Web based Computer- Adaptive Multiple Choice Assessment (CAAS) Using the New NRET Scoring Method

Sie Hoe, Lau[1]; Kian Sam, Hong[2];Ngee Kiong, Lau[3]; Hasbee, Usop[4]

lausiehoe@sarawak.uitm.edu.my: hksam@fcs.unimas.my; plnk@sarawak.uitm.edu.my:
hasbee@fcs.unimas.my

[1]Universiti Teknologi MARA Sarawak
[2]Universiti Malaysia Sarawak
[3]Universiti Teknologi MARA Sarawak
[4]Universiti Malaysia Sarawak

Abstract: *Multiple-Choice (MC) item is an important component of testing. MC items have been heavily used in computer-based assessment due to the ability of computer to be programmed to enable automatic and easy retrieval of test scores and feedback. Past studies have shown that feedback on degree of realism essential to support metacognition, which empowers students to regulate and control their own learning. Unfortunately this feedback is rarely given especially to young children. Furthermore, whether the students value this feedback remains an open question. Effort to provide this feedback through computer-based assessment by incorporating Probability Measurement (PM), Confidence Weighting (CW), the Conventional Number Right (NR), and Elimination Testing (ET) seem not possible especially for young schooling children. The main weaknesses of CW, PM and ET are their complex or confusing test instructions. With the advent of the internet, thus it is proposed in this study to develop and evaluate a web-based Computer-Adaptive Multiple Choice Assessment (CAAS) that has the ability to provide feedback on the student's knowledge state as an alternative to feedback on degree of realism. NRET scoring method which is a hybrid of NR and ET scoring method is used in CAAS to provide feedback to students. This is an on-going project and this paper presents some initial findings of CAAS. The initial results indicated that CAAS has the potential to assist students in learning but students may require more training to get used to the NRET scoring procedure.*

1. Introduction

Over the past decade, the internet has become an increasingly important environment to promote learning in the schools. Computer-based assessment becomes increasingly attractive as education faced pressure such as increasing number of students and workload, and the need to reduce cost but yet improve the quality and accountability of the assessment process [33]. Among the various format of test item, Multiple-Choice (MC) items have been heavily used in computer-based assessment due to the ability of computer to be programmed to enable automatic and easy retrieval of test scores and feedback [6,23,33].

One of the most promising aspects of computer-based assessment was instant feedback [29]. Feedback on degree of realism is useful in diagnosing content misconceptions, reasoning errors, and learning difficulties [28]. Degree of realism is a type of outcome feedback, which indicates how aware a student is of his or her own knowledge states. For example, in a ten item MC test, Student who answered confidently all the items, only to discover that all his answers are wrong show that he overvalued his knowledge and was having serious misconceptions. On the other hand, a student who was not sure (less confident) of his answers, yet getting all the answers correct is undervaluing his knowledge. Both students indicated lack of realism but on two extreme ends. Accurate self-assessment is also essential to support metacognition, which empowers students to regulate and control their own learning and for which laboratory studies show positive correlations with studying and achievement [1,27]. From educational point of view, accurate self assessment of students' own knowledge will enable them to plan study activities and deciding what topics to pay attention to

while studying. However, this feedback is rarely given especially to young children. Furthermore, whether the students value this feedback remains an open question.

2. Statement of the Problem

In an attempt to provide feedback on degree of realism, Dirkwager [10,11] develop a TestBet, computer-based testing with automatic scoring based upon Probabilities Measurement (PM). Paul [28] designs an interactive response system called Computer-Based Alternative Assessment (CBAA) Confidence Weighting (CW). Under probability measurement, students are instructed to indicate the probability of correct for each option. Confidence weighting on the other hand requires students to choose an answer and indicate how confident (or the probability) the chosen answer is correct. In both cases, students need to translate their degree of correctness of the chosen option into a fairly correct numerical scale. This involves the understanding of the concept of probability. If the students do not have a well-developed concept of probability, it is just like asking them to complete an impossible task [22]. Unfortunately it is found that even older students (junior high to college students) have difficulty employing probabilistic reasoning [15,30,14]. In addition to this, the calculation of degree of realism involves complex computation of student's assigned probabilities to the correct answer and a fairly large number of items are needed to obtain accurate measures of degree of realism. Thus, TestBet and CBAA may not be practical for young schooling children.

Number Right (NR) scoring method is a conventional scoring method, where students need to pick one option as the correct answer. Unfortunately, there is no information on how the students think with respect to each alternative: complete certainty, vacillation between several alternatives, blind guess, or even more or less serious fallacy [16]. Thus, feedback on how realistic or how aware the students of their knowledge state could not be provided.

Coombs [7] proposed Elimination Testing (ET), a procedure where students must cross out all alternatives that they consider incorrect. Students' knowledge for a given item can be categorized as: full knowledge (identify all the incorrect options), partial knowledge (identify some of the incorrect options), partial misinformation (identify the correct answer and some incorrect options), full misconception (identify only the correct answer), absence of knowledge (either the item is omitted or all options are identified).

Although the students' response pattern under ET can provide a rough indication of students' knowledge state, its test instructions are confusing despite prior practice [19]. It is conflicting where students are taught to solve for the correct answer but being assessed on their ability to identify the wrong answer. Dirkwager [11] say that based on students' partial knowledge, they can identify that some option which is more likely to be the answer than others. However, under ET, students are denied the opportunity to indicate their preferred option as the possible answer. Thus this portion of partial knowledge is not captured.

At this moment, it seem likely that, feedback on degree of realism for young schooling children which are often tested using MC items are not possible in normal educational setting such as in classroom. Thus, any new scoring method that is simple and familiar to young children yet able to give a rough indication of students' knowledge state will be useful. Therefore, a web-based Computer-Adaptive Multiple Choice Assessment (CAAS) using the new Number Right Elimination Testing (NRET) scoring method was proposed.

3. Purpose of the Study

The purpose of the study is to develop and evaluate an assessment tool that is able to give feedback on student's knowledge states an alternative to feedback on degree of realism. CAAS is a

web-based assessment tool for learning mathematics using the new NRET scoring method. This study addresses the following research questions.

1. Do the students accept CAAS as a learning tool?

2. What are the students' perceptions on the NRET test instructions?

3. What are the students' perceptions on the feedback on their knowledge state?

4. CAAS using the New NRET Scoring Method

CAAS is a web-based computer-adaptive assessment using the new Number Right Elimination Testing (NRET) scoring method. It is currently available at http://caa.bestservices.com.my. NRET is a hybrid of two existing scoring method, Number Right (NR) and Elimination Testing (ET). Under NRET, students need to eliminate alternative(s), which is/are incorrect and based on the remaining alternatives, choose one as the answer.

An item is presented on the screen, for each response option in a MC item, students should indicate on each option how they think about that response; "Correct", "Wrong" or "Not Sure". Students must choose one option as "correct". However, they have the flexibility in choosing none, one, two or three as "Wrong" or "Not Sure". Figure 4.1 shows the tools used by students.
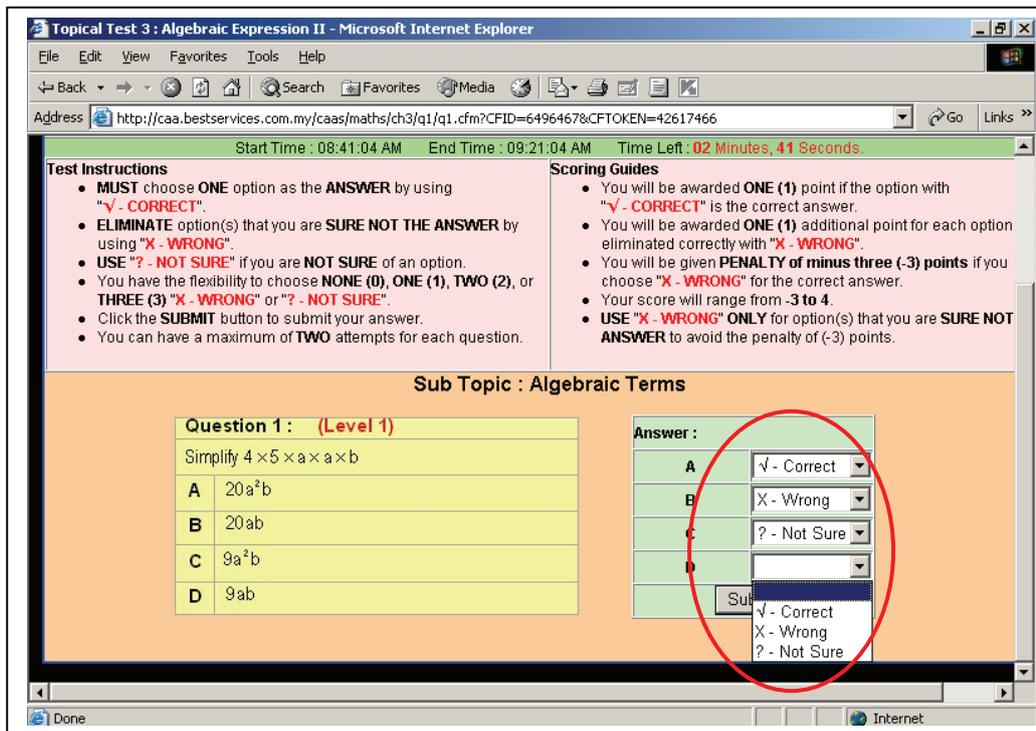
Based on the students' response pattern under NRET, the students' knowledge state can be identified. Table 1 below shows the classification of students' knowledge state for all possible response patterns under the NRET scoring method for a four alternative MC item.

Table 4.1: Classification of students' knowledge level based on the NRET response pattern.

Knowledge state	Response pattern
Full knowledge	Answer correct
	Identify 3 options as definitely incorrect
Partial knowledge	Answer correct
	Identify 2 options as definitely incorrect
	Answer correct
	Identify 1options as definitely incorrect
	Answer wrong
	Identify 2options as definitely incorrect
	Answer wrong
	Identify 1 options as definitely incorrect
	Answer correct
	Identify no option as definitely incorrect
Absence of knowledge	Answer wrong
	Identify no option as definitely incorrect
Partial misconception	Identify answer as definitely incorrect
	Identify 2 options as definitely incorrect
	Identify answer as definitely incorrect
	Identify 1 option as definitely incorrect
Full misconception	Identify only the answer as definitely incorrect

CAAS is adaptive in nature where the selection of the successive item will depend on the student's performance of the item. Thus, students do not waste their time attempting items that are too hard or trivially easy. In addition to this, hints are given and students are also given a second chance to correct their thinking if their first response is incorrect. Full solution will be given as reinforcement for student having partial knowledge and misconception. At the end of the exercise, student will not only obtain their scores, but also feedback on their knowledge level for each item. For items with misconception and absence of knowledge, students can review to check where they are wrong. On top of that, students can view their ranking against all participating students in their school as well as against all participating students. Example of feedback output on students' knowledge level is as shown in Figure 4.2.

Figure 4.2: Feedback on students' knowledge state

5. The Sample

CAAS is currently implemented in 37 secondary schools throughout Sarawak, Malaysia. At the moment, 630 students and 59 teachers had registered and used this assessment tool. It is an ongoing project and the data collection is expected to be completed in early November 2008. However, only 73 students from three participating schools are used as the sample for the preliminary finding. They were requested to complete a questionnaire upon completing five MC exercises. The questionnaire consisted of general background questions regarding the students' gender and the computer experience. The questionnaire contained a further ten questions relating directly to the three research questions in the study.

6. Preliminary Results and Discussions

Profile of the sample. The samples for the preliminary findings consisted of 73 students from three secondary schools in Kuching and Samarahan divisions, Sarawak, Malaysia. There were 34 male and 39 female students. A total of 35 students or 47.9% had computers at home. Out of 35 students having computers at home, 31 students spent at least one hour on learning using computer in a week. All the three schools had computers and 51 students out of 73 students spent at least one hour learning using computer at their schools per week. Thirty one out of 73 students surveyed stated that they had been exposed to other learning software besides CAAS.

Research questions 1: Do the students accept CAAS as a learning tool?

The findings showed that 60.3% or 55 out of 73 students surveyed agreed that CAAS helped them in learning mathematics. Only one student disagreed while another 17 students were undecided. Sixty four point four percent or 57 out of 73 students stated they would always use the CAAS if it was available on the web as compared to only two students who say otherwise. The rest of the students were undecided. A further 61.6% suggested that CAAS with NRET scoring method be extended to other subjects such as science. Thus, the majority of the students accepted CAAS as a learning tool and was willing to use it if it was available on the web.

These findings were generally consistent with past studies which reported the benefits of using computer-based assessment. According to Fleischman [13] and Lockwood [21], web based assessment where students have more opportunities for practice, self-testing, self-evaluation and self motivation will promote learning. Bennett [2] and Bransford, Brown and Cocking [4] says that instant feedback enable students learn about their mistake or misconceptions thus enabling them chance to readjust their thinking instantly. Laborde [20] stresses the need of student being given another chance to answer the same item to confirm their adjustment. Paul [28] says a good assessment tool for learning should include: establishing and revealing status ("knowing what you know"), diagnosis of weaknesses ("Knowing what you don't know") and comparative assessment with larger populations (answering "where do you stand?"). The design of CAAS takes all the above factors into consideration. Thus, as shown by the findings, CAAS can be regarded as an acceptable learning tool.

Research question 2: What are the students' perceptions on the NRET test instructions?

Addressing the issue on providing feedback on students' knowledge state requires going beyond conventional NR multiple-choice scoring techniques. This means introducing new test instructions. To ensure practically, the new tests instruction must be simple and familiar to the students. Review of literature has shown that the new NRET test instructions where students need to eliminate alternative(s), which is/are incorrect and based on the remaining alternatives, choose one as the answer is actually a test taking strategy commonly used by students while answering MC items. For example, in a survey on test taking strategies employed by students, Oh [25] reported that out of 264,900 students who took the Fall 2001 administration of the SAT, 60% or roughly 159,940 students first eliminate incorrect alternative before selecting one alternative as the answer. Dirkwager [11] also discover that most students used this strategy.

Consistent with past finding, only 16.4% or 12 out of 73 students stated that they did not use the strategy of eliminating the wrong options before choosing the answer. Although majority of the students used strategy similar to the NRET test instructions, it was surprising that 35.6% or 26 out of 73 students still regarded NRET test instructions as confusing despite prior practice. One possible explanation to this was that under the NR scoring, students have the flexibility of using more than one strategy in answering MC item. Orth [26] listed 11 most popular test taking strategies. However, once the strategy of eliminating the wrong option before choosing the answer strategy is formalized as the scoring method, student have to consistently use this strategy through out the whole test. Perhaps, some students might have difficulty in consistently using this strategy thus finding it confusing. In addition, this was also the first time they were exposed to the new test instructions beside the conventional NR. Nonetheless, according to Shepard [31,32] and Nitko [24] test taking strategies such as the NRET test instructions were skills that can be learned and there have been numerous studies that indicate that both children and adults can be taught to use the test taking strategy consistently. Thus, five sessions of training in this case could be insufficient. It was inappropriate to assume that students would be able to follow the new test instructions consistently though it seemed familiar to the students. It was because they had been so conditioned to the conventional NR scoring where they only needed to pick one option as the answer.

Research question 3: What are the students' perceptions on the feedback on their level of knowledge?

The finding showed that 42.5% perceived feedback on the knowledge level as useful as compared to only 13.7% who said otherwise. The remaining 43.8% were undecided. Fifty seven point 5 percent the students agreed that feedback on knowledge level enabled them to identify the area they

were weak in as compared to only 1.4% who disagreed. Forty seven point nine percent reported that these feedbacks helped them in planning what to study while only 5.4% disagreed. The remaining 46.7% of the students were undecided. This finding is consistent with suggestion by Barnett and Hixon [1], Paris and Winograd [27], and Paul [28] that accurate self-assessment is useful and essential to support metacognition which empowers students to regulate and control their own learning. Students value feedback on their knowledge level. By reviewing their knowledge state, students will be more aware of whether they overvalue or undervalue their knowledge and help them in planning their learning.

Although the majority of the students viewed feedback on knowledge state as useful, more than half of the students (50.7%) stated they were more interested in their marks than feedback on their knowledge level. For items having misconception or partial knowledge, CAAS gave opportunity for students to review the questions, their responses and the actual answer to determine where their mistake are. Unfortunately only 35.6% of the students say that they always get back to check why they get the item wrong. This finding is consistent with finding by Wotjas [34]; Hounsell [17]; Crooks [8] and Jackson [18] who reported that students seem to be only concerned with the final grade or mark. They furthers found that most of the students did not read the feedback if the dislike their grade or mark. According to Ding [9] even if students read feedback comments, they do little with them. In contrast Brookhart [5] found that successful students use both marks and feedback and actively self-assess, learn and to direct their future learning. Black & William [3] suggested that in the absence of marks, students will read feedback much more carefully and use it to guide their learning.

7. Conclusion

A good assessment tool should enable students to evaluate their understanding of a topic before it is too late. CAAS using the NRET scoring method has the potential to assist students to evaluate their understanding by providing feedback on the knowledge state. Generally the majorities of the students surveyed accept CAAS as the learning tool for mathematics and are willing to use it if it is available on the web. Although NRET test instructions seem to be familiar to students, the fact that 35.6% of students felt that it is confusing after 5 training sessions is worth noting. Perhaps, more training should be conducted to enable students to consistently follow the NRET test instructions. Students perceived feedback on knowledge level as useful in learning. However finding on what student do with feedback is rather discouraging. Suggestion by Black & William [3] should be given consideration for effective feedback that promotes learning. This study has produced some encouraging evidence to support the use of CAAS using the NRET scoring method as assessment tool.

References

[1]   Barnett, J.E., & Hixon, J.E. (1997). Effects of grade level and subject on student test score prediction. *The Journal of Educational Research, 90*. 170-174.
[2]   Bennett, R.F., (2001). *How the Internet will help large-scale assessment reinvent itself. Education Policy Analysis Archives*. 9(5). online. Retrieved on April 24, 2002 from http://epaa.asu.edu/epaa/v9n5.html.
[3]   Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1). 7-74.
[4]   Bransford, J.D., Brown, A.L., & Cocking, R.R. (1999). *How people learn*. Washington, D.C: National Academy Press.

[5]   Brookheart, S.M. (2001). Successful students' formative and summative uses of assessment information. *Assessment and evaluation in higher education 8(2)*. 154-169.

[6]   Conrad, R.M., & Donaldson, J.A. (2004). *Engaging the online learner: Activities and resources for creative instruction*. San Francisco: Jossey-Bass.

[7]   Coombs, C.H. (1953). On the use of objective examinations. *Educational and Psychological Measurement, 13*, 308-310.

[8]   Crooks, T.J., (1988). The impact of classroom evaluation practices on students, *Review of educational research. 58(4).* 129-138.

[9]   Ding, L. (1998). Revisiting assessment and learning: implications of students' perspectives on assessment feedback. Paper presented to Scottish Educational Research Association Annual Conference, University of Dundee, Sept 25-26.

[10]   Dirkwager, A. (1975). Computer-based testing with automatic scoring based on subjective probabilities. In Lecarme, O. & Lewis, R. (eds), *Computer in Education*. North-Holland Publishing Company, 305-311.

[11]   Dirkwager, A. (1996). Testing with personal probabilities: eleven year olds can correctly estimate their personal probabilities. *Educational and psychological measurement*, *56*, 957-971.

[12]   Farrell, G., & Leung, Y.K. (2004). Innovative online assessment using confidence measurement. *Education and Information Technologies 9:1*, 5-19.

[13]   Fleischman, J. (2001). Approaches and strategies for assessment: Project-based learning using technology. *Converge Journal,* (Feb. 2001) 38-40.

[14]   Green, M.G., Clark, T.B., & Holland, L. (1982). *Empirical induction and the transition between concrete and formal operational probability concepts*. Paper presented at the Twelfth Symposium of the Jean Piaget Society, Philadelphia.

[15]   Hale, G.A., Miller, L.K., & Stevenson, H.W. (1967). Developmental changes in children's concepts of probability. *Psychonomic Science, 9*, 229-230.

[16]   Holmes, P. (2002). Multiple evaluation versus multiple choice as testing paradigm- feasibility, reliability and validity in practice. *Phd. Dissertation*. ISBN : 90-3651757-5

[17]   Hounsell, D. (1987). Essay writing and the quality of feedback. In J.T.E. Richardson, M.W. Eysenck & D. Warren-Piper (eds.). *Student learning: research in education and cognitive psychology,* Milton Keynes: Open University Press and Society for Research into Higher Education.

[18]   Jackson, M. (1995). Making the grade; the formative evaluation of essays (online), UtiliBASE: available from http://ultibase.rmit.edu.au/Articles?jacks1.html.

[19]   Jaradat, D., & Tollefson, N. (1988). The impact of alternative scoring procedure for multiple-choice item on test reliability, validity, and grading. *Educational and Psychological Measurement*, *48*, 627-635.

[20]   Laborde, C., (1991). The computer as part of learning environment: The case of geometry. In C. Keitel & K. Ruthven (Eds.), *Learning from Computers: Mathematics Education and Technology* (pp.48-67). Berlin Heidelberg, Germany: Springer-Verlag.

[21]   Lockwood, F., (2001). Innovation in distributed learning: creating the environment. In F. Lockwood and A. Gooley (Ed.) *Innovation in open and distance learning: Successful development of online and Web-based learning* (pp.1-14). London: Kogan Page Limeted.

[22]   Lukin, L.E., (1989). *The Psychometric Properties of Confidence Testing as a Function of Computerized Administration, Training, and Development of Probabilistic Reasoning*. Doctorate dissertation. University of Nebraska. United States of America

[23] Lynch, M.M. (2001). Evaluating student mastery and program effectiveness. DEOSNEWS Archives, 11(12). Retrieved July 10, 2004 from http://www.ed.psu.edu./acsde/deos/deosnews/deosnews11_12.asp

[24] Nitko, A.J. (1983). *Educational tests and measurement: An introduction*. New York, New York: Harcourt, Brace Jovanovich.

[25] Oh, H.J. (2004). *Reasoning Test Taker' Guessing Strategy and Their Understanding of Formula Scoring.* Paper presented at annual meeting of the American Educational Research Association (AERA) held between April 12 to 16, 2004, in San Diego, CA.

[26] Orth, J.A. (1995). The effect of test taking strategies for multiple choice tests. Doctoral dissertation, University of Pittsburgh. United State of America.

[27] Paris, S.G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B.F. Jone, and L. Idol, Eds. Hillsdale, NJ: Erlbaum, *Dimensions of Thinking and Cognitive Instruction*. pp. 15-52.

[28] Paul, J. (1994). Improving education assessment by incorporating confidence measurement, analysis of self-awareness, and performance evaluation: The Computer-Based Alternative Assessment (CBAA) project. Retrieved July 16, 2008, from http://www.jodypaul.com/ASSESS/CBAA.pdf

[29] Peat, M., & Franklin, S. (2003). Has student learning been improved by the use of online and offline formative assessment opportunities? *Australian Journal of Educational Technology 2003, 19(1)*, 87-99.

[30] Ross, B.M., & DeGroot, J.F. (1982). How adolescents combine probabilities. *Journal of Psychology, 110*, 75-90.

[31] Shepard, J.F. (1990). *College study skills* (4th ed.). Boston, Massachusetts: Houghton Mifflin Company.

[32] Shepard, J.F. (1992). RSVP: The college reading, study and vocabulary program (4th ed.). Boston, Massachusetts: Houghton Mifflin Company.

[33] Taniar, D. & Rahayu, J.W. (2003). On-line Multiple Choice Testing. *Web-Powered Database.* Idea Group Publishing.

[34] Wotjas, O. (1998). Feedback? No, just give us the answers, *Times Higher Education Supplement.*