

Identifying SQL Misconceptions to Improve Database Education

Daphne Miedema¹
Eindhoven University of Technology

Efthimia Aivaloglou
Leiden University, Open Universiteit

George Fletcher
Eindhoven University of Technology

Abstract

SQL is the most commonly taught database query language. While previous research has investigated the errors made by novices during SQL query formulation [1, 2, 4, 5, 6], the underlying causes for these errors have remained largely unexplored. Understanding the basic misconceptions held by novices that lead to these errors, can give us pointers on how to improve SQL education.

In a recent paper, we identified the misconceptions that might be the causes of documented SQL errors that novices make [3]. We gathered information on the thinking process of university students solving query formulation problems through think-aloud experiments. With the queries in hand, we analyzed the underlying causes for the errors made by our participants.

Our findings include four top-level SQL misconception categories: misconceptions based in previous course knowledge, generalization-based misconceptions, language-based misconceptions, and misconceptions due to an incomplete or incorrect mental model. Further research on these misconceptions, and the design of interventions, can lead to more efficient SQL education.

References

- [1] A. Ahadi, V. Behbood, A. Vihavainen, J. Prior, and R. Lister. Students' Syntactic Mistakes in Writing Seven Different Types of SQL Queries and its Application to Predicting Students' Success. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 401–406, 2016.
- [2] A. Ahadi, J. Prior, V. Behbood, and R. Lister. Students semantic mistakes in writing seven different types of SQL queries. In *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, pages 272–277, 2016.
- [3] D. Miedema, E. Aivaloglou, and G. Fletcher. Identifying SQL Misconceptions of Novices: Findings from a Think-Aloud Study. In *ICER 2021 - Proceedings of the 14th ACM Conference on International Computing Education Research*, 2021.
- [4] A. Migler and A. Dekhtyar. Mapping the SQL learning process in introductory database courses. In *SIGCSE 2020 - Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 619–625, 2020.
- [5] P. Reisner. Human Factors Studies of Database Query Languages: A Survey and Assessment. *ACM Computing Surveys*, 13(1):13–31, jan 1981.
- [6] T. Taipalus and M. Siponen. Errors and Complications in SQL Query Formulation. *ACM Transactions on Computing Education*, 18(3), 2018.

¹Contact author, d.e.miedema@tue.nl