

**Qualifying Examination
Part I
Question from Dr. R. Bias**

Don Drumtra
University of Texas at Austin
16 January 2004
Revised

Approved by the Examining Committee
February 27, 2004
Dr. Fran Miksa, Chair
Dr. Randolph Bias
Dr. Larry Browning
Dr. Glynn Harmon

Qualifying Examination
Part I
Question from Dr. R. Bias

Question

In your (Mr. Drumtra's) research concept paper, you say, "The structuring of databases and their contents are important if they are to be useful tools for business decision-making." Your paper goes on to suggest that the design of data structures is catch-as-catch-can, with no accepted, systematic approach to guarantee, or even maximize, success. When the field of user interface design was in a similar state, it benefited from a turn to a user-centered design approach, where user data are gathered to inform (early on) and validate (in later stages) user interfaces.

Discuss any research on database structure design that involves users of the ultimate databases in an empirical, iterative design approach. If this corpus of research is thin, summarize some of the literature on user-centered approaches to user interface design, or to other design, that might be applied to the design of database structures.

Response

There has been significant database research conducted in the three decades since the 1970s when database research was identified as a basic science research field. "Achievements in database research underpin fundamental advances in communications systems, transportation and logistics, financial management, knowledge-based systems, accessibility to scientific literature, and a host of other civilian and defense applications" (Silberschatz, Stonebraker, & Ullman, 1991, p. 111). The following provides a summary of that research, identifies the dearth of interest in the database user community, and suggests future user oriented research.

Research Prior to 1990

In 1990 Avi Silberschatz, Michael Stonebraker, and Jeff Ullman¹ reported on behalf a workshop convened by the National Science Foundation. The purpose of the workshop was to identify “the *technology pull* factors that will serve as forcing functions for advanced database technology and the corresponding basic research needed to enable that technology.” The resulting report provided “a brief description of some of the important achievements of the database research community” as well as the information of those factors (Silberschatz et al., 1991, p. 111).

The authors report that there are three basic areas in which research was conducted: relational databases, transaction management, and distributed systems. In the relational database area investigations included: (pp. 111-112)

- High-level relational query languages to ease the use of the DBMS by both end users and application programmers. These investigations concentrated on theory of higher level query database language constructs.
- Theory and algorithms necessary to optimize queries.
- The theory of database design normalization—mathematical approaches to eliminating redundancies data.
- Allocation of tuples in pages in files on secondary storage to reduce access costs.
- Buffer management algorithms.
- Indexing techniques to provide fast associative access to records.
- Prototype relational database management systems.
- In the transaction management area, investigations included
- Locking data items to prohibit conflicting access.
- Time stamping accesses.
- Keeping multiple versions of data available.
- In the distributed database area, investigations included:
- Algorithms for location transparency.
- Concurrency control algorithms.
- Failure recovery.
- Distributed queries.
- Update consistency.

¹ At the time the report was written, Avi Silberschatz was with ATT, Bell Labs, Michael Stonebraker with UC-Berkeley, and Jeff Ullman with Stanford.

As may be seen except for the first item there is little effort to study the user. And the studies on the first item seemed to relate to the structured query language as an application.

For the future the report suggested the following research topics:

- Management and query of large volumes of image data (such as data being returned from space exploration).
- Management of computer aided design data with multiple organization access.
- DNA research data structures.
- Large department store data from product-code scanning.
- Insurance firm data management.
- Large record management.
- Data type management.
- Logical rule processing for criteria enforcement.
- Data modeling primitives and implementation
- Multiple dimensional space data management.
- Time based and time interval data structures.
- Uncertainty principles.
- Scaling.
- Parallel processing
- Tertiary storage for long file locking.
- Configuration and Versioning.
- Multi-project collaborative database management.
- Inter-company database management.
- Heterogeneous database to database management.
- Database browsing.
- Incomplete and inconsistent data.
- Mediators.
- National name service
- Security.
- Transaction management

It is clear by scanning the above items, (and by reading the more detailed description in the paper that there are no suggestions for research involving users.

Database Research 1990-1995

In 1995, the National Science Foundation reconvened the workshop to repeated the workshop and tasked the participants to update their previous report (Silberschatz, Stonebraker, & Ullman, 1996). The report indicated a substantial change in computer technology (including

the Web) in the preceding five years since the last report. The workshop participants recommended research in the following areas:

- “Problems associates with putting multimedia objects info DBMSs.
- Problems involving hew paradigms for distribution of information.
- New uses of databases.
- New transaction models.
- Problems involving ease of use and management of databases.” (p. 8)

To support the area of Ease of Use, the authors wrote,

The next generation of database management systems needs improved interfaces, not only for the end user, but for the application programmer and system administrator. A goal for user support is database system with the ease of use of spreadsheets, which are today often used as rudimentary database systems. (p. 16)

Clearly, through their recommendation that NSF fund research the workshop, the participants finally recognized the importance of user research on and regarding improved user interfaces.

Business School Database Research through 1995

Although the NSF sponsored reports do not specify that the research evaluated is restricted to that conducted at computer research facilities or not, one might wonder if research in non-computer oriented journals was included. Vincent Lai², of the Business School at the University of Alabama at Huntsville, conducted a study that compared database research conducted by Management Information System (MIS) researchers compared to computer scientists. Lai reported that, “Over the 15 year period from 1980-1995 a total of 725 research articles appeared in the 13 journals selected for this study. However, only 159 (21.9%) of these

² At the time he conducted the research, Vincent Lai was an Assistant Professor, in MIS, at the Business School, University of Alabama at Huntsville.

articles were the work of MIS researchers” (Lai, 1996, p. 39). Lai also reports that the majority of research articles are published in the *ACM Transactions on Databases*. Data from the tables in Lai’s article also show that the majority of MIS users published their articles in ACM SIGMIS journal, *Database*.

Of more importance, Lai lists the research by Topics and research approaches. The 14 topics include no mention of research using end users (unless such research be one of the 18 “other” research articles). The research approaches include “Conceptual Studies” for over half of the research articles, followed by Laboratory Experiment, Field study/survey, Review/Tutorial, Case studies, and Field Experiment (Lai, 1996, p. 41). Some of these might include end users but it is difficult to see from the summary.

Database User Research Articles available from the ACM Digital Library

Based on the Lai’s publication information, I searched the ACM digital database not only to look for research but to confirm that there were no database user studies prior to 1995. The results did not support my conclusions drawn from the three surveys; there definitely has been done previously using users. There were 185 articles in the ACM database under the general search terms “database user research” in article abstracts including one article with the terms in the title. Of those 91 were dated before 1995. When the search was restricted to database design articles three emerged dated 1980, 1983 and 2003. All three were thought pieces with no actual research.

Although I have not completely reviewed of all the articles the following is evident from the set of documents I have reviewed:

- Ninety percent of the 100 papers I reviewed are not actual research. Rather they are conceptual or information papers suggesting new ideas for research, research tools, conference research proceedings, and related items. should be conducted.
- Some of the papers dealt with research unrelated to database structures or users. For example:
 - Barbara M. Wildemuth, Charles P. Friedman, Stephen M. Downs conducted research on access to biomedical information using Hypertext versus Boolean access (CODASYL--The Stored-Data Definition and Translation Task Group, 1977)
Although it provided good user data on the interfaces, there was no intent to look at the underlying database structure.
- In some of the papers, researchers state that the results of their work or the concepts they tested were beneficial to users, but users were not involved in the research. For example:
 - Randy Katz' research states, "A heterogeneous database management system combines multiple dissimilar models of data within a single integrated system. The objective is to allow a user to access data independently of how it is actually organized" (Katz, 1980, p. 135). The research however explored the ability of database management systems to accommodate dissimilar data. No users were involved.
 - Helen Klein did some excellent work for her dissertation at University of Michigan exploring modeling of real world objects in object-oriented systems (Klein, 1998). The rationale underlying her research was based on Elaine Rosch's category theory on how humans use basic-level categories to simplify their understanding of things. Klein's argument is that using notions of basic class will allow database structures to

- more closely match users' mental concepts and therefore provide more friendly access to the data. Her conceptual work is therefore important for my research. She discusses the importance of adequately modeling end user concepts to reduce system complexity. Her research however, does not involve the user. Rather it measures the goodness of the database structure in terms of complexity of the software as measured by software metric suits accepted in the computer industry. Her findings indicate that systems that reflect a higher proportion of commonly used business domain concepts tend to have better structure. But the results, using software metrics were not significant.
- In some of the papers, user research is done through the user interface; and although the relationship of the interface to the database structure is recognized as important, it is not addressed in the research. For example:
 - Stephen Curl and his colleagues investigated “the relationship between the representation of a database view and individual differences as measured by spatial visualization ability. It reports a laboratory experiment where users were asked to make queries based on a database view that was represented either spatially or nonspatially” (Curl, Olfman, & Satzinger, 1998, p. 51). The database usability framework the researchers use shows that database structures have a controlled effect on the question formulation, but in the experiment only the views of the database structure were varied. So the results do not reflect database structure impacts on the user.
 - Occasionally a study does address the core problem and provide data that will be useful for my research. For example:

- Chan, Siu, and Wei recognize that the database design structure can impact the user through even though it is seen through the cloud of the user interface. And they correctly recognize that the interaction between the database structure and the interface may affect what the user sees regardless of the intent of the interface designer. Also significant are their comments on the importance of user research and how little prior research has been done.

With the proliferation of end-user computing and empowerment of end users in the organizations, the user-database interface is increasingly being recognized as an important research area. A user-database interface consists of two key components: the data model and the query language. Database management system designers can manipulate these two components to produce an effective database interface to maximize user task performance. However, in order to do this, the designers first need to know the effect of the two components on end users and the interaction effect of the two components. Despite the acknowledgment of user-database interface as a cornerstone in database utilization, there is still a paucity of research in this area. (Chan, Keng, & Wei, 1998, p. 31)

The research itself is also significant. The approach used three factors—data model (reflecting the database design), system characteristics (reflecting the interface), and task complexity. The data model and interface factors were crossed with repeated complexity in order to isolate the effects of each. Ninety subjects participated providing a high level of confidence. The result of the study was that for simple database queries (using both SQL and QBE) the effects of the interface and database structure were not significant. But for complex queries the effects were significant.

- Although it did not involve users, Helen Klein's research, discussed above, will assist with my research particularly in the conceptual area.

Conclusions

With this limited review, I found only one research report of a study directly related to assessing the impact of database design structures on user access to data. With additional time, I suspect that perhaps I could uncover more examples of research. For example, Chan, Siu, and Wei provide some 70 citations in their report (1998, pp. 43-46), some of which might refer to research that has been conducted.

Chan, Siu, and Wei's report provides excellent data and it is absolutely clear that additional research is needed showing the various aspects of database design on the user. For example, Klein's concepts could be developed into a study testing whether databases designed around basic classes would actually improve user performance at database information seeking tasks. I believe that such research would have a positive impact on future database design activities.

References

- Chan, H., Keng, S., & Wei, K.-K. (1998). The effect of data model, system and task characteristics on user query performance: An empirical study. *Database*, 29(1), 31-49.
- CODASYL--The Stored-Data Definition and Translation Task Group. (1977). Stored-data description and data translation: A model and language. *Information Systems*, 2(3), 95-148.
- Curl, S. S., Olfman, L., & Satzinger, J. W. (1998). An investigation of the roles of individual differences and user interface on database usability. *Database*, 29(1), 50-64.
- Katz, R. H. (1980). Hetrogeneous databases and high level abstraction. In *Proceedings of International Conference on Management of Data and Symposium on Principles of Database Systems, Pingree Park, CO* (pp. 135-137). New York: ACM Press.
- Klein, H. A. (1998). *Does modeling real world objects in object-oriented systems result in well-structured systems?* Unpublished Dissertation, University of Wisconsin, Ipselanti.
- Lai, V. S. (1996). An assessment of database research interest in MIS. *Database*, 27(2), 47-43.
- Silberschatz, A., Stonebraker, M., & Ullman, J. (1991). Database systems: Achievements and opportunities. *Communications of the ACM*, 34(10), 110-120.
- Silberschatz, A., Stonebraker, M., & Ullman, J. (1996). *Database systems: Achievements and opportunities* (No. STAN-CS-TR-96-1563). Stanford, CA: Department of Computer Science, Stanford University.