

# The Effects of Database Complexity on SQL Query Formulation (journal-first)

Toni Taipalus  
Faculty of Information Technology  
University of Jyväskylä  
Jyväskylä, Finland  
<https://orcid.org/0000-0003-4060-3431>

**Abstract**—The learning of practical Structured Query Language (SQL) skills often takes place in digital environments, where the learner writes queries against an exercise database. The exercise database is usually designed and implemented by the teacher, and populated with makeshift data. Although this approach is common, and SQL taught in almost all database courses, little scientific attention has been given to the nature of the exercise database.

**Index Terms**—Structured Query Language (SQL), database, database complexity, education, student learning

## I. OVERVIEW

**Motivation.** There is growing support for using more natural environments – i.e., environments that more accurately reflect future work – in teaching SQL. In terms of SQL education, a more natural learning environment may entail ambiguously phrased data demands, anomalous data, students not knowing what data the correct query should return, and more complex database structures and data [1]. However, the effects of more complex database structures on SQL query formulation have not been studied.

**Approach.** By utilizing a previously reported query concept framework [2], and an error categorization [2], [3], we formulated three databases of varying logical complexity with as similar SQL exercises as possible. We analyzed SQL queries from 744 students using the error categorization, and marked whether a student was able to complete an exercise. In case a student failed, we analyzed what types of errors halted the query formulation process.

**Results.** The results indicate – with a statistically significant effect – that it is more difficult for students to successfully write SQL queries against a more complex database structure (Fig. 1). Furthermore, we observed that the number of syntax, semantic, or logical errors committed did not increase with database complexity. Unnecessary complications in queries, however, increased, implying that students are perhaps less able to refactor their queries when the database structure is complex.

**Implications.** Based on our results, we caution educators to also consider the negative effects of more complex exercise databases. Even though more natural environments have been shown to increase student engagement [4], a more natural exercise database is not necessarily more complex. Finally, if a teaching setting requires the use of more complex databases,

students should be steered towards planning queries before writing by, for example, visualization [5].

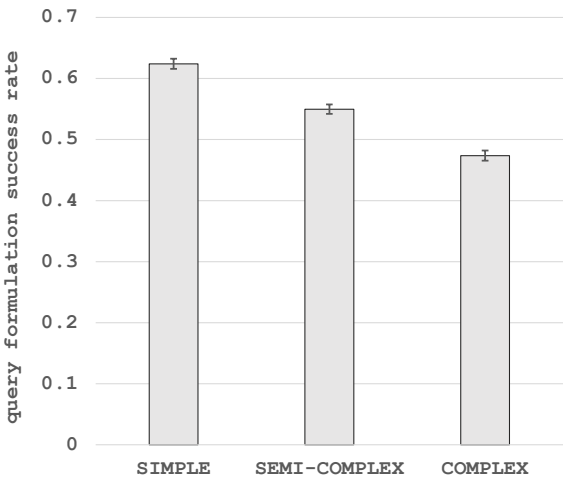


Fig. 1. Success rates (mean ± SEM) for each database [6]

## II. DISCLAIMER

This is a journal-first presentation of the paper “The Effects of Database Complexity on SQL Query Formulation”, published in the Journal of Systems and Software [6].

## REFERENCES

- [1] P. J. Wagner, E. Shoop, and J. V. Carlis, “Using scientific data to teach a database systems course,” in *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education*, ser. SIGCSE ’03. New York, NY, USA: ACM, 2003, pp. 224–228.
- [2] T. Taipalus, M. Siponen, and T. Vartiainen, “Errors and complications in SQL query formulation,” *ACM Transactions on Computing Education*, vol. 18, no. 3, pp. 15:1–15:29, Aug. 2018.
- [3] T. Taipalus and P. Perälä, “What to expect and what to focus on in SQL query teaching,” in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE ’19. New York, NY, USA: ACM, February–March 2019, pp. 198–203.
- [4] K.-B. Yue, “Using a semi-realistic database to support a database course,” *Journal of Information Systems Education*, vol. 24, no. 4, pp. 327–336, 2013. [Online]. Available: <http://jise.org/Volume24/n4/JISEv24n4p327.html>
- [5] T. Taipalus, “Teaching Tip: A Notation for Planning SQL Queries,” *Journal of Information Systems Education*, vol. 30, no. 3, pp. 160–166, 2019. [Online]. Available: <http://jise.org/Volume30/n3/JISEv30n3p160.pdf>
- [6] —, “The effects of database complexity on SQL query formulation,” *Journal of Systems and Software*, vol. 165, p. 110576, 2020.