The Detection of Data Errors in Computer Information Systems: Field Interviews with Municipal Bond Analysts

BARBARA D. KLEIN, University of Michigan-Dearborn, USA

There is strong evidence that data stored in organizational databases have a significant rate of errors. As computerized databases continue to proliferate, the number of errors in stored data and the organizational impact of these errors are likely to increase. The impact of data errors on business processes and decision making can be lessened if users of information systems are able and willing to detect and correct data errors. However, some published research suggests that users of information systems do not detect data errors. This paper reports the results of a study showing that municipal bond analysts detect data errors. The results provide insight into the conditions under which users in organizational settings detect data errors. Guidelines for improving error detection are also discussed.

There is strong evidence (e.g., Laudon, 1986; Morey, 1982; Redman, 1992) that data stored in organizational databases have a significant rate of errors. Between one and ten percent of data items in critical organizational databases are estimated to be inaccurate (Laudon, 1986; Madnick and Wang, 1992; Morey, 1982; Redman, 1992). This estimate is based on the findings of several studies such as those by Laudon (1986) and Morey (1982). Inaccurate data have been reported in a student loan database maintained by the U.S. Department of Education (Knight, 1992), in records maintained by the U.S. Department of Agriculture (“Dead Farmer,” 1992), and in records maintained by credit reporting bureaux (“Consumer Enemy,” 1991).

As computerized databases continue to proliferate, and as organizations become increasingly dependent upon these databases to support business processes and decision making, the number of errors in stored data and the organizational impact of these errors are likely to increase. Indeed, Mason (1986) argues that the scope of this problem is such that data quality will become an important issue facing MIS managers. Redman (1992) argues that inaccurate and incomplete data may adversely affect the competitive success of an organization. For example, strategies such as total quality management may be difficult to implement if the data needed to support the decisions required by the strategy are not of adequate quality (Fox et al., 1993; Madnick and Wang, 1992; Redman, 1995). Errors in data can have a significant financial impact on organizations. For example, Dun & Bradstreet paid $350,000 to a construction company after they incorrectly reported that the company was bankrupt because a Dun & Bradstreet employee had entered inaccurate data into their credit reporting system (Percy, 1986).

Two main approaches to managing data errors in organizations are: (1) validating data as they are input to or stored in databases (e.g., Morey); and (2) depending on users to detect and correct errors. While useful, automated approaches to data validation do not generally yield completely accurate data. The best approach to reducing data errors in most organizations will include both automated data validation and user detection of errors.

Early research suggested that users are poor at error detection (Davis et al., 1967; Laudon, 1986; Ricketts, 1990). The editor’s executive overview of one study on user error
The remaining sections of this paper present (1) a review of prior research bearing directly on the question of the conditions under which individuals detect errors in data, (2) a theory of error detection, (3) the design of the field interview study, (4) the results of the study of the municipal bond analysts, and (5) a discussion of the implications of the results for information system managers and users of information systems.

BACKGROUND

In a broad sense, this investigation falls in the literature on data quality. Several general conclusions can be drawn from the existing research on data quality. First, while no single definition of data quality has been accepted by researchers working in this area, there is agreement that data accuracy, currency, and completeness are important areas of concern (Agnon and Ahituv, 1987; Davis and Olson, 1985; Fox et al., 1993; Huh et al., 1990; Madnick and Wang, 1992; Wand and Wang, 1994; Zmud, 1978). Second, while it is difficult to compare error rates across studies, rates substantially greater than zero have been found in all of the studies addressing the extent to which data errors exist in databases (Ha1985; Johnson et al., 1981; Knight, 1992; Laudon, 1986; Stone and Bublitz, 1984). Third, there is disagreement about the extent to which efforts to purge all errors from databases should be attempted. Some researchers propose methods designed to completely rid databases of errors (Janson, 1988; Svanks, 1988; Naus, 1975; Parsaye and Chignell, 1993), while others propose tools for determining how to best allocate limited resources to controlling the level of data errors (Ballou and Pazer, 1987; Ballou and Tayi, 1989; Ballou et al., 1987; Bowen, 1992; Paradice and Fuerst, 1991). Fourth, many researchers argue that users need not discard data containing errors. A variety of approaches for using imperfect data have been suggested (Ballou and Pazer, 1985; Ballou and Pazer, 1995; Ballou and Pazer, 1987; Bansal et al., 1993; Gaba and Winkler, 1992; Garfinkel et al., 1986; O’Leary, 1993; O’Neill and Vizine-Goetz, 1988).

This study fits into the literature on data quality that describes ways in which users of information systems might modify their use of data if they are aware that errors exist in data. In general, the early literature on data quality argues that users are not very capable of finding errors in data and then altering the way in which they use the data. More specifically, there is considerable evidence of poor user performance in detecting data errors. Davis et al. (1967) conducted a field experiment in which individuals were mailed banking confirmation statements with imbedded errors. The individuals were asked to verify their account information, and approximately half failed to detect important errors. Laudon (1986) found that users of criminal information systems rarely detected errors in these records even though information provided to police departments by the FBI is accompanied by a warning stating that the user should verify that the information is accurate. Ricketts (1990) conducted a laboratory experiment in which over ninety percent of the subjects failed to detect a substantial data error in production planning reports. The failure of humans to detect errors in data is also assumed in much of the literature on data quality in which it is argued that resources should be devoted to the up-front improvement of the quality of data in organizational databases (Redman, 1992; 1995).

While prior findings (e.g., Davis et al., 1967; Laudon, 1986; Ricketts, 1990) suggest that error detection is not a frequently occurring phenomenon, the findings of a field study of actuaries suggest that users in at least one professional domain detect data errors (Klein, 1997). This finding raises the question of whether error detection in organizational settings is a rare phenomenon occurring only under a limited set of conditions or a more frequently occurring phenomenon. This paper addresses this question by investigating the phenomenon in another professional domain: municipal bond analysis.

THEORETICAL FRAMEWORK

A theory of individual task performance and theories of effort and accuracy in decision making underlie this research.

Theories of Individual Task Performance

Theories of individual task performance provide some general guidance for identifying conditions under which users detect errors in data. For example, experience seems to affect performance in general and may affect performance in error detection. One theory we could use is Campbell’s (1990; Campbell and Pritchard, 1976) theory of individual task performance (depicted in Figure 1). This theory suggests that experience (e.g., Weber et al., 1993), knowledge, and effort (e.g., Payne, 1982; Payne et al., 1988) all affect error detection. Campbell’s (1990) theory argues that performance on a