

Performance Evaluation of Students in Adaptive E-Assessment using Structural Equation Modeling

V. Geetha, V. Jalaja Jayalakshmi, R.Vivek



ABSTRACT-- In today's competitive world, students have to learn many skills and assessments are designed to judge their abilities in a more scientific way. Adaptive E-Assessment is a popular mode of evaluating the performance of the learners. It is a reliable approach and is commonly used by educational institutions around the world. An adaptive strategy using multiple choice questions for conducting E-assessment has been formulated for assessing the knowledge level of students. Various statistical measures were used to compare the performance of the students. This study uses structural equation modelling (SEM) to identify the relationship between the various factors contributing to the test score and to assess the reliability of the relationships between these factors.

Keywords - Adaptive E-Assessment, Multiple Choice Questions(MCQ), Degree of Toughness (DT), Structural Equation Modeling (SEM)

I. INTRODUCTION

The growth of information technology has increased the popularity of web based education; particularly E-Assessment. Assessment is the process of measuring learning and is one of the key issues in modern education. It is an integral part of the learning process which is related to learning outcomes. The goal of assessment is the estimation of the knowledge that has been acquired by the students via learning. Automated testing systems produce great deal of information which can be used to analyze the performance of the learners. These assessments are created according to teachers' specifications and are adaptive, that is, the questions are selected wisely to fit the student's level of knowledge. Different sets of questions are generated for different students keeping their zeal to face the test progressively. In this way, more accurate estimations of student's knowledge with significantly shorter tests are obtained. In adaptive assessment, the difficulty level of questions adapts dynamically based on the answering pattern of the students and the questions are generated based on the individual ability of the students.. The students are

assessed with fewer questions and accuracy of assessment is an important criteria. It is also easy to discriminate the performance of the students using adaptive strategy and this will enhance the effectiveness of the assessment.

A real time online system has been designed for selecting and admitting candidates to professional education. They have associated a toughness factor on a five point scale, but the questions generated for each student is of the same standard [1]. Ang Tan Fong et al [2] have also included the concept of difficulty level of question in the evaluation system. They have designed an intelligent online system in which a student is given a pretest with different difficulty levels. The scores of this pretest are used as the starting estimated ability value of the student. When the student takes up the actual test, questions are generated based on the estimated ability. This ability gets updated based on the student's performance in the test. Item Response Theory (IRT) has been applied for student ability estimation in E-learning and K-means clustering algorithm has been used to group the learner's abilities [3]. Catherine C. Marinagi et al [4] have proposed architecture for an adaptive assessment tool using client server platform. Bayesian decision model was used to select the questions for the test. They have also demonstrated a framework [5] for adaptive multiuser web based assessment using Bayesian decision theory and then to classify the students into categories such as "masters" and "non-masters". Mariana Lilley [6] in her PhD thesis has investigated the issues on the use of computers in assessment in higher education of U.K using IRT to estimate students' knowledge. Thomas Marie Huffmaster [7] has compared hierarchical linear modeling and structural equation modeling for modeling changes in response latency (timing) patterns on a high-stakes adaptive test.

Structured Equation Modeling (SEM) is a multivariate analysis technique visualized by a path diagram. It can be viewed as a combination of factor and path analysis. Structural equation modeling was used to develop separate models for the two sets of students and to assess the reliability of the model. SEM not only estimates multiple interrelated dependence relationships but also has the ability to handle latent variables [8]. Structural equation modeling was chosen because it examines a series of dependence relationship simultaneously [9]. It is a straight forward method of dealing with multiple relationships while providing statistical efficiency. SEM uses path diagrams to depict not only the predictive relationships but also associative relationships among the variables [10].

Manuscript published on 30 August 2019.

* Correspondence Author (s)

V. Geetha, Professor, Dept. of Computer Applications, Kumaraguru College of Technology Coimbatore, T.N, India. (E-mail: geetha.v.mca@kct.ac.in)

V. Jalaja Jayalakshmi, Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, T.N, India. (E-mail: jalajajayalakshmi.v.mca@kct.ac.in)

R.Vivek, Student, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, T.N, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

The main contribution of the paper is to provide an insight into the process of applying statistical techniques and structured equation modeling to analyze the results of E-Assessment and determine the factors that contribute to the final result of the student. An adaptive testing strategy has been formulated to test the skill of students in 'C' programming language .to classify the students according to their ability and Intelligence Quotient (IQ). A great number of multiple choice questions were gathered from many course experts. The questions have been categorized into different difficulty levels and tests were conducted for a large set of students[11][12].

The organization of the paper is as follows: Section 2 discusses about the adaptive strategy developed for E-assessment. Experimental results and analysis are discussed in Section 3. Section4 concludes the work and discusses the scope for future work.

II. ADAPTIVE TESTING STRATEGY

This section describes the strategy used for adaptive E-Assessment and the creation of a Question bank with multiple choice questions using PHP software as the front end and MySQL as the back end[11][12].

A. Question bank creation

The success of any adaptive assessment generally depends on the quality of the question bank. The total number of questions in the bank must be sufficient to supply informative items throughout the assessment session. A question bank relating to 'C' programming language was formed by gathering questions from course experts. A standard test (where all the questions in the question bank have to be answered by all the students) was given to a group of students. Assigning the difficulty level of questions was done with the fraction of the students who answered each question correctly to the total number of students who took the initial test, based on which the questions were primarily classified into various levels of DT ranging from 1 to 5 (1–very easy and 5 – very difficult) as shown in TABLE I.

TABLE I. Initial Classification of Questions

DT	Percentage Answered correctly	Number of Questions
5	0 – 10	76
4	11 – 29	80
3	30 – 49	97
2	50 – 69	87
1	70 – 100	98

Each question in the question bank is allocated with a DT and it has to be updated from time to time, after a large number of students undergo the tests and the question has been asked adequately a large number of times. A difficult question is allotted a higher weightage than an easy question.

B. The Adaptive Strategy Model

The procedure for conducting online tests using adaptive strategy is discussed below. The important aspect of this model is that it allows the student to primarily opt for the

DT of the questions as soon as he/she logs into the system of assessment. If he opts for the i^{th} DT ($k=1, 2, 3, 4, 5$) the system will start displaying the questions randomly.

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

Case 2: In case the candidate answers all the three questions of the i^{th} DT incorrectly, the system will automatically shift to $(i-1)^{\text{th}}$ DT provided $i \neq 1$. When $i = 1$, the system continues to ask questions from the same level.

Case 3: This case relates the situation where the examinee answers either one or two questions correctly out of the first three questions from the i^{th} DT. The system displays one more question from the same DT. Thus the examinee has to answer a total of four questions. A total of three correct answers shifts to $(i+1)^{\text{th}}$ DT, provided $i \neq 5$; otherwise to $(i-1)^{\text{th}}$ DT provided $i \neq 1$.

Case 4: In case the examinee answers two questions correctly out of the first four questions from the i^{th} DT, one more question from the same DT is given. A total of three correct answers out of five given questions, shifts to $(i+1)^{\text{th}}$ DT; otherwise to $(i-1)^{\text{th}}$ DT. However changing to a higher DT does not take place when $i=5$ and changing to a lower DT is not done when $i=1$.

The test will get terminated either on the completion of the time frame or the student has attempted the questions for the prescribed maximum marks, whichever occurs first.

C. Evaluation Procedure

The marks for a question in each difficulty level are given in TABLE II. It can be seen that as the DT level increases, marks also increase. The maximum marks and the duration of the examination can be set according to the needs of the subject. The DT-wise number of questions attempted and answered correctly and its marks and the total score are displayed at the end of the test.

TABLE II. Marks for each DT

DT	1	2	3	4	5
Mark	0.2	0.4	0.6	0.8	1.0

The adaptive test was administered to different sets of Under Graduate (UG) and Post Graduate (PG) computer science students in an educational institution and the data relating to the time taken in each difficulty level for correctly and wrongly answering the questions, the marks scored at each level and the total time taken to complete the test were recorded.

III. RESULTS AND DISCUSSION

This section describes about the use of statistical tools and SEM on the results of adaptive E-Assessment.

A. Statistical Tools

The statistical tools such as mean, standard deviation and coefficient of variation are obtained DT-wise for scores,

correctly answered time, wrongly answered time and the total time of the test mainly for finding the average and the uniformity/consistency in the pattern of answering. Further t-test was also employed to assess whether there is any significant difference between the two sets of students in their performance relating to the score and answering behavior.

TABLE III describes the average score, time taken to answer the questions correctly and wrongly, standard deviation and coefficient of variation for UG and PG students. Further t-values and p-values are also provided in order to test the hypothesis that “There is no significant difference between the UG and PG students on their scores, the time taken to answer correctly and wrongly at each level”. It is understood from Table-3 that the average time of

UG students for correctly answering the questions in DT5 is less (5.62 seconds) followed by DT1 (22.82 seconds), DT4 (25.72 seconds), DT2 (30.82 seconds) and DT3 (44.62 seconds). The average time of UG students in DT5 is less (6.1 seconds) followed by DT1 (28.69 seconds), DT2 (29.42 seconds), DT4 (47.14 seconds) and DT3 (56.23 seconds) for wrongly answering the questions.

Similarly for PG students the average time in DT1 is less (27.89 seconds) followed by DT3 (34.07 seconds), DT4 (40.08 seconds), DT2 (40.15 seconds) and DT5 (58.54 seconds) for correctly answering the questions. Also, the average time of students in DT4 is less (18.68 seconds) followed by DT2 (30.4 seconds), DT3 (45.54 seconds) and DT1 (52.52 seconds) for wrongly answering the questions.

TABLE III. Descriptive statistics

Attributes	UG			PG			t-value	p-value
	Mean	SD	Coefficient of variation	Mean	SD	Coefficient of variation		
Mark	7.17	1.00	14.016	7.50	2.02	26.95	-1.294	0.198
DT1_CA_Avg_time	22.82	21.56	94.49	27.89	18.94	67.91	-1.264	0.209
DT2_CA_Avg_time	30.82	18.86	61.19	40.15	50.37	125.44	1.957	0.053
DT3_CA_Avg_time	44.62	23.59	52.87	34.07	31.53	92.55	-1.777	0.078
DT4_CA_Avg_time	25.72	38.05	147.92	40.08	45.12	112.58	-5.575	0.000
DT5_CA_Avg_time	5.623	20.38	362.44	58.54	66.07	112.88	-2.769	0.007
DT1_WA_Avg_time	28.69	30.83	107.45	52.52	54.67	104.10	-0.135	0.893
DT2_WA_Avg_time	29.42	21.49	73.05	30.40	48.78	160.43	1.164	0.247
DT3_WA_Avg_time	56.23	37.6	66.87	45.54	55.47	121.80	2.990	0.003
DT4_WA_Avg_time	47.14	52.54	111.46	18.68	45.72	244.75	2.211	0.029
DT5_WA_Avg_time	6.10	20.28	332.33	-	-	-	-1.078	.283
DT1_mark	0.92	0.94	102	1.04	0.72	69.69	-0.742	0.460
DT2_mark	2.32	1.25	53.58	2.41	1.25	51.78	-0.374	0.709
DT3_mark	2.88	1.45	50.5	2.46	1.49	60.65	1.476	0.143
DT4_mark	0.92	1.18	128	1.17	1.33	114	-1.025	0.308
DT5_mark	0.13	0.48	364.9	0.43	.96	226.22	-1.989	0.49

(* DT1 - Degree of Toughness 1, * DT2 - Degree of Toughness 2 , * DT3 - Degree of Toughness 3 , * DT4 - Degree of Toughness 4 , * DT5 - Degree of Toughness 5 , * CA_Avg_time – Average time taken for correctly answering the questions , * WA_Avg_time – Average time taken for correctly answering the questions)

Further the coefficient of variation is least with DT3 (52.87) in answering the questions correctly and with DT3 (66.87) in answering the questions wrongly indicating more uniformity in the pattern of answering among UG students. The coefficient of variation is least with DT1 (67.91) in correctly answering the questions and with DT1 (104.1) for wrongly answering the questions indicating more uniformity in DT1 in the case of PG students.

The hypothesis is rejected in respect of DT4 (p-value <0.05) in answering the questions correctly and in DT3 and DT4 in answering the questions wrongly. It clearly indicates that UG students have taken significantly lesser time in answering the questions correctly in DT4 than PG students and UG students have significantly taken more time in answering wrongly in DT3 and DT4 than PG

students, which indicates that UG students take more time in analyzing certain questions.

IV. STRUCTURED EQUATION MODELING RESULTS

A path model was developed using the Visual PLS (Partial Least Squares) path modeling tool in SEM. Fig. 1 and TABLE IV describes the results of SEM for UG students and Fig. 2 and TABLE V shows the results of SEM for PG students respectively. Initially the test of reliability has to be done on the constructs of the model because the reliability of the construct (variable) refers to the accuracy with which the construct repeatedly measure the same phenomena without much variation.

The reliability of the constructs used in the path model are calculated using Cronbach alpha values (acceptable value ≥ 0.7) and composite reliability (acceptable value ≥ 0.7). According to Chin [11], “Alpha tends to be a lower bound estimate of reliability whereas composite

Performance Evaluation Of Students In Adaptive E-Assessment Using Structural Equation Modeling

reliability is a closer approximation under the assumption that the parameter estimates are accurate". Therefore, in this model composite reliability was used instead of Cronbach Alpha.

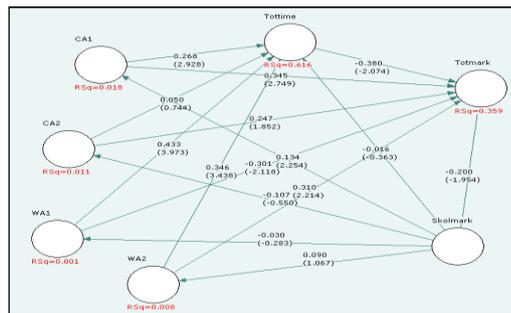


Fig 1. Path diagram for UG

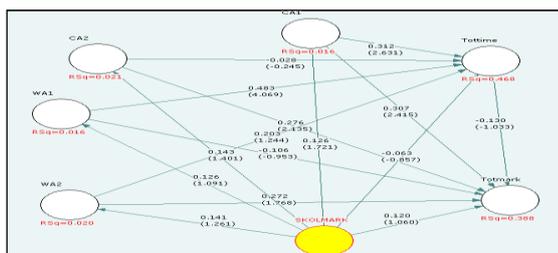


Fig 2. Path diagram for PG students

After performing reliability tests, the validity of the model has to be checked using convergent validity. The

convergent validity of each construct is measured by examining the "Average variance extracted (AVE)". Average Variance Extracted was proposed by Fornell and Larcker [12] as a measure of the common variance in a construct in relation to the amount of variance due to its measurement error. Constructs which have AVE greater than 0.5 are said to have convergent validity [11]. TABLE IV and TABLE V show the reliability and construct validity for UG and PG students respectively. It can be seen that the constructs are reliable and AVE have good convergent validity.

Hypothesis testing is done to test each relationship in the hypothesis using bootstrapping. It is technique in which samples are randomly selected from the original data set. The size of each sample is the same as the number of samples collected for this study. Bootstrap is performed by collecting a large number of such resamples and their means are used to test the hypothesis using t-test. The cutoff value for t-coefficients depends on the assumed significance level. A commonly assumed significance level is a 2-tailed 5% significance level that is indicated by t=1.96. If the value of t-statistic is higher than this cutoff value, it implies the path being considered is significant. The bootstrap summary for UG and PG student models are given in TABLE VI and TABLE VII respectively.

TABLE IV. Reliability and Construct validity for UG students

Construct	Indicator	Factor Loading	AVE	Composite Reliability
CA1	DT1_CA_Avg_time	0.819	0.63	0.78
	DT2_CA_Avg_time	0.773		
CA2	DT3_CA_Avg_time	0.848	0.66	0.79
	DT4_CA_Avg_time	0.770		
WA1	DT1_WA_Avg_time	0.828	0.68	0.81
	DT2_WA_Avg_time	0.821		
WA2	DT3_WA_Avg_time	0.890	0.70	0.83
	DT4_WA_Avg_time	0.784		
Schoolmark	SSLC	0.989	0.98	0.99
	HSC	0.991		
Totmark	TotMark	1.000	1.00	1.00
Tottime	Tot_time	1.000	1.00	1.00

TABLE V. Reliability and Construct validity for PG students

Construct	Indicator	Factor Loading	AVE	Composite Reliability
CA1	DT1_CA_Avg_time	0.809	0.68	0.81
	DT2_CA_Avg_time	0.837		
CA2	DT3_CA_Avg_time	0.400	0.50	0.71
	DT4_CA_Avg_time	0.910		
WA1	DT1_WA_Avg_time	0.438	0.50	0.702
	DT2_WA_Avg_time	0.853		
WA2	DT3_WA_Avg_time	0.308	0.52	0.75
	DT4_WA_Avg_time	0.974		

Schoolmark	SSLC Mark	1.000	0.90	0.95
	HSC Mark	1.000		
Totmark	TotMark	0.971	1.00	1.00
Tottime	Tot_time	0.925	1.00	1.00

(* SSLC – Secondary School Leaving Certificate , * HSC – Higher Secondary School Certificate, * Totmark – Total mark scored in the test , Tottime – The total time taken to complete the test)

TABLE VI. Bootstrap summary for UG students

Hypothesis	Path	Path Coefficient	T-Statistic	Results
H1	CA1→Tottime	0.268	2.928	Significant
H2	CA1→Totmark	0.345	2.749	Significant
H3	CA2→Tottime	0.050	0.744	Not Significant
H4	CA2→Totmark	0.247	1.852	Not Significant
H5	WA1→Tottime	0.433	3.973	Significant
H6	WA1→Totmark	-0.301	-2.118	Not Significant
H7	WA2→Tottime	0.346	3.438	Significant
H8	WA2→Totmark	0.310	2.214	Significant
H9	Skolmark→CA1	0.134	2.633	Significant
H10	Skolmark→CA2	-0.107	-0.550	Not Significant
H11	Skolmark→WA1	-0.030	-0.283	Not Significant
H12	Skolmark→WA2	0.090	1.067	Not Significant
H13	Skolmark→Totmark	-0.200	-1.954	Not Significant
H14	Skolmark→Tottime	-0.016	-0.363	Not Significant
H15	Tottime→Totmark	-0.380	-2.074	Not Significant

The marks scored in the school and the total time taken by the students in answering the questions has an impact on the performance in adaptive testing. In this study, the total time taken and school marks are taken as the variables in adaptive testing strategy and various hypotheses have been formulated using these variables.

It is found from Table -6 that the average time taken to answer DT1 & DT2 questions (represented by the construct CA1), the average time taken to answer DT1 & DT2 (represented by the construct WA1) wrongly and the average time taken to answer DT3 & DT4 questions wrongly (represented by the construct WA2) have significant impact on the total time taken for the test and the total marks scored. The more the time they spend in answering the questions at each level, they tend to score better marks. The marks scored by the student in HSC and SSLC also have a significant impact on the average time taken to answer DT1 & DT2 questions correctly for UG students. This indicates that average students do not spend more time in analyzing the questions for which they don't have answers readily available.

In case of PG students(TABLE VII) , it is found that the average time taken to answer DT1 & DT2 questions correctly (represented by the construct CA1) have significant impact on the total time taken for the test and the total marks scored in the test and the average time taken to answer DT3 & DT4 questions correctly (represented by the construct CA2) have a significant impact on the total mark . The average time taken to answer DT1 & DT2 questions wrongly (represented by the construct WA1) have significant impact on the total time taken to complete the test.

The results show that the construct CA1 influences the total mark scored and the total time taken in the test both UG students and PG students. This indicates if the students spend more time on answering DT1 and DT2 questions, they score more marks. The mark scored in school does not have a significant impact on the test for PG students because of diversified background of subjects in their UG degree.

TABLE VII. Bootstrap summary for PG students

Hypothesis	Path	Path Coefficient	T-Statistic	Results
H1	CA1→Tottime	0.312	2.6308	Significant
H2	CA1→Totmark	0.307	2.4153	Significant
H3	CA2→Tottime	-0.028	-0.2446	Not Significant
H4	CA2→Totmark	0.276	2.1347	Significant
H5	WA1→Tottime	0.483	4.0688	Significant
H6	WA1→Totmark	-0.106	-0.9527	Not Significant

Performance Evaluation Of Students In Adaptive E-Assessment Using Structural Equation Modeling

H7	WA2→Totttime	0.203	1.2444	Not Significant
H8	WA2→Totmark	0.272	1.7679	Not Significant
H9	Skolmark→CA1	0.126	1.7205	Not Significant
H10	Skolmark→CA2	0.143	1.4011	Not Significant
H11	Skolmark→WA1	0.126	1.0914	Not Significant
H12	Skolmark→WA2	0.141	1.2613	Not Significant
H13	Skolmark→Totmark	0.12	1.0605	Not Significant
H14	Skolmark→Totttime	-0.063	-0.8571	Not Significant
H15	Totttime→Totmark	-0.13	-1.0329	Not Significant

V. CONCLUSION

A strategy for conducting adaptive E-Assessment was formulated and tests were conducted for UG and PG students of an educational institution. The empirical results were tested using statistical tools and structured equation modeling. Results indicate that students who spend more time in the basic DT levels tend to score better marks. The adaptive assessment score of UG students is dependent on their school performance whereas it does not have a significant impact on the performance of PG students. This approach can also be used to evaluate the performance of students in other courses.

ACKNOWLEDGEMENT

The authors duly acknowledge the UG and PG students and their tutors for their cooperation and support in conducting this adaptive assessment.

REFERENCES

1. George Washington , V.Rhymend Uthariaraj: A Soft Real Time online Examination System.: Asian Journal of Information Technology, (2006)
2. Ang Tan Fong , Hu Heng Shew, Por Lip Yee , Liew Chee Sun: IOAS : An Intelligent Online Assessment System.. Journal of WSEAS Transactions on Computers, (2007)
3. Wen-Chih Chang, Hsuan-Che Yang: Applying IRT to Estimate Learning Ability and K-means Clustering in Web based Learning., Journal of Software,(2009)
4. Catherine C Marinagi, Vassilis G.Kaburlasos ,Vassilis Th. Soukalas: An Architecture for an Adaptive Assessment Tool .. 37th ASEE/IEEE Frontiers in Education Conference,(2007).
5. Catherine C Marinagi and Vassilis G.Kaburlasos: Work in Progress: Practical Computerized Adaptive Assessment based on Bayesian Decision Theory, 36th ASEE/IEEE Frontiers in Education Conference,(2006)
6. Mariana Lilley: The Development and Application of Computer Adaptive Testing in a Higher Education Environment, PhD Thesis, University of Hertfordshire, UK, (2007)
7. Thomas, Marie Huffmaster, Modeling Differential Pacing Trajectories in High Stakes Computer Adaptive Testing Using Hierarchical Linear Modeling and Structural Equation Modeling. PhD Thesis , University of North Carolina at Greensboro ,(2006)
8. Hair.J.F,Jr, Anderson.R.E, Thatham.R.L & Black.W.C , Multivariate Data Analysis, 5th Edition , Pearson Education.
9. J.J.Hox and T.M.Bechger , An Introduction to Structural Equation Modeling, Family Science review.
10. Chin, W. The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), Modern

methods for Business Research, Mahwah, NJ: Lawrence Erlbaum Associates, 295-336 (1998)

11. Geetha, V., Chandrakala, D., Nadarajan, R. and Dass, C.K. "A Bayesian Classification Approach for handling Uncertainty in Adaptive E-Assessment", International Reviews on Computers and Software, Vol.8, No.4, April 2013.
12. Geetha, V., Surendiran, B., Nadarajan, R. and Nandakumar, G.S. "An Adaptive E- Assessment Grading (AEAG) Model for Performance Evaluation", International Journal on Computational Sciences & Applications (IJCSA) Vol.2, No.4, pp.25-38.2012.