

## USE OF BLOOM'S TAXONOMIC COMPLEXITY IN ONLINE MULTIPLE CHOICE TESTS IN GEOMATICS EDUCATION

H. Bâki Iz and H.S. Fok

Department of Land Surveying and Geo-Informatics  
The Hong Kong Polytechnic University

### ABSTRACT

*Tests are the most common method to evaluate teaching effectiveness throughout the period of instruction (formative) and assess student learning outcomes at the end of the instruction period (summative). Teaching geomatics is not an exception to this rule. The recent proliferation of computer hardware and software technologies now makes Multiple Choice (MC) type assessment methods more attractive. MC tests provide more objective measures on assessing student learning with efficiency over the Internet. In this study we show, using in-house software, that multiple choice tests when combined with technology, especially multimedia, can be flexible and easy to use. The interface has two unique features: 1) it accommodates questions of various multimedia formats, diagrams, tables, images, or equations that can be conveniently incorporated into the examination; 2) it tracks the students' time spent on each question during the examination. We found that 79 percent of the variation in the average time students spent on each question can be explained by Bloom's taxonomy (with a correlation coefficient on 0.89). Hence by grouping exam questions at Bloom's taxonomic levels we can detect student learning difficulties effectively within each level.*

KEYWORDS: Bloom's taxonomy. Education. Multiple-choice. Online tests.

### INTRODUCTION

Tests are necessary to evaluate the teaching effectiveness throughout the instruction period (formative) and to assess student learning outcomes at the conclusion of the instruction period (summative). The Constructed-Response (CR) questions (such as short answer, essay, or problem-solving questions) are often used by educators and these require students to both organize their ideas and create their own answers.

However, CR questions sometimes tend to be highly subjective and unreliable measures of student performance [3]. They are also limited to assess a wide range of knowledge in short time.

Moreover, because of the ever-increasing teaching/research/service loads at all levels of education it is difficult for educators to strike a balance between time consuming and labor-intensive continuous assessment imposed by CR questions, research, and service activities, especially if they are teaching large classes [18].

Alternatively, the proliferation of computer hardware and software technology makes Multiple Choice (MC) type of assessment methods more attractive. These tests provide more objective measures on assessing student learning with efficiency over the Internet [11] [19] [21]. MC based assessment methods are usually regarded as the most valuable and the most applicable test form to measure educational objectives such as "inferential reasoning, reasoned understanding, sound judgment and discrimination" on the part of the student [12].

MC tests are sometimes criticized as testing only the knowledge level (rote facts) of Bloom's taxonomy. However they can also measure higher taxonomy levels [12]. Students can be tested to discriminate among alternatives that can require a level of mastery that a CR test may not be able to detect. The best answer type MC questions can measure the degree of understanding of abstract concepts [12]. As for all the other

assessment methods, the measurement value of MC tests depends more on the careful selection of questions with appropriate distracters (choices) than any other factor.

In this study we show that multiple choice tests when combined with technology, especially multimedia, can be made flexible and easy to use. With the advancement of information technology (IT), most introductory courses could be taught on the Internet by providing a rich resource of material in an timely manner, such as text, graphics, animation and visualization [16] [17]. Assessment could also be done in the same manner using multimedia [5] [8] [9] [22]. Currently, a number of the educational and professional bodies, such as TOEFL and Medical Council of Canada (MCC) [15], adopted this new technology for assessment since online testing provides an efficient way to collect and grade large numbers of student learning [7]. As a result, educators give immediate feedback to the students and monitor their learning difficulties [6]. Moreover, internet-based MC tests provide great flexibility for question randomization, editing, pre-testing, storage and recycling of questions [10] [14].

Technology also helps avoid inconsistency in grading essays due to the subjectivity of different educators and provides accuracy in marking tests hence, enabling fair assessment of students [13]. In addition MC tests also help students avoid losing points for poor spelling and writing skill [4] [23] especially for second language learners. As a result, it reduces student anxiety [20] and enables educators to test students' learning on a wide range of subject materials [1].

In the following sections we will present first our online multiple choice test interface which makes use of multimedia technologies through power point slides over the internet. The software also monitors how long the students work on each question during the test. We postulate that the average time a student spends on a multiple choice item is related to its taxonomic complexity. We use Bloom's taxonomy to quantify the complexity of the MC items and compare them to the average time students spent on answering each question. We show that there is a significant correlation between these two variables. As a result, instructors can identify those questions on which students spend more time in answering rather than predicted to identify topics or issues in which they have difficulties, at a given confidence level.

#### OVERVIEW OF THE INTERNET-BASED MULTIMEDIA MULTIPLE CHOICE TEST INTERFACE

We used the ASP .NET web programming language to develop an Internet-based multimedia Multiple Choice (MC) test program, "Online MC Test". The interface has two modules; administrator module and student module which are accessible through the Internet using Internet Information Service (IIS) in a Windows Server. The operational environment of the program is displayed in Figure 1.

##### *Administrator Module*

The Administrator module is an application program used for administering general settings for the Online MC test. The toolbar menu consists of settings for configuring the assessment (Figure 2), setting the answer key, viewing students' test results and calculating descriptive statistics based on the test results.

Once the administrator password is entered and submitted, the assessment settings page appears as shown in Figure 2. On this page, the assessment title, examination date, access code to enter to the test, number of MC questions (items), number of distracters for each item and media for the delivery of question (power point, etc.) are defined. Randomized question ordering is offered as a setting. If selected, questions for each student would appear in different chronological order during the test in order to avoid

students sitting next to each other from cheating during the test. Another choice box allows or denies students getting access to their exam results and the key (Figure 6) after the test.

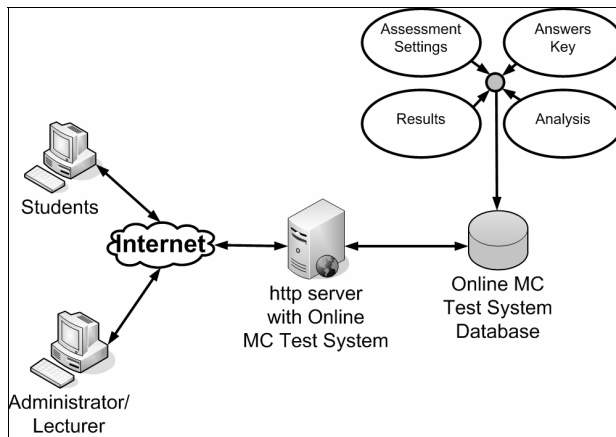


Fig. 1. Online Multiple Choice Test Environment

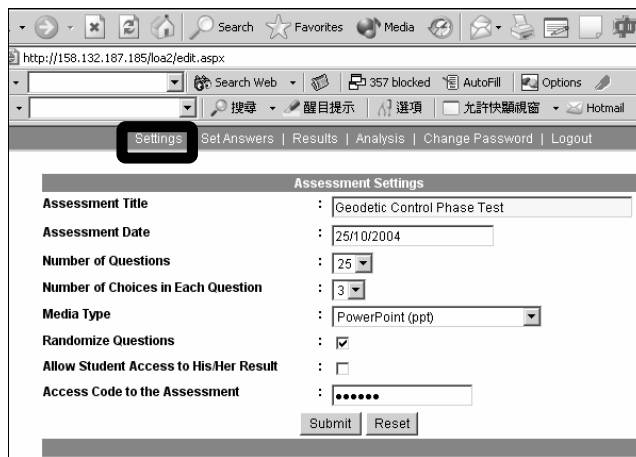


Fig. 2. Administrator module page: Assessment Settings (accessed from the circled toolbar menu item)

“Set Answers” page enables instructor to set the key for questions. Instructor could set exactly one answer or any possible combination. In this case, any possible combination within three answers (i.e. 8 combinations) could be an answer so that students could not easily guess the answer by chance without prior knowledge (Fig. 3).

Selecting “Results” option in the top menu bar opens the page where the percentage score for the correct answers is accessed by the instructor once the test is finalized (Figure 4). The page displays student IDs, their names, start time, and percentage scores. Instructor views test results of a particular student by clicking on the “Details” button which opens another page (Figure 5). The display on this page is quite similar to the original test screen with an additional column side next to the students’ answers. The incorrect answers are highlighted in red. Finalized results can be downloaded into an Excel file to be used for the analysis of results and record keeping.

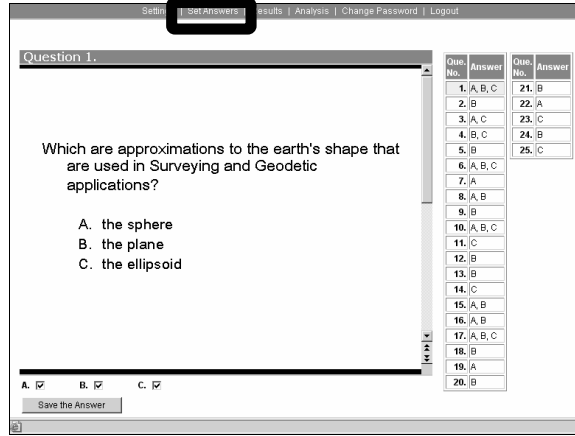


Fig. 3. Administrator module page: Set Answers



Fig. 4. Administrator module page: Results

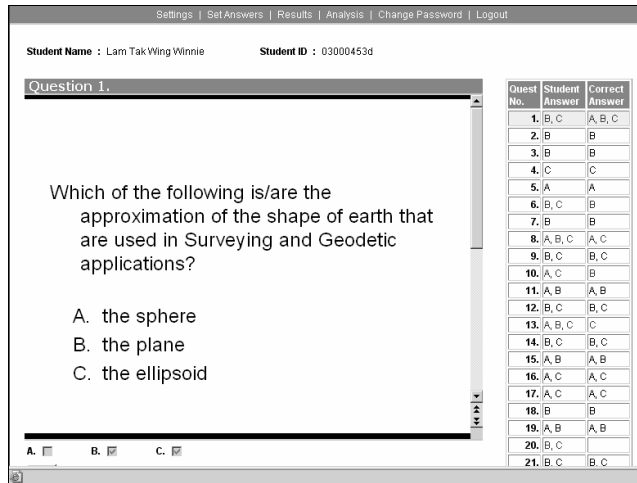


Fig. 5. Administrator module page for displaying details of a particular student's performance

The *analysis* tab on the tool bar menu opens a page where the mean and median values of the students' time utilization are displayed for each question (Figure 6). A histogram on this page shows the distribution of time student spent on each question. The location estimates and histogram information are collected and displayed for further analysis to detect and monitor students' learning difficulties using one of the methods that will be discussed in the next section.

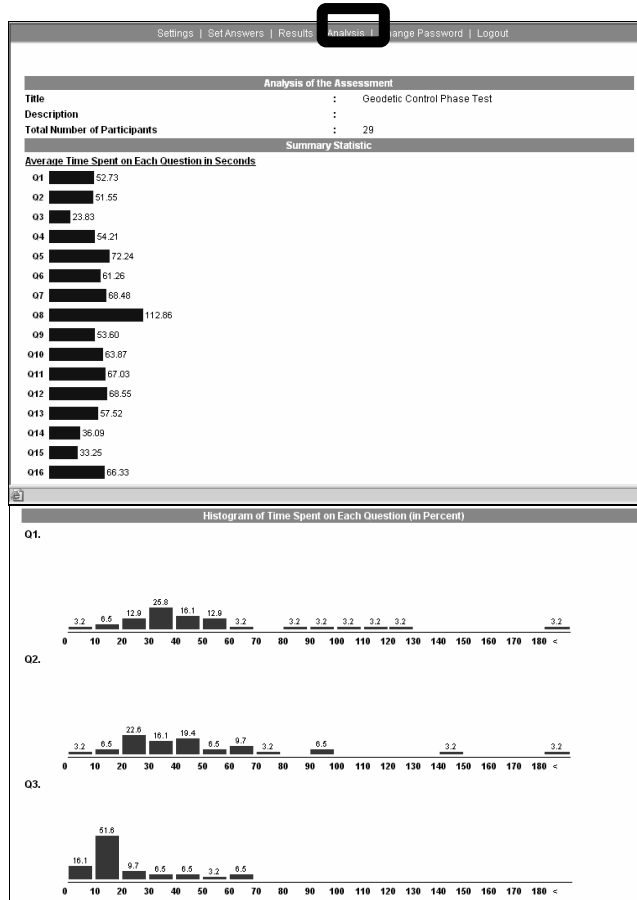


Fig. 6. Administrator module page for "Analysis"

*Student module*

The Student module is an application program used to allow students to enter the test. The first page requires students to enter their names, IDs, a password for checking test results later and access code to enter the test (Figure 7). Students are required to read the instructions and then enter the question and answer pages (Figure 8) once the information is submitted.

**Geodetic Control Phase Test**

Student ID :

Last Name :

First Name :

Password :

Access Code :

Fig. 7. Student module pages: Personal information page

Assessment Title: Geodetic Control Phase Test

**INSTRUCTIONS**

**LSGI3541 Geodetic Control Phase Test**

- There are 25 questions in the test. Mark the correct answer(s) for each question.
- There are 3 choices for each question.
- Please be reminded that there may be one, two or three correct answers, or none of them are correct for each question.
- The phase test will last for 30 minutes. You can leave the classroom in the first 20 minutes.
- GOOD LUCK ☺

---

Student ID : 04458437d      Student name : Fok, Hok Sum

**Question 1.**

The radius of curvature in the normal section azimuth is given by

A.  $N = \frac{a}{\sqrt{1 - f(2 - f)\sin^2 \phi}}$

B.  $M = \frac{a(1 - e^2)}{(1 - e^2 \sin^2 \phi)^{3/2}}$

C.  $R_n = \frac{MN}{N \cos^2 \alpha + M \sin^2 \alpha}$

A.     B.     C.

Mark this question to revise it later

Que. No.	Your Answer	Marked	Que. No.
1.			21.
2.			22.
3.			23.
4.			24.
5.			25.
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			

© 2005 Hong Kong Polytechnic University. All Rights Reserved.

Fig. 8. Student module page: Instruction, and question-and-answer

TAXONOMIC CLASSIFICATION OF MULTIPLE CHOICE QUESTIONS

It is reasonable to assume that students will spend more time on a multiple choice question for which they do not have a clear answer compared with the other questions that they studied well. This extra effort can take place by visiting a particular question more often and spending more time on the average. On the other hand, students' time utilization on MC questions is also compounded by the *difficulty level* of

questions. This relatively vague concept of *difficulty* has been examined by Bloom et al. [2] resulting in *rank ordering of cognitive educational objectives* known as *Bloom's Taxonomy* (see Simkin and Kuechler [18] for an updated discussion on Bloom's Taxonomy). In other words, Bloom's taxonomy is a classification of the level of difficulty of different types of questions to be tested on a particular student's understanding on the taught subject. This classification has influenced the development of educational curricula and evaluation measures since its realization. The taxonomy consists of six hierarchical levels (Figure 9); *knowledge, comprehension, application, analysis, synthesis and evaluation*. The levels are assumed to be interdependent generating prerequisite for each subsequent level.

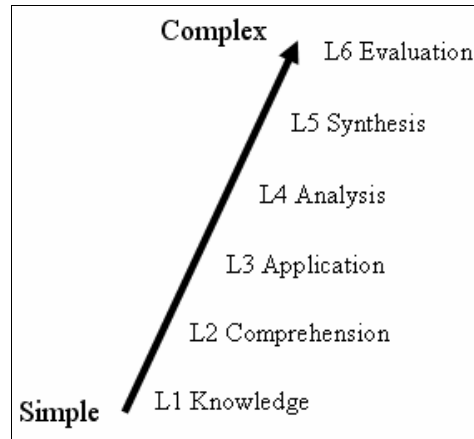


Fig. 9. Hierarchical levels of Bloom's Taxonomy

*Knowledge-level* objectives are the knowledge of specific facts, terminology, date, persons. For instance answering the following question requires very little reasoning, only recalling the simple knowledge of the definition of the deflection of the vertical.

- Q. Which of the following statement(s) about the deflection of the vertical is/are correct?
- It is the angle between the ellipsoid and geoid.
  - It is the angle between the ellipsoidal normal and the plumb line.
  - It is the angle between the ellipsoidal normal and the direction of gravity.

Although knowledge is the essential part of a cognitive process it is not sufficient. For instance, testing skills in translating mathematical and verbal material into symbolic statements or visa versa, or the ability to grasp the principals behind the work as a whole at any desired level of generality are also essential. The *comprehension level* provides a basis for measuring the aforementioned skills [2].

In the following example, *knowledge-level* background about solar radiation and atmospheric drag that causes perturbing acceleration on a satellite is required. But understanding that these effects vary with satellite orbit altitudes and shapes, and the formulation of the acceleration which leads to a small area to mass ratio for reducing the effect of perturbations is essential to find the correct choice.

- Q. Small area-to-mass ratio of an artificial satellite
- reduces the impact of the atmospheric perturbations on a low orbiting satellite
  - reduces the impact of the solar radiation pressure

- perturbations on a low orbiting satellite
- c) reduces the impact of solar radiation pressure or atmospheric drag on a satellite.

As an example for the *application-level*, consider the following question where identifying the correct answer requires a degree of knowledge and transfer from instruction to practice. Mastering the concept of leveling is needed, in particular about the “line of collimation” or the “rise and fall” method, and its application in an unknown situation.

- Q. If the height of a benchmark A is 128.480 m and the staff readings from a level located in-between benchmarks A and B are 3.222 m and 1.414 m respectively, then the height of the benchmark B is
- a) 126.672 m
  - b) 130.288 m
  - c) 129.894 m

*Analysis-level* questions require to distinguishing facts from hypotheses and recognizing unstated assumptions. The solution to the following example requires a number of steps. The most sophisticated approach to the problem requires fitting a straight line to the slope data using a least square solution followed by the calculation of standardized observation residuals to detect the observation which is subject to error. However a simpler approach is first to plot the elevation data against the distance and to examine the graph for data that deviates from a straight line more significantly than other data points keeping in mind that the random errors are proportional to the square root of the distances between benchmarks.

- Q. Suppose the following levelled points on a slope have measured elevations as follows. Which bench mark is more likely to have a significant measurement error?

From-To	Leveled point	Distance of the leveling run (km)	Elevation (m)
P-A	A	3	105.43
A-B	B	4	103.77
B-C	C	4	103.21
C-D	D	3	98.43
D-E	E	1	93.41

A. B. C. D. E

*Synthesis* involves putting together elements and parts in a meaningful way and requires skills for organizing ideas and statements. In the following question the physical process of precession necessitates the understanding of the role of each item and further the comprehension that *all* choices are part of the precession phenomenon.

- Q. Precession happens because?
- a) The is a mutual attraction among the Sun, Moon and the Earth
  - b) Earth’s rotation axis is tilted with respect to the ecliptic plane
  - c) Earth is not perfectly spherical (flattening)

*Evaluation-level* questions are to test the ability to indicate logical fallacies, evaluate the value of judgments and actions critically (Bloom, 1956). The following example carries all the signatures of the analysis, application, comprehension and knowledge levels. Despite its simple set up, its solution requires additional skills in recognizing the right formulation that leads to the calculation of the correlation between measurements which is rather difficult at elementary level. Moreover the calculated result must be examined and weighted critically to assess the significance of the correlations.

Q. Suppose three angles that close the horizon are measured at a station. Their values and standard deviations are:

Angle	$\sigma$
134 <sup>0</sup> 38' 56"	6.7"
83 <sup>0</sup> 17' 35"	9.9"
142 <sup>0</sup> 03' 14"	4.3"

- A. These measurements are correlated because -----
- B. These measurements are not correlated because -----

EXPERIMENTAL TESTING AND RESULTS

An online MC test that consists of 25 questions with three distracters for each is administered for a class of 85 Higher Diploma students. As mentioned before, if there can be more than one correct answer or no correct answers at all, three distracters would generate 8 possible combinations of answers so that students could not guess the right combination easily without prior knowledge. One could also set only one answer to a question with four or five choices, which is a normal practice for most instructors. Out of 25 questions 9 questions are classified L1, 6 of them L2, 4 question at L3, and 2 of them at L4 level. L5 and L6 level questions are lumped into one group (L5) since they were only few in numbers due to the level of the subject and are difficult to discriminate. The black circles in Figure 10 indicate the distribution of the questions in each level. During the 25 minute test, the students' time spent on each question is automatically monitored by the program. The average time students spent on each question are displayed against the taxonomy level in Figure 10.

We investigated the correlation between taxonomy levels and the average time spent on each question through a regression model in which the variability in the *dependent variable* "average time" is a linear function of the *independent variable* "Bloom's taxonomy level" in the following regression equation,

$$Average\ Time_i = a + b * Taxonomic\ Level_j + Random\ Variability$$

In the above *equation*  $i = 1 \dots n$ ,  $n$  is the total number of questions,  $a$ , and  $b$  are the intercept and slope of the regression line,  $j = 1 \dots 5$  are the indices of Bloom's taxonomic levels.

The least squares solution to this problem shows that Bloom's taxonomy level can explain over 50 percent of the variability,  $r^2$ , in average time spent on each question which is statistically significant at 95 percent confidence level (Table 1). The estimated regression line is shown in Figure 10. It turns out that one of the questions is an outlier (which is exactly what we will be looking for the next level of this analysis but a nuisance in this case) and unduly influences the results. If this question is excluded from the solution, the  $r^2$  value increases to 0.8 which corresponds to a 0.9 correlation coefficient ( $r$ ) between the average time students spent on each question and Bloom's taxonomy levels. This is a very significant indicator of a statistical relationship between these two variables; taxonomic levels and time (Table 1).

With this result we can now test the following null-hypothesis:

$H_0$ : The average time spend for a question within a Bloom's taxonomy level is significantly different than that for the others in the same level.

Table 1. Regression of average time spent on each question to Bloom's taxonomy level.

Solution	Solution statistics of the regression model
All Data	$r^2 = 0.5140$ ; $r = 0.72$ ; Time = $36.494 + 9.802 * \text{Level}$
Outlier Removed	$r^2 = 0.7915$ ; $r = 0.89$ ; Time = $33.413 + 10.231 * \text{Level}$

A two tailed null-hypothesis test can be carried out for each question in a given taxonomic level and the questions that are found significantly different from the mean of the level are either too easy (shorter than the mean) for the corresponding level or too difficult (longer than the mean). If a question is found to be statistically too difficult, then this question either belongs to a different level of complexity or students are having learning difficulties in this particular area.

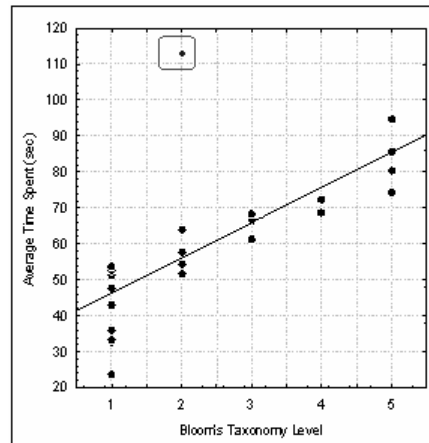


Fig. 10. Distribution of the average time students spent on each question monitored during the test as a function of Bloom's taxonomic hierarchy.

Operationally, there are various approaches to carry out the null-hypothesis testing. However having already established a strong statistically significant relationship between the average time spent and taxonomic complexity, the following simple testing procedure can be carried out not only for the above example but for all multiple choice tests.

1. Calculate the mean values of the average time spent within each taxonomic level.
2. Calculate the corresponding standard deviation of the means for each level.
3. Using the mean values and their standard deviations establish confidence intervals (at 1, 2, or 3 standard deviation levels) for each taxonomic level. If average time spent on an individual question within a particular level falls outside the confidence interval, then scrutinize the question as below.
4. If the average time spent is less than the average of time of the Bloom's taxonomic level then this is almost an easier question and may belong to a lower taxonomic level or it is a question with little value for assessment. If the average time spent is more than the level average, then this is either a question that belongs to an upper level or students are having learning difficulties in this particular area.

We did not find any problematic question in our MC test except the one question which is already obvious from the plot using this approach. It turns out that this question is indeed problematic, because the related material was not discussed in detail during the lectures. However our conclusion is not statistically sound in this test because there was insufficient number of questions for some of the taxonomic levels (degrees of freedom).

Alternatively, we can use the predicted average time spent using the regression model we established (Table 1) instead of the mean average time spent at a given Bloom's taxonomic level. This approach is more appropriate and sensitive because it accounts for the variability *among* taxonomic levels. Nevertheless it is more complicated because of the requirement of solving a regression model for each MC test. Using this approach one of the questions in Level 1 is identified as an easy question for the students (the one which took less than 25 seconds on the average for the students to answer) and therefore, its contribution to our MC test is questionable.

#### CONCLUSION

The average time students spend on each multiple choice question can be analyzed using the *on-line MC testing* software that is capable of monitoring the time spent on each question. We can then attribute teaching and learning difficulties to the questions on which students spend most of their time. The timing of student responses will also help to identify trivial questions; the ones that took significantly less time than the others.

On the other hand, the simple-to-complex hierarchy of Bloom's taxonomy (Figure 9) suggests that students may spend more time on a particular question simply because it is a complex one. We therefore need to separate the influence of the taxonomic level on the time spent from the learning difficulty students might have encountered for different multiple choice questions.

We expect students to spend less time on simple questions at the *knowledge-level* that require only a simple recall by the student resulting in shorter time utilization for L1 type questions. More complex questions such as L3 and L4, *Application and Analysis* levels, where the student will have to go through a series of additional time consuming steps such as calculations to arrive at the correct answer, will take more time utilization on the average. Although the statistical relationship between the time spent on each question and the taxonomic complexity is intuitively evident, it must be demonstrated and quantified for identifying students' learning difficulties. This was the motivation for the experiment in this study and we have demonstrated that average time students spent in answering multiple choice questions and Bloom's taxonomy are two important parameters which can be effectively used in identification of the students' learning difficulties.

#### ACKNOWLEDGEMENTS

The authors acknowledge the support received from various sources in carrying out this software development. They are: Online Program Development for the web-assisted teaching Analysis of Observations I, Online Program Development and LSGI Departmental Teaching and Learning Projects for web-assisted teaching of *Site Surveying* subject and *web-assisted teaching for Geomatics*. Mr. A. Tsang coded the software. We thank to the reviewers who carefully evaluated the manuscript.

## References

1. Becker, W.E. and Johnston, C., 1999. The relationship between multiple choice and essay response questions in assessing economics understanding. *Economic Record*, **75** (231): 348-357.
2. Bloom, B., Englehart, M., Furst, E., Hill, W. and Krathwohl, D., 1956. *A taxonomy of educational objectives. Handbook I: Cognitive domain*. New York, NY: David McKay Company.
3. Bradbard, D.A., Parker, D.F. and Stone, G.L., 2004. An alternative multiple-choice scoring procedure in a macroeconomics course. *Decision Sciences Journal of Innovative Education*, **2** (1): 11-25.
4. Carey, J., 1997. Everyone knows that  $E=MC^2$  now, who can explain it? *Business Week*, **3547**: 66-68.
5. Chappidi, S., Patel, K.C. and Lawhead, P.B., 2003. Online evaluation management system for e-learning. *Annual Conference on Web-based Teaching and Learning*. NAWeb 2003.
6. Delgado, A.R. and Prieto, G., 2003. The effect of item feedback on multiple-choice test responses. *British Journal of Psychology*, **94** (1): 73-85.
7. Dufresene, R.J., Leonard, W.J. and Gerace, W.J., 2002. Making sense of students' answers to multiple-choice questions. *The Physics Teacher*, **40**: 174-180.
8. El-Darahali, S., 2002. Instructors quiz preparation software: A customized evaluation software. *Annual Conference on Web-based Teaching and Learning*. NAWeb 2002.
9. Greer, L. and Heaney, P.J., 2004. Real-time analysis of student comprehension: An assessment of electronic student response technology in an introductory earth science course. *Journal of Geoscience Education*, **52** (4): 345-351.
10. Haladyna, T.M. and Downing, S.M., 1989. A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, **2** (1): 37-50.
11. Holder, W.W. and Mills, C.N., 2001. Pencils down, computer up: The new CPA exam. *Journal of Accountancy*, **191** (3): 57-60.
12. Hopkins K.D., 1998. *Educational and Psychological Measurement and Evaluation*. Allyn and Bacon Pub, 8<sup>th</sup> Edition. 486 pages
13. Kniveton, B.H., 1996. A correlational analysis of multiple-choice and essay assessment measures. *Research in Education*, **56**: 73-84.
14. Kreig, R.C. and Uyar, B., 2001. Student performance in business and economics statistics: Does exam structure matter? *Journal of Economics and Finance*, **25** (2): 229-241.
15. Miller, D.R., 2002. Moving from a paper-and-pencil to a computer-based examination: challenges and lessons. *Annual Conference on Web-based Teaching and Learning*. NAWeb 2002.
16. Shortis, M.R., Leahy, F.J., Ogleby, C.L. and Kealy, A., 2000. Self-paced learning of plane surveying concepts and field procedures. *International IT Conference on Geo-spatial Education*, The Hong Kong Polytechnic University, Hong Kong, pp 171-180.
17. Shortis, M.R., Leahy, F.J., Ogleby, C.L., Kealy, A. and Ellis, F.G., 2004. Web-based learning of spatial design and analysis concepts using simulations and visual feedback. *Survey Review*, **37** (291): 344-359.
18. Simkin, M.G. and Kuechler, W.L., 2005. Multiple-choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, **3** (1): 73-97.
19. Smith, M.J. and Roberts, G., 1996. Computer aided learning in surveying. *Survey Review*, **33** (262): 546-552.
20. Snow, R.E. (1993). Construct validity and constructed-response tests. In E. B. Randy & C.W. William (Eds.), *Construction versus choice in cognitive measurement*. Hillsdale, NJ: Lawrence Erlbaum, pp. 45-60.
21. Snyder, A., 2003. The new CPA exam: Meeting today's challenges. *Journal of Accountancy*, **196** (6): 11-120.
22. Tahir, A. and Tanalol, S.H., 2003. Web-based marking system: design and implementation analysis. *Annual Conference on Web-based Teaching and Learning*. NAWeb 2003.
23. Zeidner, M., 1987. Essay versus multiple-choice type classroom exams: The student's perspective. *Journal of Educational Research*, **80** (6): 352-358.