

Improving Learning Throughput in E-learning using Interactive-Cognitive Based Assessment

Melvin Ballera
Computer Science Department
Sirte University – Libya
maballera@yahoo.com.ph

Ismail Ateya Lukandu
Faculty of Information Technology
Strathmore University – Kenya
iateya@strathmore.edu

Abdalla Radwan
Computer Science Department
Sirte University – Libya
radwan2004@hotmail.com

ABSTRACT

An e-learning website is not sufficient to fully attain the results of online education. There also is a need to align the educational objectives into the design of the assessment to improve and develop cognition, critical thinking and problem-solving skills. Previous studies have explored the potentials of the assessment models but few ventured into their implementation. Others only proposed and introduced conceptual frameworks. The implementation of these proposals, however, revealed that the question type in the assessment phase neglected to align their questionnaire formats into a cognitive schema. At present, the standard multiple-choice question is the most frequently used of the question type of e-learning assessments. However, if this type is the only format adopted by e-learning developers, then the potentially rich and embedded assessment of the computer platform will be given up. This paper focuses on the design of assessment questions, which is created and guided by the hierarchical Bloom cognitive taxonomy and by utilizing rich media formats. Results conducted for eighteen weeks show a dramatic increase in the academic performance of the students. Likewise, digital transcripts converted from the collected perceptions after training undergoes sentiment analysis have correlated with the student improved academic throughput.

KEYWORDS

Bloom cognitive taxonomy, e-learning, interactive-cognitive assessments, learning throughput, sentiment analysis

1 INTRODUCTION

Assessment is defined as “a device or procedure used for evaluation by obtaining a sample of a learner’s behavior in a specified domain and scoring this behavior in a standardized process [1]. It constitutes a vital part of web-based learning instruction. Through assessment, educational strategists can determine how effective their lessons are in teaching students the intended facts and skills. To effectively assess students, educational strategists must not simply relegate assessment at the end of the learning process or training. This must be also fully integrated into the process of educating students [2]. Assessment designs can greatly influence the learning of the students. It can also be a tool for data gathering and the results gathered can help teachers decide on the performance of the students [3]. At present, many e-learning assessments used the standard multiple-choice questions. However, it can be argued that if e-learning developers adapt only this type of assessment, then the potentially rich and embedded assessment of the computer platform will not be totally utilized [4].

Today, the question type currently dominating many e-learning assessments is the standard multiple-choice question. It is necessary for assessment practices to reflect the combinations of acquired skills and knowledge. The complexity and use of these combinations will enable students to interpret, analyze, evaluate problems and explain their arguments. These assessments, which should be fully integrated into the learning process, provide information

about the learner’s progress and support them in selecting appropriate learning tasks. The consistency of the content, methodology, and the manner of assessment will make teaching become more effective. Therefore, it is a worthy undertaking to invest in the design of performance assessments because it provides multidimensional feedback for fostering learning [5].

The objective of this paper is to present an assessment questionnaire format by adapting a number of the assessment designs which were investigated and discussed in the related literature. These designs are redesigned and realigned into the Bloom Cognitive Theory Schema and is presented in a more interactive way to suit the computer science curriculum at tertiary level. The paper is organized according to the discussion of related literature, methodologies, results and analysis and lastly, the conclusion and future works.

2 RELATED LITERATURE

For the alignment of an effective assessment design, three components are investigated in this section: 1. searching and adapting the existing assessment design literature that can be useful in the computer science curriculum; 2. redesign the assessment according to Bloom Cognitive Taxonomy schema (manner or way

to present question in accordance to cognitive prescription) and; 3. the incorporation of simulations and interactivity in the assessment process.

2.1 E-Learning Assessment Questions

There are several ways to make assessment items innovative and creative. The use of technological enhancements of sound, graphics, animation, video or the incorporation of media can be utilized also for e-learning assessment designs [6]. Figure 1 below shows the summary of the 13 questions types collected from 15 various sources of scientific research and publications and used as guiding tool in designing questionnaires. Each question type has a different cognitive level and requires demonstration of varying skills from the learners during assessment process. Although there are many existing assessments that used various question types, few have been tested to computer science perspective and at the level of tertiary education. Majority of these assessments were implemented in a pencil-paper test and few transformed these assessments into digital form [7], [8], [9], [10], [11], [12], [13], [14]. If such assessment features will be implemented fully into the e-learning, the system will hypothetically deliver cognitive gain among students.

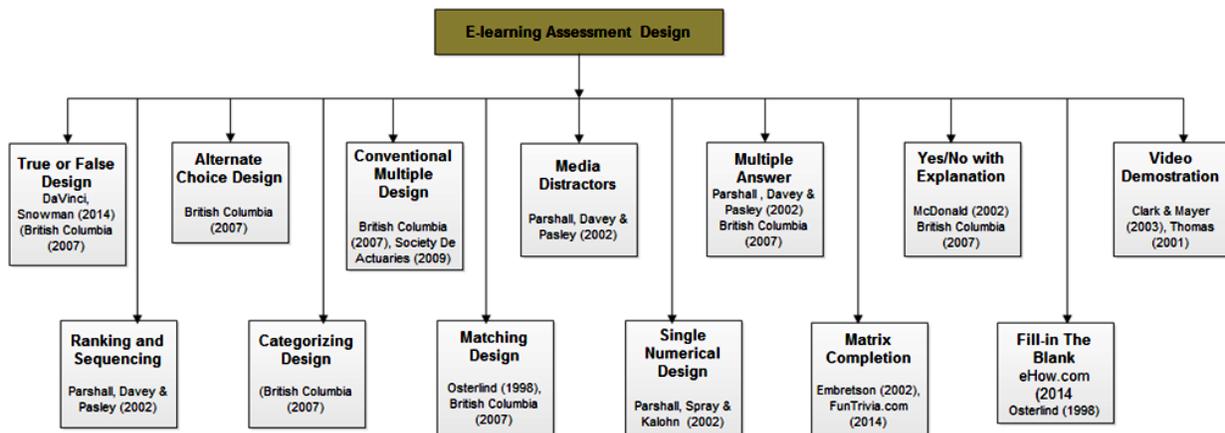


Figure 1: Question Types for E-Learning Platform

2.2 Bloom Cognitive Schema

The cognitive domain of Bloom Taxonomy involves knowledge and the development of intellectual skills, therefore it is necessary to align assessment according to this schema. This schema includes recall or recognition of specific facts, procedural patterns and concepts that help in the development of intellectual abilities and skills [15]. There are six major categories, starting from simplest behavior to the most complex. The categories can be viewed as degrees of difficulties [16]. Layer one is, “Remembering”, which entails establishing definitions, creating facts, charts, lists or oral activities. Layer two, “Understanding”, includes producing drawings or summaries. “Applying” is layer three, and used for learning models, presentations, interviews or simulations that are applied to new situations. “Analyzing” is layer four which includes “distinguishing” between the parts, creating spreadsheets, surveys, charts, or diagrams. “Evaluating”, which is layer five involves critiquing, recommending, and

reporting. Putting the parts together in a novel and unique way falls in the sixth layer which is “Creating” [17]. At present, this model becomes a basis in developing e-learning by transforming its contents, instructional delivery and most importantly the assessment. The layers represent the levels of learning and increasing complexity.

Figure 2 shows the cognitive levels in Bloom’s original taxonomy, arranged in ascending order. Each step suggests activities for the specific level. A list of verbs which are commonly used to create learning objectives can be found below each step. When Bloom created this cognitive schema, he intended to use this in assessing the expertise in order to develop new ways in measuring what college students learn. At present, this model becomes a basis in developing e-learning by transforming its contents, instructional delivery and assessment to suit the learners’ needs. His work contributed greatly in shifting the focus of educators to learning from teaching.

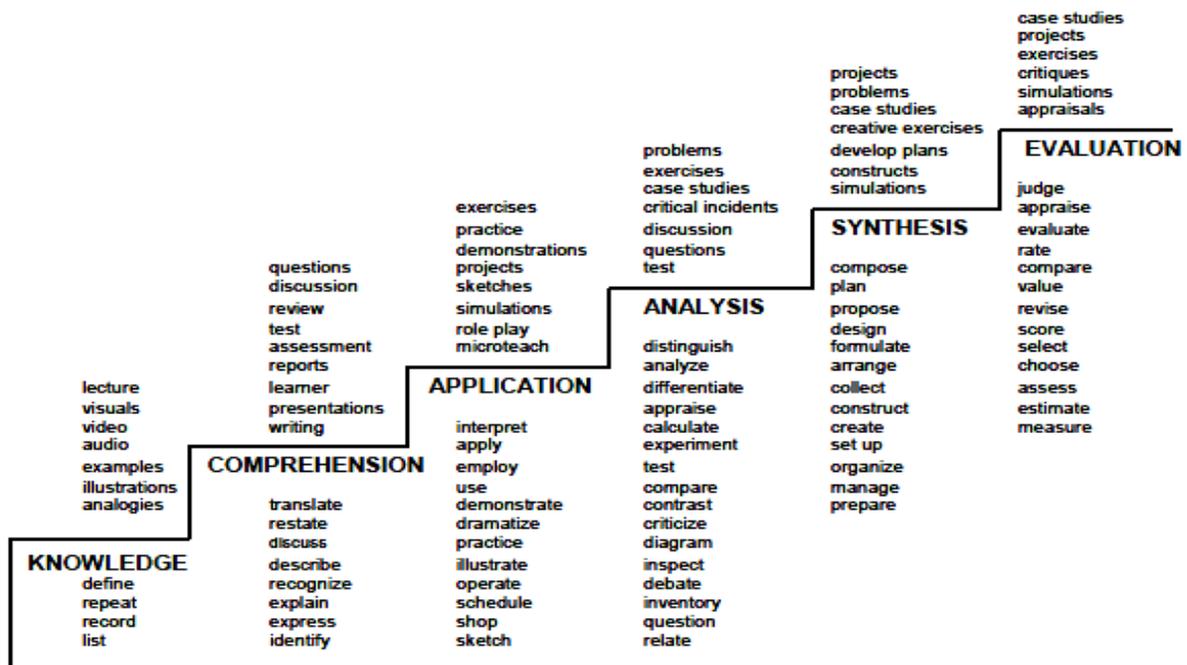


Figure 2: Bloom’s Taxonomy Staircase [17]

2.3 Multimedia and Interactivity

Many educators believe that interactive e-learning assessments allows “learning by doing”, arouses interest and generates motivation. Interactivity leads to a more meaningful learning because students are able to test their comprehension, learn from their errors and make sense of what is unpredictable. It can also improve the students’ knowledge and performance during the assessment process [18]. Simulations and modelling tools are the best examples of complex, meaningful interactivity in assessments. Such applications models represents a real or theoretical system, and allows users to manipulate input variables, change the system’s behavior and view the results. With such applications, learners can construct and receive feedback as a result of their actions. Inclusion of interactive simulations in e-learning assessment improves the quality and outcomes of e-learning. Simulations and visualization tools make it possible for students to bridge experience and abstraction which help to deepen their understanding of ambiguous or challenging content [19]. Interactivity, when used in assessment, is a factor that has the biggest impact on cognitive learning and is the most powerful model of instruction [20].

The use of multimedia in assessment such as graphics refers to the variety of illustrations that include line drawings, charts, photographs, motion graphics such as animation and video. These multimedia can indeed increase learning. Research shows that graphics improve learning through cognitive exercises, storing and retrieving ideas. Mayer claims that a student who practices on assessment with text and graphics is claimed to gain an average of 89% on transfer text as compared to those students who rely on texts alone. It is also found out that the integration of text near the visuals during assessment yielded an average improvement of 68%. Furthermore, explaining

graphics with audios followed by a question improved learning by almost 80% [21].

Studies have compared the effect of multimedia-based learning with traditional classroom-based learning. Allen discusses the effect of multimedia-based training. He claims that a good multimedia training is not only faster than classroom training, it is also better [22]. People remember and retain longer in memory what they learn more accurately and use what they learn to improve their performance. Adams reviewed six studies that carefully compared multimedia training to classroom instruction: Learning gains were up to 56% greater while consistency of learning" (variance in learning across learners) was 50-60% better and content retention was 25-50% higher [23]. Brett claims that multimedia-based learning is more motivating and exciting than the more traditional educational methods [24]. It can also be claimed that using multimedia increases learning effectiveness and cognitive skills.

Adapting question types from different researches and re-aligning its questionnaires or small tests into cognitive model and presenting it in an interactive and simulative manner can thus hypothetically guarantee learning.

3 METHODOLOGY

3.1 Respondents

The study is organized within the context of Design and Analysis of Algorithms class which is taught at Sirte University, Libya. The entire data collection and training lasted for 18 weeks. All learning materials and online assessments were conducted in the e-learning portal prototype at <http://maballera.byethost7.com/elearning/>. The learning materials is composed of 12 chapters, L₁ to L₁₂. All students are familiar with the use

of electronic materials and had seen the implementation of the e-learning system. Students were given one week familiarization of the system flow and navigation. Diagnostic examinations were conducted every four weeks while formative examinations were administered at the end of each training module. At the end of the training, summative examination is given to determine the overall learning throughput or learning benefits in using the prototype. The passing mark is 75.

Prior to implementation, students were informed about the research and the task involved. Students were given time to navigate the e-learning system so that they would be familiar and be directly involved in the learning process. If issues arise during the learning process, the researcher provided necessary assistance in support for blended learning.

3.2 Data Collection

In this study, primary data were collected in two ways. The first was the experimental collection where various tables were populated dynamically, manipulated and extracted to generate several reports such as practice examination results, cognitive performance graphs and frequency of the practice examination or trials. The second was the survey which was divided into two parts. The first part was the measure of the internal reliability of all the questionnaires stored in the Item Bank. The second part was the acceptability of assessment design factors and the collection of student's perceptions and experiences. Factors that affect the assessment design were content and visual design (colors, balance, readability), accessibility (links, feedback and explanation facilities), assessment types (difficulty, bloom level), navigation (transition of questionnaires, pop-up windows, reminders), learning support (specific part of the lesson, additional references) and

interactivity. To measure the internal consistency of the questionnaires, the Cronbach alpha was used while z-test was used to evaluate the acceptability of the assessment design factors.

To correlate the results of the Bloom Taxonomy, theme extraction using a special software called Semantria [25] was used to analyze the digital transcripts of the students. The student were requested to write a report in one or two sentences about their experiences and perceptions in using the system and the new learning delivery per se. In particular, the respondents did the following: gave simple summary of actions they had done as part of their participation, proposed and discussed some strategies that could be applied in a situation, stated the topics for which they got assistance, examples and topics that were products of their work and finally provided their personal reflection and experiences in participating in the exploratory study. Semantria software extracts themes using the digital transcript of the students taken from the survey to determine and follow themes that appear over a period of time.

3.3 Degree of Difficulty

The 13 question types investigated and presented in the literature were categorized according to the Cognitive Bloom Taxonomy. Table 1 shows the question types description and the degree of difficulty df , for each type in different assessment formats. In formative assessment, the df is 1 for reviewing purposes and practice examination at the end of each lesson. The df of Bloom Cognitive examination (diagnostic) on the other hand is also 1, to measure the cognitive improvements of the learner which is usually administered every two weeks of the training.

Table 1: Questions Types and their Degree of Difficulty (*df*)

Bloom Taxonomy	Question Types	Description	Degree of Difficulty		
			Summative	Formative	Diagnostic
REMEMBER	MATF	Multiple True or False Questions	1	1	1.5
	MTCQ	Matching and Categorizing Questions	1	1	1.5
	TOFQ	True or False Question	1	1	1.5
UNDERSTAND	MCMA	Multiple Choice Multiple Answer	1	1	1.5
	MCID	Multiple Choice with Illustrative Diagram	1	1	1.5
APPLICATION	CSMA	Complex Single Multiple Choice Questions	1.5	1	1.5
	SNQC	Single Numerical Construction Questions	1.5	1	1.5
ANALYZE	SMCQ	Situational Multiple Choice Questions	1.5	1	1.5
	SAMC	Single Answer with Enumeration Questions	1.5	1	1.5
EVALUATE	MSOQ	Matrix Completion Questions	2	1	1.5
	MALT	Multiple Alternative Questions	2	1	1.5
CREATE	FIBE	Situational Fill-in the Blanks and Enumeration	2	1	1.5
	DSVQ	Video Simulation with Audio Play Questions	2	1	1.5

The *df* of summative assessment differs accordingly since it is the most important performance matrix. As the Bloom category goes to the bottom of the table, the more difficult the questions and the deeper the cognitive development become. Every question has a level of difficulty, and this level is also

utilized upgrading the students' performance matrix. Higher ability is demonstrated when a student answers harder questions that correctly answering an easier question. The *Remember* category has *df* 1 while the *Understand*, *Application*, and *Analyze* categories have a *df* of 1.5 while *Evaluate* and *Create* have *df* of 2.

After aligning the collected question types as shown in Table 1, questionnaires underwent formatting using the guided cognitive verbs schema as presented Figure 2 and then the interactivity and simulation to the question were added. The use of graphs, videos and other media formats, and required-response questions was incorporated in the system. The Item Bank is currently a repository of different questions types with varying difficulty level. It contains 280 questions with explanation facilities divided among thirteen (13) questions types and were used to produce the Bloom Cognitive Taxonomy examination, the random formative examination and the random summative examination.

3.4 Question Item Design and Interactivity

For brevity, three live illustrative question types were extracted from the system prototype for the purpose of illustration. Alternate choice items are somewhat similar to true/false type of

questions. However, rather than letting the students determine whether a single statement is correct or not, this type of questions asks the student to select the better answer between two choices. Choices are often scenarios or cases, as shown in Figure 3 below.

MALT2. Evaluate the 2 codes below and determine which one is better? Used 99, 6, 0, 89, 30 as test data.

ALGORITHM 1

```

for i ← 1 to n - 1 do
  v ← A[i]
  j ← i - 1
  while j ≥ 0 and A[j] > v do
    A[j + 1] ← A[j]
    j ← j - 1
  A[j + 1] ← v
                    
```

ALGORITHM 2

```

for i ← 1 to n - 1 do
  j ← i - 1
  while j ≥ 0 and A[j] > A[j + 1] do
    swap(A[j], A[j + 1])
    j ← j - 1
                    
```

Algorithm 1 runs at $O(n^2)$ with 7 lines, the criteria of analyzing algorithm in this case uses simplicity and readability. Tracing back using the test data is straightforward however line 3 and line 6 repeatedly executed.

Algorithm 2 runs at $O(n^2)$ with 5 lines. It eliminates redundancy and straightforward mechanism. Thus it is better to implement and simple.

Figure 3: Alternate Choice Example

In this type, students were shown two possible algorithmic models for computing their running time complexity and must choose the most accurate response option. In this case, the correct answer was the second option due to its simplicity. Innovations in the multiple-choice category for online settings can include new response actions not commonly found in paper-and-pen settings. This entails clicking on an area of a graphical image. It can also include new media, such as sound clips which can be considered as distractors. Such new media innovations are represented in Taxonomy as Multiple Choice with Illustrative Diagrams. An example is given in Figure 4.

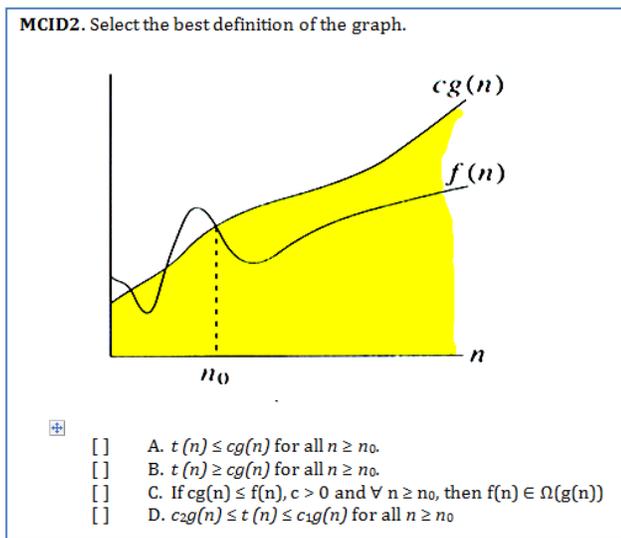


Figure 4: Multiple Choice with Illustrative Diagrams Example

In this example, respondents must select one of the four choices that corresponded to the meaning of the graph. There were four choices to choose from. This is similar to the standard multiple choice question but aside from choosing from the four possible answers, this method of response involves also analysis.

MCOQ1. The Tower of Hanoi is a mathematical game or puzzle. It consists of three rods, and a number of disks of different sizes which can slide onto any rod. The puzzle starts with the disks in a neat stack in ascending order of size on one rod, the smallest at the top, thus making a conical shape. The objective of the puzzle is to move the entire stack to another rod, obeying the following simple rules:

- Only one disk may be moved at a time.
- Each move consists of taking the upper disk from one of the stacks and placing it on top of another stack.
- No disk may be placed on top of a smaller disk.

Complete the table from given $n = 7?$

n	1	2	3	4	5	6	7
M(n)	1	3	7	15	31	63	

127

Figure 5: Matrix Completion Example

The matrix completion format, presents a matrix of patterns with one or more cells left blank. Respondents were asked to fill the empty cells from a set of supplied answers. Matrix completion has an extensive history in intelligence measurement and has been used in various tests of pattern recognition, correspondence and generation. The matrix is a table or spreadsheet of correct patterns, which can be in the form of graphics, words or numbers, as well as sound clips, film clips and animations. These are dragged to the appropriate empty cells. The item type allows for a great deal of flexibility in the task assignment, openness of response and media inclusion, and is readily computer-scorable, making it potentially a powerful item type in computer environments. It can be seen that depending on what is called for by *in* matrix completion, the matrix type can fall into a number of categories. These are reordering, substitution and construction, as well as simple completion. Thus, this type blurs the lines of the constraint-based item taxonomy. Domain-specific matrix completion tasks may be among the families of innovation most ripe for computer based applications such as shown in Figure 5.

Many interactive activities were included in the assessment design to give learners the “personal touch and control” in the assessment process. Student could write their answer using the fill-in the blank question types, compute the next sequence and analyze the pattern in completion matrix question type. Students could also enumerate answers, view and analyze graphs and allow feedback. The explanation facilities could also derive the solution and link student’s misconception into specific part of the learning materials. To enhance the learning process further, videos, and other simulative process were incorporated into the system to allow the method of “learning by doing”. Student could view the algorithm and its simulative effects given certain inputs and variables.

4 RESULTS AND DISCUSSION

4.1 Internal Consistency and Z-test Results

Prior to the post survey for students, the survey forms were presented among the academic staff to validate the measurement scale and questionnaires. The Cronbach’s Alpha coefficient for internal consistency reliability test was used for each scale. Cronbach’s alpha reliability coefficient normally ranges between 0 and 1. It provides the following rules of thumb: $\alpha \geq .9$ – Excellent, $.7 \leq \alpha < .9$ – Good, $.6 \leq \alpha < .7$ – Acceptable, $.5 \leq \alpha < .6$ – Poor and $\alpha < .5$ – Unacceptable [26]. The results of Cronbach’s Alpha coefficients for each scale are presented in Table 2.

The results indicated that all scales satisfied the requirement for internal reliability. All Cronbach’s alphas of the scales were higher than .60. The lowest value of Cronbach’s alpha is .62 in Accessibility scale while the highest is .74 in Navigation scale. The impact of the reliability of each question in the survey can be determined by calculating Cronbach’s alpha the i th variable for each $i \leq k$ is deleted. Thus, for a

test with k questions, each score x_j alpha was calculated for x_i for all i where $x_i = \sum_{j \neq i} x_j$.

Figure 5 shows that the overall reliability for Content is .636 while individual reliability of questionnaire within the scale are: for C1 is .677, C2 is .774, C3 is .519, C4 is .337 and C5 is .457. In this scale, C4 was the most affected and could be deleted from the survey form. The Visual Design overall reliability is .650 and the most affected was V5 with Cronbach value of .457. On the other hand, Accessibility scale overall reliability is .617 and the most affected questions were A3 and A4 which both have values of .427. Similarly, with the remaining scale, questions with smaller Cronbach alpha compared to the overall scale reliability were the most affected and could be deleted from the survey form. If the reliability coefficient increased after an item was deleted, it can be assumed that the item was not highly correlated with the other items. Additionally, the decrease in reliability coefficient can lead to the assumption that the item is highly correlated with other items. [27]. As shown in the table, the omission of any single question does not affect the Cronbach’s alpha very much. Questions with low reliability compared to its overall measurement scale were not deleted because small set of questionnaires affects the reliability value [28]. In this case, five questions in each measurable scale were acceptable and there was no need to delete the item since the uniqueness of each item could easily be seen. According to Cortina [29], the uniqueness of the item can be assessed with the coefficient alpha.

Table 3 shows the results of the post survey conducted among staff members to determine the overall reliability of the software and the 280 questionnaires stored in the Item Bank. These questions were used for various assessments employed in the prototype. During the survey, random questions were shown from the system to evaluate and rate their reliability. The overall internal consistency of the software

is .81, which is considered good while the overall reliability for the 60 questionnaires for Bloom Cognitive Taxonomy is .84. Similarly,

the internal reliability for questionnaires which were used for formative and summative assessment is .72.

Table 2: Cronbach’s Alpha Coefficient for each Measurement Scale

VARIABLES	CRITERIA						
	Content	Visual Design	Accessibility	Assessment	Navigation	Learning Support	Interactivity
k	5	5	5	5	5	5	5
Cumulative Variance (Yi)	0.86	0.99	1.08	1.00	0.99	1.01	1.01
Variance	1.75	2.06	2.14	2.14	2.41	2.08	2.31
α (Alpha)	0.63	0.65	0.62	0.67	0.74	0.64	0.70

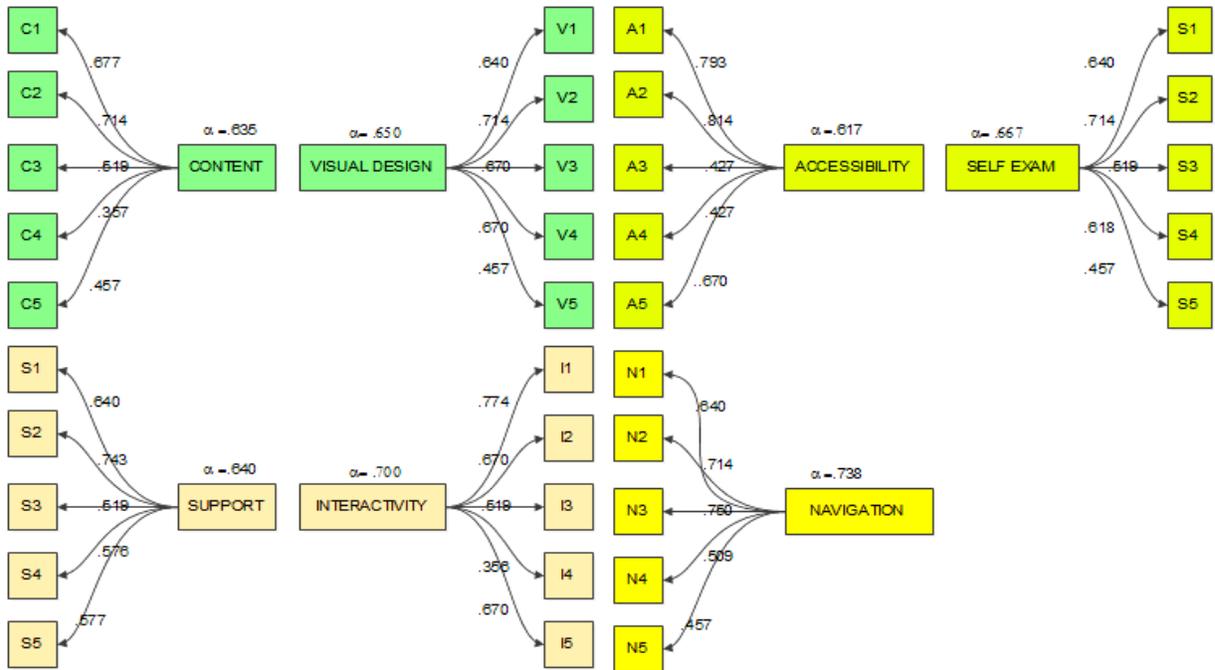


Figure 5: Reliability Coefficient after Deleting an Item

Table 3: Cronbach’s Reliability of Questionnaires and Overall Acceptability

	Software Acceptability	Bloom Taxonomy	Questionnaires
K	13.00	3.00	3.00
sum var	7.10	2.22	1.13
var	28.64	5.02	2.19
α = alpha	0.81	0.84	0.72

Table 4: Z-test of Different Measurable Scale

Criteria	Mean	Standard Deviation	z
Content	4.37	0.79	2.89
Visual Design	4.29	0.69	2.57
Accessibility	4.26	0.76	2.13
Self Assessment	4.34	0.88	2.40
Navigation	4.21	0.62	2.09
Learning support	4.29	0.93	1.92
Interactivity	4.26	0.92	1.76
Motivation	4.29	0.80	2.22

An important concept in the evaluation of assessments and questionnaires is the Alpha. It is required from the assessors and researchers that they estimate this quantity to add validity and accuracy to the interpretation of their data. A low value of alpha can mean a low number of questions and poor interrelatedness between items or heterogeneous constructs. For example, if a low alpha is due to poor correlation between items, then some items should be changed or totally eliminated. If an alpha is too high, it may suggest that some items are repetitive as they evaluate the same questions in a different manner [30]. As observed in the study, the overall alpha is not too high but still considered highly acceptable at all levels.

Table 4 shows the summary of the perception of students on the significant level of different assessments scales. The mean is given with its standard deviation. The highest mean is 4.37 from the Content scale while the lowest is 4.21 from Navigation scale. The z -values at $z_{:05} = 1.645$ makes all the critical values of measurable scale significant using one-tailed critical region. The z -values computed are greater than tabular value at alpha of .05. Based on Likert scale, the mean of each measurable variable is higher than the agreeable level which was successfully correlated by the z -test.

4.2 Experimental Results and Analysis

This section of the chapter focuses on the experimental results and its impact to the

student performance. The discussion also includes the impact of formative assessment, and other learning activities to remediate learning difficulty. The later part of this section discusses the overall effect of these various mechanisms which contributed to the tangible learning benefits.

4.2.1 Formative Examination

The purpose of formative assessments is to promote the attainment of knowledge by the students rather than testing a body of attained knowledge. Designing a curriculum that includes many rich formative assessments will result into a student-centered approach to teaching which often leads to students' success. To achieve this objective, the system employs a forced mechanism which prevented student from proceeding to the next learning materials without passing the formative assessment at hand. It is in this process that explanation facilities, lesson links and re-loading of random questions occurred in the prototype. This flexibility allowed students to suit their knowledge and exert effort to pass every formative or practice administered. As the students went through the e-learning materials, several formative assessments were implemented. This allowed student to recognize and address any misconceptions or learning difficulty they had during the learning process.

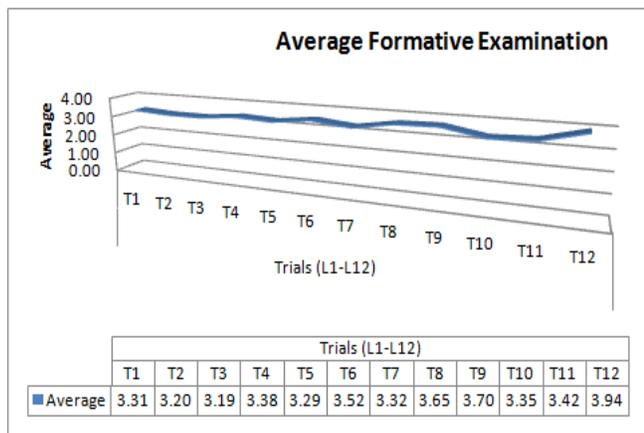


Figure 6: Average Trials of Formative Assessment

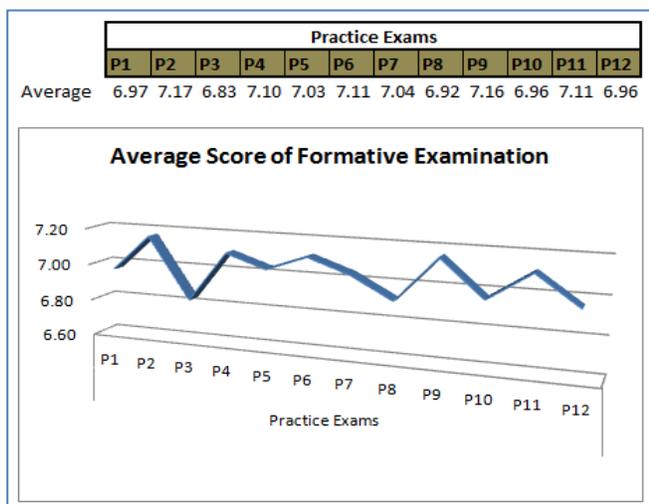


Figure 7: Average Score in Practice Examination

Figure 6 shows the average number of time spent by the 41 students in taking the formative assessment for L_1 to L_{12} . The average trials for L_1 , T_1 is 3.31 while L_{12} obtained the most number of trials with an overall average of 3.94. The students could reload as many times as they wanted the formative assessment. This allowed them to review the learning materials. The more trials were employed, the more students gained efficiency in taking a timed test in an online format. Preview questions were used to test their current knowledge and for them to be familiar with the questionnaire formats which were similar to the ones given during summative examination. By conducting

these trials, a student would understand which domains he/she was proficient in and in which areas he/she would require additional study and preparation. Each offered questions which were randomly selected from the Item Bank. At the end of each set of eight questions, student would be able to see the scores, and give feedback on the answers.

The graph in Figure 7 shows the average score of the class in taking practice or formative assessments. The highest overall average score P_2 is 7.17 for L_2 while P_3 with 6.83 for L_3 is the lowest. The e-learning prototype dictated that only 6 out of 8 score would be recorded in the database, forcing the student to study harder until a competency level was achieved. If the student failed the formative, he/needs to reload a new formative assessment. The trial table in Figure 6 was updated every time a new formative was reloaded to a particular student.

The Item Bank contains over 280 questions and distributed across three examinations extracted from 13 different question types table. These questionnaires can test students' knowledge while the answers, explanations and further reading links improve student's learning. Doing practice several times enabled students to systematically go through the Item Bank and allows them access to questions from all topics relevant to the current examination in a random pattern. These encouraged students to answer as many questions as possible, testing their knowledge on multiple topics. It followed the "practice makes perfect" attitude. The study of Walker, Brooks et. al. shows that a 4.3 and 5.7 percentage points overall is gained from practice [31]. The results indicated that online practice significantly improved student learning and examination performance. Practice testing is more powerful, useful for learners of different ages and abilities. It is more far effective than summarization, highlighting, keywords mnemonics, imagery and rereading [32]. The results of the study show how students benefited from the prototype.

4.2.2 Bloom Taxonomy or Diagnostic Examination

The Bloom Cognitive Taxonomy is a special assessments that measures the cognitive development of the student while taking the e-learning course. This 60- items assessment was specifically designed based on the Cognitive Schema and readily extracted from the Item Bank. The assessment was taken every four weeks during the experimental sessions. The assessment was equally divided to six categories specified in the Bloom Cognitive Taxonomy.

Create knowledge and experience from multiple sources into a high-order schema which will equip him/her to deal with the domain more effectively.

The graph in Figure 8 shows the overall class average of the cognitive development of students taken every four weeks during the training. It must be noted that the cognitive level of the six categories increased. The *Remember* category, for example, had an initial average of 2.5 for R1, 4.12 for R2, 6.17 for R3 and 8.6 for R4. These initial scores clearly represent 25% of the R1 followed by an increase of 16% for R2, an increase of 20% for R3 and an increased of 24.3% for R4. Similarly, as the other learning process or training neared its end, the individual average score increased. As further shown in the graph, the category with highest gain is *Remember* since it is the easiest among the six categories while *Evaluate* has the lowest learning gain. The purpose of this was just to determine whether students would improve their learning by recalling lessons that they had read and understood as they went through the sessions. As a general observation and as shown in the graph, students increases their cognitive domain at different levels. However, these results cannot be interpreted as truly cognitive gain due to the absence of a single domain during testing. The questions were defined and extracted from various topics. To compensate for this gap, the study examined the cognitive development and its relationship to the experiences and perceptions of the students in using the prototype.

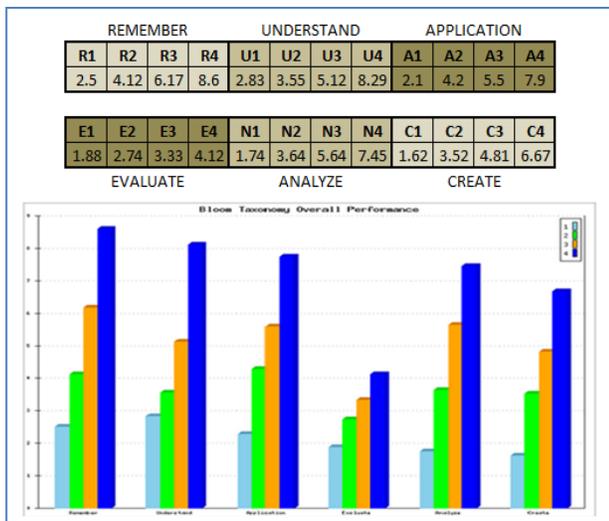


Figure 8: Average Cognitive Graph Output

For a Bloom Cognitive Taxonomy to become effective, the results must be entirely based on the use of all six levels of the pyramid. For a student to evaluate his/her cognitive development he/she needs to *Remember* the basic facts. But beyond that, the student has to *Understand* the significance of those facts, and their interrelatedness, *Apply* them to solve real life problems, *Analyze* everything from all possible alternatives and study the results. After which the student has to *Evaluate* several alternatives or solutions and which of these is most reliable. He/she has to decide which of the several alternative answers is most appropriate in a particular case. Lastly, the student has to

4.2.2 Theme Extraction from Digital Transcripts

The study employed Semantria, a special software that can compute and determine whether the coded transcripts of the student is positive, negative or neutral. During the post survey, the students were asked to write briefly their reactions, perceptions and experiences in

using the system to correlate the results of the cognitive development. Out of the 38 students, 35 wrote their reactions, perceptions or

experiences in the survey form. Their responses were coded and transformed into digital transcripts for further analysis.

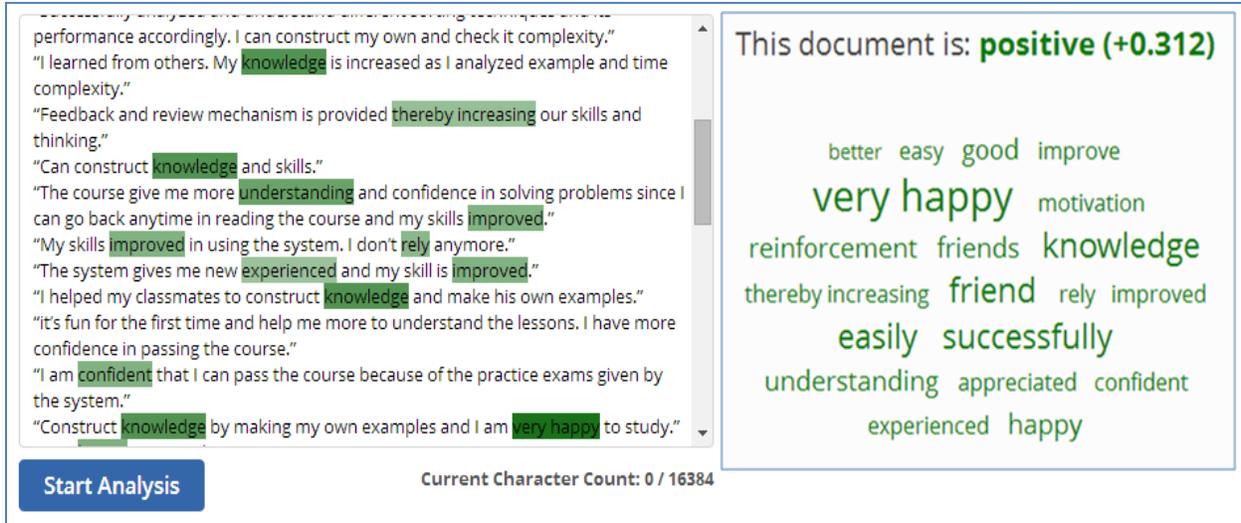


Figure 9: Semantria Analysis of the Digital Transcript

Table 5: Entity Sentiment Breakdown of the Digital Transcripts

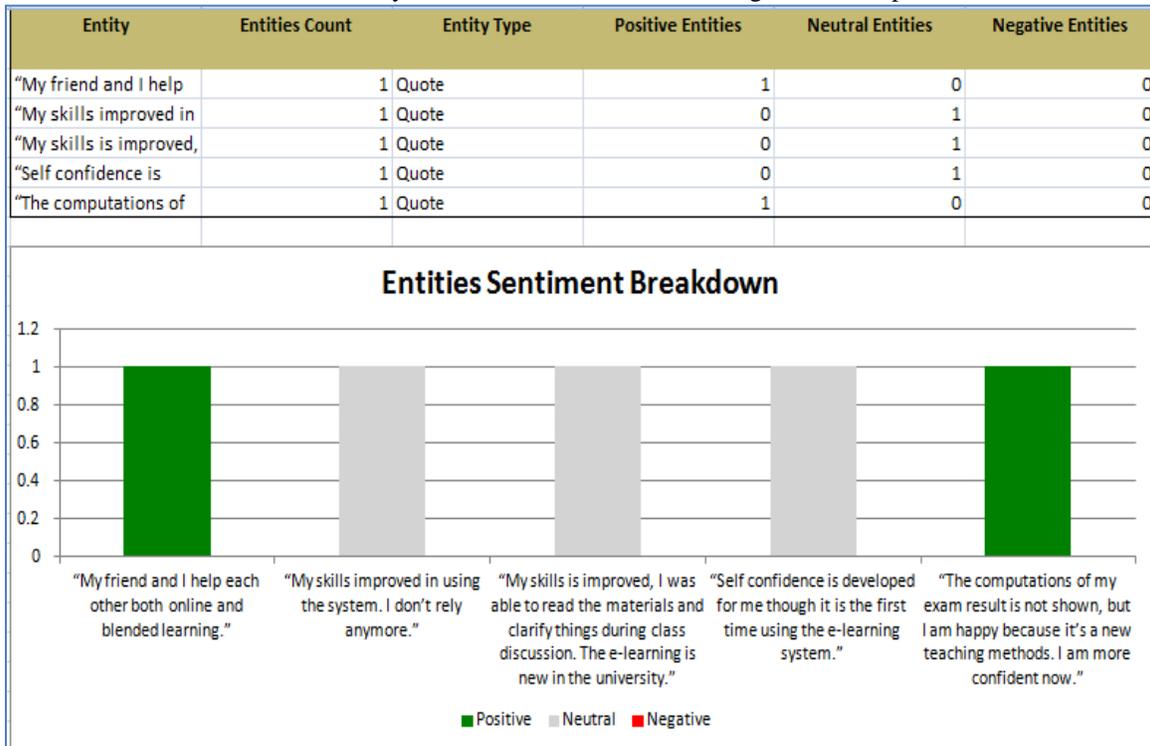


Figure 9 shows the output of the Semantria and reveals that the digital transcripts are positive with a score of +.321. Several words reveal and can attribute to this attitude. As shown in the figure, the words are *very happy, friends, motivate, improve, understanding, knowledge, good*. According to Scheve [33], students who have high cognitive benefits and self esteem will likewise reflect these in life or in their reactions to objects or surroundings. Being happy and positive increases the overall self-esteem and partly results to good school performance [34]. Thus, it can be concluded that the results coincide with the findings of Franken [35] that being happy results to "making reasonable progress towards the realization of a goal".

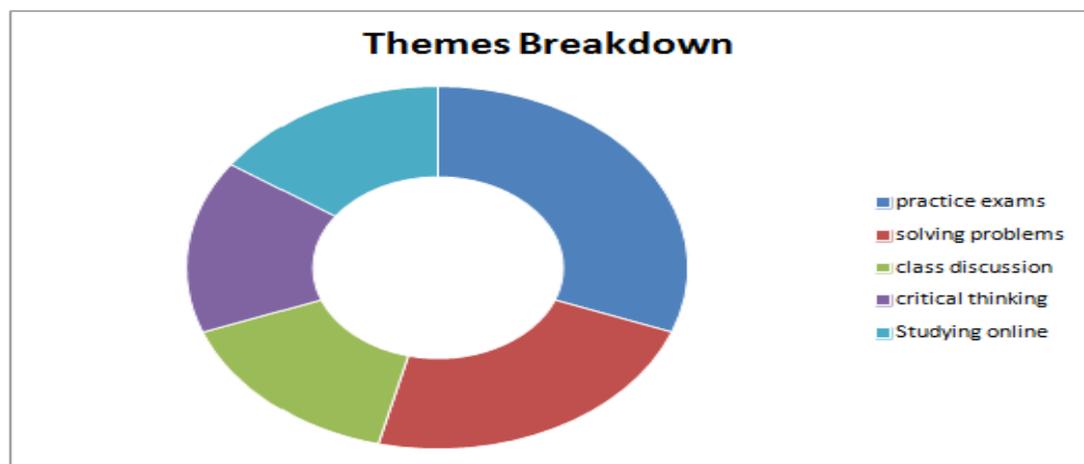
To further strengthen the findings, Semantria extracted five entities from digital transcripts and identified two positive sentiments and 3 neutral leading to positive. These results can be seen in Table 5. No negative feedback is

received from the 35 coded entities. Sentiment analysis is the process of detecting positive, negative, or neutral feelings in a piece of writing [36]. Semantria software is an information-gathering behavior that discovers what other people think [37].

Table 6 shows the five themes extracted from the digital transcript. They are *practice examinations, solving problems, class discussion, critical thinking* and *study online* with their respective themes count of 4, 3, 2, 2, 2. The theme sentiment score is between - 1 and +1 is considered neutral. The overall theme sentiment polarity is neutral. According to Koppel and Schler, neutral improves the overall accuracy and should not be considered as a state between positive and negative but as a separate class that denotes the lack of sentiment [38]. The sentence the weather is hot” for example, cannot be considered negative or positive.

Table 6: Themes Extracted from the Digital Transcript

Theme	Themes Count	Theme Sentiment Score	Theme Sentiment Polarity
practice exams	4	-0.00590241	neutral
solving problems	3	0.064992435	neutral
class discussion	2	0.313600004	neutral
critical thinking	2	0	neutral
Studying online	2	-0.057878751	neutral



4.2.3 Summative Examination

Table 7 is a chunk of the final examination results of students generated and stored in the prototype database. This live data was extracted from the database that summarized vital information including the average score of formative results, study performance, review performance, the cumulative rewards, teacher evaluation, the three scores of the summative, and the final marks. In this table, the administrator or the teacher of the class could view and analyze individually and in details all related performances of the student. The action column or Edit icon allowed the instructor to inputs additional mark to deserving students. This mechanism was a request from staff members of the faculty during initial testing and pre-survey. The F1 column shows the results of first set of examination while F2 and F3 shows the columns that stores the results of

the second and the third sets respectively after additional reading and reinforcements [39].

The results based on the implementation of the prototype shows that among the 41 students, 36 have passed the course or 88%. Previous end semester result shows a 60% passing rate and when compared to the number of student who passed the course, it is considerably higher by 28%. This achievement can be attributed to practice, and various learning strategy employed in the prototype as discussed in the initial findings published in previous papers [40]. The remaining five students or 12% out of 41 discontinued the learning process for various and personal reasons. The results of the study can greatly help improve the teaching environment.

Table 7: The Final Examination Results Module

Home Course Profiles Syllabus Lessons Activities Activated Exam Item Bank Contact Us Log Out													
Final Exam													
Stud_ID	Name	Ave_P	Study	Review	Rewards	T_Eval	F1	F2	F3	Final Mark	Remarks	View	Action
502315	Mabrouka Suleiman	7.00	4.60	0.83	5.43	0	75.01	0.00	0.00	75.01	Passed	ANALYZE	EDIT
702452	Najah Suleiman	7.00	4.40	2.04	6.44	0	78.44	0.00	0.00	78.44	Passed	ANALYZE	EDIT
802092	Rada Rabia	6.75	3.15	1.63	4.78	0	65.81	75.20	0.00	75.20	Passed	ANALYZE	EDIT
802098	Zainab Hazar	7.25	3.15	0.88	4.03	0	67.11	78.25	0.00	78.25	Passed	ANALYZE	EDIT
802137	Sherifa Abubaker	7.33	3.90	1.08	4.98	0	74.73	63.46	59.86	59.86	Failed	ANALYZE	EDIT
802144	Salhin Muftah	7.58	4.60	2.46	7.06	0	79.65	0.00	0.00	79.65	Passed	ANALYZE	EDIT
802151	Alia Sesu	6.92	5.00	1.42	6.42	7	74.93	72.60	70.63	77.63	Passed	ANALYZE	EDIT
802178	Fatma Omar	6.75	4.25	1.13	5.38	0	69.46	67.15	82.35	82.35	Passed	ANALYZE	EDIT
802197	Mabrouka Bin Sharaf	6.42	2.55	2.33	4.88	1	73.92	73.71	74.88	75.88	Passed	ANALYZE	EDIT
802236	Nora Mohamed	6.67	2.55	1.08	3.63	6	63.86	70.00	69.19	75.19	Passed	ANALYZE	EDIT
802237	Nora Nufel	7.17	4.25	0.92	5.17	0	77.14	0.00	0.00	77.14	Passed	ANALYZE	EDIT
802491	Mona Ali	6.92	3.60	1.42	5.02	6	73.68	73.47	69.67	75.67	Passed	ANALYZE	EDIT

5 CONCLUSION

This paper successfully combined 13 question types extracted from 14 publications. It also aligned the 280 questionnaires stored in the Item Bank according to cognitive schema. The cognitive schema was composed of different “verbs” words which served guide in creating questionnaires that support hierarchical cognitive development. The questionnaires were reproduced as part of the e-learning assessment with added interactivity and simulations. The questionnaires stored in the Item Bank were measured using internal reliability test and all were at acceptable level. The design factors of the assessment level were statistically significant at all assessment measurement scale. Based on the results of the study, students improved their academic performance. The number of trials in taking the practice assessment became less as the results increased. The success of the testing was attributed to the design of the assessment which allowed the students to review and reload the questionnaires several times thereby making them familiar with the graded assessment. Being interactive, the item or the question was linked to the explanation facilities, specific learning materials and review module. Document analysis using special software Semantria, the digital transcripts of the students reveal a positive sentiments towards the study as directly correlated with the increase passing rate of 28%. Although the initial results are quite convincing and acceptable, a thorough study is needed to establish the impact of the design in the diagnostic, formative and summative assessment, an intelligent profiling of the students is needed to guide them better in the learning process.

6 REFERENCES

[1] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.. *Standards for educational and psychological testing*.

Washington, DC: American Educational Research Association. 1999.

- [2] M. Birenbaum, M. New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, The Netherlands: Kluwer. (pp. 13–36). 2003.
- [3] K. Scalise & B. Gifford. Computer-Based Assessment in E-Learning: A Framework for Constructing “Intermediate Constraint” Questions and Tasks for Technology Platforms. *Journal of Technology, Learning, and Assessment*, 4(6). 2006.
- [4] E. Schreiner.. How to Design Effective Classroom Assessments. Retrieved July, 22, 2014 from http://www.ehow.com/how_7408048_design-classroom-assessments.html
- [5] J. Cowan. Designing assessment to enhance student learning. Retrieved June 10, 2014 from http://www.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/
- [6] C. G. Parshall, T. Davey & P.J. Pashley. Innovative Item Types for Computerized Testing. In W. Van der Linden, Glas, C. A. W. (Ed.), *Computerized Adaptive Testing: Theory and Practice*. Norwell, MA: Kluwer Academic Publisher. pp. 129–148, 2000.
- [7] C. G. Parshall, J. Spray, J. Kalohn, & T. Davey, T. Issues in Innovative Item Types. In *Practical Considerations in Computer-Based Testing*. New York: Springer. pp. 70–91, 2002.
- [8] S. Embretson. Generating Abstract Reasoning Items with Cognitive Theory. In S. Irvine, Kyllonen, P. (Ed.), *Item Generation for Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers. pp. 219–250, 2002.
- [9] M.E. McDonald. Developing Multiple-Choice Items. In *Systematic Assessment of Learning Outcomes*. pp. 83–120, 2002.
- [10] British Columbia. Examination Booklet. Retrieved July 7, 2014 from http://www.bced.gov.bc.ca/exams/search/grade11/english/sample/exam/0708ss_P.pdf
- [11] S. J. Osterlind. *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and*

Other Formats. Norwell, MA: Kluwer Academic Publisher. 1998.

- [12] Society De Actuaries. Construction and Evaluation of Actuarial Models. Retrieved July 8, 2014 from <http://www.soa.org/files/pdf/edu-2009-fall-exam-c-questions.pdf>
- [13] FunTrivia.com. True or False Trivia. Retrieved August 21, 2014 from http://www.funtrivia.com/quizzes/general/true_or_false.html
- [14] EHow.com.. How to Write a Fill In the Blank Questions. Retrieved August 21, 2014 from http://www.ehow.com/how_8233244_write-fill-blank-questions.html
- [15] Bloom, B. Mastery learning. New York: Holt, Rinehart, & Winston. 1971.
- [16] L. W. Anderson, & D. R. Krathwohl. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon. Boston, MA (Pearson Education Group). 2001. http://epltt.coe.uga.edu/index.php?title=Bloom's_Taxonomy
- [17] A. Churches. Bloom's Taxonomy Blooms Digitally. 2008. <http://www.techlearning.com/studies-in-ed-tech/0020/blooms-taxonomy-blooms-digitally/44988>
- [18] M. Rosenberg, M. E-Learning: Strategies for delivering knowledge in the digital. 2000.
- [19] R. Thomas.. Interactivity and Simulations. Multi-verse Solutions Ltd. 2001. <http://www.jelsim.org/resources/whitepaper.pdf>.
- [20] R. E. Clark. & T. G. Craig. "Research and Theory on Multimedia Learning Effects." In: M. Giardina (red). "Interactive Multimedia Learning Environments. Human factors and technical considerations on design issues.. NATO ASI Series. s. 19-30. 1992
- [31] Walker, J.D, Brooks, D.C, Hammond, K., Fall, B., Peifer, R., Schnell R. & Schottel, J. (2014). Practice Makes Perfect? Assessing the Effectiveness of Online Practice Examinations in Learning Biology Concepts. Retrieved August 10, 2014 from https://it.umn.edu/prod/groups/oit/@pub/@oit/@web/@evaluationresearch/documents/article/oit_article_224749.pdf
- [21] R. Mayer. The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*. Vol. 13. Pp. 125-139. 2003
- [22] Allen, I. E. and Seaman, J. (2008) *Staying the Course: Online Education in the United States, 2008* Needham MA: Sloan Consortium.
- [23] Adams, G. L. (1992). "Why Interactive?" *Multimedia & Videodisc Monitor*. New York: McGraw Hill.
- [24] Brett, P.A. (1997). *Multimedia applications for language learning - what are they and how effective are they*.
- [25] Semantria (2014). Retrieved July 29, 2014 from <https://semantria.com/features/themes>
- [26] D. George & P. Mallery,. *SPSS for Windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Boston: Allyn & Bacon. 2003.
- [27] C. Zaiontz. *Real Statistics Using Excel: Cronbach Alpha*. Retrieved January 21, 2013 from <http://www.real-statistics.com/author/>
- [28] M. Tavakol & R. Dennick. Making sense of Cronbach's alpha. *International Journal of Medical Education*. 2011; 2:53-55 Retrieved June 12, 2014 from <http://www.ijme.net/archive/2/cronbachs-alpha.pdf>
- [29] J. M. Cortina. What is coefficient alpha? An examination of theory and applications" *Journal of Applied Psychology*. pp. 98-104. http://psychweb.psy.umt.edu/denis/datadecision/front/cortina_alpha.pdf
- [30] D. Streiner. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*. 80:99-103. 2003.
- [32] Pinola, M. (2013). *Improve Your Learning with Practice Tests (and Skip Less Effective Techniques Like Highlighting*. Retrieved July 24, 2014 from <http://lifelifehacker.com/5975203/improve-your-learning-with-practice-tests-and-skip-less-effective-techniques-like-highlighting>
- [33] Scheve, T. (2014). *What Makes People Happy?*

- [34] Baumister, R.F., Campbell J., Krueger, J. & Vohs, K. (2003). Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles? A journal of the Association for Psychological Science.
- [35] Franken, R. (1994). *Human motivation* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Co.
- [36] Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques.
- [37] Turney, P. (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. . *Proceedings of the Association for Computational Linguistics*. pp. 417–424.
- [38] Koppel, M. & Schler, J. (2014). The Importance of Neutral Examples for Learning Sentiment. Retrieved August 1, 2014 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.9735&rep=rep1&type=pdf>
- [39] Ballera, M., Lukandu, I. A., & Omar A. (2014). Applying Reinforcement and Mastery Learning: How it Works Based on Personalized E-Learning Curriculum. *Proceedings of the International Conference on Computer Science, Computer Engineering and Social Media (CSCEESM2014) – Society of Digital Information and Wireless Communication (SDIWC)*. Thessaloniki, Greece. Pp 94-103.
- [40] Ballera, M., & Salih, O M. (2014). The Design of Interactive Assessment-Cognitive Schema Based System: An Exploratory Study in E-Learning Implementation. *Proceedings of the International Conference on Computer Science, Computer Engineering and Educational Technologies (CSCEET2014) – Society of Digital Information and Wireless Communication (SDIWC)*. Kuala Lumpur, Malaysia. Pp 40-50.