SQL Pattern Design and Development

Huda Al-Shuaily
Department of Information Technology
Higher College of Technology
Muscat, Oman
huda.alshuaily@hct.edu.om

Karen Renaud
School of Computing Science
University of Glasgow
Glasgow, UK
karen.renaud@glasgow.ac.uk

Abstract—**This paper presents the processes involved in the design and development of a set of Structure Query Language (SQL) patterns. The intention is to support novices during SQL acquisition. This process is grounded in the SQL learning model developed from learning theory literature and empirical investigations into SQL acquisition. One of the crucial cross-cutting factors identified during the development of this model was the quality of the instructional material provided to learners to support the acquisition process. Since patterns have been successfully deployed in other areas to support knowledge transfer, we set out to develop SQL patterns to meet this need for effective instructional material. We detail the process by which we identified the required components of SQL patterns. Our patterns were also informed based on observations of how novices and experts solved SQL queries. We conclude by presenting our proposed SQL pattern content and structure.**

Keywords-Pattern; SQL; Expert

I. INTRODUCTION

SQL is one of those languages that seem particularly challenging to master. We previously reported on an investigation which advanced explanations for this apparent difficulty [1] and developed a model of SQL learning shown in Fig.1. Our investigation uncovered four cross-cutting issues that impacted the learning process. One major factor was the quality and suitability of the instructional material to support the learning process. In providing instructional material, two specific aspects are important:

- The structuring of knowledge within the instructional material. Merrill [2] explains that the organization and representation of knowledge impacts learning. Mayer [3] makes the same argument, positing that the way in which a body of knowledge is structured determines how readily it will be grasped by learners.
- When, during the learning process, the afore-mentioned instructional material should be introduced. Learning happens in a predictable and mediated way, with subsequent knowledge and skills building on prior knowledge and understanding [4], [5]. Hence, the sequence in which we present knowledge is important. It should support learning rather than encouraging trial-and-error attempts to produce correct SQL queries without understanding the underlying principles [6].

Patterns are a widely used mechanism for supporting knowledge transfer. We set out to investigate whether patterns could meet the need for optimally-structured instructional material in this context. Schlager and Ogden [7] found that incorporating a cognitive model in the form of expert user and product-independent knowledge into novice instruction enhances learning. They concluded that such a cognitive model framework could potentially help to support knowledge acquisition.

Patterns traditionally structure knowledge in such a way that they can transfer best practice from experts to novices. We cannot assume that SQL patterns can be structured in the exactly the same way as other more well-established patterns, however, so we need to carefully align them with the SQL acquisition process.

We briefly present our previously developed model of SQL learning (Fig. 1), which is the logical place to start when identifying SQL patterns and positioning them within the learning process. This model is grounded in Bloom's taxonomy [4] and validated by studies of how novices learn to write SQL queries. The SQL learning model emerged from the analysis of the educational literature, and was augmented by the analysis of data gathered during qualitative and quantitative studies of SQL acquisition.

This model incorporates the notion of the development of mental models. We demonstrate how mental models are constructed during SQL acquisition. Learners start with the development of individual schemata, moving on towards meaningful structuring of schemata into hierarchies and constructed mental models. Existence of these models suggests that the learner will be able to solve a variety of problems of similar nature, i.e., they have abstracted. An abstraction exists and they should be able to apply core concepts in many contexts: learning has resulted in a heuristic.

The SQL acquisition process is also modeled in the diagram, showing that learners need first to have a basic knowledge of SQL concepts, and an understanding of how to use them. They then have to practice applying these concepts to a variety of problems: analyzing, synthesizing and evaluating. They ought to emerge from this stage with an appreciation of the core principles; with an ability to make judgments about strategies to be deployed. Learners who have progressed to this upper level can be considered to have mastered SQL.

	Development of Mental Models		SQL Learning Taxonomy		Cross Cutting Factors			
Learning	Meta Mental Model	Abstract Model Develop-ed	Creating	*Query modification, reflecting, making judgements*	Characteristics of Learners	SQL Characteristics	SQL-Specific Cognition	Instructional Material
	Structuring & Confirming Schema	Context Specific Applica-tion	Analysis Synthesis Application Evaluation	*Problem solving: analysis, synthesis, SQL writing, checking, SQL debugging*				
	Schemata	Building Chunks Of Know-ledge	Compre-hending	*SQL Query Reading, Interpreting, Explaining*				
			Remem-bering	*SQL Concept: Recognizing, Recalling*				

Figure 1. A Model of SQL Learning

SQL patterns' identification process is explored in Section II. The processes that involve SQL pattern identification using text mining are explored in Section III. The SQL pattern identification phase using novices' observation is presented in Section IV while expert observation is considered in Section V. Section VI discusses and Section VII concludes.

II. SQL Pattern Identification

To identify our SQL patterns, we commenced by studying other pattern identification methods and procedures. Patterns are not an optimistic or ephemeral collection of ideas. They encapsulate specific tried-and-tested best practice techniques specific to a particular field [8]. Patterns do not state obvious solutions to trivial problems nor do they cover every possible eventuality, but they do capture important "big ideas" [9]. A pattern should explain *how* a problem should be solved and *why* the presented solution is appropriate in a particular context. Alexander [8] points out that patterns may be discovered by identifying a problem and later finding a solution or by seeing a positive set of examples and abstracting a common solution. Coad and Mayfield [10] suggest that patterns are based on a designer's experience of the area.

The SQL pattern development process needed to focus on both the behavioral and the cognitive aspects of SQL acquisition. Understanding learner ability to perform different cognitive tasks such as query formulation, translation, and writing is essential to be able to design this new SQL instructional material.

The SQL patterns we derived emerged from an iterative research process, which involved a review of educational research, uncovering relevant human factors related to SQL usability as well as psychology-related research. The process aimed to accommodate the nature of SQL acquisition. We explored this process by conducting a general overview of the literature about educational theory and cognitive psychology research [4], [5], [11] and instructional design research. The next step narrowed to cover Computer Science (CS) educational research [12] and focused on fostering of problem solving skills.

Having confirmed the possibility of using patterns to support SQL acquisition, we proceeded to identify and define the patterns using text mining followed by observation of novices and experts.

III. Pattern Mining

According to Bruner [15] new instructional methods need first to apply what educators know about how students learn, remember, and use related skills. A text mining procedure was therefore used to extract this information from texts, to discover patterns from existing knowledge repositories, solutions, or designs. This process captures practice that is both good and significant [14].

This identified common knowledge arguably represents the core concepts and practices of SQL query writing. There is a strong precedent for this approach [13].

The mining process was carried out manually based on the dimensions defined in the SQL Learning Model (Fig 1), and is depicted in Fig. 2.

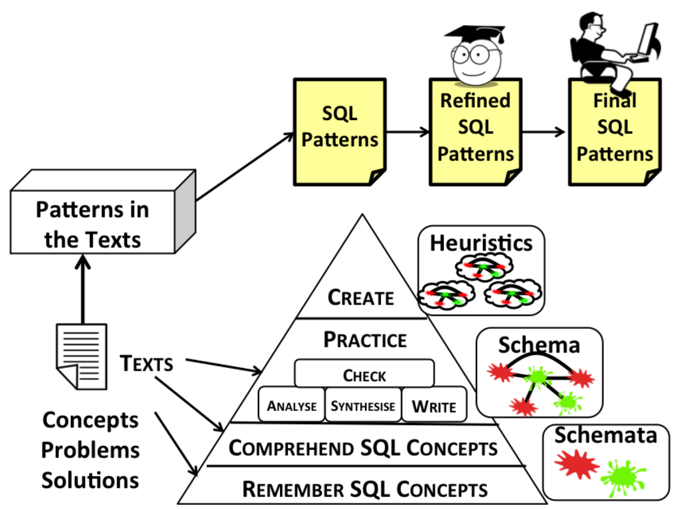


Figure 2. Mining Process

The following steps were followed during the mining process:

1. Identifying Data:
    a. Identify expected SQL knowledge from database texts and categorize the knowledge into the SQL learning model categories.
    b. Identify the declarative or "remembering" SQL concepts. Here, we mined data such as SQL facts or concepts. For example joining, aggregation, sub-query.

 ISBN: 978-1-61208-343-8

2. Identifying Information:
    a. Identify the procedural or "comprehension" SQL concepts, and how they are used in a particular context.
    b. Identify the "practice" skills by considering how concepts should be applied in solving problems. For example, show a context scenario and explain how the relevant syntax and rules are applied. Also illustrate the scenario with appropriate examples which show, step-by-step, how such a concept should be applied.
3. Identifying Knowledge:
    a. Investigate the "creating" activity. For example, find evidence of generic principles being applied in particular contexts.
4. Identify the SQL misconceptions that could potentially be corrected by the patterns.

The mining process was informed by knowledge management research that distinguishes between data, information and knowledge [16].

The process of knowledge extraction and categorization led to an initial set of patterns. The mining process provided a starting point, delivering a static understanding of how SQL core knowledge is presented in textbooks and commonly-used texts. How such SQL concepts are applied in reality, by query writers, could not be gauged without a field investigation. The next step was to observe and analyze novice SQL problem-solving behavior.

IV. Observing Novices

Researchers in the field of pattern identification agree on the value of direct observation in arriving at patterns. "*In order to discover patterns which are alive we must always start with observation*" [8] (p.254). Furthermore, [17] points out that "*Patterns are not created or invented; they are identified via an invariant principle*".

Strategy identification, by means of learner observation, helps determine what gaps "best practice" SQL patterns should fill. Cognitive science suggests giving learners a problem and observing everything they do and say while attempting a solution. Unstructured observations were thus conducted over a period of two semesters.

The focus of the observation was on the following particular aspects:

- Remembering:
    - When they remembered the required concepts, were they correct?
- Searching (Not Remembering):
    - How were the required, but forgotten, concepts obtained? For example, did they refer to textbooks or teaching materials, or did they search the Web to find similar problems and related solutions.
- Problem Solving:
    - Were the required concepts identified correctly?
    - Were the gathered concepts correctly matched to the given problem context?
    - Did they search for examples on the Web?
    - Did they try different possible solutions? If so, why was a particular solution selected?
    - How did they react to errors?

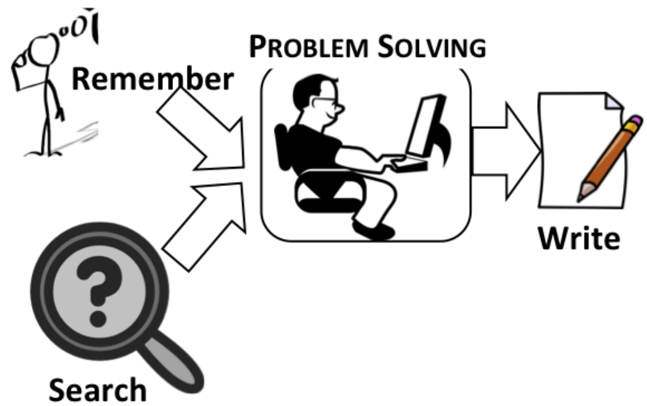


Figure 3. Novice Observation

The observation data, based on observation of 63 students (see Table I), revealed that many students lacked problem solving skills. Students often started to write SQL queries without taking the time to consider different approaches. There was no attempt to choose an optimal approach from a number of candidate approaches. They behaved tactically and did not take time to analyze the problem description and to consider what they should do before attempting to write the query. This tendency confirms previous research findings [18]. Students spent the bulk of their time solving syntax and semantic errors and assessing the correctness of the generated results.

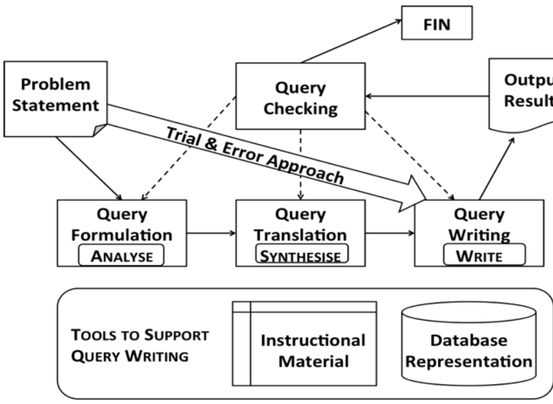
TABLE I: STUDENTS' DEMOGRAPHICS INFORMATION

Time	Participants	Participants
2009/10	Students registered for 2nd year course	17
2010/11	Students registered for 2nd year course	21
2010/11	Students registered for third year course	15

Furthermore, novices lacked the ability to sub-divide the problems into sub-problems or to identify the specific knowledge required to solve individual sub-problems [19]. If they *do* divide and conquer, they then have to synthesize identified sub-solutions to design a complete solution to the problem. This, too, seems to be a skill that novices lack (Fig. 4).

 ISBN: 978-1-61208-343-8

Less searching behavior than anticipated was observed and when it did take place it was often unproductive. Students searched for similar problems on the Web or spent some time looking at the lecture notes, trying to understand different concepts. This was often unproductive since they wasted time searching for irrelevant concepts.



Analogy	Sources of analogy
A1→ Formulating the problem "number of loan …"	Schemata
D1→ Dividing the problem into sub problem or sub goals	Advanced knowledge
A2:→ Analysis : Joining two tables "joining the copy with titles"	Basic knowledge
R1→Reasons why *D1* " to get single table"	Basic knowledge
A3→ writing: Write SQL syntax	Basic knowledge
D2→ use subquery	Advanced knowledge
A4→Evaluation: Execute the *A3* without applying *D2* and checking the results	Basic knowledge
D5→to apply Self join for the table	Advanced knowledge
A12→ Writing: Modifying the query	Basic knowledge
A13→Evaluation: Modifying the query with no clear decision	Schemata
A14→Writing Applying D5	Schemata
A15: Analysis of the problem	Schemata
Applying aggregation	Advanced knowledge
A16: Writing: Iteration of changing the query	Advanced knowledge

Figure 4. Problem Solving Stages, and Novice Strategies

Observation of novices was invaluable in understanding how to design supporting instructional material. However, we needed also to understand which particular strategies were deployed by SQL experts since this was the behavior we wanted to guide the novices towards. What emerged from this analysis was the fact that an intervention was required to support students during problem solving, where they apply the basic SQL concepts and principles.

The initial set of SQL patterns were refined based on the observation process.

V. EXPERT OBSERVATION

Professionals have acquired knowledge and skills through study and practice over the years and are termed "expert". Patterns are means of codifying experts' knowledge and expertise to facilitate knowledge transfer. Pattern content must be informed by experts' actual practices. This section presents a description of problem solving strategies deployed by two individual expert SQL query writers. The experts were master students at Glasgow University who had experience using SQL. The tasks they solved are shown in Fig. 5. The aim of this observation was to determine how experts solved these problems, as opposed to novices.

Q1: Give the titles of books that have more than one author. Q2: Display the names of borrowers who have never returned a book late

Figure 5. Expert Observation task

The cognitive activities performed during problem solving of two tasks were recorded by employing a "talk-aloud" protocol [20].

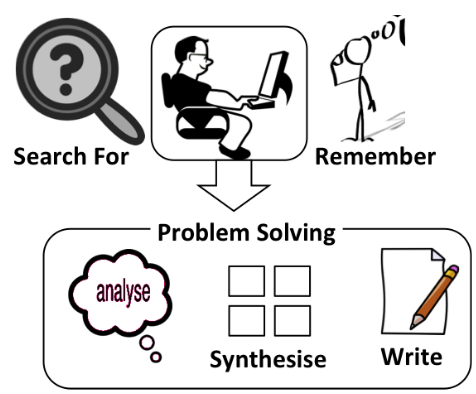


Figure 6. Expert Observation process

The observation process recorded all cognitive activities (see Fig. 7), such as schemata retrieval (Remembering) and Searching (Not Remembered). The collected information was categorized as conceptual (basic building blocks from Fig. 1), schemata (knowledge of how concepts are used), or rule (abstract heuristic knowledge) as recommended by [7], [15], [21].

Experts, after reading the problem description, made an initial decision about the type of technique that had to be applied. They then looked at the provided data model and verbally listed the possible approaches to solving the problem that they could deploy. After mulling it over, they settled on one particular approach and provided reasons for discarding the other options. Both experts used a divide-and-conquer approach and sub-divided the problem: they did not attempt to write the whole query at once. They wrote and tested the commands related to the sub-queries and then synthesized the sub-solutions to arrive at the final complete solution.

The most interesting part of this observation was the fact that the experts applied an implicit pattern matching approach to their assessment of the problem. They clearly tried to match a number of different learned heuristics to the problem before settling on the best approach. One can only assume that they had internalized a number of abstract heuristics which they tried to match to the given problem before settling on a "best-fit" approach.

1-Reading and understanding the problem 2-Search for more information from the Internet "Googling" 3-Problem Solving: a. Analysis ✓ Consulting ER model ✓ Identify the available table holding the required data. ✓ Rereading the problem. b. Synthesis ✓ Deciding which concepts to apply. ✓ Searching for SQL syntax or relevant examples. c. Writing ✓ Start writing the first SQL query in the tool. d. Checking: ✓ Evaluate the result of the first attempt. ✓ Manipulate the query with some justification (fixing the errors). This is done iteratively until they are satisfied. 4-Repeat sub-steps in number 3 until satisfied.	The participant broke the overall problem into a number of sub-problems. He first started by joining Book-Copy and Book-Title tables. At this stage a few actions (A1-A3) and a decision (D1) were performed and other decisions were pending. The participant was happy with his performance at this stage. He then applied another decision, i.e., to use sub-queries The participant then exercised the decision to apply the self-join technique. However, he failed to apply it correctly. The participant then deployed aggregation and was satisfied that he had solved the problem.

Figure 7. The cognitive activities experts deployed

The analysis acted to inform research into the type of approach that ought to be nurtured in novices. The next section discusses how the reported results contributed to the SQL pattern identification.

VI. DISCUSSION

The upper half of Fig. 8 shows how experts solve a task using an analogical approach. The model is based on the ideas of [23], which depicts how scientists think and solve physical problems. The bottom half shows how this model reflects the SQL acquisition process. This model presents the different sources of knowledge and strategies experts deploy.

Observation of expert activities showed that they divided the problem into sub-problems. Then, for each sub problem, different relevant knowledge is applied to arrive at a sub-solution. When experts solved the first part they applied basic knowledge. Then, as the problem requirements required more understanding they applied advanced knowledge which was sometimes obtained by searching. They then applied problem solving strategies such as incremental development, division into sub-queries, consideration of a number of different ways of solving the problem, and choice of the optimal strategy.

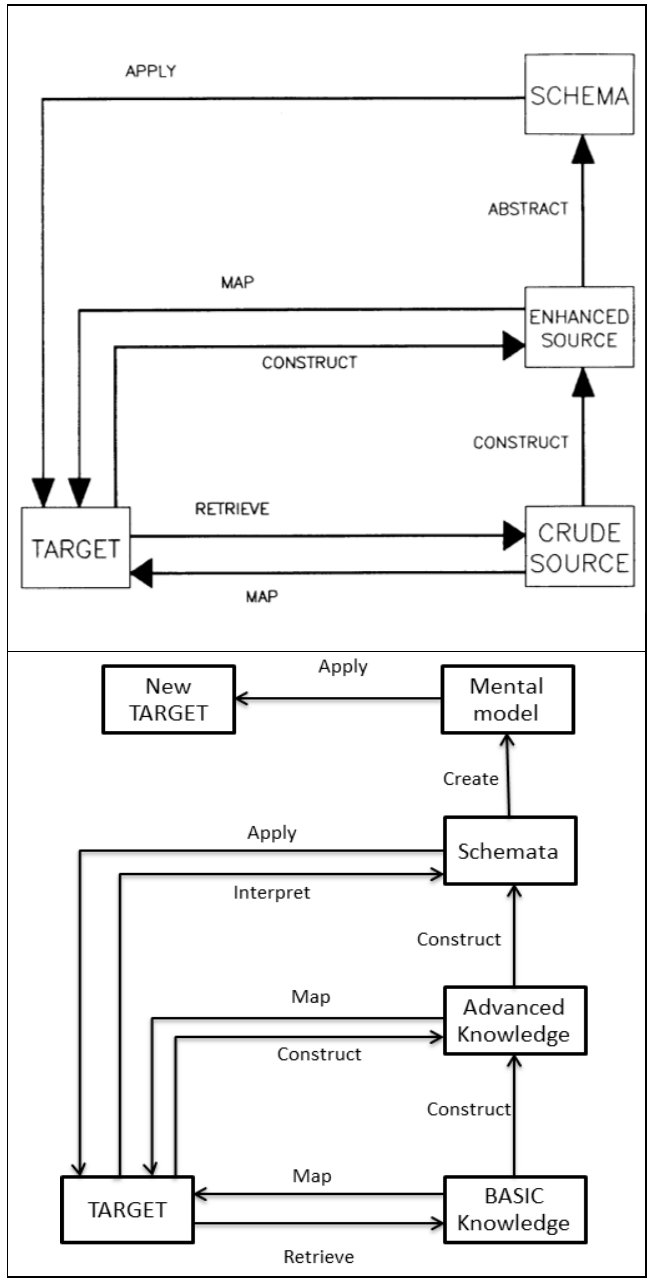


Figure 8. Expert Problem Solving [23](top) and SQL Acquisition on Expert Model (bottom)

 ISBN: 978-1-61208-343-8

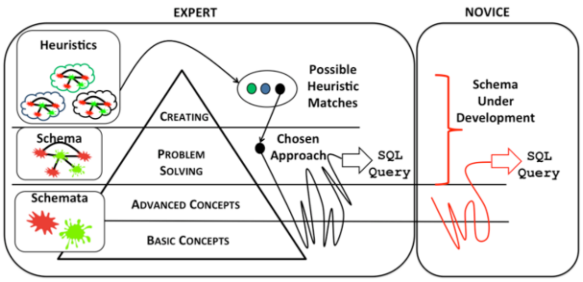


Figure 9. Typical Expert Actions (left) and Novice Actions (right)

Observing experts and novices solving led us to visualize both the expert and novice actions in solving SQL queries (Fig. 9).

Considering how experts approach the task and relating it to novices' actions indicates the nature of the gap between expert and novice. There was no evidence that novices struggled with basic knowledge of SQL syntax. They also knew how the SQL constructs ought to be used. However, novices clearly lacked the knowledge and skills required to solve novel problems. This is related to the "PROBLEM SOLVING" stage as seen in Fig. 9. This, then, is where more effective instruction material needs to assist the process.

This analysis allowed us to determine what type of knowledge and skills are required to solve the SQL problems. We were also able to determine how the information should be presented to learners, i.e., what the optimal sequence of information. The results suggest that:

- Experts start solving the problem by re-formulating the problem statement and determining its context using the data model.
- Expert knowledge is structured, connected and abstract. They have:
  1- Basic knowledge about SQL syntax and semantics "SQL Syntax and Semantics";
  2- Advanced knowledge about the meaning of SQL concepts "SQL Query comprehension" and about how to apply SQL concepts in the given context;
  3- Heuristics:
    - Knowledge about the wisdom of SQL applicability in a certain context "problem-context-solution". This is a high level of knowledge that novice lack as was discussed.
    - Knowledge about the consequences of applying SQL concepts "impact-of-solution". This is a skill of evaluating SQL concepts, which is a high level of knowledge that novices lack.

Observation made it clear that instructional materials, such as notes, did not nudge students towards productive activities or support effective problem solving. To help novices to achieve a measure of SQL expertise we propose that the SQL patterns should include components shown in Table II.

TABLE II. PROPOSED SQL PATTERN CONTENTS

Provide students with data models to help them understand the context of the problem	Schema Formation
The impact of applying the pattern in such a problem context.	Schema Formation
Support for matching a problem to a solution in a simple format such as a checklist	Schema Formation
A section which includes the basic knowledge required to solve the problem.	Schemata
Step-by-step SQL visual examples of the pattern being applied	Schemata
This should be augmented with a step by step plan to train students to deploy effective problem solving strategies, as suggested by (Mayer, 2008).	Encourage Engagement with Analysis and Synthesis Phases during Problem Solving

VII. CONCLUSION AND FUTURE WORK

The successful implementation of instructional materials (SQL patterns) will depend on the pattern writer understanding all the different factors that influence SQL learnability, such as: learner characteristics, SQL language specifications, human cognition and instructional material. The pattern writer must align with established wisdom about human cognition. Our study has provided the guidance to inform SQL pattern content, which should ultimately serve as the link between the task requirement and the generic pattern.

REFERENCES

[1] Al-Shuaily, H., SQL pattern design, development & evaluation of its efficacy. PhD Thesis. University of Glasgow. 2013

[2] Merrill, M.D. Knowledge objects and mental models. in Advanced Learning Technologies, 2000. IWALT 2000. Proceedings. International Workshop on. 2000. pp. 244-246.

[3] Mayer, R.E., Learning and instruction 2003: Prentice Hall.

[4] Bloom, Benjamin Samuel, and David R. Krathwohl. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. 1956.

[5] Gorman, M.E., Types of Knowledge and Their Roles in Technology Transfer, in The Journal of Technology Transfer Springer Netherlands. 2002, pp. 219-231.

[6] Al-Shuaily., H. Analyzing the Influence of SQL Teaching and Learning Methods and Approaches. in

 ISBN: 978-1-61208-343-8

10th International Workshop on the Teaching, Learning and Assessment of Databases. UK, London. 2012.
[7] Schlager, M.S. and W.C. Ogden, A cognitive model of database querying: a tool for novice instruction. SIGCHI Bull. 17(4) 1986, pp. 107-113.
[8] Alexander, C., The timeless way of building. 1979, Oxford, UK: Oxford University Press.
[9] Winn, T. and P. Calder, Is This a Pattern? IEEE Softw. 19(1) 2002, pp. 59-66.
[10] Coad, P. and M. Mayfield, Object model patterns: workshop report. SIGPLAN OOPS Mess., 5(4) 1994, pp. 102-104.
[11] Anderson, et al., A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives (Complete edition) 2001. New York: Longman.
[12] Bower, M., A taxonomy of task types in computing. SIGCSE Bull. 40(3) 2008, pp. 281-285.
[13] van Welie, M., K. Mullet, and P. McInerney. Patterns in practice: a workshop for UI designers. in CHI'02 extended abstracts on Human factors in computing systems. 2002. ACM.
[14] Fincher, S. and I. Utting. Pedagogical patterns: their place in the genre. in ACM SIGCSE Bulletin. 2002. ACM.
[15] Bruner, J.S., Toward a theory of instruction. Vol. 59. 1966: Belknap Press.
[16] Tian, J., Y. Nakamori, and A.P. Wierzbicki, Knowledge management and knowledge creation in academia: a study based on surveys in a Japanese research university. Journal of Knowledge Managemen. 13(2) t, 2009, pp. 76-92.
[17] Fincher, S. Patterns for HCI and Cognitive Dimensions: two halves of the same story. in Kuljis, J., Baldwin, L., Scoble, R., Proceedings of the Fourteenth Annual Workshop of the Psychology of Programming Interest Group. 2002.
[18] Ramalingam, V., D. LaBelle, and S. Wiedenbeck, Self-efficacy and mental models in learning to program. SIGCSE Bull 36(3) 2004, pp. 171-175.
[19] Lahtinen, E., K. Ala-Mutka, and H.M. Järvinen. A study of the difficulties of novice programmers. in ACM SIGCSE Bulletin. 2005. ACM.
[20] Dunbar, K., How scientists think: On-line creativity and conceptual change in science. Creative thought: An investigation of conceptual structures and processes. 1997, pp. 461-493.
[21] Ogden, W.C., Implications of a cognitive model of database query: comparison of a natural language, formal language and direct manipulation interface. ACM SIGCHI Bulletin. 18(2) 1986, pp. 51-54.
[22] Reisner, P., Human Factors Studies of Database Query Languages: A Survey and Assessment. ACM Comput. Surv. 13(1) 1981, pp. 13-31.
[23] Nersessian, N.J., How do scientists think? Capturing the dynamics of conceptual change in science. Cognitive models of science 15 1992, pp. 3-44.