A comparative analysis of student SQL and relational database knowledge using automated grading tools

BOISVERT, Charles <http://orcid.org/0000-0002-3069-5726>, DOMDOUZIS, Konstantinos <http://orcid.org/0000-0003-3679-3527> and LICENSE, Joshua

Published version

BOISVERT, Charles, DOMDOUZIS, Konstantinos and LICENSE, Joshua (2018). A comparative analysis of student SQL and relational database knowledge using automated grading tools. In: Proceedings of the international symposium on computers in education (SIIE) 2018. IEEE.

A Comparative Analysis of Student SQL and Relational Database Knowledge Using Automated Grading Tools

Charles Boisvert, Department of Computing
Sheffield Hallam University, City Campus, Howard Street, Sheffield, S1 1WB, United Kingdom
Email: c.boisvert@shu.ac.uk
Konstantinos Domdouzis, Department of Computing, Sheffield Hallam University
Email: k.domdouzis@shu.ac.uk
Joshua License, BSc Graduate (2017) in Computing, Sheffield Hallam University
Email: joshua.license@gmail.com

Abstract

This paper evaluates a blended learning methodology for Relational Database Systems. Our module offers students a range of interconnected tools and teaching resources. Among them is `TestSQL`, a query tool giving the students automated feedback on SQL query exercises; but we do not use it to assess the students. Instead assessment is through a range of questions which test not only SQL writing skills, but also other aspects of the field, including questions on optimisation, physical modelling, PL/SQL, and indirect questions on SQL knowledge, such as processing order.

The effectiveness of the approach is investigated through a survey of student attitudes', and assessment data. Our analysis shows, unsurprisingly, that the students' use of more resources correlates significantly with better results; but also that success at the different sub-topics tested is not at all well correlated, which shows that students can master some topics while remaining weak at others; and finally, that indirect SQL questions is best predictor of success at each of the other sub-topics. This last result confirms our choice to broaden the testing of SQL skills, and has implications for the use automated SQL assessment tools: we recommend that in automated testing for Database Systems, SQL writing tests be complemented with indirect questions on keyword use, parsing, or error recognition aimed at revealing broader abilities of learners.

Index Terms

Computer Science Education, Structured Query Language, Automated assessment, Relational Databases

I. Introduction

At Sheffield Hallam University, computing students learn about Relational Database Systems through a module that incorporates many parallel approaches and teaching aids over an academic year. In this paper we use data on student assessment and on their attitudes to evaluate the complementarity of the multiple topics addressed and tools used in the first semester. Of particular interest is the comparison of the students' skills in the Structured-Query Language (SQL) with other aspects of Database Systems.
We refer to our approach as blended learning, because we deliberately mix on-line interactive, print, and face-to-face delivery methods, though blended learning is ill-defined and arguably a misnomer [1]. More clearly, we follow the recommendations of variation theory [2]: that to support the learners in identifying defining characteristics of the object of study, we should provide varying material which confronts them to the same objects in different forms and contexts.

II. A wide range of tools

The module is divided in two semesters. Here, we evaluate the first one, which focuses on developing proficiency with physical database implementation using an Oracle relational database management system. This includes the SQL language, understanding a Physical Model, Query Processing and Optimisation, programming in Procedural-SQL (PL/SQL).
Teaching methods follow a traditional pattern of lectures and seminars, but use a range of interrelated learning aids to introduce the different topics at hand. As well as lecture materials, seminars are supported

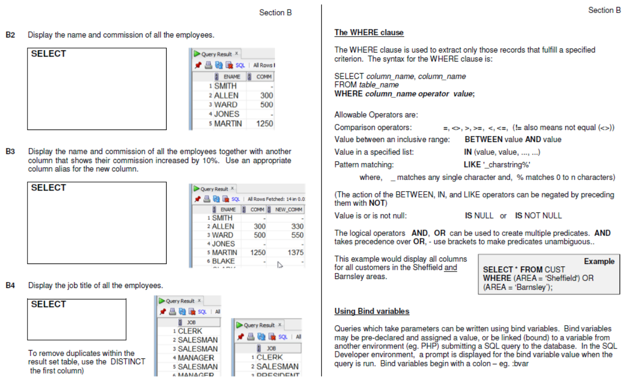

Fig. 1. SQL exercises from the Workbook

by a series of exercises on SQL and PL/SQL topics, which are presented in a workbook and distributed to each student at the beginning of the academic year. Summaries of key points in support of the same material are also offered in video form, and much of the SQL practice is available with automated feedback via our SQL test tool, `TestSQL`.

A. Lecture material

The lectures are divided into different sections and each section describes a separate characteristic of SQL. Examples are the description of the different types of constraints and JOINS in the development and operation of relational database. Through the whole semester, students gain theoretical knowledge combined with practical experience in all the aspects of the language and also, in PL/SQL.

B. Workbook

An SQL workbook combines practice exercises and pointers to key information. Its learning material is deliberately limited to short syntax reminders and to references to lecture and video materials.
This both dedicates the workbook to practice, and encourages the students to refer to more complete information; but navigation - section naming, titles, order of topics addressed - is kept consistent with the lectures described above and with other learning material.
Students can practice the same exercises either in print or electronically at the interface. Working at the interface is very common but we encourage the written media as it also helps learners reflect to work out their solutions.

C. TestSQL

Learning SQL needs to be supported by more than pen and paper practice [3], [4]. Our interactive web application was adapted to particularly facilitate this work.
`TestSQL` was initially written and released as Open Source software in 2017 [5]. The system can use any relational database, imported in `.sqlite` format, support querying that database in SQL and dynamically construct questions for that database. The user's query is checked first testing for inclusion of keywords, then the query results are compared against those of a model answer for total number of rows and columns, and finally to verify that the result set is identical.
More importantly, because `TestSQL` checks the student's answers against a browser-embedded database. This architecture gives a system that can be deployed with very little configuration effort.
This simplicity was put to use to adapt the tool to this module by developing a set of prepared questions to match the exercises and example data used in the workbook.
Not all workbook exercises can be tested through `TestSQL`. Some of them require answers *about* SQL, rather than *in* SQL; also, `TestSQL` does not yet support checking DDL statements (this is considered for the future, by embedding queries in pre-test and post-test operations to maintain the test database, as shown by [6]). Finally, the `TestSQL` architecture does not support control and programming statements.

D. Video material

Parallel to the lecture material and practice workbook, a set of videos was recorded and made available –on a Youtube channel– which emphasises, and delivers asynchronously, key elements and summaries of the material.

The material is carefully selected to emphasise key points, then scripted to make sure that each film is short, focused on a single point, and clear. This keeps each video under 10 minutes. In keeping with the approach to vary the points of view on subject information, the topics addressed are segmented to match the sections in the workbook, and the lectures.

E. Facilitating access to other materials

Since the start of the module, students are encouraged to practice with the use of SQL through the provision of practice material found by different sources on the Internet (eg. similar modules from other UK, European or American Universities) or through the provision of a bibliography which the students can use to access relevant material from specific books.

III. Study Methodology and Data Collection

A. Study hypothesis

Automatic grading of SQL has become common, for example in [3], [7], [6], [4], [8]. It offers multiple advantages: testing the students' skills in a realistic environment; giving rich feedback in the form of debugging information and automated test results; allowing for questions that truly test the students' problem solving abilities.

Yet, considering whether to make `TestSQL` a central element of our assessment strategy, we remained unpersuaded that writing SQL statements, and automatically grading them, captures the students' skills completely enough. Rather, it is our conviction that

> *Many Database Systems skills, including SQL skills, can't be assessed through SQL code alone (e.g. through testSQL)*

This hypothesis forms the basis of our choice of assessment strategy in the semester under scrutiny here. To try to corroborate it, we will use student assessment data.

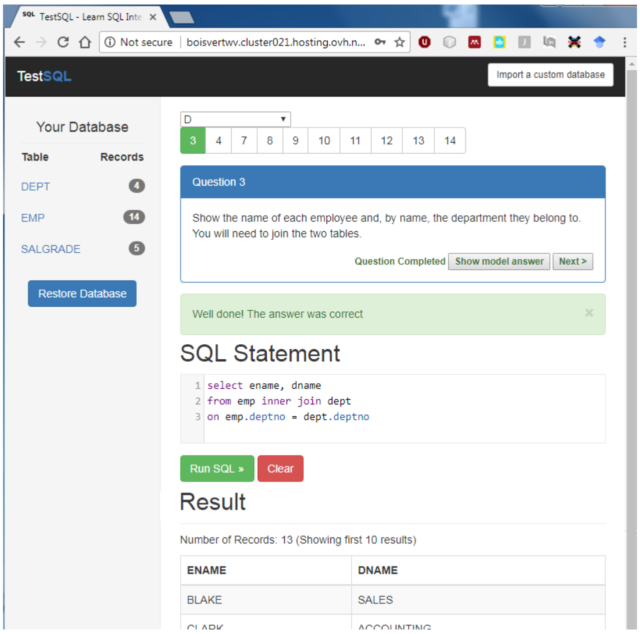

Fig. 2. A `TestSQL` session: automatic feedback gives the student immediate information about their work, and model solutions are also available.

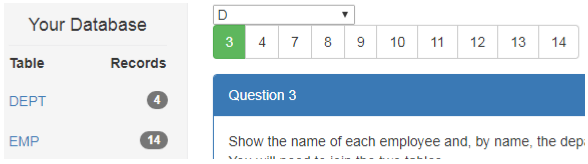

Fig. 3. The `TestSQL` progress bar. Questions are numbered with reference to the workbook contents.

B. Student assessment

Our assessment method in the first semester also uses automatic grading, but rather than SQL queries, relies on pre-written, multiple choice questions. As [9] makes clear, multiple choice questions can, when well written, be used to assess higher order thinking. We also use True-False questions, to assess essential points of knowledge, and multiple answer questions, which also help measure higher order skill and are not open to guesswork. delivered through the University's VLE, Blackboard.

The test comprises of 28 questions taken from a range of subtopics:

- SQL: the majority of questions are about SQL proficiency, requiring to read and interpret SQL statements, or choose statements that answer correctly the question. Some are also indirect questions, e.g.

  What is the difference between having and where?

  which involves SQL, but doesn't require reading or selecting an SQL statement as such.
- Database management: questions on database management include problems of security and access control, as well as optimisation and the workings of a DBMS.
- Physical modelling: as the students work on SQL and practice with database system, this is an opportunity to reify modelling concepts, and we test their ability to relate these concepts to the use of a system, such as the use of constraints or the interpretation of `NULL` values.
- PL/SQL: questions on the language are kept simple, following an introduction in the first semester, but test their understanding of syntax and of common uses of the language.

The test is composed by drawing questions randomly from a series of pools; in all, the pools contain 56 questions, of which each student sees 28.

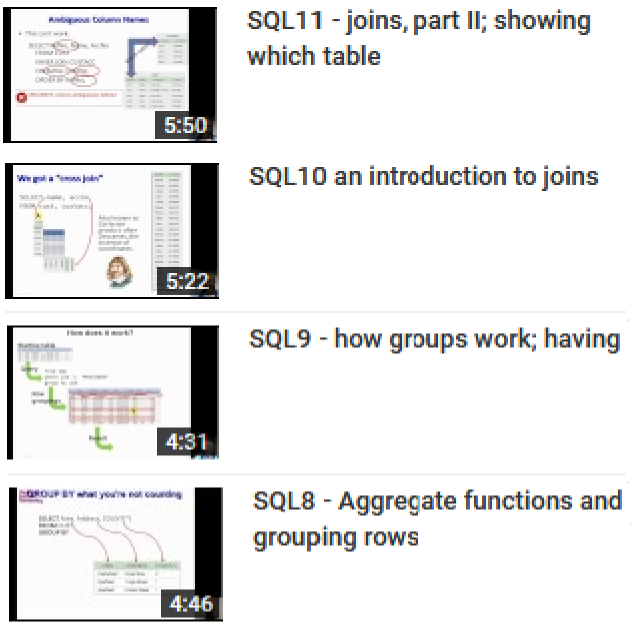

Fig. 4. A subset of the video material. The few minutes' duration of each recording is visible by each thumbnail, resulting in multi-part topics

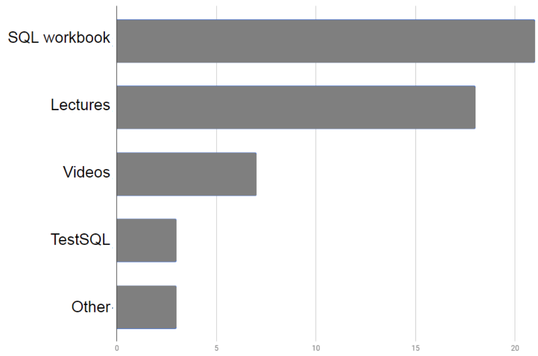

Fig. 5. Number of survey respondents using different types of resources

C. Data collection method

Our data to analyse this work has two sources: the assessment record, and a survey of the students' views.

1) Assessment record: The record of the assessment provides a complete trace of each student, questions asked answer given. This provides a means to compare the performance for each student, but also each question, and according to question type.

The question information also includes a 'discrimination' coefficient, calculated by Blackboard, estimating the question's ability to distinguish between the better and worse students. The company [10] specifies that it is the Pearson's correlation coefficient between the students' overall test scores and their scores for this question.

2) Student survey: After the test was ended, we contacted the students to ask them their views on the resources available. The purpose of the survey was to ask them to confirm:

- Which materials they had used
- Whether they had found each material (those they used) useful
- Whether they had found the materials they used technically correct –that is, whether they were able to access electronic resources, whether any the paper materials free of printing errors, if software running correctly, accessible...
- Specific points pertaining to each type of materials.

To keep the survey anonymous, while retaining the ability to analyse individual test answers and scores in comparison to survey responses, the test records was first anonymised, replacing the student individual IDs with a random number for the data assembled for this research. Each student was then given an individual link to their survey encoded the random number that matched their result data.

IV. Results and Analysis

The multiple choice test was administered in December 2017 and was taken by 103 Students. Of those 103, 25 (24.3%) also answered the voluntary survey.

A. Student use of resources

From the responses, we can observe that not all students used all resources available to them. Some resources, namely the workbook and the lecture material, were consulted by nearly all students, while others attracted much less interest: only 3 respondents (12.5%) used the TestSQL application, as few as chose external resources. Students using external resources indicated publicly available websites —Code School and w3schools— as their material of choice.

The distribution of how many were used, by how many respondents, shows similar disparity. Some students used very few resources (just one or two), while a minority made use of all that was available to them.

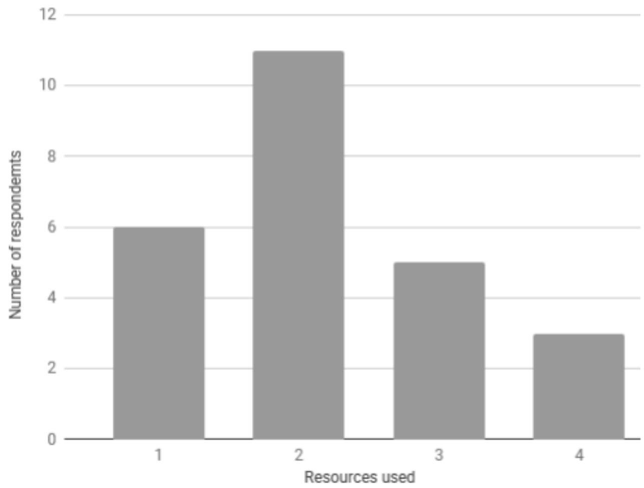

Fig. 6. How many different resources were used by respondents

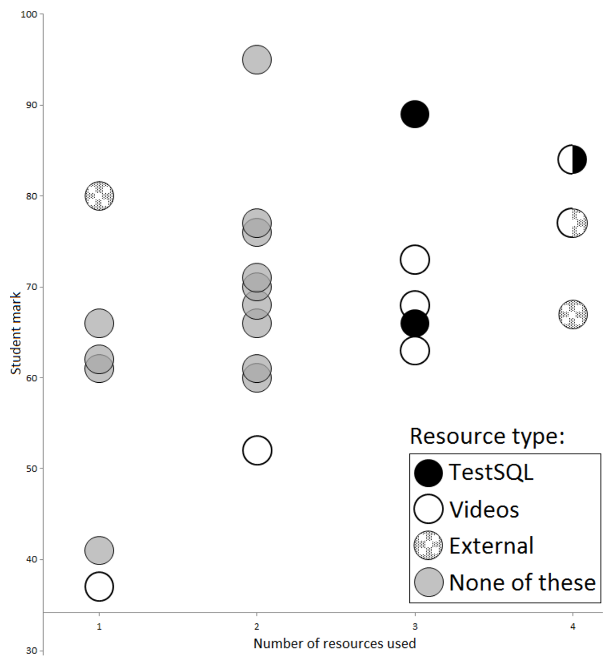

Fig. 7. Use of more resources by students correlates with better marks, but types of resources chosen has little visible effect (based on 24 usable responses)

Note that as the survey is voluntary, the respondents' sample is likely to be biased towards more engaged students. So while three of our 25 respondents (12.5%) have used all resources available, it likely that a smaller proportion of all students have done so.
However, the same number of resources used by respondents can also be used as an proxy to measure their engagement in the module. The use of more of the tools available does correlate to better results. The correlation is significant (P=0.034), although of course, correlation is not causation: this is no indication that engaging with more resources results in better test scores. But this gives an indication of validity of the test: more engaged students do obtain better results.
Fig. 7 illustrates this correlation. The link between the number of resources used by each respondent and their mark is clearly seen.
The same figure shows little effect of the types of resources chosen by our students. Some specific resources coincide with better marks, but they also coincide with accessing more diverse sources of information overall, and so there is no means, from this survey, to pinpoint the value of specific media to

Question type	Database management	PL/SQL	Physical modelling	SQL (easier)	SQL (trickier)	SQL (indirectly)	All but ...
Database management		-0.17	0.02	-0.09	0.03	0.14	-0.04
PL/SQL	-0.17		0.09	0.11	0.15	0.16	0.11
Physical modelling	0.02	0.09		0.16	0.06	0.11	0.14
SQL (easier)	-0.09	0.11	0.16		0.23	0.27	0.20
SQL (trickier)	0.03	0.15	0.06	0.23		0.23	0.23
SQL (indirectly)	0.14	0.16	0.11	0.27	0.23		0.33
All but ...	-0.04	0.11	0.14	0.20	0.23	0.33	

Fig. 8. correlation matrix of question types. Darker cells correspond to higher correlations; the three SQL types of questions are visible on the right.

our students. Rather, we find again, to paraphrase [11] that *how* students use learning resources has a greater impact on effectiveness than *what* those resources show them.

B. Analysis by types of questions

This section analyses the assessment to further understand what gives it validity, and to investigate our hypothesis that we should not assess through SQL code alone. To do this, we separate the assessment questions in several groups, dividing more precisely the earlier (section III-B) set of sub-topics:

- SQL is by far the larger set with 41 questions, but some of the questions do not involve complete statements, and we separate these as *indirect SQL questions*. We also separated those questions directly involving SQL statements in two equal sets: easier and harder SQL questions, according to the percentage of correct answers. These three groups form 13 for easier SQL (the half of the questions which were answered correctly by more students), 12 for harder SQL (answered correctly by fewer), and 16 for indirect SQL questions;
- Database management: 3 questions;
- Physical modelling: 4 questions;
- PL/SQL: 8 questions.

Totalling 56 questions in the whole question bank.

Comparing the question groups, we find that some groups are not at all well correlated to each other, as can be seen fig. 8 (next page).

Scores on Questions about Database Management, PL/SQL and Physical Modelling, are not well correlated to any of the others. This indicates that each of these sub-topics form a distinct group where a good score at questions in one topic is not an indicator of success in another.

The three SQL subtopics show a greater correlation. These correlations are unlikely to be coincidental, as at N=103 subjects, correlations above 0.194 are statistically significant, with a 5% probability that they are the result of random chance.

We also evaluated the score of students at all other questions, for each group under consideration, and this is where the highest correlation is found. The *indirect SQL questions* are very well correlated (0.33) with the scores at all other questions, forming the best predictor of the students' overall scores.

The result is significant with a probability p=0.000664 - less than one in a thousand.

The size of the indirect SQL questions bank could be a factor in that difference. By way of comparison we computed the correlation between marks at all 25 direct SQL questions, compared to all others: it is much lower at 0.204, with a p-value of 0.0039, comparable to the correlations of SQL groups of questions to each other. So the high correlation between marks on indirect SQL questions and marks at others, is not only due to the large number of questions of that type.

This appears to support our initial hypothesis, that some skills can't be assessed through SQL code alone; more precisely, it shows that other questions, which do not exclusively focus on writing SQL clauses, are better predictors of success on a wider range of sub-topics.

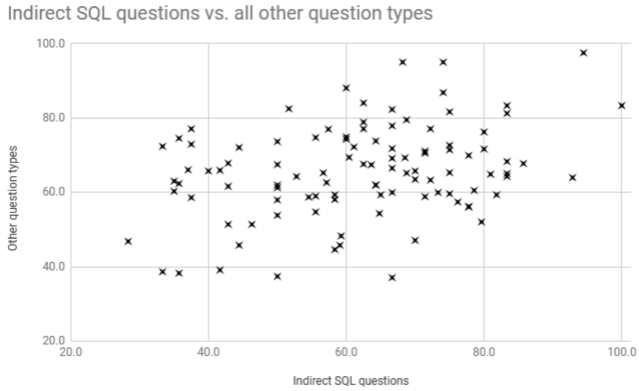

Fig. 9. Scatter plot of scores at indirect SQL questions vs. all other question types

V. Conclusion and Further Work

More research is needed to replicate this result and to identify more precisely what kind of questions form the best summative assessment in SQL.

It could also be that carefully designed SQL problems test the students' higher-order skills and combine the benefits of practical relevance and openness to alternatives of automated SQL testing, with the best of carefully chosen multiple-choice questions.

Yet it appears that as well as solving SQL questions with full SQL clauses, it is helpful to ask alternative questions, questions *about* code rather than questions answered *in* code.

A future development for automatic practice and grading tools reliant on SQL testing like `TestSQL`, maybe to incorporate alternative question styles - multiple choice, true-false, multiple answer for which answers are predetermined.

References

[1] M. Oliver and K. Trigwell, "Can blended learningbe redeemed?" *E-learning and Digital Media*, vol. 2, no. 1, pp. 17–26, 2005.

[2] M. Ling Lo, "Variation theory and the improvement of teaching and learning," 2012.

[3] S. I. Mehta and N. W. Schlecht, "Computerized assessment technique for large classes," *Journal of Engineering Education*, vol. 87, no. 2, pp. 167–172, 1998.

[4] J. R. Prior, "Assesql: an online, browser-based sql skills assessment tool," in *Proceedings of the 2014 conference on Innovation & technology in computer science education*. ACM, 2014, pp. 327–327.

[5] J. License, "testsql: Learn sql the interactive way," in *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, ser. ITiCSE '17. New York, NY, USA: ACM, 2017, pp. 376–376. [Online]. Available: http://doi.acm.org/10.1145/3059009.3072991

[6] A. Abelló, M. E. Rodríguez, T. Urpí, X. Burgués, M. J. Casany, C. Martín, and C. Quer, "Learn-sql: Automatic assessment of sql based on ims qti specification," in *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 592–593.

[7] C. Kleiner, C. Tebbe, and F. Heine, "Automated grading and tutoring of sql statements to improve student learning," in *Proceedings of the 13th Koli Calling International Conference on Computing Education Research*. ACM, 2013, pp. 161–168.

[8] I. Štajduhar and G. Mauša, "Using string similarity metrics for automated grading of sql statements," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*. IEEE, 2015, pp. 1250–1255.

[9] M. E. Piontek, "Best practices for designing and grading exams," *Occasional Paper*, no. 24, 2008.

[10] "Blackboard help: Item analysis," https://help.blackboard.com/Learn/Instructor/Tests_Pools_Surveys/Item_Analysis, accessed: 2018-01-21.

[11] C. D. Hundhausen, S. A. Douglas, and J. T. Stasko, "A meta-study of algorithm visualization effectiveness," *Journal of Visual Languages & Computing*, vol. 13, no. 3, pp. 259–290, 2002.