“A Database is Like a Dresser With Lots of Sorted Drawers”: Secondary School Learners’ Conceptions of Relational Databases

Torsten Brinda
University of Duisburg-Essen
Computing Education Research Group
Essen, Germany
torsten.brinda@uni-due.de

Thorsten Terjung
Grillo Gymnasium
Gelsenkirchen, Germany

ABSTRACT

Latent pre-instructional conceptions on a certain topic, which learners bring into the classroom, can have a significant influence on the learning success to be achieved there. In science, therefore, the model of educational reconstruction was developed, which regards the comparison between technical domain concepts and existing learner conceptions as an essential and iterative element within instructional design. In order to support this process, the conceptions of learners and teachers on different topics of science were investigated in thousands of works so far. In the area of computer science, however, comparable research has so far only been conducted to very little extend – concerning e.g. conceptions of the Internet, the computer and the way programs work. Within the paper at hand, the authors intend to add to this research by contributing learner conceptions of (relational) databases. A semi-structured online questionnaire was used to collect the perspectives of 193 German learners. It contained open questions on the conceptions of databases and their everyday applications as well as on the explanation of computing-related phenomena with an indirect relation to databases. The data analysis was carried out using the method of qualitative content analysis according to Mayring and provided three different category systems for various aspects of the subject. As a major result it was found that the conceptions of learners regarding relational databases can be structured according to the ANSI/SPARC acrchitecture model for DBMS, however, completed by an everyday perspective on its three layers (external, conceptual, and internal). Finally, the identified learners’ perspectives were compared with the associated scientific concepts for similarities and differences with the method of reciprocal comparison and guidelines for structuring database teaching were derived.

CCS CONCEPTS

• **Social and professional topics → Computing literacy**; **K-12 education**; *Computational thinking*;

KEYWORDS

Learner conception, K12, computing education, database, phenomena, empirical study, questionnaire, categorization

ACM Reference format:
Torsten Brinda and Thorsten Terjung. 2017. “A Database is Like a Dresser With Lots of Sorted Drawers”: Secondary School Learners’ Conceptions of Relational Databases. In *Proceedings of WiPSCE ’17, Nijmegen, Netherlands, November 8–10, 2017,* 10 pages.
https://doi.org/10.1145/3137065.3137074

1 INTRODUCTION

In times of continually increasing digitalization and dissemination of computing systems in all areas of private and professional life, computing education as a part of modern general education has the goal to explain school learners the computing-related aspects of the world surrounding them and to prepare them for comprehensive participation in the so-called "digital world". Adolescents get in contact with computing systems earlier than ever [3] and often before such systems or phenomena occurring in relation with them are dealt with in class at school. In this way, they can develop individual conceptions about the structure and the functioning of such systems, which do not necessarily match the underlying scientific concepts. According to the model of educational reconstruction [11], which was originally developed for the field of science education, but transferred to and extended for computing education [6], such pre-instructional conceptions should be aligned with the scientific domain concepts in order to develop approaches for teaching design, like, for example, leading error-containing conceptions to contradictions or by picking up and extending incomplete conceptions. There is a rich body of research on primary and secondary school pupils’ conceptions regarding science education matters [10], but so far only few studies in this research area have been conducted regarding computer science topics, e. g. concerning conceptions about search engines or the Internet (cf. e. g. [7], [26]) or the computer (e. g. [25]). For learner-centered computing education, which incorporates existing learners’ conceptions into the learning and teaching processes of scientific principles and concepts, there is a need to investigate further curricular-relevant concepts with regard to existing learners’ conceptions and to make such conceptions available for the instructional design.

Because of the increasing economic exploitation of personal data and the resulting societal impacts, learning and teaching of automated data processing by means of databases and data mining technologies are considered to be of high educational value

WiPSCE ’17, November 8–10, 2017, Nijmegen, Netherlands

ACM ISBN 978-1-4503-5428-8/17/11…$15.00
https://doi.org/10.1145/3137065.3137074

[13]. Up to now, this relevant field has been mainly related to relational databases in German and international computing curricula at secondary schools. Therefore, in the study reported in this paper, secondary school learners' conceptions of relational databases and their applications as well as learners' explanations of computing-related phenomena with an indirect relation to databases were investigated using a semi-structure online questionnaire.

The further text is structured as follows. Section 2 anchors the work in the research background. In section 3 the research method is described, section 4 presents the results of the study, which are interpreted afterwards in section 5. Section 6 completes the paper with conclusions and outlook.

2 RESEARCH BACKGROUND

Since this study is about the investigation of database-related learner conceptions, the belonging scientific concepts are sketched in section 2.1 to be able to compare them with the learners' conceptions (see section 5). From educational research, the conceptual change theory and the model of educational reconstruction provide a theoretical framework for the research presented here and are therefore sketched in section 2.2.

2.1 Database Management Systems and Data Warehouse

2.1.1 Database Management Systems. In the digital age, enormous amounts of data are generated everywhere, which must be organized, processed and stored by appropriate computing systems. Database management systems (DBMS) usually perform this task today. The architecture of most commercial DBMS is structured according to the *ANSI-SPARC database architecture model* [30]. This layer model provides three levels of data abstraction: the *internal*, the *conceptual* and the *external* level (see Figure 1). The lowest internal or physical level describes, how data is actually stored on storage devices using efficient data and file structures as well as data compression and encryption techniques. It provides an internal view of physical storage of data. The next higher conceptual or logical level describes which kinds of data are stored in the database and which relationships among them exist. It also hides the complexities of physical storage from the user. The highest level of abstraction, an external view, provides only the part of concern of the entire database to an end user.

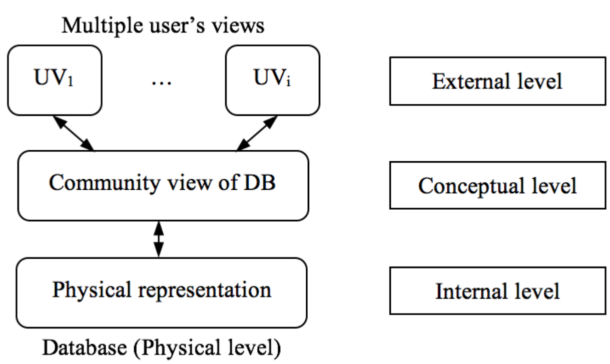

Figure 1: The ANSI/SPARC 3-level architecture

Particularly common[1] are *relational databases*, which go back to Codd [4]. Their data model is based on the fundamental assumption that all data can be stored using tables. This view also allows a clear separation of the three layers according to the ANSI/SPARC 3-level model. To achieve this, Codd developed a list of eight *capabilities*, which a full-scale DBMS must fulfill (cf. [5, p.114]):

(1) data storage, retrieval, and update;
(2) a user-accessible catalog for data description;
(3) transaction support to ensure that all or none of a sequence of database changes are reflected in the pertinent databases [...];
(4) recovery services in case of failure (system, media, or program);
(5) concurrency control services to ensure that concurrent transactions behave the same way as is run in some sequential order;
(6) authorization services to ensure that all access to and manipulation of data be in accordance with specified constraints on users and programs [...];
(7) integration with support for data communication;
(8) integrity services to ensure that database states and changes of state conform to specified rules.

2.1.2 Data Warehouse. Databases play an important role in the evaluation of personal data, e. g. for public and economic purposes. In order to implement the respective analysis and evaluation possibilities, the concept of *data warehouse* [18] is used. In addition to general operating data, personal data are captured for the purpose of statistical analyses. Though in separate databases, user data is collected in order to find interrelationships between different characteristics (for example, suggestion systems for products in Amazon or music in Spotify). This analysis of data is also referred to as *data mining*, which can be characterized according to Vossen by the following three objectives [31, p. 703f]:

(1) finding connections, trends and behavior patterns;
(2) clustering of data into disjoint sets;
(3) automated recognition of useful information.

While the first objective is aiming at analyzing consumer behavior in order to increase sales through advertising, the second is seen as a primary objective, for example, for the risk analysis of banks or insurance companies. The third objective is important for both, since it is not possible to analyze an entire data stock and to discover connections in the data without automation.

2.2 Conceptual Change Theory and Educational Reconstruction

Dealing with learner conceptions about scientific concepts in class should be performed in a way guided by theory. The theories of conceptual change and educational reconstruction provide a framework for this attempt.

[1] cf. https://db-engines.com/en/ranking

2.2.1 Conceptual Change. The term *conceptual change* is the major term to theoretically frame learning processes within constructivist oriented approaches [11, p. 28], however it can be found in the literature with different meanings (cf. [29]). In the so-called "cold" theory of conceptual change, the dissatisfaction of the learners with their own ideas is just as decisive as a new concept, which appears to be meaningful and well applicable, to activate the process of conceptual change [23]. "Hot" theories of conceptual change emphasize the importance of motivation, social status and the boundary conditions for conceptual changes [22, p. 261]. The origin of the various approaches can be traced back to the 1970s, where observations in the field of science education led to the conclusion that learning new technical concepts was partly unsuccessful, because the learning processes did not appropriately change existing, inadequate ideas and hindered the development of appropriate conceptions [22, p. 260]. Thus, according to [9], learners are guided by their already existing conceptions in conducting and evaluating scientific experiments, and these individual perspectives are also used to interpret the results of experiments. Duit et al. [11, p. 28] state that it should not be a purpose of teaching to "exchange" wrong conceptions just by the right ones, since the conceptions of pupils are not necessarily wrong, but can also prove themselves in scientific contexts. That is why terms like *conceptual growth* or *conceptual enrichment* have been proposed instead to avoid misunderstandings. Pupils must be made aware that in some contexts, the scientific concepts are superior to their everyday concepts. This change of individual conceptions to scientific concepts however is suitable as a teaching goal [16].

2.2.2 Educational Reconstruction. The "hot" approach of conceptual change was the basis of the model of *educational reconstruction*. Its goal is to develop teaching guidelines by aligning existing scientific concepts on a certain topic with the conceptions of the learners [11]. In the course of planning instruction, it is therefore necessary to clarify the scientific background of the topic in order to be able to compare it with existing learner conceptions. As a result of this comparison, appropriate everyday phenomena can be selected and elaborated for teaching in class. It is necessary to embed them in environmental, societal or individual matters to illustrate their significance for the life of the individual. For computing education, Humbert and Puhlmann described three *categories* of *computing-related phenomena* [17, p. 68]:

(1) Phenomena that are directly related to computing systems (conscious use of computing systems, such as the direct use of a computer or a smartphone)
(2) Phenomena that are indirectly related to computing systems (rather unconscious use of computing systems; e. g. when paying at the supermarket with a credit card, a computing system is used without the customer having to worry about it)
(3) Phenomena that are not associated with computing systems (algorithms such as searching or sorting methods are phenomena that are not associated with computing systems, but nevertheless constitute important foundations of computer science)

In their *adaptation of the model of educational reconstruction for computing education*, Diethelm et al. [6] brought computing-related phenomena to the center of consideration, where they are subject to reciprocal influence with the scientific concepts, pupils' perspectives and teaching guidelines. In addition, their extended model for computing education includes two new aspects: on the one hand, the teachers' views, who select computing phenomena to be dealt with and include their own conceptions into instructional processes (and thus interact with the pupils' perspectives) and also, on the other hand, in addition to the technical also the clarification of societal demands concerning computing education at school.

2.3 Computing-Related Learner Conceptions

According to the model of educational reconstruction (see 2.2.2), learner conceptions should play a central role in learner-centered instructional design.

Computing-related research about learner conceptions was conducted to some extent in the area of higher education, mostly concerning the field of programming (e.g. [27], [28], [32]), but so far only to little extend concerning secondary education. Available research focuses on conceptions about the Internet, the computer and programming concepts.

Concerning the *Internet*, for example, Diethelm et al. investigated middle school pupils' conceptions about its structure and functioning, in particular about e-mail, chat, and video streaming [7] using semi-structured interviews. For the structure of the Internet, for example, 40% of the participants had the idea of a central computer which differs from the real situation, about half of them considered a structure with several computers. Seiffert et al. investigated middle school learners' conceptions on how the search engine Google works [26] also using interviews. According to their participants, the ranking of websites in the result lists was determined by the number of visits on the respective site, fees paid to Google or ratings of the respective websites.

For the concept *computer*, there is already a literature review available by Rücker and Pinkwart [25]. From their analysis, they derived five categories of conceptions, namely "computers are intelligent", "computers are omniscient databases", "computers are mechanical", "computers are wire networks" and "computers are programmable". Grover et al. investigated 117 middle and high school learners' conceptions about the question "What is a computer?". They come to the conclusion that their "[...] findings indicate that the question (a) places a misplaced focus on the computer (rather than computing or computation) and (b) triggers confusion when students are expected to apply the definition to other devices and appliances" [15, p. 569].

In the field of *programming*, for example, Rabel investigated pupils' conceptions about basic concepts of object-oriented programming [24]. He surveyed 22 learners in computing classes twice within nine months by means of the assignment to describe the concepts class, object, attribute, method, and constructor as clearly as possible. Based on literature analyses, the class conceptions as a blueprint, as a factory, as a set and as a prototype were expected. Except for the factory metaphor, all the other conceptions were found among the participants. Furthermore, it was found that the answers of the pupils differed considerably from the first to the second survey. This confirms results, according to which at the beginning of a new learning process existing phenomena from the everyday world are used for model formation and that subsequent teaching is suitable for developing correct conceptions [12].

Existing papers on school learners' conceptions about different computing topics (as sketched above), on the one hand, show that research in this field is also regarded as relevant by other researchers. On the other hand, they also indicate that earlier conception-based research focussed on other topics and – as the authors of this paper did not find any work concerning learners' conceptions about databases – therewith that related research closes a research gap.

3 RESEARCH METHOD

3.1 Research Questions

Empirical findings on pupils' conceptions can be a useful aid for planning instruction in the respective subject area, since it is not possible for every teacher to carry out such investigations in the respective group of learners. Based on the theoretical background set out in section 2, which came to the conclusion that research on learners' conceptions in computer science has so far only been conducted to little extend, but is a research desiderate for improving learner-centered computing education at secondary schools, and that respective research especially on the database concept is not yet available, the following research questions were formulated:

- *What are secondary-school learners' conceptions of databases? (RQ1)*
- *Which application examples of databases from their everyday lives can learners give? (RQ2)*
- *Which characteristics of relational database are known by them and considered to be important? (RQ3)*
- *Can learners explain everyday phenomena indirectly associated with the database concept using this concept? (RQ4)*

3.2 Data Collection

Conceptions, i. e. existing cognitive structures, are not objectively observable [2]. It is therefore necessary to stimulate the learners to articulate their conceptions. Since no empirical findings concerning learners' conceptions about relational databases were found, a qualitative, explorative research design was chosen. The qualitative paradigm of empirical social research is primarily aimed at a comprehensive, interpretive reconstruction of social phenomena in their respective contexts [1], e. g. to reconstruct patterns of thought, the comprehension of concepts and constructions of one's own reality or theories of everyday experience.

A semi-standardized online questionnaire with a number of open questions was created to be able to include a larger and more locally distributed group of participants than would have been possible in interviews. The semi-standardized questionnaire for semi-structured written surveys is the counterpart to the interview guide for semi-structured interviews [1]. When answering the questions, the participants were to express themselves freely about databases and write down their thoughts and ideas, but this usually leads to limitations in the scope and depth of the answers: very extensive and complex answers are not to be expected in writing; one must confine oneself to a few and clearly defined questions [8, p. 398]. The online survey method was chosen as a compromise between reach and practicability.

The final online questionnaire was structured as follows: In addition to sociodemographic questions concerning e. g. the federal state a participant lives in, the attended school type, school year, gender, and experiences with databases at school, the participants were asked three open questions about their conceptions of and their experiences with databases (the questions were asked in German language):

- *Certainly, you've already heard of databases. How do you imagine such a database actually? Describe your conception in a few sentences. (cf. RQ1)*
- *Databases play an important role in everyday life. Please give some examples, where you encounter databases in your everyday life. (cf. RQ2)*
- *Databases store all kinds of data. Which characteristics of a database do you consider to be especially important? (cf. RQ3)*

In addition, *six different everyday scenarios* (concrete scenarios listed in section 4.3) with a relation to databases were presented to the participants and they were asked for their explanation (cf. RQ4).

With these questions, the learners were asked for their own database-related conceptions. It was *not* the goal to have them reproduce memorized definitions learned at school, but instead to find out, how learners conceptualize and perceive databases, their characteristics and uses in every day life as constructed from hearing or reading about them, own experiences or lessons about databases at school.

3.3 Implementation of the Survey

The online questionnaire was created with the survey software *LimeSurvey* on a server of the University of Duisburg-Essen. An invitation letter was sent out to computing teachers to motivate them to take part in the survey at their schools in the respective computing classes. The letter was distributed together with a link to the online version and with a pdf version of the questionnaire attached via mailing lists for computing teachers at the universities of Duisburg-Essen (North Rhine-Westphalia) and Oldenburg (Lower Saxony), representatives of a computing teachers' working group of the German Informatics association in Saxony as well as to individual further teacher contacts. Pupils from all three federal states of Germany participated in the study.

In total, 223 answers were received. At first, a selection took place. All empty questionnaires were sorted out, also those, which obviously did not contain any meaningful answers by the learners. In part, the questionnaires were incompletely filled in or did not contain serious information on all questions. However, if at least single responses were relevant for later evaluation, the questionnaires were retained. Finally, 193 questionnaires were evaluated using an inductive qualitative content analysis according to Mayring [21]. Its goal is to summarize the available written results in such a way that it is possible to classify their meaning in the form of categories and subcategories. Classification is intended to mean the ordering of data material according to certain empirical and theoretically appearing principles of order, thus enabling a more structured description of the collected material. This classification of the data is conducted by reducing the content and increasing the degree of abstraction.

3.4 Evaluation and Categorization of Learners' Conceptions

Quantitative aspects were of almost no relevance in the evaluation of this study, since basic concepts and mental models of learners were to be determined. Following the approach of educational reconstruction, statements were to be made about the structure and quality of scientific and everyday conceptions, and not in which quantities certain individual conceptions occur deprived of their context [19, p. 101]. The conceptions in the learners' answers were paraphrased and then used to create categories. This procedure is also described for the model of educational reconstruction [20]: in order to be able to comprehend and compare learner conceptions, the concepts that are recognizable in them are dissected and generalized. The goal is to identify domain and topic specific ways of thinking. The generalization takes place in the form of category formation.

The following example of an exemplary learner conception taken from the data may illustrate the concrete categorization procedure carried out. The task of describing the own conception of a database was e. g. answered by *participant 1 (p1)* as follows: *"A database is a collection of data that is mostly used to manage information. Using databases, information can be structured and stored systematically."* (all examples translated into English by the authors). At the *first abstraction level (A)* three basic conceptions were extracted:

- *management of information/data (A1)*
- *structured storage of information/data (A2)*
- *systematic storage of information/data (A3)*

At the *second abstraction level (B)*, these concepts were combined in a common category of *electronic data processing/ data storage (B1)*. This category includes conceptual models, in which the idea of a database consists mainly of possible applications for the administration, organization, structuring and respectively or storage of data. At the *third level of abstraction (C)*, the B categories were abstracted to find common main categories in order to be able to structure the basic conceptions of the pupils as compactly as possible with a small number of categories. For this purpose, finally three main categories were found. It turned out that these categories correspond with the three layers of the ANSI/SPARC 3-level architecture (see Figure 1) namely the *external view (C1)*, the *conceptual view (C2)* and the *internal view (C3)* on databases. Thus, the category *Electronic data processing/data storage (B1)* could be assigned to the main category *Conceptual view (C2)*.

4 RESULTS

4.1 Sociodemographic Data

The $N = 193$ participants were from three federal states of Germany (North-Rhine Westphalia (88), Lower Saxony (90) and Saxony (15)), different school forms and different classes. 28.5% of them (55 pupils) were females. Table 1 shows the age distribution of the participants. 13 students said they had not yet attended any computing class. 67 students stated that databases had been covered in their computing classes yet for more than two hours. Since the participants came from a wide age spectrum and from different federal states within Germany with different school curricula for each school type, it would have been speculative to reconstruct more precise statements on the expected computing competencies of the participants according to their respective curriculum. Therefore, only the database-related learning time at school was further considered.

Table 1: Distribution by age

Age	13	14	15	16	17	18	19	20	>20
#	6	42	36	31	33	29	5	9	2
%	3.1	21.8	18.7	16.1	17.1	15.0	2.6	4.7	0.9

4.2 Questions Concerning Basic Conceptions

The qualitative content analysis for the evaluation of the three questions (RQ1, RQ2, RQ3) on the basic conceptions of the pupils took place according to the procedure described in section 3.4 and led to the results presented subsequently.

Question 1: Certainly, you have already heard of "databases". How do you imagine such a database actually? Describe your conception in a few sentences.

In the answers to this question, 77 different conceptions on abstraction level *A* could be identified, which were grouped into 31 categories. These 31*B* categories could then be classified at the highest abstraction level into six main categories, which are presented in Table 2 . It became clear that many learners did not present a well-elaborated conception of the database concept, but simple obvious aspects instead. For example, there were a number of mentions, a database was "for storing data" or "a collection of data". Physical

Table 2: Categorization of pupils' conceptions of relational databases (categories, descriptions, anchor examples)

Scientific view	
C1 – External view	*A special user view on a database* "To do this, you can set up suitable forms and queries to display only the requested data and enter it" (p4)
C2 – Conceptual view	*The logical concept of a database* "A database is a systematically structured, long-lasting collection of data" (p201)
C3 – Internal view	*The physical storage of data* "Databases are large hard disks on which a lot of information is stored" (p108)
Everyday view	
C4 – External view	*Everyday storage of data* "I imagine a database like a blood donor card index" (p133)
C5 – Conceptual view	*Structuring of data* "I imagine it is for example on a card where the respective data of the person are stored" (p161)
C6 – Internal view	*Storage and safety of objects* "Like a dresser with lots of sorted drawers" (p26)

characteristics were also present to the pupils, but they showed a view, which focused on the pure storage of data. In the case of the more elaborate answers, it was often apparent that memorized definitions were reproduced, because several participants answered in partly the same or very similar wording. There were also pupils, who gave *MS Excel* or similar programs as database examples.

In the investigation of the influence of the measured sociodemographic variables on the mentioned learners' conceptions, no relevant relation with sex, age or the school type was found, however, a clear relation with the learning experience concerning databases, see Table 3.

Table 3: Relation between database-related learning time at school and learners' conceptions of databases

Learning time (in hours)	C1	C2	C3	C4	C5	C6	Learners
0	0	11	**34**	4	7	**12**	53
up to 2	0	8	11	3	3	6	18
2 to 10	1	24	15	7	8	8	49
more than 10	5	**36**	16	9	**15**	10	55

While the answers of the pupils, who had not yet dealt with databases in class, could be assigned predominantly to the category *C3*, the classification shifted towards the category *C2* with an increase of learning time on databases in class. Even though everyday conceptions were expressed more rarely, a similar shift from the physical *(C6)* to the conceptual view *(C5)* could also be observed there.

Question 2: Databases store all kinds of data. Which characteristics of a database do you consider as particularly important?

In the data on the characteristics of databases, which were regarded as particularly important by the pupils, the answers were grouped into ten main categories (*D1* to *D10)*, see Table 4 . It is worth noting here that many conceptions were assigned to several categories. In some cases, several aspects were mentioned, in other cases the specified properties could be assigned to different categories. An example for several different statements is the answer "Security against intruders of any kind, clarity, efficiency & speed" (p86), which can be classified into the categories *D7* (protected), *D3* (simple, clear), and *D10* (qualitative claims). The influence of learning experience with databases seems also to be most significant in the case of the stated characteristics of databases, see Table 5. With rising database experience a shift towards the categories *D1* and *D3*, i. e. the structure of the database, and a little to category *D10*, which brings together qualitative aspects of data processing, was found in the data.

Question 3: In everyday life, databases play an important role. Please give some examples, where you can find databases in your daily life.

Table 4: Database characteristics considered important by the pupils (categories, descriptions, anchor examples)

D1 – structured	*Data is stored in a structured way* "It is structured and clear" (p135)
D2 – filterable	*Data can be filtered; different views on data are possible* "I find filters good to access data quickly" (p167)
D3 – simple and clear	*Easy access to data is made possible; data is stored clearly* "It is particularly important for me that they are simple-structured and clear" (p205)
D4 – safe	*Data is stored securely; no data loss* "Secure storage, no data loss [...]" (p45)
D5 – available	*Data is always available* "Continuous availability [...]" (p180)
D6 – independent of location	*Access to data is not local, decentral access is possible* "It may also be a disadvantage, but an important property would be that it can be accessed from anywhere" (p103)
D7 – protected	*Privacy – access to data can be regulated* "That they are protected from external attacks" (p189)
D8 – large	*Databases store large amounts of data* "That you can store and retrieve a large amount of information" (p46)
D9 – reality-related	*There are application possibilities in real life* "For me, databases are important to store the data about life, general knowledge to pass it on" (p59)
D10 – qualitative requirements	*Quality standards are met – fast, reliable, efficient, error-free* "Speed, safety, operability, readability" (p95)

Table 5: Relation between database-related learning time at school and learners' conceptions of relevant characteristics of databases

Learning time (in hrs.)	D1	D2	D3	D4	D5	D6	D7	D8	D9
0	4	2	9	**20**	5	**15**	7	6	14
up to 2	1	1	5	4	0	4	1	3	1
2 to 10	16	6	24	7	2	3	4	3	10
more than 10	**21**	3	**29**	4	0	0	5	4	9

In this case, categorization was very difficult to carry out since the given answers were very individual. A division into relatively broad categories led to the results presented in Table 6.

The pupils mostly gave examples, which clearly originated from their everyday lives. Again, the extent of learning databases in class is of interest. Without such experience especially *smartphone*, *computer* and *Internet* were mentioned. This shifted with more

Table 6: Examples of databases in real life (categories, descriptions, anchor examples)

E1 – school	*School, computing education, cafeteria* “At school and at work” (p8)
E2 – work	*At work / later job* “At work” (p6)
E3 – smartphone	*Mobile phone / smartphone* “Within my smartphone, for example the stored contacts [...]” (p93)
E4 – computer	*Computer, PC, laptop, navigation system* “On the computer, where everything you do is stored by anyone [...]” (p69)
E5 – devices	*Other electronic devices or subsystems* “I am quite, quite sure that my alarm clock contains a database which stores my alarm times” (p130)
E6 – Internet	*Internet, social networks, social media, online gaming* “I would say the Internet is one of the largest databases [...]” (p118)
E7 – shopping	*When shopping, businesses, companies* “For companies, e. g. the customer data is stored in an Access table” (p9)
E8 – authorities	*Authorities, public institutions* “[...] inhabitants registration office [...]” (p51)
E9 – bank	*Bank, account* “In school, in a bank, in a hospital” (p76)
E10 – health	*Doctor, hospital, blood donation* “Blood donor card index” (p125)
E11 – cards	*Credit cards, customer cards* “Health insurance card, bank account card, [...]” (p153)
E12 – management of digital media	*Management of music video files, iTunes, Spotify, ...* “iTunes, Spotify, PlayStation Network, [...]” (p55)
E13 – analogue media	*Books, tables, lists* “Books, e. g. for cooking, or logbooks” (p135)
E14 – Intelligence services	*NSA, BND*[2] “[...] NSA, BND, [...]” (p75)
E15 – brain	*Brain as a database* “The database of a human being is the brain [...]” (p145)
E16 – universal	*Almost always and everywhere* “Post office, doctor, school, almost everywhere” (p214)

database experience to real-world applications such as *shopping*, *Internet use*, *school* – that is, a shift from a primarily hardware-oriented perspective towards real-world application possibilities, see Table 7.

Table 7: Relation between database-related learning time at school and learners’ everyday examples of databases

Learning time (in hrs.)	**E1**	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**
0	10	0	**20**	**20**	2	**35**	4	2
up to 2	6	0	2	7	0	8	1	6
2 to 10	14	4	11	18	2	24	15	15
more than 10	**15**	8	5	10	4	**19**	**20**	12
Learning time (in hrs.)	**E9**	**E10**	**E11**	**E12**	**E13**	**E14**	**E15**	**E16**
0	13	2	3	1	0	1	0	1
up to 2	2	1	1	0	1	0	0	0
2 to 10	9	9	2	2	4	0	0	1
more than 10	2	10	1	3	3	1	0	1

4.3 Questions About Phenomena

The questions about the computing phenomena in everyday life and the explanatory approaches of the pupils (RQ4) were influenced by the so-called “halo effect”. The students were asked in the first three questions explicitly about their conceptions of databases and after that for their explanations for phenomena, which were indirectly related with the database concept. Some responses suggested that these students were aware that they were expected to respond referring to the database concept. However, it is important to note that most of the learners did not explicitly link the phenomena to databases.

In the following, due to the diversity of the answers given, only a selected, but quite representative part of the explanatory approaches obtained is presented for the individual phenomena.

Phenomenon 1: You get a message from your mobile service provider unasked with a great, new contract offer. The offer seems to be tailored to your personal needs.

Possible explanation: The mobile service provider stores the user behavior in a database and analyzes the data according to the data mining principle.

Apart from the fact that many pupils could not express a concrete idea to explain this phenomenon, it was possible to identify three different conceptions:

(1) *User data are stored and evaluated by the provider:* “The service provider stores information about my mobile usage in his database.” (p62)
(2) *User data are collected by third parties – e. g. by Google – and sold to the mobile service provider:* “Apparently, I have been sucked out by ‘data octopuses’ that have sold the information to the mobile service provider [...]” (p188)
(3) *The smartphone automatically analyzes and evaluates the data:* “The phone captures your activities and concludes what you need.” (p122)

Phenomenon 2: You want to buy a new smartphone cover at an online shop (e. g. Amazon). At the bottom of the web page you see the message: "Customers, who bought this cover, also bought ...".

Possible explanation: All customer data are stored in a database and the data are evaluated (data warehouse), when an item is selected.

The data warehouse concept is not familiar to the majority of the participants. It was often stated that the phenomenon was known, but an explanation could often not be given. However, many answers contain the appropriate explanatory approaches such as data storage, linking of databases, cookies, etc. Here are some sample answers ...

- *... for correct explanation:* "It is captured in an anonymous database, which products have been bought together with what. If there are enough common purchases, they will be presented under the described communication." (p49)
- *... for a technically incorrect conception:* "If you buy something at Amazon, in the section 'customers who bought this cover, also bought ...' products are shown that can be used well with this device in combination." (p63)
- *... for no idea, but personal experience:* "This situation has often happened to me. I then often looked at the articles, but I never bought them." (p148)

Phenomenon 3: In the contact list of your smartphone a contact suddenly appears twice.

Possible explanation: The contact must have been mapped to two different primary keys and therefore appears twice.

Mostly, there was no meaningful explanation for this phenomenon. Matching ideas here concern either the hardware or the software:

- *Hardware*
  - *Contact was saved on SIM card and SD card:* "It could be, because the number was stored on the memory card as well as in the phone memory, so the number appears twice." (p16).
  - *Hardware error:* "It happens, because the mobile phone has an error and accesses the database incorrectly or doubly." (p168)
- *Software*
  - *Contact has been saved by different apps:* "Because there are several links to the same contact. For example, WhatsApp and Skype." (p106)
  - *Software error:* "Programming errors in the database" (p50)

Phenomenon 4: You are on a trip in a city you have not been to before. Your smartphone shows you where the next ice cream parlor is near to your current position.

Possible explanation: The software locates the smartphones' position using GPS, analyzes the weather and the user's preferences and searches in database for the next ice cream parlor.

The linking of GPS location data with a database, which stores different shops and restaurants together with their position, is quite clear to the learners. The aspects of weather and behavioral analysis however were hardly mentioned. Positive examples:

- "I will be located and maybe the weather will be analyzed. Through this connection this proposal comes about." (p85)
- "The GPS data of your mobile phone is forwarded to a server and, depending on the weather as it is there, you will be made suggestions for the correct situation." (p185)
- "Because I've been looking for ice cream more often and my smartphone knows I like ice cream" (p38)

Phenomenon 5: In the mailbox of your home you find an advertising letter with the inscription "To all inhabitants of the house ...". Residents in the same district have also received this letter, friends in another district however did not.

Possible explanation: Advertising companies analyze databases with social and economic background information of persons and cluster residential areas with inhabitants according to socio-economic status.

In the case of phenomenon 5, there was the least number of reasonable explanations. Frequently, pragmatic reasons were given, which do not relate to the intention of district clustering.

- *Good approaches:* "The advertising letters are tailored exactly to the different areas, depending on whether this area is prosperous or not" (p201); "It was determined that 'my' neighborhood might be 'susceptible'." (p100)
- *Pragmatic approaches:* "It is perhaps a letter from the building cooperative for a reconstruction in the respective quarter" (p151); "A bomb is defused in the area" (p208).

Phenomenon 6: You enter the phone number of a new friend on your smartphone. Automatically, your friend will also appear in your *WhatsApp* contacts.

Possible explanation: When installing *WhatsApp*, the permission is granted to access the contact list of the smartphone and to compare it with the *WhatsApp* customer database.

The automatic recognition of the smartphone contacts by the messenger service *WhatsApp* was common to almost all participants and was often explained in a detailed and correct way, as well as the fact that during the installation process the app referred to this fact and a user permission was given. Examples:

- "WhatsApp's database is matched with the phone book on the mobile phone, but the user has given permission to install and create the account. Is quite handy." (p25)
- "WhatsApp accesses the user's data and compares, whether a particular phone number can be found with WhatsApp." (p99)

5 INTERPRETATION

In order to interpret the results, the empirically collected learner conceptions were compared with the scientific concepts presented in section 2.1 with the method of reciprocal comparison. The reciprocal comparison includes both the basic scientific concepts and the pupils' conceptions [14]. This is done in instructive intent [20]. The aim is to work out similarities and differences in order to develop instruction from the empirically collected conceptions as a status quo, which regards the scientific concepts as a teaching target to be reached. In principle, however, the scientific concepts and the conceptions of the pupils are regarded as equivalent and should be analyzed from the perspective of each other. This comparison is structured by comparing the concepts in the following four categories [20]:

- *Peculiarities:* Are conceptions more or less related to science or are they experiences from everyday life?
- *Commonalities:* Are there same or similar concepts in science and the pupils' conceptions?
- *Differences:* Where are differences between the scientific theories and the conceptions of pupils?
- *Limitations:* To what extent do learners' conceptions show the limits of the technical concepts and vice versa?

The primary objective of the empirical investigation of pupils' conceptions is not to replace them with professional concepts, but rather to reconcile them and to construct a coherent overall picture, which can present and explain learning content from different perspectives. Thus, existing pupils' conceptions must be seen as an aid in the planning and structuring of teaching, not as an obstacle to be overcome [14, p. 15f].

5.1 Peculiarities

The pupils' conceptions were assigned to six categories which, on the one hand, represent the ANSI/SPARC 3-level model, whereas a real-world analogy to this model could be derived on the other. The individual statements often contained different aspects of several categories, also aspects from scientific and real-world perspectives. In particular, the pupil's answers concerning everyday encounters with databases show that these are strongly influenced by their own experiences. However, since some of these experiences were mentioned often within the same computing course, it cannot be ruled out that these examples were from the classroom and not personal experience.

Although a number of scientific concepts were used in the explanation of phenomena, it is clear that in-depth knowledge was often lacking, but attempts were made to explain the phenomena with spontaneous considerations. In this case, a very pragmatic view was used to explain well-known phenomena without actually referring to the underlying technical concepts.

5.2 Commonalities

The influence of database instruction is clearly visible in the answers of the pupils. As the learning experience increases, the basic technical concepts are becoming more and more apparent in the answers. The focus of the conceptions is then on structuring and efficient storage of data, whereby, apart from the concept of indivisible transactions, all properties of a DBMS required by Codd [5] were mentioned. The explanations of the phenomena also show that this knowledge could not always be used in a meaningful way.

The observed conceptions referring to the ANSI/SPARC model must not be missing here: the interpretation of the model as a whole from an everyday life point-of-view can also be understood as a commonality between conceptions and scientific concepts.

5.3 Differences

First of all, one of the most important findings of this research is that pupils, who have not yet or just little learned about databases in class, often reduced databases to the function of storing and collecting data. The focus initially is on physical storage (ANSI/SPPARC level 3), structuring and ordering of data stocks (ANSI/SPARC level 2) play a rather subordinate role. The more about databases was learned in class, the more the pupils moved towards level 2 and the concrete DBMS concepts according to Codd. ANSI/SPARC level 1, the view of the data structure according to the needs of an application, was hardly addressed.

The often observed conception of reducing databases to the pure task of storing data clearly shows that the learners' views do not correspond with the scientific concepts regarding this aspect. Also the comparison of databases with spreadsheet programs like MS Excel points in this direction.

5.4 Limitations

Important basic concepts of DBMS and relational databases were not mentioned, for example, that transactions are carried out completely or not at all, or that a consistent state must always be maintained when accessing a database in parallel. Also, the elimination of anomalies and the consequent concept of normalization, which are important for a good structure of a database, seem to be less common to the pupils. Especially the explanations given to the phenomena reveal the limitedness of the pupils' conceptions. The learners were often unable to grasp the entire complexity of the database concept, and the multifaceted interweaving of different aspects was not recognized.

6 CONCLUSION AND OUTLOOK

As the most important result the individual learner conceptions could be classified into a manageable number of categories. In particular structuring the categorization in analogy to the ANSI/SPARC 3-level model is to be mentioned here as a central result.

In view of the results presented in section 4 and their interpretation in section 5, the authors suggest the following guidelines for structuring relational database teaching in class:

(1) **Use the ANSI/SPARC 3-level model from the start** The study reported in this paper has shown that learners' conceptions with regard to the model presented above form a broad basis for the educational reconstruction of database learning and teaching. It could therefore be useful to make the different perspectives on databases visible also in initial examples. A simple starting example could be, for example, to choose a combination of pants, shirt and shoes from a very large clothing stock for a specific occasion as efficiently as possible. This can be explored through different models: one wardrobe with no sorting (physical aspect, *C6*), several wardrobes (clothing assigned to them e. g. by category, color; conceptual view, *C5*), or several pharmacist's racks with many drawers to support different views on each rack's content (e. g. just white shirts). A next step would to transfer the everyday example into a database scheme, whereby the above-mentioned levels can be transferred from the real world to the model. In this way, the different pupils'

conceptions could be applied at different points of the model, thus linking their conceptions with the computing concepts without having to completely replace the conceptions with specialist concepts.

(2) **Focus on structuring and modeling** Many students have shown the conception during the survey that a database merely serves as a memory for large amounts of data. Here the important aspects of structuring, efficient search and individual views on the database should be focused more strongly.

(3) **Phenomena should be tangible** The explanations of the pupils about the presented phenomena have shown that a more intensive analysis of everyday uses of databases seems quite reasonable. Especially the linking of databases to the creation of user or customer profiles requires a more intensive analysis in class. A suitable phenomenon as a starting point in the sense of educational reconstruction should be selected in such a way that it originates from the everyday lives of the learners. The functionality of apps on a smartphone and the storage and use of data on the WWW could be promising examples.

According to the extended model of educational reconstruction by Diethelm et al. [6] further research is necessary in this domain. The views of teachers concerning databases have so far been disregarded and also require an empirical study to complete the overall picture. Furthermore, it is necessary to examine whether the created category system proves itself in further investigations or whether the system needs to be expanded if new or different conception or experiences are mentioned. Interview studies could also help to further refine the results.

REFERENCES

[1] P. Alasuutari, L. Bickman, and J. Brannen. 2008. *The SAGE Handbook of Social Research Methods.* SAGE Publications, London.

[2] J. R. Anderson. 2009. *Cognitive psychology and its implications* (7th ed.). Worth publishers.

[3] Child Trends. Data Bank. 2015. Home Computer Access and Internet Use. Indicators of Child and Youth Well-Being. (2015). https://www.childtrends.org/wp-content/uploads/2015/12/69_Computer_Use.pdf

[4] E. F. Codd. 1970. A Relational Model of Data for Large Shared Data Banks. *Commun. ACM* 13, 6 (June 1970), 377–387. https://doi.org/10.1145/362384.362685

[5] E. F. Codd. 1982. Relational Database: A Practical Foundation for Productivity. *Commun. ACM* 25, 2 (Feb. 1982), 109–117. https://doi.org/10.1145/358396.358400

[6] I. Diethelm, P. Hubwieser, and R. Klaus. 2012. Students, teachers and phenomena: educational reconstruction for computer science education. In *Proceedings of the 12th Koli Calling International Conference on Computing Education Research.* ACM Press, New York, USA, 164–173.

[7] I. Diethelm, H. Wilken, and S. Zumbrägel. 2012. An investigation of secondary school students' conceptions on how the Internet works. In *Proceedings of the 12th Koli Calling International Conference on Computing Education Research.* ACM Press, New York, USA, 67–73.

[8] N. Döring and J. Bortz. 2016. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* Springer, Berlin Heidelberg.

[9] R. Duit. 1996. *Lernen in den Naturwissenschaften.* IPN, Kiel, Germany, Chapter Lernen als Konzeptwechsel im naturwissenschaftlichen Unterricht, 145–162.

[10] R. Duit. 2009. Bibliography – STCSE. Students' and Teachers' Conceptions and Science Education. (2009). http://archiv.ipn.uni-kiel.de/stcse/

[11] R. Duit, H. Gropengießer, U. Kattmann, M. Komorek, and I. Parchmann. 2012. *Science education research and practice in Europe.* Sense publishers, Chapter The Model of Educational Reconstruction – a Framework for Improving Teaching and Learning Science, 13–37.

[12] S. Dutke. 1994. *Mentale Modelle: Konstrukte des Wissens und Verstehens. Kognitionspsychologische Grundlagen für die Software-Ergonomie.* Verlag für Angewandte Psychologie, Göttingen.

[13] A. Grillenberger and R. Romeike. 2015. Bringing the Innovations in Data Management to CS Education: an Educational Reconstruction Approach. In *Proceedings of the 10th Workshop in Primary and Secondary Computing Education*, J. Gal-Ezer, S. Sentence, and J. Vahrenhold (Eds.). ACM, New York, USA, 88–91.

[14] H. Gropengießer. 1997. *Didaktische Rekonstruktion des Sehens: wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung.* Ph.D. Dissertation. University of Oldenburg, Didaktisches Zentrum, Oldenburg.

[15] S. Grover, D. Rutstein, and E. Snow. 2016. What is a Computer: What do Secondary School Students Think?. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education.* ACM Press, New York, USA, 564–569.

[16] P. W. Hewson and M. G. Hewson. 1992. *Theoretical issues and empirical studies.* IPN, Kiel, Chapter The status of students' conceptions. Research in physics learning, 59–73.

[17] L. Humbert and H. Puhlmann. 2004. Essential ingredients of Literacy in Informatics. In *Concepts of Empirical Research and Standardization of Measurement in the Area of Didactics of Informatics*, J. Magenheim and S. Schubert (Eds.), Vol. 1. Köllen, Bonn, 65–76.

[18] W. H. Inmon. 1996. *Building the data warehouse* (2 ed.). Wiley, New York, USA.

[19] U. Kattmann. 2007. *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden.* Springer, Berlin, Heidelberg, Chapter Didaktische Rekonstruktion – eine praktische Theorie, 93–104.

[20] U. Kattmann, R. Duit, H. Gropengießer, and M. Komorek. 1997. Das Modell der Didaktischen Rekonstruktion. Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften* 3, 3 (1997), 3–18.

[21] P. Mayring. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution.* Klagenfurt.

[22] K. Möller. 2007. *Handbuch Didaktik des Sachunterrichts.* Klinkhardt, Bad Heilbrunn, Chapter Genetisches Lernen und Conceptual Change, 258–266.

[23] G. J. Posner, K. A. Strikeand P. W. Hewson, and W. A. Gertzog. 1982. Accommodation of a scientific conception. Toward a theory of conceptual change. *Science education* 66, 2 (1982), 211–227.

[24] M. Rabel. 2013. Grundvorstellungen zu Klassen und Objekten.. In *Informatik erweitert Horizonte – INFOS 2013*, N. Breier, P. Stechert, and T. Wilke (Eds.). Köllen, Bonn, 57–66.

[25] M. Rücker and N. Pinkwart. 2016. Review and discussion of children's conceptions of computers. *Journal of Science Education and Technology* 25, 2 (2016), 274–283.

[26] O. Seifert, T. Sauck, M. Schwarzbach, C. Lerch, M. Weinert, and M. Knobelsdorf. 2013. "Ich glaube, Google ist so was wie eine Vorhalle des Internets" – Erste Ergebnisse einer qualitativen Untersuchung von Schülervorstellungen von der Suchmaschine Google. In *Informatik erweitert Horizonte. 15. GI-Fachtagung "Informatik und Schule" – INFOS 2013*, N. Breier, P. Stechert, and T. Wilke (Eds.). Köllen, Bonn, 45–46.

[27] I. Stamouli and M. Huggard. 2006. Object oriented programming and program correctness: the students' perspective. In *Proceedings of the Second International Workshop on Computing education research (ICER '06).* ACM, New York, USA, 109–118.

[28] E. Thompson, L. Hunt, and K. Kinshuk. 2006. Exploring learner conceptions of programming. In *Proceedings of the 8th Australasian Conference on Computing Education (ACE '06)*, D. Tolhurst and S. Mann (Eds.), Vol. 52. Australian Computer Society, Inc., Darlinghurst, Australia, 205–211.

[29] D. F. Treagust and R. Duit. 2008. Conceptual change: a discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education* 3, 2 (2008), 297–328.

[30] D. Tsichritzis and A. Klug. 1978. The ANSI/X3/SPARC DBMS framework report of the study group on database management systems. *Information Systems* 3, 3 (1978), 173–191.

[31] G. Vossen. 2000. *Datenmodelle, Datenbanksprachen und Datenbank-Management-Systeme* (4th ed.). Vol. 1. Oldenbourg, München.

[32] S. Xinogalos. 2015. Object-Oriented Design and Programming: An Investigation of Novices' Conceptions on Objects and Classes. *ACM Transactions on Computing Education (TOCE)* 15, 3 (July 2015), 13.