Improving Database Learning with an Automatic Judge

Enrique Martin-Martin
Adrián Riesco
Manuel Montenegro
Rubén Rubio

July 8, 2022

Universidad Complutense de Madrid

Databases

- Essential in software engineering.
- Taught as basic course in technological studies.

- Learning to write queries in SQL and also programs in the programming language of the DBMS (PL/SQL...).
- Training and practice is required to master these skills.

Online judges

- Web applications offering a collection of exercises.
- Students submit solutions that are automatically checked.
- Immediate feedback.

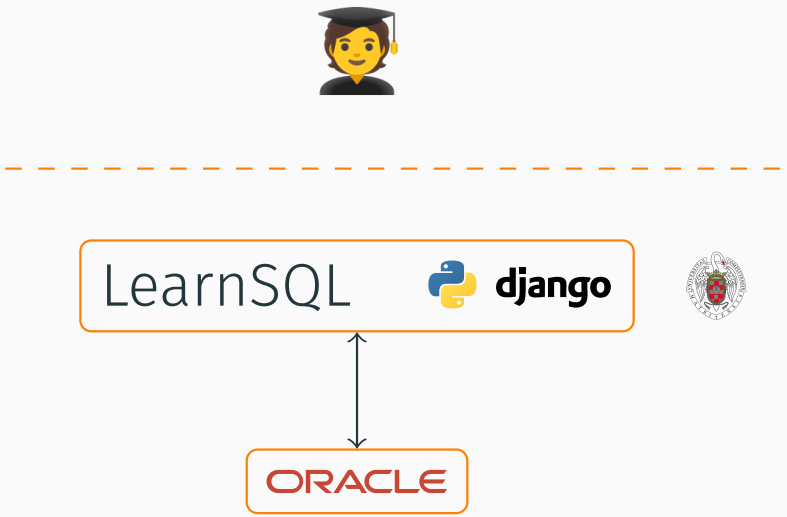
- Widely used in general-purpose programming courses.
- No open source and adaptable online judges for DB.

LearnSQL

LearnSQL: an online judge for database learning

https://github.com/emartinm/lsql

- Open source
- Easy to use
- Learning-centered
- Open to ludification
- Exercises: queries, DML, PL/SQL, triggers...



Select rows

Consider a table defined as follows that contains data about football clubs:

```
CREATE TABLE Club(
    ID CHAR(9) PRIMARY KEY,
    Name VARCHAR(40) NOT NULL UNIQUE,
    Location VARCHAR(30) NOT NULL,
    No_Members INTEGER NOT NULL,
);
```

Write a SQL query that returns all data about clubs with a number of members between **70,000** and **80,000** (both included). The schema of the result should be the following:

```
(ID, Name, Location, No_Members)
```

Example problem II

Database

Download script

CLUB

ID	NAME	LOCATION	NO_MEMBERS
11111111X	Real Madrid CF	Concha Espina	70000
11111112X	Futbol Club Barcelona	Aristides Maillol	80000
11111113X	Paris Saint-Germain Football Club	Rue du Commandant Guilbaud	1000

Expected result

ID	NAME	LOCATION	NO_MEMBERS
11111111X	Real Madrid CF	Concha Espina	70000
11111112X	Futbol Club Barcelona	Aristides Maillol	80000

Detailed feedback

```
1 select *
2 from Club
3 where Num_Socios < 75000
```

Send solution

Feedback

There are some rows that are wrong. All rows are shown below, highlighting those that contain incorrect values in any column or that should not appear.

CIF	NOMBRE	SEDE	NUM_SOCIOS
11111111X	Real Madrid CF	Concha Espina	70000
11111113X	Paris Saint-Germain Football Club	Rue du Commandant Guilbaud	1000

Evaluation

Course context

- Databases, 2nd year, CS, UCM, Madrid, Spain.
- Academic year 2021-2022, first term.

- Syllabus: relational model, SQL queries, procedural SQL, triggers, transactions.

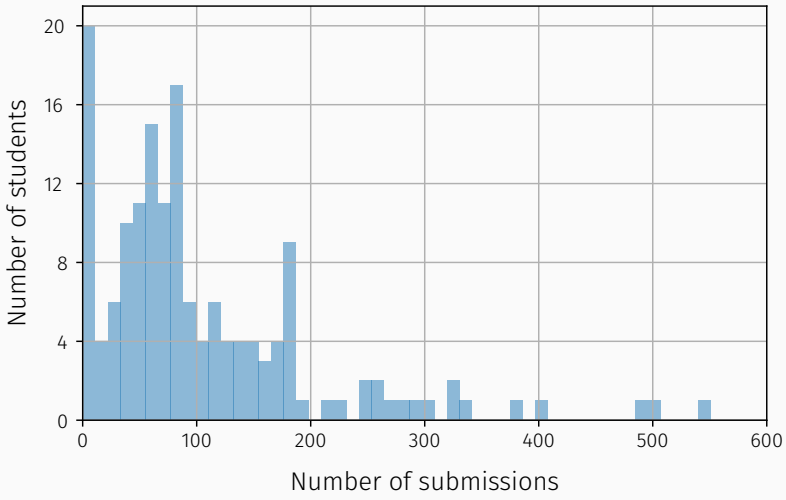
- 30 in-class lectures, 100 minutes each, 50% practical in lab.
- 70% of the grade is the final handwritten exam.

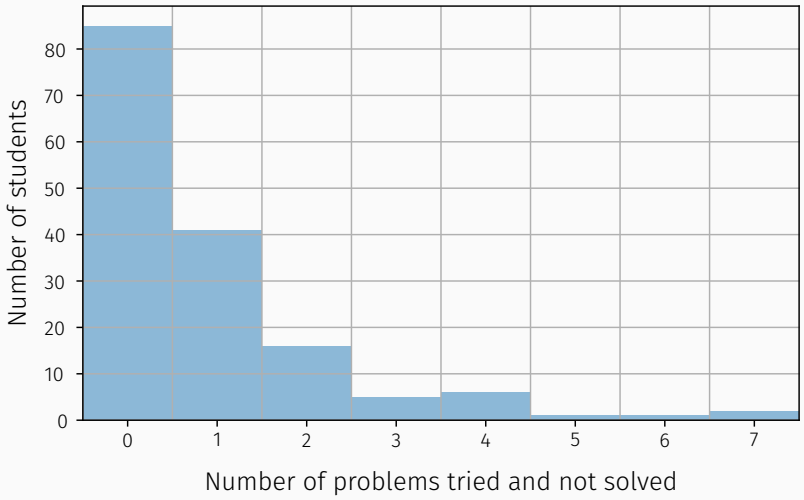
Evaluation design

- 3/6 groups.
- 157 students (130 attended the final exam).

- Submission logs.
- Marks by exercise of the final exam.

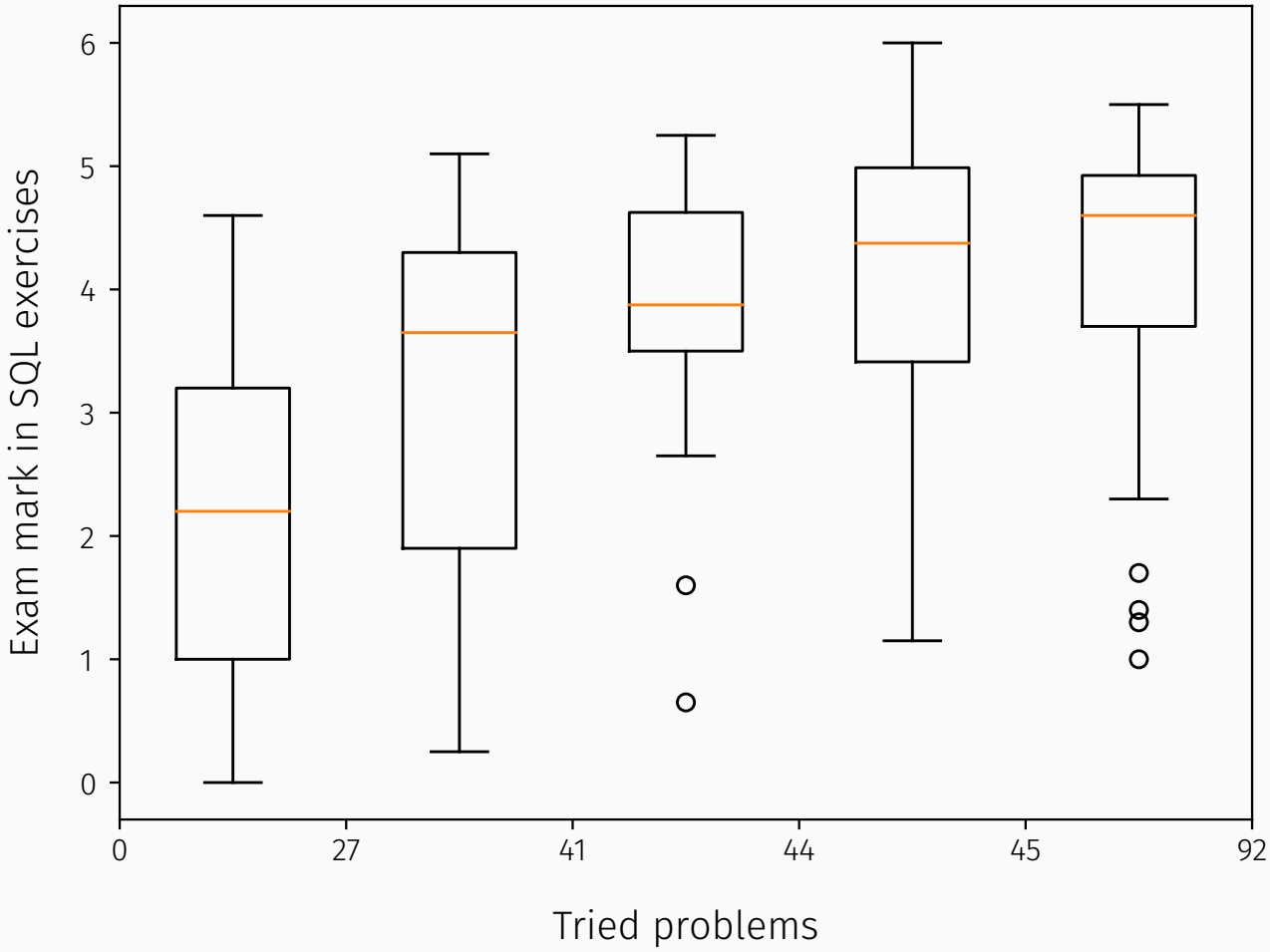
Evaluation results





- Students tried $\sim$ 40 exercises each.
- Problems solved in $\mu = 2.57$ attempts ($\sigma = 3.88$).
- Engagement: 85.91% of submissions outside class hours.

Evaluation results



Significant correlation between tried problems and marks in SQL exam exercises (0.497, $p = 1.8 \cdot 10^{-9}$).

Conclusions

Conclusions

- LearnSQL is a learning-oriented open source online judge.
- Evaluated in a introductory course on databases.
- There is statistical evidence that using the judge improves the final score obtained by the students.

Future work

1. Improvements on the tool (other DBMS, NoSQL).
2. Integrating the judge into Moodle.
3. Repeating the evaluation next year with a control group.

Thank you!