QUALITY MANAGEMENT APPROACH TO KEYSTROKE VERIFICATION

Robert F. Bailey, Robert Miller, and Howard Speizer, National Opinion Research Center
Howard Speizer, NORC, 1155 E. 60th St., Chicago, Illinois

KEY WORDS: Data entry, keystroke verification, total quality management

Keystroke verification detects errors caused by data entry operators misreading and mis-transcribing source documents. A number of approaches have been used over the years to verify the accuracy of keying in computer-assisted data entry (CADE) and key-to-disk applications at NORC. Recently, we developed an automated system based on quality management principles. This new verification system uses process control techniques to measure keystroke error. It provides analytical tools that enable supervisors to identify and address significant problem areas. This paper reports on the design of the system, its benefits to the quality of the product, and its limitations both in design and in how it was implemented.

The general method of verifying the accuracy of data entry is to key data twice, compare results, calculate error rates, and provide feedback to data entry operators. Error rates can be calculated by having the data keyed a third time and allowing a computer program to choose the correct answer based on a majority-rules algorithm or, more commonly, by having a supervisor adjudicate discrepancies between two data sets. Generally, independent re-keying by a second, different operator is preferred, although a dependent, mass-inspection approach to verification persists.

Regardless of the data entry verification method used, little work has been done in the past to determine the best approach to analyzing error rate statistics. Often the data entry supervisor provided feedback to operators about all errors detected in an arbitrary percentage of cases selected for verification. Usually, only mean error rate was used to monitor and report on the overall accuracy of keyed data. Few questions have been asked about the right number of cases to verify, the efficacy of the feedback, and the best approach to improving the data entry process across different applications.

We re-designed NORC's keystroke verification system to take advantage of process control techniques. The five main objectives we defined for the system were that it must 1) provide data on the most important problems on a timely basis to improve the data entry process, 2) create an accurate picture of the quality of the keyed output, 3) be cost-effective to operate and maintain, 4) be easy for the data entry supervisor to learn, use, and manage, and 5) be a standard procedure which can be used with little modification for all data entry projects.

New System

The keystroke verification system is used for all NORC CADE and key-to-disk applications. Cases are chosen at random for inclusion in the verification sample (usually 10% of the cases are selected). The initial operator does not know which cases will be re-keyed. The data entry supervisor operates all components of the system. The supervisor assigns a second, different operator to do the re-keying. The second operator does not have access to the data keyed by the first.

After the case is keyed a second time, the system automatically compares the two sets of data on an item-by-item basis and produces a discrepancy report. The supervisor adjudicates any discrepancies by referring to the source document and determines the number of errors made by the initial data entry operator. The supervisor uses observations made during adjudication to provide immediate feedback to operators.

The verification system automatically produces a database entry for every case in the verification sample. The database record includes a reference to the survey, instrument, and identification number of the case that was re-keyed. The record also includes the date the case was keyed, the initials of both data entry operators, the number of opportunities for error, and the number of errors made by the first operator.

The notable feature of this system is the way in which these data are analyzed. With a simple command the system produces a control chart for each data entry application. The control chart maps all of the error rates calculated for verified cases in a time series. The program calculates and plots on the chart both the upper control limit and the mean error rate. The upper control limit is based on a function of the standard deviation around the mean and provides an upper limit under which most normally-distributed observations should fall.[1] See Figure 1 for an example of an error-based control chart.

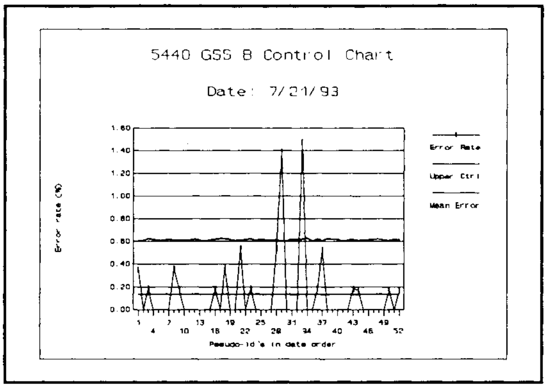

Figure 1

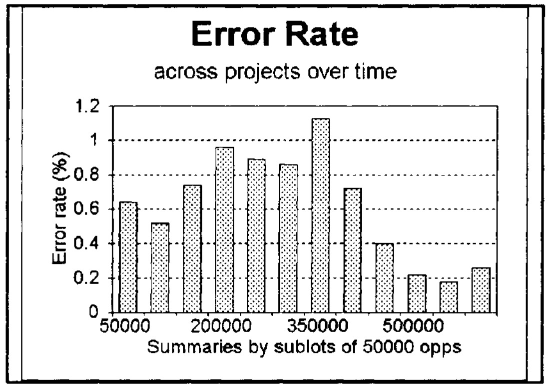

Figure 2

The control chart is designed to provide the CADE supervisor with more insight into the quality of the keying process than can be obtained by simply considering the mean error rate. The control chart differentiates between random error and special-cause error. The cases with error rates below the upper control limit contain random rather than special-cause errors. Tampering with the keying process in response to random error may only introduce more variation. When observations fall above the upper control limit, an in-depth investigation is undertaken to bring the keying process into statistical control. After the process is in statistical control, a design change would be required to lower the mean error rate and reduce the amount of random error.

Implementation Results

The verification system quickly became a standard used by all data entry projects. Supervisors were able to learn the mechanics of the system and to produce the control charts with relative ease. Programming steps were simplified and the cost of installation became less than it had been when a variety of customized approaches were used. The simplicity of installation led to the more timely delivery of verification applications. Having the verification system in place at the beginning of data entry production improved our ability to provide feedback to operators at this critical time. Verification systems used in the past were more difficult to install, and were often operational only some time after the beginning of production. Feedback to operators was delayed and errors caused by operator misunderstanding became ingrained behaviors that were more difficult to correct. Errors were repeated which might have been avoided.

The new keystroke verification system provided a consistent method for measuring errors introduced by data entry operators. For the year-long period starting in March 1992, the error rate averaged .6 percent and showed improvement over time. See Figure 2 for a summary of error rates over time.

Analyses of special-cause errors provided supervisors with insights about problematic sequences on specific projects. They were able to suggest modifications to data entry programs to eliminate ambiguity, and provide feedback and retraining to operators about important, non-random errors.

Examining the special-cause errors obtained on a number of projects enabled managers to identify a few "quick-fix" changes to CADE application programming which could be implemented to the benefit of all projects. One example is worth noting. Operators are more likely to lose their place in long series of similarly-formatted questions with identical response categories than in other question series. They do not realize that they are out of synchronization with the data entry application until they make an out-of-range entry, when the format and response categories have changed. In some instances, this realization occurs only after a long sequence of incorrect entries. Even after recognizing the problem, operators may have difficulty finding the point at which the initial error occurred. By examining a few similar situations across data entry projects, the CADE supervisors and programmers were able to design a solution which provides operators with intermediate checkpoints within these long series of similar questions. The checkpoints not only limit the number of incorrect entries before detection, but also allow operators to return to the previous correctly-keyed checkpoint and re-attempt the problematic sequence more carefully, without having to reconstruct their original error.

Two Shortcomings of Our Approach

The control chart was designed as an analytical tool that would help data entry supervisors do their jobs more effectively and improve the performance of operators. It became apparent, however, that the supervisors produced the control charts for the authors rather than for their own use, they did not use the charts to make decisions about providing feedback, and they did not entirely trust the data presented on the charts. Decisions about providing feedback to operators were still made during the adjudication of discrepancies, before control charts had been produced. Feedback to operators was not informed by the distinction on the control charts between important, special-cause errors and less-significant, random errors.

The analysis of errors led to fewer improvements to data entry in general than we expected. We had hoped that the identification and analysis of special-cause errors would provide more insights about general changes to the data entry process and produce a marked improvement in keying accuracy on all projects.

Where We Went Wrong

Our mistake quickly became apparent during discussions with the data entry supervisors. We had not fully involved them (or the operators) in the design of the system, and so failed to benefit from their first-hand experience. Instead of including them as a part of the design team, we relied on our own understanding of their jobs and so failed to engineer the system to their requirements. We gave the supervisors a system to use rather than one that responded to their needs. Because they were not familiar with statistical process control techniques, supervisors did not use the system exactly as intended and were not entirely convinced that using control charts to drive feedback would result in the right type or amount of feedback. Their main concern was that complete reliance on control charts might cause them to overlook isolated, non-random errors that revealed a lack of understanding on the part of the operator when these errors did not happen to result in an error rate above the upper control limit.

Discussions with the data entry supervisors revealed why we did not improve data entry accuracy as much as we had hoped. First, we discovered that data entry operators make very few errors. The error rate (although too high to ignore) does not leave much room for improvement. Eventually, it becomes too expensive to significantly improve keying accuracy.

Second, and most important, we found that we could not rely on supervisors to contribute to the identification and resolution of cross-project data entry problems. Individual supervisors were more concerned with monitoring and ensuring accuracy on their own data entry projects than with identifying higher-level problems and discovering solutions that would benefit the data entry effort in general. We had not established a forum to analyze special-cause errors and tended to work with the supervisors individually. We failed to develop a team approach to the identification and elimination of special-cause errors common to many projects.

The third reason we did not make as much progress with reducing error rates as we had anticipated was more obscure. We discovered that supervisors subjectively suppress certain classes of error, preventing them from being captured in the verification database and contributing to the broader analysis of the quality of the process. During adjudication, the supervisor examines all discrepancies between the data keyed by the original operator and the verification operator and assigns an error whenever a mistake was made by the first operator. The error assignment process is somewhat subjective. If, for example, the supervisor felt that the first operator did not accurately key a handwritten numeric response, but that this was due to illegible handwriting, the supervisor might be inclined to decide that the original operator was not to blame and not assign an error. Similarly, errors could be omitted from the error count if the operator had keyed data exactly as recorded on the source material, but had failed to correctly interpret a marginal note or notice and correct an error made earlier in the process (by the respondent, interviewer, coder, or editor). The result is that some errors are masked in this verification procedure, get suppressed from the database, and cannot be used to improve the larger process of accurately capturing data from respondents.

Item Discrepancy Data

Because the adjudication process masks some errors and takes time to perform, we considered using the item discrepancy data directly to produce control charts prior to adjudication. A discrepancy occurs whenever one operator keys a value differently than the other operator. The verification process remains unchanged with a second verification operator independently keying the data and the computer automatically comparing the two data sets. Instead of having the supervisor assign errors by examining each discrepancy, the system calculates the ratio of the

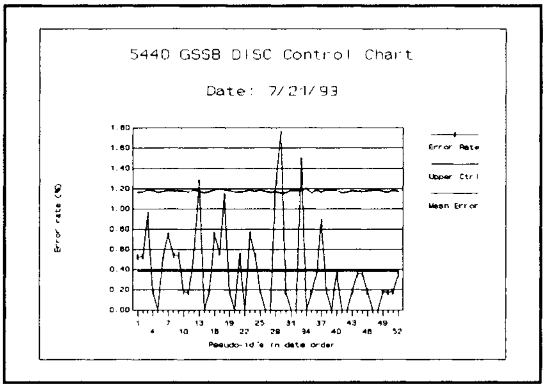

Figure 3

number of discrepancies to the total number of opportunities for discrepancies. Except for instances in which supervisors do not assign errors because of "outside" factors, discrepancies should occur twice as often as errors. Everything else about their distribution should be similar to the distribution of errors. Plotting discrepancies on a control chart using the same formulas for the mean and upper control limit produces a chart which, although amplified one order of magnitude from the error chart, can be used to reach the same conclusions as those reached by analyzing the error chart. See Figure 3 for an example of a control chart using discrepancy data.

Using discrepancies to inform supervisors about the quality of the data entry effort eliminates the time and expense required for full-scale adjudication of discrepancies. Adjudication can, and should, be performed on a small set of cases to determine an error rate and confirm that it is at a tolerable level. Adjudication should also be performed whenever the discrepancy rate exceeds the upper control limit. Discrepancy control charts can be used to decide when, whether, and about what to provide operator feedback. One other advantage of using discrepancies in the analysis, as mentioned before, is that special-cause errors resulting from operations outside of data keying will be captured and, if systematic, can be teased out from the data and used to improve the process.

Conclusions

Analyzing and using discrepancies instead of errors as a basis for making decisions about operator feedback is an idea worth pursuing. Using discrepancies instead of errors eliminates from the process the primary feature supervisors currently rely on to determine the quality of the work performed by their operators. Supervisors need to learn more about the theory and ideas on which the verification system is built before moving forward. They need to be a part of the analysis and design team and become owners instead of users of the verification system. Once this happens, and the supervisors trust the control charts to guide their decisions, we can consider applying discrepancy data analysis more broadly.

We have shown that NORC's new verification system provides an effective method for improving the data entry process. The system is designed to eliminate the cost of investigating and providing feedback to operators about random errors that occur when the process is in statistical control. We need to refine our approach so that, with the help of the data entry supervisors, we can identify problems which cut across different applications and provide insight about other factors that contribute to data entry error. These factors include questionnaire design and formatting, software functionality, and ergonomic considerations such as keyboard design, lighting, seating, and work schedules. We expect that identifying the effects of these factors will enable us to continue to modify the process and improve data entry accuracy.

1. Ishikawa, Kaoru: *Guide to Quality Control*, 2nd Edition, p 79-81. Tokyo, Asian Productivity Organization, 1985.