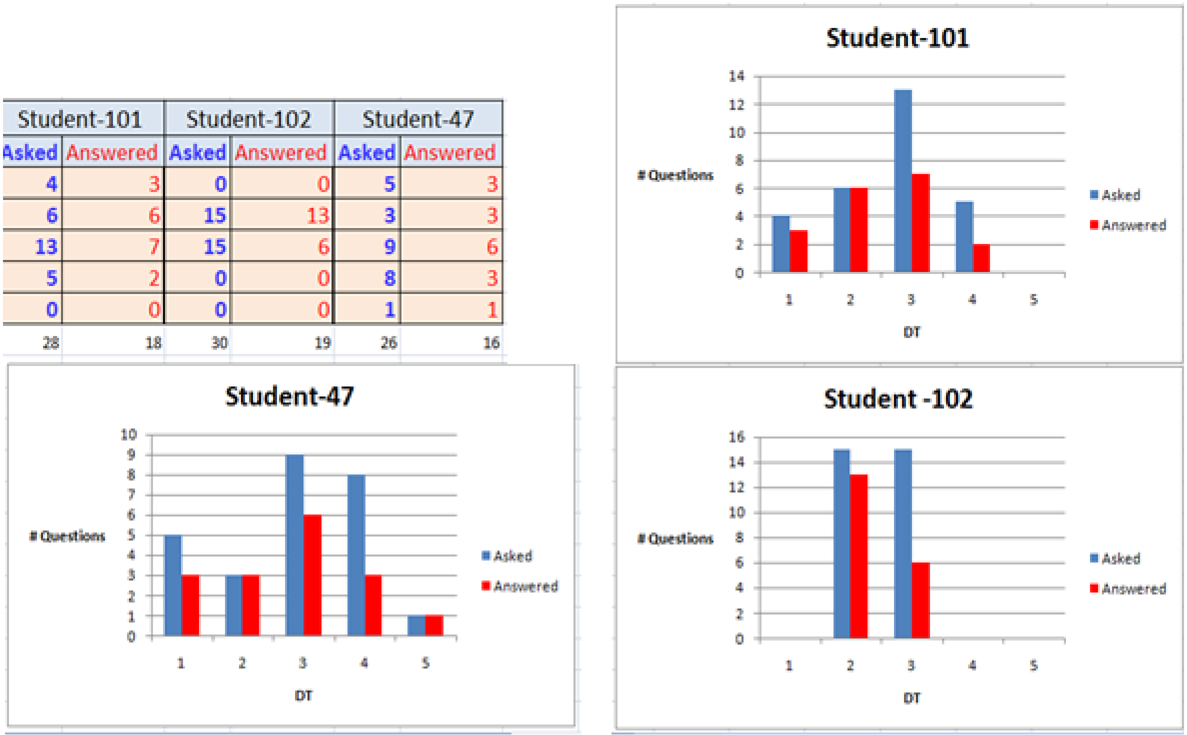


Figure 4: Sequence of questions posed and scores obtained by the candidate No.102



Student-101		Student-102		Student-47	
Asked	Answered	Asked	Answered	Asked	Answered
4	3	0	0	5	3
6	6	15	13	3	3
13	7	15	6	9	6
5	2	0	0	8	3
0	0	0	0	1	1
28	18	30	19	26	16

Figure-5: Comparison of scoring patterns by the top three scorers

CONCLUSION

The adaptive strategy made use of to assess the proficiency of several batches of students in one institution has brought out the effectiveness of such strategy for on-line assessment. It gives flexibility to students and provides good discrimination over a range of proficiency levels. Among the score calculation options, the score based on proportion of successful answers to each class of questions is found to be more dependable than the conventional scores with or without negative markings to wrong answers. The present adaptation strategy has been found successful in retaining the interest of the entire range of learners during the test. However the question bank size, quality and classification over difficulty levels need lot of attention. This needs to be done in each of the diverse courses offered in massive online programs. Also the best adaptation strategy might vary with the diversity in the backgrounds of learners for the same course and the subject matter for the course.

REFERENCES

Elena Papanastasiou, (2015), *Adaptive Assessment*, Encyclopedia of Science Education, pp 18-19 Retrieved 12-03-2016 from http://link.springer.com/referenceworkentry/10.1007/978-94-007-2150-0_3

John Dermo, (2009), *E-Assessment and the student learning experience: A survey of student perceptions of e-assessment*, British Journal of Educational Technology. Vol.40 No.2, p-203-214

Nandakumar. G.S., Thangasamy. S, Geetha.V, & Surendiran. B, (2014). "*A Rough Set Based Classification Model for Grading in Adaptive E-Assessment*", International Review on Computers and Software (IRECOS) Vol. 9, No 7

Shijuan Liu, "*Factors In Design Of Assessment For Online Courses: Instructors' Reflections*", The Online Journal of Distance Education and e-Learning Vol. 2, Issue 3. Retrieved March 12, 2016 from http://www.tojdel.net/pdf/v02i04/v02i04-11.pdf

SIR, U. S. A. OSUJI, (October 2012). "*The Use Of E-Assessments In The Nigerian Higher Education System", Turkish Online Journal of Distance Education*-TOJDE, Vol. 13 Number: 4 Article 9

 Copyright © The Online Journal of Distance Education and e-Learning