ISSN:2320-0790

An International Journal of Advanced Computer Technology

Computer Assessment for Secondary School Tests

Queen M. Sello[1], Dimitar M.Totev[2], Rose Kgosiemang[3], Yaqiang Liu[4]

[1]Lecturer Computer Science Dept., University of Botswana
[2]Lecturer Dept. Maths-Science Education, University of Botswana
[3]Senior-Librarian, University of Botswana
[4]Lecturer Dept. Maths-Science Education, University of Botswana

Abstract: Secondary school tests are very important component of the student's assessment process and their frequency is directly related to the success of year examinations. On the other hand, the number of students could restrict the frequency of tests that is not desirable in terms of quality educations. Obviously such a dilemma is a challenging methodological and technological problem. In such cases a Multiple Choice test paper could be the only feasible solution, taking into account staff and time constraints. There are a few ways of marking such papers, mainly: a fully manual procedure with preset answer sheets (punched templates); fully automated marking process, applying optical character recognition (OCR) and scanners; the use of special pens and answer sheets could reduce to a certain extent the human factor involvement. The first case is time consuming, error prone and the stress under which the staff involved is working, contributes additionally to the relatively low quality of marking. The other two techniques require additional investment and technological infrastructure, but the marking process is significantly improved (Mogey, N. Watt, H. 2009). Similar results could be obtained with the demonstrated software project, whose main features are as follows:

- No additional investment involved.
- Significantly improved accuracy of mark calculations – fully automated marking procedure.
- Reduced stress factor – the computer keyboard is used in the most convenient way.
- Better synchronization between Marking and Quality Assurance staff.
- Significantly improved accuracy of the moderated scripts.
- Improved record keeping of grades that is individualized and easy to track and manage.
- Immediate statistical analysis of results.
- Improved record keeping with respect to the school archives.
- Registered error tolerance less than 1.5%.
- Ability for marked and moderated work to be accessed based on users pre-determined rights.
- Multi-level hierarchical approach to data security and staff responsibility structure (www.tcexam.com, 2009).

Keywords: Computer based grading, Multiple-choice, questions, Student, assessment, Secondary schools, Manual Marking

1. Introduction

Testing of students in schools is a widely used method that is geared towards assessing some form of understanding of course material. Various institutions use various forms of assessment tests. In some states various tests are used. The Norm Referenced Tests (NRTs) are used to compare individual performance of a representative national sample where national averages are used as bases for comparison and as such they are designed to represent curricula nationwide rather than a single course of study (National Association of State of Education, p.1, 2001) Also, the NRT are predominantly multiple choice and are considered efficient, economical and require the least amount of subjective involvement by scorers or those setting performance standards Other states use Criterion Referenced Tests (CRTs) to compare student performance to clearly defined standards. Assessment results would be reported according to a level of performance (e.g.) "not proficient", "proficient", "exceeded standards") or a numerical score (National Association of State of Education, p.1, 2001).

a. A. Use of Multiple Choice Assessment Internationally

The CRT requires the development of meaningful learning objectives that are keyed to assessment items and they assess what students know and can do rather than how students compare with their peers (National Association of State of Education, p.1, 2001). It is possible for every student to meet a high standard on a CRT whereas the NRTs classify half the students as "below average" rather than giving all of them the chance to succeed (National Association of State of Education, p.1, 2001). It is observed that for teaching and testing to be effective, it is important for the teacher to acquire skills in developing criterion referenced tests that are valid and reliable (Botswana. Ministry of Education p.3, 1994).

The other types of tests that are used to assess performance are the Performance Assessments and the Multiple-choice assessment. Performance assessments require students to formulate an original response to a question and communicate that response through the performance of some act producing a written essay, a diagram or a persuasive speech, while multiple-choice on the other hand requires students to select their responses from among a set of specific choices. Like most of the tests used in schools, these forms of assessment also have their strengths and weaknesses.

b. Use of Multiple Choice Assessments in Botswana Secondary Schools

Although little research has been done on the use of the multiple–choice assessment in Botswana secondary schools, however, there is enough evidence that suggests a wide usage of multiple-choice assessment together with other forms of assessment is in practice. In a study investigating attributes of teacher-designed tests in Botswana Junior Secondary Schools, Ranku (2001) reveals that multiple-choice test is indeed widely used in conjunction with other tests. The study focuses on analyzing Form 2 topic test for year 2000 obtained from junior secondary schools using descriptive statistics. The study also revealed that a greater proportion of questions in most of the tests were short answer and multiple-choice questions in the knowledge, recall and comprehension categories of Bloom's Taxonomy. A summary of the proportion of question of question types given in all schools revealed that a greater proportion of marks in the tests by almost all schools are given to multiple-choice and short answer questions. Short answer questions carry the most marks followed by multiple-choice questions while a smaller proportion goes to the remaining types of questions. The findings of Ranku's study further revealed short answer questions and multiple-choice questions were preferred because of large class sizes that teachers handle which average to 40 students per class, teachers it cumbersome to assess in the higher order categories as this would require them to supervise and mark for these large classes. Due to limited resources, time and lack of quality control measures the teachers tend to do what is convenient to them most.

The three year junior secondary school syllabi show use of multiple-choice assessment on several subject areas such as Science, Social science, Moral education, Design and technology and Business studies. The structure of the examination and the objectives tested differ from subject to subject. For instance, the structure of Business Studies Exam consists of four (4) papers with 50 multiple-choice questions in paper 1 testing knowledge and understanding of basic business and office concepts, terminology, principles, procedures and computational skills; paper 2 tested application of principles, procedures and processes to business as well as analysis and evaluation of business issues; paper 3 consists of CA based on a group project while paper 4 assessed practical keyboarding skills (Botswana. Ministry of Education Department of Curriculum Development and Evaluation. Moral Education, p.111, 1998). In the Home Economics assessment procedure on the other hand, it is clearly stated that there are three papers and paper 1 consists of multiple-choice questions derived from all taught modules; paper 2 consists of short answer questions also derived from all taught modules where as is the case with most subjects Paper 3 consists of CA (Botswana Ministry of Education Department of Curriculum Development and Evaluation. Home Economics, p iii, 1996).

The use of multiple-choice assessment tool is not only limited to the three year junior secondary schools. It is also applied in the Botswana General Certificate of Secondary Education subjects. The Examination structure for Agriculture for example covered multiple choice, short answer questions, essay and project (Botswana Ministry of Education Department of Curriculum Development and Evaluation. Botswana general certificate of secondary education teaching syllabus, Agriculture, 2001). Multiple-choice assessment tool is currently used for some of the courses offered at the University of Botswana, e.g. Computing and Information Skills course.

2. Evaluation of Multiple - Choice Assessment Tools

In most of the countries where multiple-choice assessment tool is applied in schools there are other tests that are applied in same exam. According to Rosa, et. al … (2001) if the collection of items is sufficiently well represented by a unidimensional item responses theory (IRT) model, scale scores may be a viable plan for scoring such a test. This is the trend even in the Botswana secondary schools where multiple-choice is used with other forms of tests. Therefore, it is important to here what people say about multiple-choice when compared to other forms of assessment tools. It is said that for any assessment to be effective it must balance validity, reliability and efficiency**.** Haladyna (1994) is of the opinion that multiple-choice items are difficult to prepare than essay items. The wording of the stem, the identification of a single correct answer, and writing of several plausible choices is challenging). The same author is of the opinion that since the 50-60 multiple-choice items set cannot be remembered they can be reused and this is an advantage of multiple-choice format over the essay format. Regarding the administration of essay test format it is felt that essay require more time because people tested have to write the response which might also take a much longer time, that way students tend to prefer multiple-choice format which is far less demanding.

With regards to scoring, Chase (1986) describes essay tests as judgmentally scored with a number of biases existing in scoring the essay as compare to multiple-choice which are said to be objectively scored. According to him there are several studies showing the existence of racial and gender biases in scoring which might pose very serious threats to the validity and interpretations and uses of essay test scores. With multiple-choice on the other hand, one can use a key, a scoring template which identifies the right answer and a multiple-choice answer sheet, or an optical scanning machine, which provides a total score for each test taken with a higher degree of accuracy. The scanner also provides an electronic file that can be used to analyse characteristics of the total test scores and the items.

Regarding analysis and evaluation of test items, Chase observed that essay items are not easily analysed and devaluated. Not only is an ambiguous essay test item is difficult to detect. It is observed that what constitutes effective and ineffective essay is not clearly discernible. However, with multiple-choice items there are many standard computerized item analysis programs that provide complete summaries on item and test characteristics. When comparing essay test to multiple-choice for reliability, the essay test is found to yield lower reliability than multiple-choice version.

3. Manual Marking of Multiple - Choice Assessment Papers

The manual marking of multiple choice exam papers could be mainly done in two different ways:

- Comparing every student’s answer with the provided marking scheme.
- Using a model-template of the correct answers.

The first mode requires a lot of concentration because the marker’s attention should be split between the exam paper and the marking scheme. It is very tiresome and for large classes is practically inapplicable.

The second technique needs a model-template of the correct answers to be prepared in advance. The template is superimposed on an exam paper and the marker counts the matching entries (Figure 1). The accuracy of the template is crucial and this is why handmade temples allow counting errors to be made when the actual marking starts. Another serious setback of this method is the cases when a student answers a question with more than one option, which invalidates the answer of the question. Unfortunately, the marker cannot identify such a situation because the original is underneath the marking template and only the correct answer, if indicated, is visible. In general, the method is error prone and requires a lot of manpower when it comes to large classes.

4. Automated Marking Techniques

General purpose scanners could be used to automate the marking process of multiple choice or short answer tests and exams given to large classes. Scanners do not require any specialised skills of the staff involved in the marking exercise and the technological procedure is simple, but reliable. Such favourable features are only related to the hardware component of the scanning system. The software component that controls the output of the system plays a crucial role and is known as Optical Character Recognition (OCR). The development of OCR is a very sophisticated piece of work, because it requires some elements of Artificial Intelligence (AI) to be incorporated as well in order human intervention to be eliminated during the processing of printed materials (e.g. multiple choice papers). The development of proper AI procedure is the most difficult part of the design that determines the accuracy of the whole system. The following example illustrates the above point. A student ticked an answer as correct one, but later on another option was selected to replace the first choice that was cancelled. In general, OCR cannot accurately identify which answer is the valid one – the initial mark/tick or its cancellation.

Such situations could be eliminated to a very high extent if special stationery (paper, pens, etc.) and certain simple filling-in rules are used by students when answering multiple choice papers. Of course, the above mentioned improvement comes at a cost that could be of significant value when it comes to large classes or frequent assessments. A typical example of such an approach is a testing, assessment and reporting system of Scantron Corporation. This system provides forms for a variety of test formats (883-E, 888-E, etc.). Using Scantron’s test scoring machines (TSM), the forms reduce the necessary time teachers have to spend on marking and validating assessment papers. More, Item Analysis forms could be run through the scanner to “reveal how many students missed each question” (http://en.wikipedia.org/wiki/Scantron#column-one 2009).

Fortunately, those convenient features could be easily emulated by a computer with very minimal human intervention and without all costly stationery, maintenance, logistic restrictions of such systems, apart from the initial investment cost. The essence of the proposed emulation is the customisation of a standard computer keyboard that allows student’s answers to be entered into a computer in the most convenient way; practically without errors (demo software is available). The developed software could be run on as many computer as it is necessary and profitable, a feature that makes such emulations highly time competitive with fully automated scoring systems. It should be emphasised that each particular case, whether fully automated systems or emulators should be applied, is a subject to detailed budgeting analysis, taking into account various restrictions (manpower, funds, deadlines, etc.).

5. Analysis

Assessment plays a very crucial role in the decision making of grading learners and in many occasions the meaning of scores given to these learners are either not well understood or interpreted. It is therefore essential to use assessment tools that not only allow the examiner to mark/grade the assessment but also allow him/her to make an analysis of the grades so as to derive meaning and understanding of examinees performance, understanding of context examined and also verify the validity, reliability and usability for the present and future uses.

In the modern today Computer Adaptive Testing [CAT] measures are put in use for testing of various things especially where large numbers are involved for example the electronic ELS or TOFEL language tests that are often required of learners wishes to study at institutions where the medium of instruction language is their second language[L2]. Not only do the CAT make quick delivery of results to the examinee but it also saves time and to a large extend eliminate the human errors that would otherwise occurred if the test where graded manually. However this type of assessment measure is only affordable where there are no financial constrains.

Below are snap shot diagrams of an assessment tool/software the Multiple Choice Calculator[MCC] developed by University of Botswana[UB] lecturers teaching a General Education Course[Computing and Information Skills- CIS] and their observations over the one and half academic years they have used it to grade the multiple choice exams for CIS. CIS is a cross curricular course offered to all year 1 students. From the use of MCC the lecturers observed that though the MCC unlike CAT is not fully automated it is however, cost-effective when compared to marking the papers manually and it also helped to address the problem of manpower constrains already stated. It is also faster allowing those lectures with large numbers of papers to grade to meet the set target dates for mark submissions. The MCC is user friendly for example it allows the user to customise the keyboard for the own comfort and is keeps record of persons marking and the point to which the marking has been done hence unlike in the manual grading human errors of grading wrong student responses or missing out on responses is not possible. For example there are numerous situations where students have responded giving two or more options for the same question thus increasing their luck of getting the correct answer when the script is marked manually.

Figure 2 below shows the marking temple where the student answers are entered. The examiner selects the script group that is to be marked, enters student ID then commences the marking. At the end of marking the script the student exam score auto-calculated and stored on the database as shown in Figure 4. When compared to manual marking where paper marking templates are used that require a lot of manpower that are not durable and also leads to a lot of human errors in calculations, using MCC for grading reduces all this. Manual marking often give rise to errors on the marking template e.g. cutting out the wrong answers or using the wrong group script template for marking it also reduces the fatigue and stress caused by marking these large groups. Below is observation table from the faculties of Engineering Technology & Humanities for the academic year 2004/05 final exam grading. Where in the before when the scripts were graded manually the manpower deployed to do the marking was enormous and this did not only add to cost but also increased the human errors. Therefore making the interpretation and understanding of the grades very difficulty or very unreliable to assess student performance. Not only that but also to verify the validity and usability of the exams. Figure 3 shows the statistics analysis feature of MCC that can be utilised syllabi and assessment paper improvements. From the statistic the lecturers observed that it was possible to infer the reliability, usability, validity and accountability and maximise security and privacy of record keeping and management. Table 1and Table 2 make a comparison of using MCC for grading compared to manual marking in the two semesters of 2005/06 for the two CIS courses [GEC121 &122]. The statistics button can be utilised to make statistic analyses of student responses per question so as to infer their performance. It reflects the Mean Standard Deviation and correct responses made on the question against the total scripts marked. Hence by so doing it assists the examiner to response to answering question on constructing validity of the exam.

NB: the ratio 1:3 (1:2) is quality assurance agreement for the course moderation. It means that if there is counting error or incorrect input errors of more than 3 the whole paper should be remarked.

Table 1: Comparison of using MCC and manual marking -2005

Faculty	Number of scripts	Time taken/per script	Manpower & utilities	Moderation report	Human Error Ratio
Engineering & Technology. Semester 1	250 150 (MCC grading)	5 minutes manual and 3 minutes MCC	3	5 scripts with errors	1:2
Humanities GEC121 semester 1	899 269[MCC grading]	2-5minutes but speed changed as users got accustomed to keyboard. Total Time 3days time 2wks	3person & 2 computers.	Generally counting errors over 1:3 including incorrect grading	Over1:3
	6 30[manual grading]		5persons		

Table 2: Comparison of using MCC and manual marking-2006

Faculty	Number of scripts	Time taken per script	Manpower & utilities	Human error
Engineering Technology	253 (100% electronic grading)	3 minutes	1 person for 172scripts & 2 for 81 scripts	2 scripts with errors. Detected human error in Moderation
Humanities	779 (100% electronic grading)	30 seconds	2 persons and 1computer	1 script with input error.

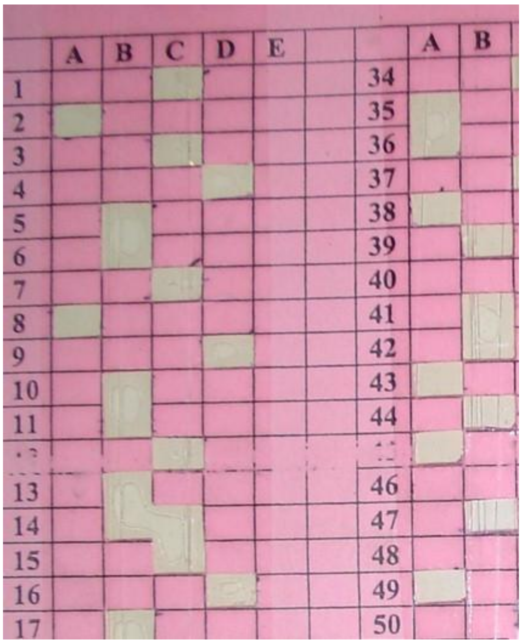
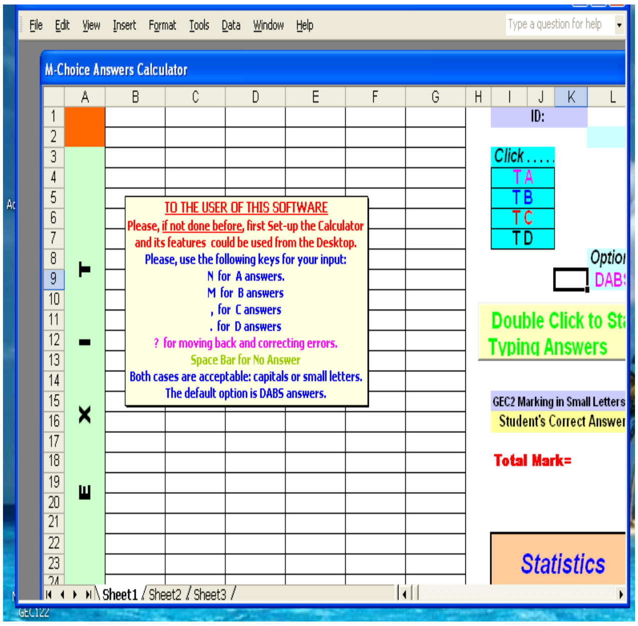


Figure: 1 Multiple choice Manual grading template



TO THE USER OF THIS SOFTWARE

Please, if not done before, first Set-up the Calculator and its features could be used from the Desktop.

Please, use the following keys for your input:

N for A answers.

M for B answers

, for C answers

. for D answers

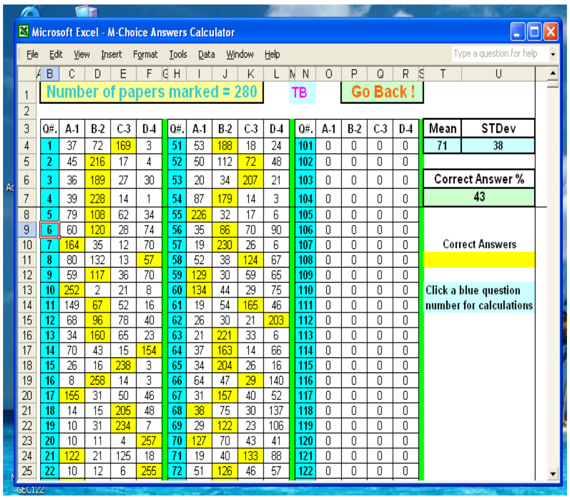
? for moving back and correcting errors.

Space Bar for No Answer

Both cases are acceptable: capitals or small letters.

The default option is DABS answers.

Figure 2: MCC Marking Template



Number of papers marked = 280

Q#.	A-1	B-2	C-3	D-4	Q#.	A-1	B-2	C-3	D-4	Q#.	A-1	B-2	C-3	D-4
1	37	72	169	3	51	53	188	18	24	101	0	0	0	0
2	45	216	17	4	52	50	112	72	48	102	0	0	0	0
3	36	189	27	30	53	20	34	207	21	103	0	0	0	0
4	39	228	14	1	54	87	179	14	3	104	0	0	0	0
5	79	108	62	34	55	226	32	17	6	105	0	0	0	0
6	60	120	28	74	56	35	86	70	90	106	0	0	0	0
7	164	35	12	70	57	19	230	26	6	107	0	0	0	0
8	80	132	13	57	58	52	38	124	67	108	0	0	0	0
9	59	117	36	70	59	129	30	59	65	109	0	0	0	0
10	252	2	21	8	60	134	44	29	75	110	0	0	0	0
11	149	67	52	16	61	19	54	165	46	111	0	0	0	0
12	68	96	78	40	62	26	30	21	203	112	0	0	0	0
13	34	160	65	23	63	21	221	33	6	113	0	0	0	0
14	70	43	15	154	64	37	163	14	66	114	0	0	0	0
15	26	16	238	3	65	34	204	26	16	115	0	0	0	0
16	8	258	14	3	66	64	47	29	140	116	0	0	0	0
17	155	31	50	46	67	31	157	40	52	117	0	0	0	0
18	14	15	205	48	68	38	75	30	137	118	0	0	0	0
19	10	31	234	7	69	29	122	23	106	119	0	0	0	0
20	10	11	4	257	70	127	70	43	41	120	0	0	0	0
21	122	21	125	18	71	19	40	133	88	121	0	0	0	0
22	10	12	6	255	72	51	126	46	57	122	0	0	0	0

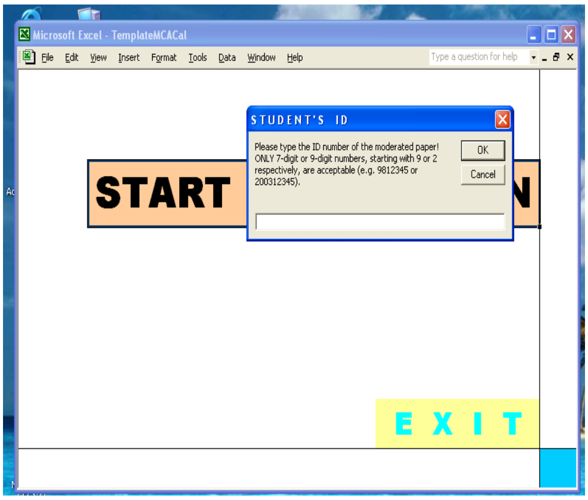
Mean	STDev
71	38

Correct Answer %: 43

Correct Answers

Click a blue question number for calculations

Figure3: MCC statistical analysis



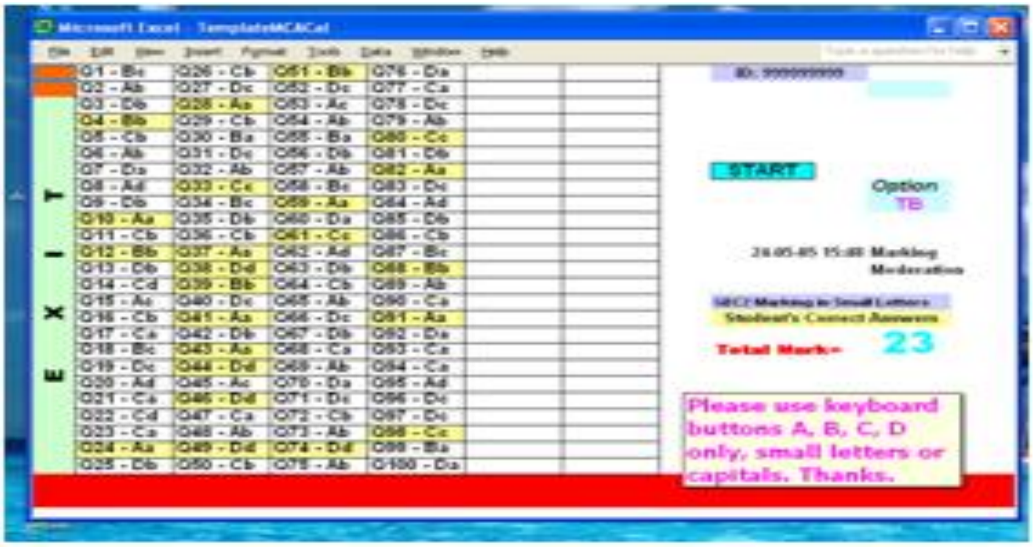
STUDENT'S ID

Please type the ID number of the moderated paper! ONLY 7-digit or 9-digit numbers, starting with 9 or 2 respectively, are acceptable (e.g. 9812345 or 200312345).

OK

Cancel

Figure 4: MCC Moderation Opening Screen


Microsoft Excel - TemplateMCACal
File Edit View Insert Format Tools Data Window Help
Type a question for help
E X I T
START

Q1 - Bc	Q26 - Cb	Q51 - Bb	Q76 - Da
Q2 - Ab	Q27 - Dc	Q52 - Dc	Q77 - Ca
Q3 - Db	Q28 - Aa	Q53 - Ac	Q78 - Dc
Q4 - Bb	Q29 - Cb	Q54 - Ab	Q79 - Ab
Q5 - Cb	Q30 - Ba	Q55 - Ba	Q80 - Cc
Q6 - Ab	Q31 - Dc	Q56 - Db	Q81 - Db
Q7 - Da	Q32 - Ab	Q57 - Ab	Q82 - Aa
Q8 - Ad	Q33 - Cc	Q58 - Bc	Q83 - Dc
Q9 - Db	Q34 - Bc	Q59 - Aa	Q84 - Ad
Q10 - Aa	Q35 - Db	Q60 - Da	Q85 - Db
Q11 - Cb	Q36 - Cb	Q61 - Cc	Q86 - Cb
Q12 - Bb	Q37 - Aa	Q62 - Ad	Q87 - Bc
Q13 - Db	Q38 - Dd	Q63 - Db	Q88 - Bb
Q14 - Cd	Q39 - Bb	Q64 - Cb	Q89 - Ab
Q15 - Ac	Q40 - Dc	Q65 - Ab	Q90 - Ca
Q16 - Cb	Q41 - Aa	Q66 - Dc	Q91 - Aa
Q17 - Ca	Q42 - Db	Q67 - Db	Q92 - Da
Q18 - Bc	Q43 - Aa	Q68 - Ca	Q93 - Ca
Q19 - Dc	Q44 - Dd	Q69 - Ab	Q94 - Ca
Q20 - Ad	Q45 - Ac	Q70 - Da	Q95 - Ad
Q21 - Ca	Q46 - Dd	Q71 - Dc	Q96 - Dc
Q22 - Cd	Q47 - Ca	Q72 - Cb	Q97 - Dc
Q23 - Ca	Q48 - Ab	Q73 - Ab	Q98 - Cc
Q24 - Aa	Q49 - Dd	Q74 - Dd	Q99 - Ba
Q25 - Db	Q50 - Cb	Q75 - Ab	Q100 - Da

ID: 999999999

Option
TB

28-05-05 15:48 Marking
Moderation

MCC Marking in Small Letters
Student's Correct Answers
Total Mark= 23

Please use keyboard buttons A, B, C, D only, small letters or capitals. Thanks.

Figure 5 MCC Moderation templates with grade scores

6. Conclusion

The multiple-choice assessment is used internationally. In most of the institutions where it is applied, it is used in conjunction with other test formats. It has been observed that in most places where this assessment tool has been applied the tested objectives included knowledge, recall and comprehension categories. Therefore, from the statistics analyses available on the MCC assessment tool, it is possible to see if the test objectives were meet thus giving room for developing in both teaching and learning of the course assessed.

Assessment methods based on ICT are quite popular today because of their data processing advantages, if silently assumed the required infrastructure is in place– a workstation for every student/pupil and reliable educational networks at various national levels. Lee and Weerakoon (2001) suggest that for purpose of student ranking, computer based assessment could be used with confidence even though there is need to take care when grading with it. Hence for the time being if the above requirements are not met then reason a compromising solution is offered with the use of MCC.

7. References

[1]. Botswana. Ministry of Education Department of Curriculum Development and Evaluation. (2001). *Botswana general certificate of secondary education teaching syllabus Agriculture*. Gaborone

[2]. Evaluation. (1996). Botswana. *Home Economics: three year junior certificate programme*. Gaborone.

[3]. Botswana. Ministry of Education Department of Curriculum Development and Evaluation. (1998). *Moral Education: three year junior certificate programme*. Gaborone.

[4]. Botswana. Ministry of Education. (1994). *Teachers handbook on criterion-referenced testing and continuous assessment*. Gaborone.

[5]. Chase, C.I. (1986). *Essay test scoring: interaction of relevant variables.* Journal of Educational Measurement

[6]. www.scantronform.com, (2009), *Home Page*, Scantron Forms, Optical Mark Scanner.

[7]. e-Government Research Group, (2004), University of Botswana (2004), *Design and Development of e-Education Materials*, 4th International Conference on On-line Learning - ICOOL 2004, p. 23, 25.

[8]. Gronlund, N. (1968). *Constructing achievement tests.* Englewood cliffs, N.J.Prentice-Hall.

[9]. Haladyna, T. M. (1994). *Developing and validating multiple-choice test items***.** Hillsdale, N.J: Lawrence Erlbaum Associates

[10]. Lee, G. And Weerakoon, P. (2001). *The role of computer aided assessment in health professional education: a comparison of student performances in computer-based paper and pen multiple-choice test.* Retrieved September 26, 2009, from Medical teacher 2, 152-157 Vol. 23,p v2.

[11]. Lukhele, R. Thissen D. and Wainer, H. (1992). *On the relative value of multiple-choice, constructed-response, and examinee-selected items on two achievement tests.* Paper presented at the Educational Research Association, San Francisco. p.27

[12]. Mogey, N. Watt, H. (2009) *the use of computers in the assessment of student learning, Learning Technology Dissemination Initiative.* Retrieved September 25, 2009, HTML by Phil Baker, p1-11.

[13]. National Association of State Boards of Education, (2001*). A primer on state accountability and large-scale assessments.*

[14]. Ranku, G. (2001**).** *An exploratory survey of teacher-designed tests used in Junior Secondary Schools in Gaborone***.** Unpublished theses. P. 24

[15]. Rosa, K. (2001). *Item response theory applied to combinations of multiple-choice and constructed-response items – scale scores for patterns of summed scores***.** In: Thessen, David, Wainer, Howard (2001). Test scoring. Mahwah, N.J.Lawrence Erlbaum Associates.p.253.

[16]. http://www.nasbe.org, (2009), National Association of State Boards of Education, *Home Page*, Educationa_Issues/Reports/Assessment.pdf 2009

[17]. www.tcexam.com, (2009), *Home Page*, CBT - Computer–Based Testing, CBA – Computer-Based Assessment.