Coverage of Course Topics in Learnersourced SQL Exercises

Nea Pirttinen
University of Helsinki
Helsinki, Finland
nea.pirttinen@helsinki.fi

Arto Hellas
Aalto University
Espoo, Finland
arto.hellas@aalto.fi

Juho Leinonen
Aalto University
Espoo, Finland
juho.2.leinonen@aalto.fi

CCS CONCEPTS

• **Social and professional topics** → *Computing education*; • **Applied computing** → *Interactive learning environments*; • **Information systems** → *Crowdsourcing*.

KEYWORDS

learnersourcing, crowdsourcing, SQL, databases, topic coverage

1 INTRODUCTION

Traditionally, learnersourcing has been used, for example, for creating multiple-choice questions (MCQs) [1, 4, 6, 9]. In computing classrooms, learnersourcing is increasingly used for the creation of programming exercises [3, 7]. In our recent prior work, we introduced a system for crowdsourcing SQL exercises [5], which to the best of our knowledge is the first such system.

One concrete goal of learnersourcing systems is to create a vast repository of exercises that could complement or even replace instructor-created exercises. An important question related to this is whether student-created exercises cover all the course topics, which would be necessary to allow students to practice all learning objectives of the course. Prior work has studied the coverage of course topics of learnersourced MCQs and found that student-created exercises did cover all course topics [2].

In this work-in-progress paper, we report preliminary results related to how well learnersourced SQL exercises cover different SQL concepts. Our goal is to examine whether – by using mostly learnersourced exercises – students will get to practice the concepts they are expected to learn during the course.

2 TOOL AND CONTEXT

2.1 SQL Trainer

SQL Trainer is a learnersourcing system for practicing SQL queries [5]. When using the system, a student is presented with a list of topics in the order of appearance in the course material. Students can choose to either create exercises or practice a specific topic. If the student chooses to create an exercise for a particular topic, the system will ask which database the student wishes to use (selection is made from a pre-defined list of databases created by the course instructor). The student is then required to give a name for the exercise, an exercise description, and a sample solution for that exercise. Once the exercise is completed, it is added to the pool of exercises for that topic. If the student chooses to practice a topic, they will be given a randomly selected exercise that they have not completed yet from the pool of created exercises.

A more detailed description of SQL Trainer and its usage, as well as the description of the teacher view, is presented in [5].

2.2 Context and Research Questions

For the present study, we use data collected from three introductory database courses offered by the University of Helsinki, where SQL trainer was used to support SQL practice. In the courses, SQL trainer had 11 topics (briefly outlined in Table 1). Approximately 10% of the course grade was based on completing at least four exercises per course topic and creating at least a single exercise per course topic. Students were free to complete and create more exercises than required for the points, however.

Our research questions are as follows:

RQ1. Which topics do students create exercises for?
RQ2. How well do the exercises created by students cover the course topics?

To answer research question 1 and 2, we partially replicate earlier work by Denny et al. [2] and Purchase et al. [8] by studying to what extent the exercises that students complete and create cover the course topics. In particular, we study how many exercises are created by students per course topic, and how the used SQL concepts match the topics.

3 RESULTS AND DISCUSSION

Table 1: Topics

#	Topic
1	Select from a table
2	Filter and order data from a single table
3	Select from multiple tables
4	Select from even more tables
5	Other types of joins
6	Adding and removing tables
7	Adding data to a database
8	Using update and delete
9	Using functions
10	Using group by and functions
11	Functions, group by, having, order

3.1 Created Exercises

In total, 1569 students entered the system. Out of these, a total of 1187 created at least one exercise of their own. In total, students created 11247 exercises for 11 different instructor-defined topics. The number of created exercises for each topic can be found in Table 2 (topic number references listed in Table 1).

From Table 2, we can see that there is a clear downwards trend in the number of created exercises per topic. One possible reason for this is that the latter topics come later in the course and thus fewer students actively participate at that point in the course. Another

Table 2: The number of exercises created per topic, 11247 in total. Arranged in the order the topics appear in the course material.

Topic	Exercises created
1	1301 (11.6 %)
2	1205 (10.7 %)
3	1113 (9.9 %)
4	1048 (9.3 %)
5	666 (5.9 %)
6	1063 (9.5 %)
7	1047 (9.3 %)
8	1037 (9.2 %)
9	1028 (9.1 %)
10	948 (8.4 %)
11	879 (7.8 %)

potential explanation for the trend is that the latter topics are related to somewhat more difficult topics and thus students might not be as inclined to create exercises for those. One clear outlier is topic 5, which was related to "other types of joins". Why this topic elicited fewer learnersourced exercises requires further research.

3.2 Topic Coverage

	1	2	3	4	5	6	7	8	9	10	11
SELECT	1282	1197	1107	1043	643	2	3	8	1014	945	875
FROM	1282	1195	1106	1043	644	4	3	171	1015	926	871
ORDER BY	30	530	212	204	209	0	0	2	4	106	443
ON	4	13	1070	1046	626	0	0	6	59	796	766
JOIN	4	11	1061	1024	102	0	0	5	55	509	672
LEFT JOIN	0	2	7	2	476	0	0	0	5	275	82
CREATE	14	4	1	1	7	690	21	2	1	1	1
TABLE	13	4	2	1	9	1047	24	6	1	1	1
DROP	0	0	1	0	2	355	1	2	0	0	0
INSERT	4	0	0	3	7	8	1013	24	0	0	0
INTO	3	0	0	3	7	8	1017	23	0	0	0
UPDATE	0	1	0	0	5	1	4	831	2	0	0
SET	0	1	0	0	5	1	4	829	2	0	0
DELETE	0	0	0	0	0	3	0	163	0	0	0
GROUP BY	0	0	1	3	3	0	1	1	10	888	812
HAVING	0	0	0	0	0	0	0	0	0	14	514
DISTINCT	0	1	49	161	122	0	0	1	3	7	15

Figure 1: The number of SQL concepts used per topic. Color legend: concept used in over 80 % of the created exercises in dark blue, over 40 % light blue, over 10 % yellow, under 10 % gray, and 0 occurrences white.

Figure 1 presents the most common SQL concepts for each topic (topic number references listed in Table 1). Only concepts that were relatively common (occurred in over 10% of created exercises in at least a single topic) are included. In the vast majority of cases, students are able to identify relevant concepts for the topic, and utilize them when creating an exercise.

Interestingly, based on the SQL concepts used in topic 5, students seem to majorly prefer creating exercises related to left joins compared to right or inner joins (which occurred rarely enough that they were not included in the figure). Similarly, for topic 6, we can see that students are more inclined to create exercises related to creating tables than dropping tables.

3.3 Future Work

Altogether, these preliminary results support the use of learnersourcing for creating a large exercise pool with a good coverage of course topics. This provides the opportunity for students to practice a wide variety of SQL concepts using the learnersourced exercises. In our future work, we are interested in extending this research to similarly examine how much students practice the different concepts using the system. In addition, we are keen to study whether the exercises created by students cover varying levels of difficulty and if the exercises are effective. Lastly, we are exploring whether there are differences between demographic groups in to what extent they participate in learnersourcing activities.

REFERENCES

[1] Paul Denny, John Hamer, Andrew Luxton-Reilly, and Helen Purchase. 2008. PeerWise: students sharing their multiple choice questions. In *Proceedings of the fourth international workshop on computing education research.* 51–58.

[2] Paul Denny, Andrew Luxton-Reilly, John Hamer, and Helen Purchase. 2009. Coverage of course topics in a student generated MCQ repository. *ACM SIGCSE Bulletin* 41, 3 (2009), 11–15.

[3] Paul Denny, Andrew Luxton-Reilly, Ewan Tempero, and Jacob Hendrickx. 2011. Codewrite: supporting student-driven practice of java. In *Proceedings of the 42nd ACM technical symposium on Computer science education.* 471–476.

[4] Hassan Khosravi, Kirsty Kitto, and Joseph Jay Williams. 2019. RiPPLE: A Crowdsourced Adaptive Platform for Recommendation of Learning Activities. *Journal of Learning Analytics* 6, 3 (2019), 91–105.

[5] Juho Leinonen, Nea Pirttinen, and Arto Hellas. 2020. Crowdsourcing Content Creation for SQL Practice. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education.* 349–355.

[6] Steven Moore, Huy Anh Nguyen, and John Stamper. 2021. Examining the Effects of Student Participation and Performance on the Quality of Learnersourcing Multiple-Choice Questions. In *Proceedings of the Eighth ACM Conference on Learning@ Scale.* 209–220.

[7] Nea Pirttinen, Vilma Kangas, Irene Nikkarinen, Henrik Nygren, Juho Leinonen, and Arto Hellas. 2018. Crowdsourcing programming assignments with CrowdSorcerer. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education.* 326–331.

[8] Helen Purchase, John Hamer, Paul Denny, and Andrew Luxton-Reilly. 2010. The Quality of a PeerWise MCQ Repository. In *Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103* (Brisbane, Australia) *(ACE '10).* Australian Computer Society, Inc., AUS, 137–146.

[9] Anjali Singh, Christopher Brooks, Yiwen Lin, and Warren Li. 2021. What's In It for the Learners? Evidence from a Randomized Field Experiment on Learnersourcing Questions in a MOOC. In *Proceedings of the Eighth ACM Conference on Learning@ Scale.* 221–233.