# IMPROVING THE STUDENT LEARNING EXPERIENCE FOR SQL USING AUTOMATIC MARKING

Dr Gordon Russell
*Napier University*
*10 Colinton Road, Edinburgh*
*g.russell@napier.ac.uk*

Andrew Cumming
*Napier University*
*10 Colinton Road, Edinburgh*
*a.cumming@napier.ac.uk*

**ABSTRACT**

An online interactive learning environment which offers immediate feedback to student learning SQL was constructed. This has supported the gathering of interesting statistics on student behaviour. Changes to the environment targeted at improving student behaviour can be made, and the impact of these changes measured statistically on real students. This paper considers a cycle through measuring student behaviour, implementing methods to try and improve the behaviour, and the result of these changes.

**KEYWORDS**

e-learning, SQL, automatic marking, learning environment., Web based learning

## 1. INTRODUCTION

The benefits of computer-based learning have been established by, among others, a working party of the Committee of the Scottish University Principals (1992). This report highlights the effectiveness and the cost effectiveness of such an approach. The working party suggest that CBL is 20%-30% more cost effective than conventional teaching. However, the idea of web-based learning is often interpreted as a website with online notes, quizzes, and a discussion forum. Such components, although valuable in themselves, only skim the surface of what lecturers see as their inputs into the practical session experience.

This paper considers a learning environment constructed by Database System’s lecturers. Although the environment has the usual notes, quizzes, and forums, the main focus is to provide an online tutorial environment that makes significant steps towards being a traditional practical lab replacement. In this particular class, the main topic taught is the learning and practical application of SQL.

SQL, or the Structured Query Language, is a way of querying information from a database. It is similar to a programming language, with syntax to learn and understand. The application of SQL to a particular problem requires problem-solving abilities. For any particular data extraction specification there may be many different yet valid SQL solutions.

The learning environment has been in place for a number of years. Students interactions with the system have been electronically monitored, and the information is regularly analysed. From this analysis strengths and weaknesses of the system are identified, and this information is used to inform the building process for the next version of the environment. This has given the authors an excellent insight into student learning behaviour, and how behaviour changes with learning environment changes. Some aspects of this is discussed in this paper.

This paper gives an overview of the learning system for SQL. It also considers some of the key aspects that were identified as design foci for the learning environment. Some behavioural analysis is shown, in conjunction with system changes designed to manipulate the student learning experience. Finally, the future of this system is considered.

# 2. TEACHING SQL

SQL is a popular language for manipulating databases. SQL is available for most relational database platforms such as Oracle, Microsoft's SQL Server and IBM's DB2. The tutorials focus on the data extraction components of the language. A typical question for the student could be "List the employee number and name for all employees with a salary of at least £20000."

A possible answer to this question would the SQL statement:

```
SELECT   employee.empno, surname
FROM     jobhistory JOIN employee
          ON jobhistory.empno=employee.empno
WHERE    salary>=20000
```

There are many correct variations on this answer, for example the equality may be written the other way around, the term `surname` might be replaced with the term `employee.surname` .

Prior to the online environment students would typically take a tutorial with printed questions concerning a database that had been prepared earlier. Students would have had access an Oracle database using the proprietary interface SQLPlus. This style of tutorial has proved popular and effective for many years, however by embedding the interaction into a more controlled environment we hoped to further improve the student's experience and allow us to monitor and control the classes more effectively.

## 2.1 Examining the answer

In working on a question the typical student faces two kinds of error. A syntax error occurs when the SQL statement has been incorrectly structured. The system simply returns with an error message. A logic error occurs when the syntax is correct and the system responds with a table of values – but these values are wrong.

The interface SQLPlus is ideal for responding to syntax errors, however it cannot possibly identify logic errors as the interface cannot "know" what it is that the student is trying to do. In contrast the web-based interface has access to the context and can react to both syntax and logic errors.

## 2.2 The environment

The environment may be regarded as a tutorial simulation as discussed by Larillard (2002). It offers both intrinsic feedback in the form of error messages or tables of results coupled with extrinsic feedback based on the system's ability to evaluate the quality of the submissions.

Before the creation of the online system a number of key observations of student behaviour were made.

- Tutorial questions were often completed incorrectly, due to the common misunderstanding that "if the SQL runs without errors the result produced must be right". Although the queries did run, they did not always answer the question asked.
- With large numbers students in a tutorial, it is difficult to police the work on the tutorials and identify student who need assistance but who have not asked for help. This may be because they are unaware of their problems or due to a reluctance to engage with the teaching staff.
- Student attendance at practicals was not compulsory, and the support of student working on the material yet not attending practicals needs to be considered. Similar support techniques are required for true distance learners.

These issues were the driving issues behind the creation of the learning environment. The environment offers students a web site that allows them to navigate through a number of SQL tutorial questions (currently 70 questions). Each question is presented on a page, accompanied with a box where the student can enter SQL commands which produces the data required by the question. The SQL can be executed at any time, and thus the student can work incrementally on a question until they are satisfied by there efforts. The system remembers all SQL entered, so that students can return to the question and see what they last entered.

When the SQL is executed, not only are the results of executing the SQL shown, but the system also analyses the result. It indicates how close the supplied SQL is to producing the information requested by the question. This is the SQL *accuracy*. It is given as a percentage from 0 (nothing to do with the question) to 100% (perfect answer). It also gives a row/column analysis, highlighting parts of the answer which is right, differentiating the rights aspects from the wrong.

Consider the question:

> Show the name of the country and the population for countries with a population more than 200000000 people.

If the student gave the right answer:

```
SELECT      name,population
FROM        cia
WHERE       population>200000000
```

then the accuracy would be 100%. However, the student who gave one extra column and mistyped in the number of people, entering:

```
SELECT      name,population,region
FROM        cia
WHERE       population>100000000
```

This would have an accuracy of 56% and would be shown as in **Table 1**:

Table 1. Automarking example

| NAME | POPULATION | REGION |
|---|---|---|
| Brazil | 172860370 | South America |
| China | 1261832482 | Asia |
| India | 1014002817 | Asia |
| Indonesia | 224784210 | Southeast Asia |
| United States | 275562673 | North America |

The highlighted rows and columns indicate additional data that should not have been returned. If rows or columns are missing, text is added at the bottom of the output instructing the student as to how many rows or columns are missing.

In addition to the accuracy, some additional measures are also taken. These are based on string searches of the student's SQL, and may result in marks being taken away from the accuracy. Such things include writing SQL which is significantly longer than that of the sample solution.

Formal assessments are taken online in a similar way to tutorials. Questions are asked of the students, and they have to provide SQL that answers the question. The system collates these assessment answers, and provides the lecturer with the final mark. The analysis section discusses how many questions are given, the time when the student can attempt the questions, and how much feedback students receive during the assessment from the system.

The environment also gathers a number of statistics to allow the authors to better understand the student's approach to learning SQL. These statistics are:

- How long a student spends logged on to the learning environment.
- How long each tutorial and assessment question took to get right.
- How many questions from the tutorial has been successfully completed.
- The student work pattern on a day by day basis.

Two versions of the system are discussed in this paper, year 2002/3 and year 2003/4. The current version of the environment can be found at http://db.grussell.org/sql/ .

## 3. OBSERVATIONS: 2002/3

In 2002/3 the feedback and anecdotal evidence indicated that the environment was a significant step forward in supporting students learning SQL. On analysing the statistics a number of interesting features were identified. The cohort consisted of over 300 undergraduate students, most of which were enrolled as Computing students on the second year of an undergraduate degree.

The students worked for 6 weeks on the SQL tutorials and 3 weeks on the assessment questions. The pattern of study for the majority of students was regular and incremental. However, a significant number of students did no work on the tutorials or assessments until the last week of the labs, and then produced weak or plagiarised assessment solutions. This group also produced weak results in the final written exam for the subject.

Mayes (1995) describes the learning cycle comprising of Conceptualisation/Action/Dialogue – for this to be achieved requires the students to study in this regular and incremental pattern.

It was decided to manipulate student behaviour in order to encourage regular progression through the assessment and tutorial questions.

A number of steps were taken to improve the student work pattern.

- The tutorials questions were grouped into 4 tutorial groups.
- The final assessment was replaced by 4 assessments, each one related to a tutorial group.
- Only after completing 75% of a tutorial group correctly could the assessment for that group be taken.

Without automatic marking this approach would be time intensive for the lecturer. With the automatic marking system the students gained feedback instantly on the assessments. The assessment procedure was redesigned, so that students went through a number of phases:

1. Start the questions in a tutorial group
2. When 75% of questions were correct the assessments are made available.
3. The students can complete the assessment in their own time.
4. Once satisfied, the student can "close" the assessment, using a similar metaphor to the idea of submitting their coursework to be marked.
5. Once an assessment is closed, no further changes can be made, but the student can immediately see their mark for that question.

For the first assessment, even without closing the assessment the system gave them some feedback continuously as to how many marks they would get. On harder assessments, no feedback was given until the assessment was closed. It was thought that this approach would produce the best mark distribution for the assessments.

## 4. OBSERVATIONS 2003/4

The first statistics to be measured was the number of students who now worked at a regular pace through the material. To measure this, an "average student time plan" was created, and the number of students who achieved or exceeded this plan was measured. The average student plan is such that a student can achieve the average mark of around 55%, which required (based on historical data) an average material coverage over the teaching weeks of about 65% of the material. This analysis was made on both the 4 assessment approach (2003/4) and the old approach (2002/3). The cohort involved a similar number of students as the previous year. Analysis of the cohorts (excluding this module) shows their overall averages for the year to be 52.6% in 2002/3 and 52.9% in 2003/4. This marginal improvement in the cohort is negligible, and no compensation for it has been made.

The unnormalised data extracted from the teaching environment is given in **Table 2**. Note that student numbers go up slightly as time goes on, as not all students arrive in the first week. Also of note is that in 2003/4 7 extra questions were added to the tutorial. These were in the form of a trivial introduction to SQL, and were done right at the beginning of the tutorial questions. To compensate for this it was assumed that the 2002/3 cohort automatically passed these additional questions, which in terms of the arguments presented is the "worst case" situation.

The experimental results after normalisation are shown in Figure 1

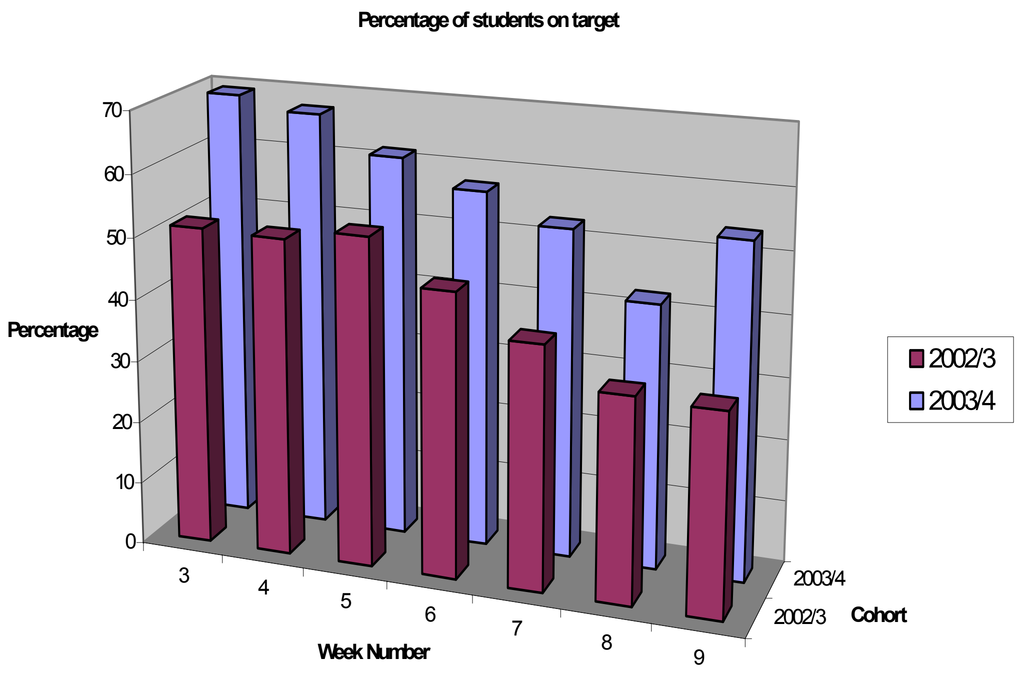


Figure 1. Average Student Target Reached

Figure 1 shows that there appears to be a 20% improvement from 2002/3 to 2003/4 for the average student target.

From the University's quality measures for the module, this has a significant impact in improving the module's perceived quality. It is also interesting to analyse the impact of this change on the better-than-average student.

Figure 2 shows a similar analysis to that in Figure 1, except that a different time plan was used. In this case a time plan was created for what the authors would consider to be the good students (the top quartile performers). This would be for students who cover almost all of the tutorial questions before the lab time runs out.

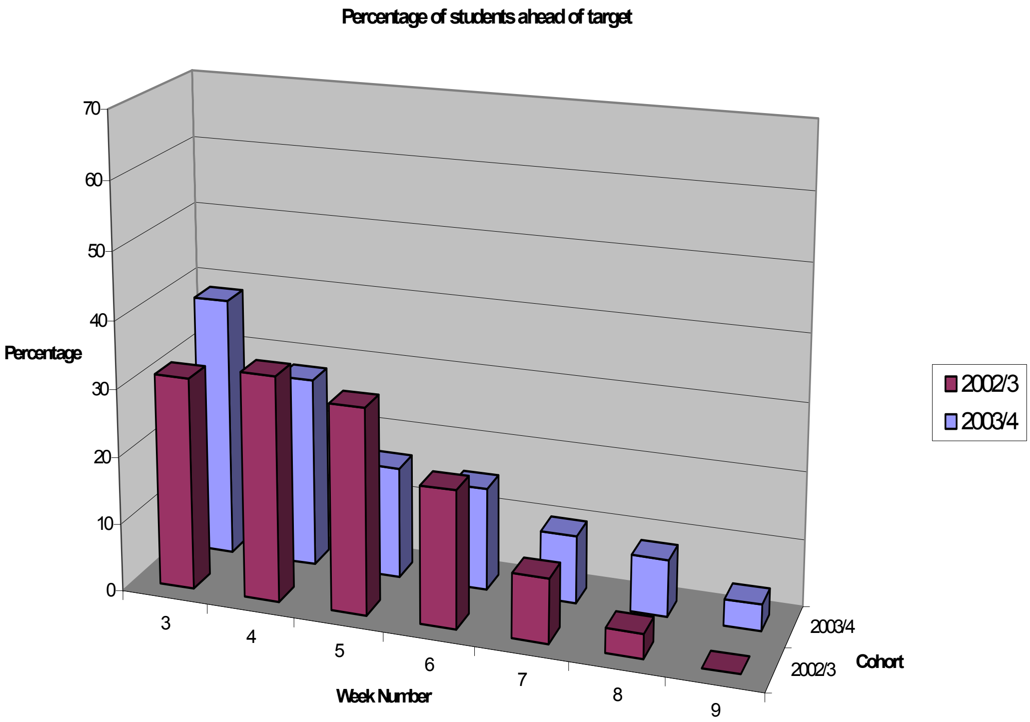


Figure 2. Good Student Targets Achieved

Figure 2 shows that there was a slight negative impact from the 2003/4 changes for good students. It indicates that slightly more students in 2003/4 worked in week 1 with the approach typical of a good student, but then in general the perception is that fewer students worked hard enough to be considered as following a "good student" time plan. Note that eventually by week 7 there was a slight improvement in the observed statistics.

It must be remembered that the main focus of this work was to improve the experience for weak and average student, but even so this impact on good students is undesirable. Some observations on this data need to be made:

- With more interim assessments there was more to do before week 7 in 2003/4 than in 2002/3. The impact of this is hard to measure.
- Interim assessments promote the tutor's perception of how time should be managed. Perhaps this was more easily accepted by weaker students and was disliked by good students.
- With clearer milestones good students may have been able to concentrate more of their time in the other modules being taught to them,
- The number of "good students" did not change, only their work pattern. However, the number of "average students" was increased and the number of weak students was reduced.
- In 2003/4 good students worked on until week 9, while average students did little work after week 7. Perhaps in the new system good student simply work over more weeks rather than faster during fewer weeks.

It should be said that in 2003/4 the pass rate was improved. This shows that the system did indeed help the weaker student. In addition, the mean and standard deviation remained healthy, and so it is likely that the good student simply worked in a different way in 2003/4 than in previous years, showing them to be good adapters to changing teaching techniques.

## 5. CONCLUSIONS

The evaluation shown here is a snapshot of the experiments currently being undertaken using this online SQL environment. Other experiments include the correlation of student marks in the exam against the coursework marks; identifying weak student early to allow tutors to target those students in the practical sessions; the detection of plagiarism and the discouragement of plagiarism.

The data that is shown here highlights that web-based teaching beyond that of simple forums, online notes, and quizzes, can have a significant positive impact on student behaviour. These changes to behaviour cannot easily be achieved through traditional practical sessions without significant cost in terms of lecturer time.

Use of the environment allows the tutors' role to evolve from "Lecturer and oracle" to "information guide and resource provider" as noted by Ryan (2000). More interestingly the students are able to move from being "passive receptacles" to "autonomous, independent, self-motivated managers of their own learning process".

In addition this work shows that it is all too easy to focus on poor performing students, and develop systems to improve their performance, but in turn to weaken the performance of the bright student. Special consideration must be given to bright students too, or there is a danger of "switching off" their brightness.

## REFERENCES


Committee of the Principals of Scottish Universities 1992. *Teaching and Learning in an Expanding Higher Education System,* SCFC, Edinburgh

Laurillard, Diana 2002. *Rethinking University Teaching (2nd Edition),* RoutledgeFalmer, London

Mayes, J T 1995 *Learning Technology and Groundhog Day 1995*. In Strang, Simpson, and Slater: Hypermedia at Work: Practice and Theory in Education, Publishers, University of Kent Press, Canterbury.

Ryan, S, Scott, B, Freeman, H, Patel, D 2000 *The Virtual University: The Internet and Resource-Based Learning* Kogan Page, London

Table 2 : No. of students who attempted a given no. of questions in a given week.

| Questions Attempted | 2003/2004 Cohort Week Number | | | | | | | 2002/2003 Cohort Week Number | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0 | 13 | 14 | 14 | 11 | 11 | 11 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 3 | 3 | 4 | 5 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 8 | 6 | 5 | 4 | 5 | 5 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 14 | 12 | 11 | 10 | 7 | 6 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 7 |
| 10 | 11 | 13 | 8 | 6 | 7 | 5 | 3 | 8 | 6 | 5 | 5 | 5 | 5 | 7 |
| 12 | 9 | 8 | 4 | 3 | 4 | 5 | 2 | 18 | 10 | 10 | 10 | 10 | 10 | 10 |
| 14 | 5 | 5 | 4 | 6 | 3 | 2 | 2 | 20 | 13 | 13 | 13 | 13 | 13 | 11 |
| 16 | 7 | 8 | 6 | 2 | 1 | 2 | 2 | 14 | 16 | 11 | 11 | 11 | 11 | 13 |
| 18 | 19 | 15 | 10 | 7 | 6 | 5 | 4 | 20 | 17 | 14 | 14 | 14 | 14 | 12 |
| 20 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 10 | 11 | 13 | 13 | 13 | 13 | 11 |
| 22 | 10 | 4 | 5 | 9 | 7 | 3 | 5 | 12 | 15 | 12 | 12 | 12 | 12 | 10 |
| 24 | 16 | 15 | 15 | 12 | 12 | 11 | 7 | 17 | 14 | 12 | 12 | 12 | 12 | 11 |
| 26 | 51 | 39 | 23 | 16 | 17 | 15 | 10 | 22 | 15 | 12 | 12 | 12 | 12 | 10 |
| 28 | 25 | 21 | 11 | 14 | 10 | 9 | 3 | 19 | 19 | 14 | 14 | 14 | 14 | 15 |
| 30 | 27 | 20 | 25 | 14 | 12 | 12 | 7 | 19 | 12 | 15 | 15 | 15 | 15 | 15 |
| 32 | 16 | 18 | 15 | 15 | 14 | 14 | 2 | 19 | 8 | 9 | 9 | 9 | 9 | 3 |
| 34 | 14 | 17 | 18 | 16 | 12 | 12 | 9 | 7 | 6 | 7 | 7 | 7 | 7 | 10 |
| 36 | 15 | 29 | 32 | 28 | 21 | 18 | 12 | 16 | 11 | 16 | 16 | 16 | 16 | 17 |
| 38 | 10 | 18 | 30 | 25 | 33 | 28 | 29 | 20 | 16 | 9 | 9 | 9 | 9 | 8 |
| 40 | 9 | 14 | 11 | 18 | 14 | 16 | 11 | 9 | 15 | 11 | 11 | 11 | 11 | 13 |
| 42 | 4 | 10 | 15 | 18 | 19 | 17 | 12 | 14 | 19 | 13 | 13 | 13 | 13 | 13 |
| 44 | 0 | 1 | 2 | 5 | 7 | 1 | 2 | 5 | 8 | 6 | 6 | 6 | 6 | 6 |
| 46 | 1 | 0 | 3 | 7 | 8 | 5 | 13 | 6 | 9 | 8 | 8 | 8 | 8 | 9 |
| 48 | 1 | 2 | 5 | 7 | 9 | 11 | 13 | 6 | 15 | 14 | 14 | 14 | 14 | 10 |
| 50 | 2 | 2 | 3 | 1 | 8 | 10 | 25 | 2 | 5 | 6 | 6 | 6 | 6 | 4 |
| 52 | 3 | 2 | 6 | 13 | 9 | 10 | 12 | 5 | 4 | 11 | 11 | 11 | 11 | 12 |
| 54 | 0 | 2 | 5 | 5 | 7 | 13 | 15 | 1 | 6 | 11 | 11 | 11 | 11 | 10 |
| 56 | 2 | 3 | 6 | 9 | 15 | 14 | 11 | 6 | 10 | 19 | 19 | 19 | 19 | 21 |
| 58 | 0 | 1 | 2 | 2 | 3 | 8 | 6 | 0 | 1 | 3 | 3 | 3 | 3 | 6 |
| 60 | 0 | 1 | 2 | 4 | 3 | 6 | 7 | 0 | 2 | 6 | 6 | 6 | 6 | 9 |
| 62 | 0 | 0 | 0 | 3 | 3 | 5 | 5 | 0 | 4 | 8 | 8 | 8 | 8 | 7 |
| 64 | 0 | 1 | 1 | 1 | 2 | 2 | 7 | 0 | 8 | 4 | 4 | 4 | 4 | 5 |
| 66 | 1 | 1 | 2 | 1 | 3 | 9 | 17 | 0 | 2 | 2 | 2 | 2 | 2 | 3 |
| 68 | 0 | 0 | 0 | 4 | 6 | 8 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 70 | 0 | 0 | 2 | 4 | 5 | 7 | 15 | 0 | 2 | 9 | 9 | 9 | 9 | 15 |
| Total Students | 301 | 310 | 310 | 310 | 310 | 310 | 310 | 300 | 304 | 307 | 307 | 307 | 307 | 318 |