AN APPROACH FOR CRITICAL EVALUATION OF EXAMINATION QUESTIONS IN AN ENGINEERING FACULTY USING THE CLASSICAL BLOOM TAXONOMY

L. Ogunwolu[1,*], E. O. Ajibola[2] and A. Sosimi[3]

[1, 2, 3,] DEPT. OF SYSTEMS ENGINEERING, FACULTY OF ENGR., UNIVERSITY OF LAGOS, AKOKA, LAGOS STATE, NIGERIA.
E-mail addresses: [1] *fogunwolu@unilag.edu.ng,* [2] *oajibola@unilag.edu.ng,* [3] *asosimi@unilag.edu.ng*

ABSTRACT
Ideally, teaching curricula are designed to exude some learning behavioural outcomes in students. When students are examined through oral, practical and or theoretical evaluation schemes in examinations, the primary objective is to measure the relative extent to which students can exude the desired behavioural outcomes. Thus it is important that appropriate means and expressions are employed in assessment to measure learning. Based on an average-occurrence-rating scale, a set of Action verbs of the classical Bloom's taxonomy were used in this study in assessing the level of learning outcomes required in each of 21 Engineering Core Course examination-questions used. Rates were computed for Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation levels of the Bloom's Taxonomy. Marginalized One-Factor Statistical experiments and randomized two-factor Statistical experiments were designed using the rates computed. Analysis of Variance (ANOVA) technique was used to test within- and between-treatments variations of the differences of means for the six Bloom's taxonomy levels using hypotheses at 0.05 levels of significance for each of the experimental design scenarios. Analyzed results show significant differences in course ratings at different Bloom taxonomy levels and significances in differences in ratings of the taxonomy for different courses. Group tests of hypotheses on differences of mean-ratings for the courses show significance. The study brings to light the need for examiners in these courses to put more effort in upper levels of behavioural outcomes especially on Synthesis and Evaluation which are germane for training of good Engineers.

Key terms: *Learning outcomes, Bloom's taxonomy, Assessment, Synthesis and Evaluation, cognitive level, hierarchical systems.*

1. INTRODUCTION
Educational evaluation is the evaluation process of characterizing and appraising some aspects of an educational process [1]. It is a continuing process which is needful for effective review and enhancement of the learning which is the goal of the classroom business. Such evaluation must be done to standard and purposefully. Evaluation goes beyond measurement and assessment. Measurement determines the attributes and dimensions of learning; assessment is a process by which information is obtained relative to some goal or objectives but evaluation determines value of learning. When evaluation is done, it necessarily embraces the two other concepts and goes further to yield information on the worthiness, appropriateness, goodness, validity, legality etc. of learning [2]. Effective learning, on the other hand, is not only a by-product of intelligence and general learning environment, but also of the ability of

the teacher to bring the students to adequately exude the desired behavioural learning outcomes expected of them for the knowledge body to which they are exposed. The commonest way that students are assessed as to the level to which individuals or group of individuals can exude the expected behavioural learning outcomes is through examination which may come in several forms: practicum, oral, written (quizzes, tests, multiple choice) or a combination of these options. It has been argued that examination is not the true test of a man's ability but the education world is yet to find a better alternative. Classroom assessment is necessary to facilitate learning [3], just as it in order to assess it. Most efficient and effective student learning will result when classroom instruction and materials align with objectives or standards [4].
[5] using an ethnographic study as basis for analysis found that assessment drives not only the students but the entire learning system including, teaching practices,

* Corresponding author, tel: +234 – 803 – 353 – 8328

designing of educational programmes and the production of educational materials.

The learning process may be complicated and its desired behavioural outcomes a tall order, at times, its measure of compliance must not be ambiguous, discriminatory, based on favouritism or below the expectation of set behavioural outcomes. Learning, teaching, identifying educational goals, and thinking are all complicated concepts interwoven in an intricate web. Out of necessity, teachers must measure their students' ability. That is where examination comes in unavoidably. [6] stated that examination arose out of the need of a teacher to give a public manifestation of the effectiveness of his teaching through a display of the talents of his students, or it arose from the compulsions of a student to win fame through public debates [7]. The examination, therefore, as an instrument of liberation either gives the individual a run for his talents and the approbation of society or it gives the country relevance in the comity of nations and enables it to take care of the problems arising out of a multiracial, technocratic or stagnant society. [3] explains the concept of examination as "a test of knowledge acquired, or more generally a means of assessing intellectual capacity or ability" [8]. According to [9], the process for testing the ability or achievement of the student in any area is called examination [10].

There are normally three types of examination: A set of questions intended to test pupil's advancement as a consequence of a course of teaching; A way of qualifying candidates for a certificate or a degree in which they are desired to achieve a definite standard for a pass or honours, and A competitive test on the power of which a scholarship or other award is made to the victorious contestant. According to [11], accurately doing so requires a classification of levels of intellectual behaviour important in learning. [8] provided the measurement tool for thinking [12]. Bloom identified three domains of educational activities or learning ([13], [11]). These are the Cognitive (mental skills) or knowledge, the Affective (growth in feelings or emotional areas and the Psychomotor (manual physical skills). The intent was to develop a method of classification for thinking behaviours that were believed to be important in the processes of learning. While it should be noted that other educational taxonomies and hierarchical systems have been developed, it is Bloom's Taxonomy which remains, even after nearly fifty years, the de facto standard.

Benjamin Bloom designed a hierarchical taxonomy of cognitive skills for the teacher who is designing curriculum and formatting educational standards and objectives. This cognitive domain is laid out in six areas now quite familiar to teachers: knowledge, comprehension, application, analysis, synthesis, and evaluation. Knowledge is memorization, the ability of the student to recall information. The concept can be found in lesson plans that require the student to define, recall, or label. Examples of knowledge as a cognitive skill include learning the alphabet or memorizing important dates in history. Once the ability to gather information at the knowledge stage is mastered the student proceeds to comprehension. At this stage the student begins to see word clues such as "estimate", "explain", and "summarize". The student is not generating anything new but is putting learned knowledge into his/ her own words. At the application stage the student learns to use the knowledge. Key words appear such as "apply", "compute", or "demonstrate". At the analysis stage the student begins to generalize information to new or different situations. The student has yet to create anything wholly new, however, the cognitive process has sequenced from basic recognition and memory skills to those tools needed for abstract thought and creation. In the next stage, synthesis, the student begins to see key words such as "compose", "create", and "modify". The pre-schooler has gone from recognizing a Lego toy to using the toys to create something new. In the final cognitive stage, evaluation, the student gains the ability to judge or critique. The student can therefore compare the creations of others and validly support, explain, or defend the work [9].

Understanding that "taxonomy" and "classification" are synonymous helps dispel uneasiness with the term. University of Wisconsin Teaching Academy Publication described Bloom's Taxonomy as a multi-tiered model of classifying thinking according to six cognitive levels of complexity [11]. The Publication further subsumed that throughout the years, the levels of Bloom's taxonomy have often been depicted as a stairway, leading many teachers to encourage their students to "climb to a higher (level of) thought." The lowest three levels are: knowledge, comprehension, and application. The highest three levels are: analysis, synthesis, and evaluation. "The taxonomy is hierarchical; in that, each level is subsumed by the higher levels. In other words, a student functioning at the 'application' level has also mastered the material at the 'knowledge' and 'comprehension' levels". Although, the Classical Bloom's taxonomy has been revised to incorporate a framework that is no longer linear but a grid, the thrust of the classical work still remains relevant. This work uses the Classical version rather than the Revised

Bloom's Taxonomy of [4] out of the conviction that the latter is more of a renaming of the six levels and amount to the same taxonomy if appropriate action verbs that connect them are identified and used ([5], [6]). In the revision, the original six components are renamed so that they still relate directly to the original taxonomy but in terms that are both more relevant to today and simplified. "Knowledge" becomes "remember", "comprehension" becomes "understand", "application" is simplified to "apply", "analysis" to "analyze", and "synthesis" becomes somewhat confusingly "evaluate" as "evaluation" changes to the more descriptive "create".

Several attempts have been made to use the Bloom's taxonomy in assessing the depth of required learning outcome afforded by examinations or students' assessment in general. For one, using the Bloom's taxonomy, [14] provides an analysis of the relationship between examination questions, learning outcomes and student performance. The paper also provides some indication of the relative changes required to move toward a more appropriate association and hence improve an assessment strategy.

This work is an attempt to assess cognitive depths of selected examination questions set to assess undergraduate students in an Engineering Faculty using the Classical Bloom's taxonomy guided by identified action verbs describing each level of the taxonomy. The rest of this paper is organized as follows. Section 2 describes the research methodology dwelling on the general approach to the work, experimental design procedure, method of analysis and selection of test problem. Section 3 exhibits the results obtained from the use of the identified methods and exhibits the analyses of same. Finally, in Section 4, the results are discussed and possible extensions to the current work identified.

2. RESEARCH METHODOLOGY
2.1 General Approach
This work is an assessment of the cognitive depth of examination questions set in some Engineering courses. It is aimed at assessing the relative differences in language of assessment of the selected course examination papers and the depth of cognitive learning demanded using the Classical Bloom's taxonomy, against the backdrop of Engineering courses which are meant to be able exude substantial learning behaviour in the upper levels of the taxonomy. A random selection of three course-examination question papers was made from seven of the programmes available in seven teaching units of the Faculty of Engineering used as a test bed. For each teaching unit, one course question paper was chosen at each out of four (200, 300, 400 and 50) levels. The Distribution of the course by teaching unit and course codes are as depicted in Table 1.

Using the action verbs for each of the 6 Classical Bloom's taxonomy levels as in Appendix Table A, Bloom's Taxonomy scores were calculated by counting the instances in which any of the words in the standard action verbs listing for each of the 6 levels of the taxonomy occur within the text strings of the question papers. The number of questions in each of the questions papers was also noted. The final ratings for each course paper and for each cognitive level of the taxonomy were then calculated as an average of the number of occurrences (of each level per question). This provided a uniform basis for comparison across all questions and across all levels of the taxonomy. The average ratings obtained (per course, per question and per Classical Bloom taxonomy level) are exhibited in Appendix Table A.

2.2 Experimental Designs
Even though the sampling from the population of course examination papers used for this course is random, it is representative cutting across all levels of programmes taught in the Faculty. The problem is a two-factor experiment as it is desired to determine whether there are statistically significant differences between (1) the ratings of the individual courses (called treatments) and (2) the ratings of the different levels of the Bloom's taxonomy (called columns).

Table 1: Summary of the courses with their corresponding host Teaching Units

Host Teaching Unit	Course Units	Course Codes by levels			
		200	300	400	500
Chemical Engineering (CH)	3	CHG202	CHG301	CHG407	
Civil and Environmental Engineering (CE)	3	CEG202	CEG306	CEG411	-
Electrical/Electronics Engineering (EE)	3	EEG201	EEG308	EEG405	-
Mechanical Engineering (ME)	3	MEG201	-	MEG413	MEG523
Metallurgical and Materials Engineering (MM)	3	-	MME311	MME405	MME505
Systems Engineering (SS)	3	SSG204	SSG312	SSG411	-
General Courses (GE)	3	-	GEG301	GEG402	GEG501

Different experimental design schemes are employed to eliminate possible sources of variation or error. These are:

Scheme 1: (Marginalized Grand Factor Scheme (MGFS) Treatment as a one-factor experiment (marginalized experimental design in which one factor is used at a time thus marginalizing the second). It is equivalent to two one-factor experiments – *The Marginal One-Factor Design for Variations within treatments (Bloom's Taxonomy levels) and The Marginalized One-Factor Design for Variations between Treatments (Courses)*.

Scheme 2: Marginalized Sector Factor Scheme (MSFS) This is an equivalent treatment of the Scheme for individual sector (programme units – e.g. Mechanical Engineering is a unit)

Scheme 3: The Two-Factor Experiment (TFE) This is a Treatment/Blocks scheme, incorporating the two factors in blocks of treatment to eliminate variability of the difference in blocks and treatments.

Scheme 4: Two-Factor Taxonomy Level Segmentation (TFLS) In this scheme, treatments were divided into two contiguous groups (e.g. Boom Taxonomy Levels 1 and 2 versus Levels 3 to 6; Levels 1 to 4 versus Levels 5 and 6; Levels 1 to 3 versus Levels 4 to 6). This is aimed at gaining understanding of the relative contiguous strengths of the levels taxonomy ratings.

Scheme 5: Two-Factor Teaching Unit/Courses Comparison among Taxonomy Levels In this experimental scheme, two-factor experiments are designed for the purpose of rating performances of Teaching Units on the basis of Bloom's Taxonomy Levels for all course questions rated in the experimental design especially along deficiency lines.

2.3 Method of Analysis

The method of analysis of variance (ANOVA) was used in analyzing results to evaluate whether there are statistically significant differences in the ratings. As explained in the preceding subsection, marginalized one-factor and two-factor experiments were designed for the purpose of analysis of differences in ratings.

For one-factor, suppose, $\bar{x}_j$ represents the treatment (row) mean of the measurements in the jth rowwhere,

$$\bar{x}_j = \frac{1}{b}\sum_{k=1}^{b} x_{jk} \; ; \; j = 1, 2, \; \ldots \; , a \qquad (1)$$

The grand mean or overall mean is the mean of all measurements, $\bar{x}$

$$\bar{x} = \frac{1}{ab}\sum_{j=1}^{a}\sum_{k=1}^{b} x_{jk} \qquad (2)$$

The total variation, v, is defined as the sum of the squares of the deviations of each measurement from the grand mean,

$$v = \sum_{j,k}\left(x_{jk} - \bar{x}\right)^2 \qquad (3)$$

By writing the identity,$x_{jk} - \bar{x} = \left(x_{jk} - \bar{x}_j\right) + \left(\bar{x}_j - \bar{x}\right)$ and summing over j and k, it is easy to show that,

$$v = \sum_{j,k}\left(x_{j,k} - \bar{x}\right)^2 = \sum_{j,k}\left(x_{jk} - \bar{x}_j\right)^2 + \sum_{j,k}\left(\bar{x}_j - \bar{x}\right)^2$$
$$= \sum_{j,k}\left(x_{jk} - \bar{x}_j\right)^2 + b\sum_{j}\left(\bar{x}_j - \bar{x}\right)^2$$
$$= v_w + v_b \qquad (4)$$

Where, $v_w = \sum_{j,k}\left(x_{jk} - \bar{x}_j\right)^2$ is called the variation within treatments while the variation between treatments is given by$v_b = b\sum_{j}\left(\bar{x}_j - \bar{x}\right)^2$.

For a two-factor experiment, assuming a treatments and b blocks, the mean of entries jth row,$\bar{x}_j$, $j = 1, 2, \ldots, a$, the mean of entries in the kth column, $\bar{x}_k, k = 1, 2, \ldots, b$ and the grand mean, $\bar{x}$,

$$\bar{x}_j = \frac{1}{b}\sum_{k=1}^{b} x_{jk} \; , \; \; \bar{x}_k = \frac{1}{a}\sum_{j=1}^{a} x_{jk} \; \; \bar{x} = \frac{1}{ab}\sum_{j,k}\bar{x}_{jk} \qquad (5)$$

As in the one-factor case, the total variation is given by,

$$v = \sum_{j,k}\left(x_{jk} - \bar{x}\right)^2 \qquad (6)$$

By writing the identity, $x_{jk} - \bar{x} = \left(x_{jk} - \bar{x}_j - \bar{x}_k + \bar{x}\right) + \left(\bar{x}_j - \bar{x}\right) + \left(\bar{x}_k - \bar{x}\right)$ and summing over j and k, it is easy to show that,

$$v = v_e + v_r + v_c \qquad (7)$$

where, v_e is the variation due to error or chance

$$v_e = \sum_{j,k}\left(x_{jk} - \bar{x}_j - \bar{x}_k + \bar{x}\right)^2 \qquad (8)$$

v_e is the variation between rows (treatments)

$$v_r = b\sum_{j=1}^{a}\left(\bar{x}_j - \bar{x}\right)^2 \qquad (9)$$

v_c is the variation between column (blocks)

$$v_c = a\sum_{k=1}^{b}\left(\bar{x}_k - \bar{x}\right)^2 \qquad (10)$$

For the One-factor experiment schemes in section 2.2, hypotheses are built to test the differences between means of the treatments. The null hypothesis tests that all means of treatments are equal while the Alternative tests the contrary.

One
$- Factor\ Hypotheses:\begin{cases}H_0: Treatment\ means\ are\ equal \\ H_1: Treatment\ means\ are\ not\ equal\end{cases}$ (11)

Using ANOVA, the statistic $F = S_b^2/S_w^2$, which is the ratio of the calculated between treatments variance and within treatments variance which has F-distribution with (a – 1) and a (b- 1) degrees of freedom respectively are used to test the hypotheses. At some specified level of significance, using a one-tailed test of the F-distribution, the Null hypothesis is rejected if F statistic is higher than the tabulated Critical F-ratio (i.e. the means are not equal), otherwise, the Null hypothesis is accepted, translating to the fact that the differences between the treatment means are Non-significant. It is also significant to note that the greater the F-statistics the more pronounced the discrepancies between the means.

For the two-factor experimental designs, two Null hypotheses are tested against an Alternative Hypothesis. The first Null hypothesis, $H_0^{(1)}$ tests that all treatment (row) means are equal, while the second Null hypothesis, $H_0^{(2)}$ tests that all block (column) means are equal. The Alternative Hypothesis tests to prove the contrary as shown in (12) below this page.

Using ANOVA, the statistic $F^{(1)} = S_r^2/S_e^2$, which is the ratio of the calculated between treatments (rows) variance and error variance which has F-distribution with (a – 1) and (a – 1) (b – 1) degrees of freedom respectively and $F^{(2)} = S_c^2/S_e^2$ is the ratio of the calculated between blocks (columns) variance and error variance which has F-distribution with (b – 1) and (a – 1) (b – 1) degrees of freedom respectively are used to test the hypotheses. At some specified level of significance, using a one-tailed test of the F-distribution, each of the Null hypotheses are rejected if its F statistic is higher than the tabulated Critical F-ratio (i.e. the means are not equal), otherwise, the Null hypothesis is accepted (i.e. the differences between the treatment means are Non-significant). It is also significant to note that the greater the F-statistics the more pronounced the discrepancies between the

means. In order to minimize the computational effort, ANOVA mode of the Statistical Package of Microsoft Excel 2007 was used to execute the designed experimental schemes.

3. RESULTS AND ANALYSES
In this section, results of the designed experiments and results of tests of hypotheses built on them are exhibited.

3.1 Experimental Design Results
Experimental Schemes 1, 3 and 4
The scenario experimental design is as depicted in Appendix Table B.
Experimental Scheme 2
In this experimental scenario, experiments are designed for ratings of questions from each of the seven programmes used. This is with a view to assessing whether there are statistical significance in differences of the treatment (row) means and in the columns also (the courses). Below is a typical design for Mechanical Engineering courses.

Table 2: Typical design average ratings for Mechanical Engineering Courses

	MEG 201	MEG 413	MEG 523	Sum
Knowledge	0.83	0.00	0.00	0.83
Comprehension	0.50	0.17	0.20	0.87
Application	0.50	1.17	0.40	2.07
Analysis	1.33	0.83	1.00	3.16
Synthesis	0.17	0.17	0.80	1.14
Evaluation	0.33	0.50	0.00	0.83

3.2 ANOVA RESULTS
3.2.1 Experimental scheme 1
The results for Marginal One-Factor Taxonomy Level (Case 1) and Marginal One-Factor Course (Case 2) are as presented in Table 3.

Table 3: Results for Cases 1 and 2 of the Marginal One-Factor Taxonomy Level

Case	Treatment	Degree of Freedom	F- Critical	Computed F-value	P- Value	Comment
1	Between Within	5 120	2.289851	9.180861	0.0002	Significant
2	Between Within	20 105	1.671357	0.624553	0.886292	Non-Significant

$$Two-Factor\ Hypothesis: \begin{cases} Null: \begin{cases} H_0^{(1)}: All\ treatment\ (row) means\ are\ equal \\ H_0^{(2)}: All\ block\ (column) means\ are\ equal \end{cases} \\ Alternative: Treatment/Block\ means\ are\ not\ equal \end{cases} \quad (12)$$

The result shows that when considered as one-factor experiment the ratings show significance in the differences in means of the different Bloom's taxonomy levels, although the observed level of significance of fit (p-value) is very low but the differences in the means for different course ratings are statistically insignificant despite averagely high level of fit to F-distribution. It is an indication that substantial differences occur in ratings for the levels of Bloom's taxonomy. This calls for in-depth study of the differences.

3.2.2 Experimental scheme 2
This scheme studied the possible differences due to different programme/unit ratings for different levels of Bloom's taxonomy and for different Courses within each teaching unit. The results of ANOVA and tests of significance for the various cases under this scheme depicted in Tables 4 and 5. For Marginal Bloom's levels variation of ratings for individual teaching units, the results of ANOVA are depicted in Table 4.

These results in Table 3 clearly indicate that the variations between ratings at the different levels of Bloom's taxonomy are not statistically significant in most of the units, except in the Civil and Environmental unit where there is slight significance. This is an indication that ratings for all courses for each Teaching Unit can be safely aggregated as blocks to study variations in the Bloom's taxonomy level ratings.

Table 4: ANOVA results for Marginal Bloom's levels variation of ratings for teaching units

Case (Units)	Treatment	Degree of Freedom	F- Critical	Computed F-value	P- Value	Comment
Chemical	Between	5	3.105875	1.35	0.309044	Non-Significant
	Within	12				
Civil	Between	5	3.105875	3.523903	0.0034269	Significant
	Within	12				
Elect/Elect	Between	5	3.105875	1.481633	0.266543	Non-Significant
	Within	12				
General	Between	5	3.105875	2.538462	0.086424	Non-Significant
	Within	12				
Mechanical	Between	5	3.105875	2.575633	0.083276	Non-Significant
	Within	12				
Met Mat	Between	5	3.105875	1.852976	0.176844	Non-Significant
	Within	12				
Systems	Between	5	3.105875	1.752496	0.197349	Non-Significant
	Within	12				

Table 5: Marginal Course Ratings Variation for individual units

Case (Units)	Treatment	Degree of Freedom	F- Critical	Computed F-value	P- Value	Comment
Chemical	Between	2	3.68232	1.261682	0.311568	Non-Significant
	Within	15				
Civil	Between	2	3.68232	0.626896	0.547677	Non-Significant
	Within	15				
Elect/Elect	Between	2	3.68232	0.015806	0.984335	Non-Significant
	Within	15				
General	Between	2	3.68232	0.686081	0.518668	Non-Significant
	Within	15				
Mechanical	Between	2	3.68232	0.373573	0.694497	Non-Significant
	Within	15				
Met Mat	Between	2	3.68232	2.90332	0,085935	Non-Significant
	Within	15				
Systems	Between	2	3.68232	0.100798	0.904723	Non-Significant
	Within	15				

Similarly, for Marginal Course Ratings Variation for individual units, the results of ANOVA are depicted in Table 4. The results are similar to what obtained when one-factor variation was sought between levels of Bloom's taxonomy.

3.2.3 Experimental scheme 3

The results of the full two-factor experiment for all course ratings at all Bloom's taxonomy levels are as depicted in Table 6. The two-factor experiment confirms the results obtained in Experimental Schemes 1 and 2; although the differences in ratings among the courses may not be statistically significant, the differences between the levels of the Bloom's taxonomy are. This informed the design of the Experimental Scheme 4, in which searchlight is beamed on the differences in ratings among the different Bloom's taxonomy level in a contiguous manner, giving the 'staircase' nature of the levels.

3.2.4 Experimental Scheme 4

The ratings for all courses are segmented into two contiguous Bloom's taxonomy level groups to study

where in particular the gaps in variation among the levels (Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation) lie. The groups are constituted as below.

Group 1: Taxonomy Levels 1 and 2 versus Taxonomy Levels 3 to 6

Group 2: Taxonomy Levels 1 to 3 versus Taxonomy Levels 4 to 6

Group 3: Taxonomy Levels 1 to 4 versus Taxonomy Levels 5 and 6

Tables 7, 8 and 9 exhibit the comparative 2-factor ANOVA results obtained for the respective groups.

The comparative results show that there are statistical significant differences in the ratings among the four upper Bloom's taxonomy levels while at the lower two the differences are Non-significant.

The comparative results also show that there are statistical significant differences in the ratings among the three upper Bloom's taxonomy levels while at the lower three the differences are Non-significant.

Table 6: Two-factor experiment outcomes for all course ratings at all levels

Treatment	Degree of Freedom	F- Critical	Computed F-value	P- Value	Comment
Rows (Bloom levels)	5	2.305318	8.969022	0.004	Significant
Column (Courses)	20	1.676434	0.861556	0.634368	Non-Significant

Table 7: Two-factor ANOVA results for group 1

Group 1	Group Mean	Treatment	Degree of Freedom	F- Critical	Computed F- value	P- Value	Comment
Levels 1 and 2	0.69	Rows (Levels)	1	4.351243	0.541932	0.47018	Non-Significant
		Columns (Courses)	20	2.124155	1.498295	0.186759	Non-Significant
Levels 3 to 6	0.60	Rows (Levels)	3	2.758078	17.90067	0.0000205	Significant
		Columns (Courses)	20	1.747984	0.886205	0.60412	Non-Significant

Table 8: Two-factor ANOVA results for group 2

Group 2	Mean	Treatment	Degree of Freedom	F- Critical	Computed F- value	P- Value	Comment
Levels 1 to 3	0.74	Rows (Levels)	2	3.231727	0.544272	0.584501	Non-Significant
		Columns (Courses)	20	1.838859	0.938563	0.547143	Non-Significant
Levels 4 to 6	0.52	Rows (Levels)	2	3.231727	27.73202	0.000278	Significant
		Columns (Courses)	20	1.838859	0.930904	0.555377	Non-Significant

These comparative results (Table 9) show that the lower four levels and also at the upper two there are no statistically significant differences in ratings. That the upper two levels show no statistically significant differences in rating now compared the cases at the upper level in Groups 1 and 2 simply shows that the significant variations in the upper levels are accounted for by the variations in levels 5 and 6. In addition, the means of the two groupings show that the two upper levels are very weak comparatively to the lower four.

3.2.5 Experimental scheme 5

For this experimental scheme attempts are made, based on the computed ratings to further establish the deductions made in the last section as to the weakness of the two uppermost levels (Synthesis and Evaluation) in the ratings of the examination questions used compared to the other four lower taxonomy levels. Cross-correlation coefficients were initially computed to see how any pair of data correlates and thus form an informed basis of comparison or non-comparison.

The cross-correlation coefficients of pairs of data obtained for all the taxonomy levels are depicted in Table 10.

Table 10 shows a lean cross-correlation (< 0.5) between the Analysis, Synthesis and Evaluation levels and the other levels. In particular, Synthesis and Evaluation generally has lean cross-correlation with each of the rest except with themselves. Again, Synthesis data explains Evaluation data well (0.789) but Evaluation does not explain Synthesis data at all. Thus, based on this table, the deductions made in Section 3.2.4 are further buttressed.

A further analysis of the data was undertaken for the different Teaching Units by examining the results of two-factor experiments at the 2 uppermost Bloom's taxonomy levels and the lower 4 levers, the former case being the region where deficiencies have been identified as in Section 3.2.4. The results are comparable in the sense that each block considered has the same degrees of freedom and hence the same Critical F-value. The lower the F-value the more insignificance is the difference in means among levels and among various teaching units.

Result in Table 11 rates the Mechanical Engineering courses highest among the 2 uppermost taxonomy levels and Chemical Engineering courses least. Mechanical Engineering courses rated were also rated best among the courses rated in the 2 uppermost Bloom's taxonomy levels and the Civil Engineering courses least. Rating in this way, especially if made a constant practice, can challenge teaching units to pay closer attention to the deficient cognitive demand of their courses.

Table 9: Comparative results of ANOVA for all levels under consideration

Group 3	Mean	Treatment	Degree of Freedom	F- Critical	Computed F-value	P- Value	Comment
Levels 1 to 4	0.83	Rows (Levels)	3	2.758078	2.204794	0.096782	Non-Significant
		Columns (Courses)	20	1.747984	0.923938	0.560627	Non-Significant
Levels 5 and 6	0.23	Rows (Levels)	1	4.351243	1.58317	0.222796	Non-Significant
		Columns (Courses)	20	2.124155	1.079338	0.433064	Non-Significant

Table 10: Cross-Correlation Coefficients for course ratings of pairs of taxonomy levels

	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Knowledge		2.58E-07	0.844	0.305	8.93E-09	0
Comprehension	0.387		0.169	0.283	9.47E-10	0
Application	0.727	3.73E-09		0.753	6.52E-07	0
Analysis	0.013	7.31E-14	0.357		5.88E-18	0
Synthesis	0.213	0.381	0.209	0.032		0
Evaluation	0.046	0.200	0.048	0.008	0.789	

Table 11: Teaching Units' Performances in the Uppermost 2 Taxonomy Levels

Uppermost 2 Bloom Taxonomy Levels				
	F-Value		F-Value	
Teaching Unit	Rows (Levels)	Rank	Columns (Courses)	Rank
Chemical	17.53516	7	1	3
Civil	1	6	7.390533	7
Elect/Elect	0.307692	2	1.461538	6
General	3.552862	4	1	3
Mechanical	0.086289	1	0.060968	1
Met Mat	0.769723	3	0.744136	2
Systems	3.552862	4	1	3
F-Critical	18.51282051		19	
Degrees of Freedom	1		2	

Table 12: Teaching Units' Performances within the Lower 4 Taxonomy Levels

Lower 4 Bloom Taxonomy Levels				
	F-Value		F-Value	
Teaching Unit	Rows (Levels)	Rank	Columns (Courses)	Rank
Chemical	0.45736	2	1.665187	6
Civil	1.216733	4	1.540735	5
Elect/Elect	0.415354	1	0.094488	2
General	1.518047	5	0.936818	3
Mechanical	3.563012	6	1.346402	4
Met Mat	3.571932	7	3.588373	7
Systems	0.479542	3	0.058605	1
F-Critical	4.757062663		5.14325285	
Degrees of Freedom	3		2	

Table 12 shows the relative performance in the 4 lower (bottom-most) taxonomy ratings where on a general basis the Engineering courses rated were adjudged strong (Section 3.2.4). In it, the Electrical/Electronics Engineering courses rated were adjudged best among the Bloom taxonomy levels while Metallurgical and Materials Engineering courses rated were ranked least. Among the courses in the lower 4 taxonomy levels and among the courses rated, Systems Engineering performed best while Metallurgical and Materials Engineering courses rated were ranked least. This again is a pointer to relative strengths of the cognitive demands of the questions rated even where Engineering courses rated were rated to have performed well.

4. CONCLUSION

In this work, random choices of examination course questions in an Engineering faculty have been rated using the Classical Bloom's taxonomy. The thrust has been to assess these questions on a rational basis as to the level of cognitive demand in the questions against the backdrop of the fact that cognitive demand for engineering courses should lean heavier on the three upper Bloom's taxonomy. The results show that the course examination questions are deficient in the crucial two uppermost levels of the Bloom's taxonomy (Synthesis and Evaluation). Furthermore, the ratings were probed to reveal relative performances of teaching units in the 2 uppermost levels where weakness was identified and the four lower taxonomy levels where relative strength has been identified. It is hoped that in the nearest future the Examination course questions will be set by the examiners with adequate attention given to Synthesis and Evaluation levels of the Bloom's taxonomy so that higher quality products may be produced by the Higher Educational Institutions, particularly in the Institution studied.

Conflict of Interest: *The paper has no conflict of interest whatsoever.*

5. REFERENCES

[1] Wikipedia (n.d.) "Educational Evaluation" https://en.wikipedia.org/wiki/Educational_eval uation, Accessed on February 4, 2015.

[2] Kizlik, Bob (2018), "Measurement, Assessment, and Evaluation in Education", www.adprima.com/measurement.htm Accessed on February 6, 2018.

[3] Aggarwal, J. C. *Essentials of Examination System: Evaluation Tests and Measurement, 1E*. Vikas Publishing House Pvt Ltd, 1997.

[4] Anderson, Lorin W., David R. Krathwohl, Peter W. Airasian, Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, and Merlin C. Wittrock. *"A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives"*, abridged edition. White Plains, NY: Longman, 2001.

[5] Havnes, A. "Examination and learning: an activity-theoretical analysis of the relationship between assessment and educational practice." *Assessment & Evaluation in Higher Education* 29, No. 2, 159-176, 2004.

[6] Mascarenhas, M. *Examinations: An Informative Update*. Doaba House, Delhi, 1991.

[7] Ayvaci, H. S., and A. Turkdogan. "Analysing "science and technology course exam questions" according to revised Bloom taxonomy." *Journal of Turkish Science Education* 7, no. 1: 13-25, 2010.

[8] Bloom, B. S. *Taxonomy of educational objectives: The classification of educational goals: Cognitive Domain*. Longman, 1956.

[9] Good, C. V., and W. R. Merkel. *"Dictionary of Education* (3rd)." New York, Mc-Graw-Hill, 1973.

[10] Bümen, Nilay T. "Effects of the original versus revised Bloom's Taxonomy on lesson planning skills: A Turkish study among pre-service teachers." *International Review of Education* 53, no. 4: 439-455, 2007.

[11] Forehand, Mary. "*Bloom's Taxonomy: Original and Revised*" http://eit.tamu.edu/JJ/DE.Blooms Taxonomy.pdf Accessed on February 4, 2015.

[12] Dave, R. H. *Developing and Writing Behavioral Objectives. (R. J. Armstrong, ed.)*. Educational Innovators Press. Tucson, Arizona, 1975.

[13] EET, Encyclopaedia of Education Technology. http://eet.sdsu.edu/eetwiki/index.php/Blooms_Taxonomy_in_learning_environments, Accessed on September 25, 2016.

[14] Jones, Karl O., Janice Harland, Juliet MV Reid, and Rebecca Bartlett. "Relationship between examination questions and bloom's taxonomy." In *Frontiers in Education Conference, 2009. FIE'09. 39th IEEE*, pp. 1-6. IEEE, 2009.

APPENDIX A

Definitions	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Bloom's Definition	Remember previously learned information.	Demonstrate an understanding of the facts.	Apply knowledge to actual situations.	Break down objects or ideas into simpler parts and find evidence to support generalizations	Compile component ideas into a new whole or propose alternative solutions.	Make and defend judgments based on internal evidence or external criteria.
Verbs	Arrange Define Describe Duplicate Identify Label List Match Memorize Name Order Outline Recognize Relate Recall Repeat Reproduce Select State	Classify Convert Defend Describe Discuss Distinguish Estimate Explain Express Extend Generalized Give example(s) Identify Indicate Infer Locate Paraphrase Predict Recognize Rewrite Review Select Summarize Translate	Apply Change Choose Compute Demonstrate Discover Dramatize Employ Illustrate Interpret Manipulate Modify Operate Practice Predict Prepare Produce Relate Schedule Show Sketch Solve Use Write	Analyze Appraise Breakdown Calculate Categorize Compare Contrast Criticize Diagram Differentiate Discriminate Distinguish Examine Experiment Identify Illustrate Infer Model Outline Point out Question Relate Select Separate Subdivide Test	Arrange Assemble Categorize Collect Combine Comply Compose Construct Create Design Develop Devise Explain Formulate Generate Plan Prepare Rearrange Reconstruct Relate Reorganize Revise Rewrite Set up Summarize Synthesize Tell Write	Appraise Argue Assess Attach Choose Compare Conclude Contrast Defend Describe Discriminate Estimate Evaluate Explain Judge Justify Interpret Relate Predict Rate Select Summarize Support Value

APPENDIX B

Courses	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	CEG 202	CEG 306	CEG 411	CHG 202	CHG 301	CHG 407	EEG 201	EEG 308	EEG 405	GEG 401	GEG 402	GEG 501	MEG 201	MEG 413	MEG 523	MME 311	MME 401	MME 505	SSG 310	SSG 312	SSG 411
Knowledge	0.38	1.50	0.80	2.50	0.67	0.33	1.00	1.60	0.20	1.50	1.33	0.50	0.83	0.00	0.00	0.60	0.25	0.40	1.25	0.00	0.25
Comprehension	0.75	1.33	2.20	1.50	0.33	0.17	0.40	0.60	0.60	0.00	0.00	0.25	0.50	0.17	0.20	1.60	0.75	0.20	0.00	0.67	1.00
Application	0.75	1.00	0.00	0.83	0.17	1.17	0.60	0.60	2.40	1.50	1.33	0.75	0.50	1.17	0.40	0.60	1.00	0.20	1.00	1.33	0.00
Analysis	1.25	2.17	0.60	0.33	0.83	0.50	1.40	0.60	1.00	0.00	3.00	1.25	1.33	0.83	1.00	1.60	1.25	1.00	0.50	1.33	1.50
Synthesis	0.13	0.00	0.20	0.17	0.33	0.17	0.20	0.80	0.00	0.50	0.33	0.00	0.17	0.17	0.80	0.40	0.25	0.40	0.50	0.33	0.00
Evaluation	0.00	0.00	0.20	0.00	0.00	0.00	0.40	0.20	0.00	0.00	0.00	0.00	0.33	0.50	0.00	1.00	1.00	0.00	0.00	0.00	0.00