The Effects of Normalisation on the Ability of Business End-Users to Detect Data Anomalies: An Experimental Evaluation

A. F. Borthick
Georgia State University
School of Accountancy
Atlanta, Georgia 30302-4050 USA
Internet: borthick@gsu.edu

P. L. Bowen
Department of Commerce
The University of Queensland
Brisbane, Queensland, Australia 4072
Internet: bowen@lorien.commerce.uq.edu.au

M. R. Liu
Optus
Level 8, 50 Miller Street
North Sydney, NSW, 2060

F. H. Rohde
Department of Commerce
The University of Queensland,
Brisbane, Queensland, Australia 4072
Internet: rohde@commerce.uq.edu.au

With the proliferation of relational database programs for PCs and other platforms, many business end-users are creating, maintaining, and querying their own databases. More importantly, business end-users use the output of these queries as the basis for operational, tactical, and strategic decisions. Inaccurate data reduce the expected quality of these decisions. Implementing various input validation controls, including higher levels of normalisation, can reduce the number of data anomalies entering the databases. Even in well-maintained databases, however, data anomalies will still accumulate. To improve the quality of data, databases can be queried periodically to locate and correct anomalies. This paper reports the results of two experiments that investigated the effects of different data structures on business end-users’ abilities to detect data anomalies in a relational database. The results demonstrate that both unnormalised and higher levels of normalisation lower the effectiveness and efficiency of queries relative to the first normal form. First normal form databases appear to provide the most effective and efficient data structure for business end-users formulating queries to detect data anomalies.

(Keywords: End-user queries, Data quality, Normalisation)

1. INTRODUCTION
Since the late 1980s, the availability and usability of relational database programs has significantly increased, especially for mini- and personal computers. This increase has led to a proliferation of databases being created and maintained by business end-users. These databases are often queried to provide information for important business decisions. The quality of these decisions depends on the