Use Of Artificial Intelligence In Highly Adaptive Exam E-Revision Systems

1Hilton Chikwiriro, 2Prudence M. Mavhemwa, 3Pharoah Chaka, 4Munyaradzi Magomelo
1Lecturer, Computer Science Department, Bindura University of Science Education, Zimbabwe
2Lecturer, Computer Science Department, Bindura University of Science Education, Zimbabwe
3Lecturer, Computer Science Department, Bindura University of Science Education, Zimbabwe
4Lecturer, Computer Science Department, Bindura University of Science Education, Zimbabwe

ABSTRACT

Effective revision is an ongoing process, not a cramming session just before the exams so as for the students to acquire knowledge because in an examination it is not about how much you know, but rather “when do you know it?” So there is great need to develop tools that can efficiently increase the knowledge retention levels.

The Intelligent Revision System (IRS) is a revision and instant marking system tool derived from intelligent tutoring systems (ITSs). A fundamental tenet of its design is that,” ONE SIZE DOES NOT FIT ALL” as the learning process varies considerably from student to student. Hence, an IRS will behave like a real teacher would, identifying your strengths and weaknesses and deploying a revision approach which will best fit your revision profile and personality.

Key Words: expert system, knowledge base, ITS, IRS

Introduction

The notion that data and research should be used to improve education policy and practice is now almost a cliché (Seashore Louis, 2003). In response to the vast technology growth and to the growing number of students entering colleges, many lecturers have turned to the use of intelligent methods for the students to retain new and old concepts in the preparation of the exams. It is widely accepted that students acquire new knowledge by engaging in a learning activity, absorbing information from textbooks and lectures and hence artificial intelligence use in e-revision ought to be useful in the form of assessment and revision of what had been learnt before.
These intelligent revision systems are in their way to solve adaptability problems among other problems currently involved in e-revision. Over the last few decades there have been significant attempts to improve the use of knowledge in creating education policy and improving education practice, but these attempts have focused on disseminating findings from research rather than using knowledge and have resulted in a series of uncoordinated activities (Seashore Louis, 2003). More recently, the standards and accountability movement has both directly and indirectly pushed the concept of knowledge utilization back into the awareness of educators, administrators, policymakers, external technical assistance providers, and researchers. The term “knowledge utilization” generally refers to the systematic application of professional wisdom and findings of high-quality research to improve educational outcomes for students.

**Why multiple choice?**

- Can quickly and accurately determine what a student knows and doesn’t know in a course.
- They can also be used to measure a great variety of educational objectives since they allow more adequate sampling of content.
- They are adaptable to various levels of learning outcomes, from simple recall of knowledge to more complex levels, such as the student’s ability to make inferences from a given data hence they properly suit to adaptive learning.

**Intelligent Learning, Revision and Tutoring Systems**

Intelligent tutoring systems (ITSs) are computer programs that are designed to incorporate techniques from the AI community in order to provide tutors which know what they teach, who they teach and how to teach it. AI attempts to produce in a computer behaviour which, if performed by a human, would be described as 'intelligent': ITSs may similarly be thought of as attempts to produce in a computer behaviour which, if performed by a human, would be described as 'goad teaching’ (Elsom-Cook, 1987).

According to Shaw (2008), “AI might be simplistically described as an attempt to use computers to mimic the functioning of human intelligence and may include knowledge acquisition, reasoning and adaptation to experience” (p. 319). Most current Intelligent Learning Systems
(ILS) have been designed to focus on knowledge acquisition and reasoning capabilities, which little work done on making them automatically adaptive.

By 1984 researchers had shown that students who were tutored 1:1 outperformed 98% of their peers (Andersen, 2011). With 1:1 teacher to pupil ratios an impossibility due to costs, ILS incorporating AI provide a cost-effective alternative for school districts to meet that goal (Andersen, 2011; Rishi & Govil, 2008). An ILS is a distributed e-learning system which can deliver a personal learning path to students based on their current skill level (Deliyska & Rozeva, 2009; Payr, 2005; VanLehn, 2006). The term originates from Intelligent Tutoring System which was coined in 1982 by Sleeman and Brown (Welham, 2008). For these ILS to be true AI systems certain criteria must be met. They must have a well defined knowledgebase, a specific pedagogical approach (e.g. constructivist vs. structuralist), they must have reasoning ability and they must be able to automatically adjust to the learner’s ability (Van den Brande, 1993).

The major strength of ILS with embedded AI agents is that they positively affect a student’s sense of responsibility and motivation for their own educational progress (Biswas et al., 2005). The engagement offered by these interfaces seems to hold the students interest and inspire them to complete their educational assignments.

Another major advantage offered by these systems is cost effectiveness. Teaching can be delivered with less manpower, in a more repeatable manner and the curriculum can be easily updated as required. As computer systems can run continuously learning is always available, and students can access this information from multiple locations. This also means that more people can be tutored simultaneously and, as is often the case with e-learning programs, less instruction is required to meet the learning objectives (Ford, 2008; Shaw, 2008). They also often offer suites of tools which make school and class administration more manageable such as automated schedules, assignment management, remediation functionality and student management tools.

Further research is needed into how Teaching Agents (TAs) and ILS can be designed so the student cannot manipulate their results, as it is often possible for students to receive high learning scores by learning how the system functions and not actually learning the educational material (Bodenheimer et al., 2009). Examples of this are systems which allow students to re-submit quiz questions until get the maximum score. Here the student simply keeps re-submitting different
answers until they get the score they want. They have not learned the material but have learned how the system functions.

Most ILS incorporating AI technology are not sophisticated enough to out-perform human teachers who naturally maintain an internal model of where a learner is along their learning path, and modify their teaching accordingly. They also have a deep understanding of the subject they teach and can engage the student’s ways that most ILS cannot (Ford, 2008). Current systems are also limited in the amount of psychological information they can collect about their students. This information is used to create cognitive models of the students and only using keyboards, mouse and question responses to create these models omits a huge volume of physiological and emotional information (Pek & Poh, 2005).

Modern ILS and agent-based AI systems have come a long way since early expert systems. Environments with teachable agents and animated interfaces are encouraging and motivating student learning at a time when it is no longer linked to survival and success. The major challenges now facing the use of pedagogical agents and systems are centred on how to design them to be useful and what is the best way to integrate them into the learning experience (Biswas et al., 2005). The cost of these agent based systems can also be prohibitive and lead to institutions using cheaper, existing technologies.

Whether the semantic web or a valuecosm ever become a reality, educators and parents must never forget that in most cases it is more important to teach students how to think and not what they should be thinking about (Boyle, 1998). Adaptive, AI based systems with teachable agents and animated user interfaces provide an unprecedented opportunity to improve the learning experience by personalizing it to the learner, and the improvement in student motivation, ownership and engagement warrants further exploration.

Adaptive learning is an educational method which uses computers as interactive teaching devices. Computers adapt the presentation of educational material according to students' weaknesses, as indicated by their responses to questions. The motivation is to allow electronic
education to incorporate the value of the interactivity afforded to a student by an actual human teacher or tutor. The technology encompasses aspects derived from various fields of study including computer science, education, and psychology.

**Expert systems**

Expert systems involves the study and design of systems or computer systems that represents, behaves and reasons with expert knowledge in some specialist subject with a view to solving problems or giving advice in areas where human expertise is falling short (Ignizio and James, 1991). These systems are centralized on the use of the knowledge base (collection of reliable expertly gathered facts; pertaining to a particular subject, which can be formally represented if form of cases, frames, patterns, rules and semantic networks) (Jackson and Peter, 1998). Fig 1 illustrates and depicts its components.

![Diagram of an Expert System Components](image)

**Figure 1:** Shows Components of an Expert System
Expert systems are applicable to various trades, professions and other sections that involve human ideas, deductions and reasoning which implies that any fields that require human expertise can use it to minimize risks associated with doing the business, or improve consistence of solutions, or improve completeness, or improve accuracy or all at once while appropriate documentation of the steps followed is compiled for reference and explanations (Darlington, 2000). In addition, today’s weather forecasting is inevitably done by expert systems. These systems do the actual prediction of the weather accurately, quickly and consistently unlike the case of human beings whose reasoning is sometimes unpredictable, slow and inconsistent (Giarratano and Riley, 2005). In spite of these areas, expert systems are an essential useful tool for instructing or training which makes them ideal to aid academic tutors to deliver quality data to their respective students.

**User Interface**

In academic expert systems, the potential users are the tutors (trainers) and the tutees (students) (Darlington, 2000). Both interact with the system via an interactive interface where user queries pertaining to a particular subject are created and the system is then commanded to compute and decide on the solution or advice to the query. It is equipped with the unique features which allow users to ask question on how, why and what format. Student’s tutorials and additional materials can be requested and passed on to the student easily over the interface. In addition, revision and self assessment is expertly conducted between the system and the student and thus better preparation for student examinations. The tutor also uses the interface to the system to create queries on what to expertly deliver to students as well as setting parameters on computer aided student assessments, tests and marking. The actual training or instructing which is supposed to be done by the tutors can easily be conducted by the expert system on the student’s pace and thus effective dissemination of data as the student interacts with the system.

**Explanation Generator**

The explanation generator clearly explains all the procedures that the system used to reach a certain decision or advice which aids the system users to keep track of the strategies being applied to arrive at certain conclusion. During the implementation of forward or backward
chaining reasoning strategies, expert systems produce permanent documentation of the decision process.

**Research Methodology**

The architecture of the system is depicted in Figure 2 below. The system consists of three units: the Identification Unit (IU), the Student Model (SM), and the Evaluation Unit (EU).

![Figure 2: System Architecture](image)

Through IU the student initially subscribes to the system. During subscription some personal settings are saved. After subscription the student can, at any time, enter the system through the IU. SM contains students’ tests results for all their tests, statistics from all the tests, their previous and current knowledge levels as prescribed by the system.

The main goal of EU is to evaluate student's progress due to his/her interaction with the system. This evaluation is achieved through testing. From the testing results the tutor is able to evaluate and watch each student's progress. The ES contributes in deciding upon the knowledge level of a student. It uses production rules described below.
Questions were first given a rating from 1-5 according their level of difficulty by the tutor basing purely on his/her expert analysis and judgment and were inserted in the database in form of pools. The difficulty levels of the questions were graded as follows:

<table>
<thead>
<tr>
<th>Category</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Difficult</td>
<td>5</td>
</tr>
<tr>
<td>Difficult</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>3</td>
</tr>
<tr>
<td>Easy</td>
<td>2</td>
</tr>
<tr>
<td>Very Easy</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Shows grading of complexity of questions

Questions were also given an area or topic basing on the concepts and topics were they originate from. The researcher also took note of the fact that a single question may originate from different topics or areas, for example to solve a particular question you would need to apply a concept from chapter 1, chapter 3 and chapter 5.
This type of question was regarded depending on the basis of which area or topic the tutor wishes to classify it depending on which concept from the topic the tutor thinks would be the most difficult to apply (major concept) for that particular question, basing purely on his expert analysis and judgment.

At each and every stage of revision every time the respondent failed a question, the question was to be pooled in a temporary storage called the buffer of failed questions although it still remained in its original pool.

**Figure 3:** Tree structure of question areas

**Figure 4:** Shows grouping the questions into pools according to their rating

No single question was allowed to be posted twice in a single test. The buffer of failed questions constituted 30% (6 questions) of the questions posted in each and every test, if questions were available in the buffer after each of the preceding test after the first test. If there were 6 questions
or less in the buffer of failed questions, then all the questions would be posted in the next test otherwise the 6 questions were to be chosen in a completely randomized manner.

At each stage of revision, each test was marked instantly and reports were generated to show the score along with any other necessary information depending on the stage of the revision. Then for each question the tutor provided brief notes to the system on why the other answers were not the most appropriate answers and why that particular answer was the most appropriate. The system provided statistics if they were available on how all the other respondents responded to that particular question(for example if the question has been answered 100 times, then 62 respondents chose A as the answer, 13 chose B, 5 chose C, 20 chose C, none chose E).

In each test at each stage of revision the number of questions was always 20. In the first stage of revision questions were selected by the system firstly from strictly pool 1 in a completely randomized manner. For the respondent to proceed to the next phase of the revision, he/she needed to score at least 16 out of the 20 questions (80%) n consecutive times (where n is an integer > 0) as specified by the student. If the respondent was to fail to score more than 80% n consecutive times as he/she would have specified, he/she had to repeat this phase until he/she managed to do so. Also he/she could choose to skip the stage at any time, as he/she wishes.

In the second phase of revision, the questions were posted in the following composition, 10 % from Pool 1, 60% from Pool 2 and 30 % from the buffer of failed questions. If there were no questions from the buffer of failed questions, the remainder of the questions was to be drawn strictly from Pool 2. At that phase of the revision, the respondent needed to score at least 75 % n consecutive times (where n is an integer > 0) as specified by the student. If the respondent was to fail to score more than 75% n consecutive times as he/she would have specified, he/she had to repeat this phase until he/she managed to do so. Also he/she could choose to skip the stage at any time, as he/she wishes, or could chose to return to the previous stage.

In the third phase of the revision the questions were posted in the following composition 10% from pool 1, 10 % from pool 2, 50 % from pool and 30 % from the buffer of failed questions. If there were no questions from the buffer of failed questions, the remainder of the questions was to
be drawn strictly from Pool 3. At this phase of the revision, the respondent needed to score at least 70% \( n \) consecutive times (where \( n \) is an integer > 0) as specified by the student. If the respondent was to fail to score more than 70% \( n \) consecutive times as he/she would have specified, he/she had to repeat this phase until he/she managed to do so. Also he/she could choose to skip the stage at any time, as he/she wishes, or could chose to return to any of the previous stages.

In the fourth phase of the revision the questions were posted in the following composition 10% from pool 1, 10% from pool 2, 10% from pool 3, 40% from pool 4 and 30% from the buffer of failed questions. If there were no questions from the buffer of failed questions, the remainder of the questions was to be drawn strictly from Pool 4. At this phase of the revision, the respondent needed to score at least 60% \( n \) consecutive times (where \( n \) is an integer > 0) as specified by the student. If the respondent was to fail to score more than 60% \( n \) consecutive times as he/she would have specified, he/she had to repeat this phase until he/she managed to do so. Also he/she could choose to skip the stage at any time, as he/she wishes, or could chose to return to the previous stages.

In the final phase of the revision the questions were posted in the following composition 10% from pool 1, 10% from pool 2, 10% from pool 3, 10% from pool 4, 30% from pool 5 and 30% from the buffer of failed questions. If there are were no questions from the buffer of failed questions, the remainder of the questions was to be drawn strictly from Pool 5. At this phase of the revision, the system used statistics from strictly the final stage of revision to determine the respondent’s areas of weakness and strength. The strength in a particular area was graded as follows:

<table>
<thead>
<tr>
<th>Category</th>
<th>Grade</th>
</tr>
</thead>
</table>
Table 2: shows grading of strength in question areas

Due to the fact that when there was very little data for the statistics (i.e. very few tests have been written in the final stage), the results from the statics were less likely to give an accurate picture on the respondent’s strengths and weaknesses, the grading of areas would only begin after at least 5 tests have been written in the final stage.

Designing the adaptive elements of the system

The system needed to be highly adaptive to student’s needs and used adaptive questioning to quickly and accurately determine what a student knows and doesn't know in a course. After 5 tests had been written in the final phase, the system needed to post more questions from the weaker areas than from the stronger areas.

Therefore to be able to do that the system now needed to use both the question’s ratings and the area of the question. After test 5 all question from Pool 1 to Pool 4 were to be drawn randomly from areas regarded as very weak areas. If there were no very weak areas, then all the questions from Pool 1 to Pool 4 would be drawn randomly from the next available weaker group in the order weak, fair, strong, very strong, otherwise they would just be drawn from the relevant pool completely at random. The composition of questions from of Pool 5 remained unchanged at 30% and questions were drawn from Pool 5 completely at random without considering the area if the question The composition of questions from the buffer of failed question also remained unchanged at 30%, and questions were to be drawn completely at random without considering the area of the question. This was very helpful in that questions were posted from all areas were taken into consideration into the new statistics, for example if an area was regarded as a very strong area it was not going to necessarily remain a very strong area after the preceding test or
tests. In the final stage of the revision, the respondents were allowed to write as many tests as they wished.

The system offered to the tutor the capability of Question management. Question management concerns the insertion, deletion or change of a topic or questions. The tutor can insert a new topic or question or delete an existing question or topic or modify a question or topic. In other words, the tutor can change the domain knowledge tree at any time according to his/her needs.

**RESEARCH INSTRUMENTS**

The author used an introductory course offered in nearly all the Universities and Colleges in Zimbabwe named *Introduction to Computer Science* which is designed to strategically introduce computer science concepts to all students in their first academic year. The research provided a platform of revision for students to use multiple choice questions extracted from the lecture notes or any other relevant materials to the course

- The operation was as follows:
  - 1st stage involved the insertion of students and tutor details together with relevant multiple choice questions into the intelligent revision system database.
  - 2nd phase involved the interfacing of users with the intelligent revision platform by accessing questions in form of tests.
  - The participants of this study were tutors (lecturers) and students within such a college. The researcher allowed students to interact with the Intelligent Revision System for two weeks after the course had been completed.
  - The tutor is able to insert, delete or modify all the attributes of a question. To insert a new question the tutor has to select the corresponding form. In that form the tutor has to fill in all the attributes of the question.
  - To delete a question, the tutor has to select the corresponding question and then click a delete button. Modification of a question requires first the selection of the question. Then
the tutor can update any part of the question. For example, the tutor can add a new possible answer or remove one or modify existing ones. Also, the tutor can see all the questions and all information concerning them.

Given the close proximity of colleges or university offering Computer science courses, the researcher engaged 52 Bindura University first year students doing Introduction to Computer Science (CS101). After implementing the Intelligent revision system and given to students doing Introduction to Computers (CS101), students used the tool to revise for their exams. The group of students was divided into two groups of 26 students each; one group using the expert system and another using a random question generator and marker that could draw questions completely at random, at topic level or from the entire course depending on the student’s preference.

Questionnaires were designed to measure people’s perception on a phenomenon and assess students ‘understanding. This could be done by comparing scores for both groups. They are used to examine the student’s understanding when using IRS as an eLearning approach from the students and tutors perspectives. In this part, rating scale questionnaires were designed with five numerical values (1 to 5) corresponding to a scaling “Strongly Agree, Agree, Neither Agree Nor Disagree, Disagree and Strongly Disagree” respectively. The scoring key for each item was taken from scale 1 to 5.

**Results and Discussion**

The questionnaire contained only closed questions. A total of eight closed questions, in the form of positive and negative statements to agree or disagree with, were asked. Each closed question used a five point Likert response scale where each scale point was defined as shown in Table

<table>
<thead>
<tr>
<th>Scale point</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>4</td>
<td>Agree</td>
</tr>
<tr>
<td>3</td>
<td>Neither agree nor disagree</td>
</tr>
<tr>
<td>2</td>
<td>Disagree</td>
</tr>
</tbody>
</table>
Table 3: Shows likert scale values

- The results drawn from the tests were as follows for the questionnaires:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage agreeing/disagreeing with the statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Using the intelligent revision system has helped me with my preparation for a multiple choice test</td>
<td>92% (Agree)</td>
</tr>
<tr>
<td>2. I like the way the intelligent revision system tells me how the class as a whole have answered the question.</td>
<td>76% (Agree)</td>
</tr>
<tr>
<td>3. Can this e-revision system quickly determine what the student knows or doesn’t know in a course and does it adapt to students learning needs.</td>
<td>88% (Agree)</td>
</tr>
<tr>
<td>4. Using the intelligent revision system was a waste of time</td>
<td>85% (Disagree)</td>
</tr>
<tr>
<td>5. I found the intelligent revision system difficult to use</td>
<td>88% (Disagree)</td>
</tr>
<tr>
<td>6. The Intelligent revision system does adapt well to students’ learning needs as students gain more knowledge.</td>
<td>81% (Agree)</td>
</tr>
<tr>
<td>7. The intelligent revision system made revision sessions more enjoyable</td>
<td>92% (Agree)</td>
</tr>
<tr>
<td>8. Would you have liked to use the revision system in all your courses</td>
<td>73% (Agree)</td>
</tr>
</tbody>
</table>

Table 4: shows results for each of the ten closed questions

In terms of student’s perception of the intelligent revision system as a helpful learning tool, 92% agreed that the Intelligent revision system helped them prepare for a multiple choice test (Question 1), with only 8% disagreeing somewhat and none disagreeing strongly. 88% agreed this e-revision system quickly determines what the student knows or doesn’t know in a course (Question 3 - 88%) and (81%) agreed that the system does adapt well to students’ learning needs as students gain more knowledge (Question 6 - 81%). In response to ‘I like the way the system tells me how the class as a whole has answered the question’ (Question 2), 76% of the students agreed with this statement.

Students perceived the system to be easy to use (Question 5 - 88%), made revision more enjoyable (Question 7 - 92%), and was not perceived as a waste of time (Question 4 – 85%).
73% of students would have liked to use the revision system in all their courses (Question 8), whereas 19% neither agreed nor disagreed with this statement and only 8% would not want to use it in all their courses.

Finally the groups were tested on the final examination performance were a set of assumptions and hypothesis were set and tested using the independent samples t-tests and one way ANOVA. Results shown using Levene’s test of equality of variance and independent sample test showed that the two groups were significantly different on their performance in the final examinations.

Both the two set of tests i.e. the tool and the use of final examination clearly showed and proved that expert systems are the best approach to students where there is a problem of shortage of experienced and expert tutors in a colleges and universities in a country like Zimbabwe or for distance education. It clearly yields a positive impact and outcomes to the student performance.

From a teaching and learning perspective, the intelligent revision system was perceived to aid understanding of a subject, help with revision purposes/preparation for exams, help students to identify their strengths and weaknesses, and provide experience of answering multiple choice questions.

<table>
<thead>
<tr>
<th>mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of Squares</td>
</tr>
</tbody>
</table>
Table 5: shows comparisons of student scores using ANOVA Report

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.96</td>
<td>26</td>
<td>8.258</td>
</tr>
<tr>
<td>2</td>
<td>65.69</td>
<td>26</td>
<td>8.475</td>
</tr>
<tr>
<td>Total</td>
<td>68.33</td>
<td>52</td>
<td>8.701</td>
</tr>
</tbody>
</table>

Table 6: Shows Average and Standard deviation results for students
With the following Hypothesis stated i.e:

H₀: Use of artificial intelligence on multiple choice e-revision and assessment has no effect to students’ understanding.
H₁: Use of artificial intelligence on multiple choice e-revision and assessment has effect to students’ understanding.

Since the P-value from the ANOVA is less than 0.051, I fail to accept H₀ and conclude that the use of artificial intelligence on multiple choice e-revision and assessment has effect to students’ understanding.

Summery and Conclusions

The tutor is able to see some statistical results over the sections the student has been examined at. The system can also provide statistical results about all students that have examined in various areas. The tutor can quickly recognize the areas that are giving students most problems. The tutor can re-visit those areas in form of lectures and even tutorials so as to make the students or the class as a whole have a broader understanding of those particular areas. The tutor also is able to identify areas that are posing difficulties to individual students and can be able to help them on an individual bases depending on the needs for that particular student.

The student also can is able to see some statistical results over the concepts and sections he/she has been examined at. He/she can quickly identify the areas that are posing much difficulty in the course and he/or she can put more effort and more time in those problematic areas. Also the system itself can also use the statistics to provide tailor made tests for individual students depending on the individual needs for that particular student.

In conclusion, expert systems for education are here to stay, and they are going to make everyone’s job (both the tutor and the student) much easier, more efficient and most of all much more precise.
References


