Need for Speed: Which Data Checking Method is Fastest?

Ryuhei Kawamoto, Maryssa Nagata, Matt Roe, Loise M. Ladrazo, and Kimberly A. Barchard

University of Nevada, Las Vegas

UNLV

ABSTRACT

Even a small number of data entry errors can significantly lower the accuracy of a data set (Day, Fayers, & Harvey, 1998). Thus, researchers have been recommending the use of data checking (Barchard & Pace, 2011). However, data enterers who need to assess a large number of people per day, such as mental health providers, are often pressed for time (Hawley, Gale, Sivakumaran, & Littlechild, 2010). Data checking can be time consuming. The purpose of this study was to compare the time it takes to use different data checking techniques, so that we can recommend which data checking techniques should be used by people who have limited time. In this study, four data checking techniques were compared: double entry, visual checking, solo read aloud, and partner read aloud. A total of 27 participants were randomly assigned to one of these four techniques. The participants then checked 20 data sheets while they were being timed. We compared the time to check the data using one-way ANOVA. There was no significant difference between the four data checking techniques in terms of the time it took to check the data. However, the fastest data checking technique (double entry) was seven minutes faster, on average, than the slowest data checking technique (visual checking). These non-significant findings contradict previous research. Barchard and Verenikina (2013) found that double entry took longer than visual checking and partner read aloud. The data collection on our study is still in progress. With a larger sample size, the differences between the groups may become statistically significant. During the poster session, a free double entry system will be available.

INTRODUCTION

Even a small number of data entry errors (5%) can lead to incorrect statistical results and research conclusions (Day, Fayers, & Harvey, 1998). Therefore, methodologists recommend the implementation of data checking techniques (e.g., Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado, Hinotsu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003; King & Lashley, 2000). Several different data checking techniques exist. In the visual checking technique, a person visually compares what is on the computer screen to the original data sheets (Barchard & Verenikina, 2013). In the partner read aloud technique, the original data sheets are read aloud by one person, while another person visually checks that they match the data shown on the computer screen (Barchard & Verenikina, 2013). In the solo read aloud technique, one person simultaneously reads the data sheets and visually checks the computer screen (Barchard & Verenikina, 2013). Lastly, in the double entry technique, data are entered a second time and then the computer identifies any mismatches between the two data sets, as well as any values that are outside of an allowable range (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003).

These data checking techniques are helpful for many fields, including mental health services where the quality of information is important. For example, Heilbrun, DeMatteo, and Marczyk (2004) note that the results of mental health assessments have a large influence the formation of patients' treatment plans, and so recommend that mental health assessment take place in undisturbed environments in order to ensure the validity of results. However, such environments are not always available for providers who have to assess high-risk clients such as patients in hospitals or prisons (Hawley et al., 2010; Heilbrun et al., 2004). In fast-paced and noisy environments, data checking can help increase the quality of the data and accuracy of the conclusions. Unfortunately, those providers who work in uncontrolled environments are often pressed for time (Hawley, Gale, Sivakumaran, & Littlechild, 2010), and the process of checking data can be time consuming. The purpose of this research is to determine which data checking technique is fastest.

METHOD

Participants

In total, 27 undergraduate students (9 male, 18 female) participated in our study in return for course credit. The participants mean age was 21.48 (SD = 6.22) years, with the lowest age 18 and highest 46. Participants reported their ethnicities as follows: Caucasian 40.7%, Hispanic 18.5%, Asian 14.8%, African-American 14.8%, Pacific Islander 7.4%, and other 3.7%.

Procedure

This study was conducted during 90-minute sessions supervised by a trained administrator. Each 90-minute session was broken into two parts. In the first part, participants learned how to use Excel via instructional videos. After being randomly assigned to one of the four data checking techniques (visual checking, solo read aloud, partner read aloud, or double entry), participants watched another instructional video about how to perform their assigned data checking technique. Participants practiced their data checking techniques using five data sheets. The administrator corrected any errors the participants were making.

In the second part of the study, participants checked the data from 20 additional data sheets. This is the data that was used to compare the speed of the four data checking methods.

Data Analysis

We calculated how long it took each participant to complete the data checking by taking the difference between the start and end time for the second part of the study. To compare the time it took for the four data checking techniques, we used one-way ANOVA.

RESULTS

There were no significant differences in the time it took to check the data using the four techniques, F (3, 23) = 1.04, p = .396. However, there was a 7-minute difference between the fastest technique (double entry) and the slowest (visual checking; see Table 1).

Table 1

Time to Complete Data Checking (in min)

Technique	Mean	Standard Deviation
Double Entry	24.57	0.0
Visual Checking	32.26	7.06
Solo Read Aloud	31.72	5.34
Partner Read Aloud	28.53	4.46

Note. The double entry technique had a standard deviation of 0 because only one participant was randomly assigned to use this data checking technique.

DISCUSSION

The purpose of our study was to determine which data checking technique is fastest. The techniques tested were visual checking, solo read aloud, partner read aloud, and double entry. After randomly assigning participants to data checking techniques, we timed how long it took them to check a set of 20 data sheets. There were no significant differences in time to complete the four data checking techniques. However, the fastest data checking technique (double entry) was seven minutes faster, on average, than the slowest data checking technique (visual checking). These results contradict a previous study in which the double entry technique took longer than the read aloud and visual checking techniques (Barchard & Verenikina, 2013). This contradiction is most likely due to the fact that we only had one participant who completed the double entry technique. The data collection on our study is still in progress. With a larger sample size, the differences between the groups may become statistically significant.

Previous research has also found that double entry is more accurate than other data checking techniques (Barchard & Verenikina, 2013; Johnson, Temple, Pearsons, & Webster, 2009; Kawado et al., 2003). Therefore, unless double entry takes significantly longer than other data checking methods (and this study found no significant differences in the time), we recommend researchers use double entry because it drastically reduces the number of errors. A free double-entry system is available and from the following website http://faculty.unlv.edu/barchard/doubleentry/ and during this poster session.

REFERENCES

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior, 27*, 1834-1839. doi:10.1016/j.chb.2011.04.004

Barchard, K. A., & Verenikina, Y. (2013). Improving data accuracy: Selecting the best data checking technique. *Computers in Human Behavior, 29* doi:10.1016/j.chb.2013.02.021

Day, S., Fayers, P., & Harvey, D. (1998). Double data entry: What value, what price? *Controlled Clinical Trials, 19*, doi:10.1016/S0197-2456(97)00096-2

Johnson, C. L., Temple, G. M., Pearsons, T. N., & Webster, T. D. (2009). An evaluation of data entry error and proofing methods for fisheries data. *Transactions of the American Fisheries Society, 138*, 593–601. http://dx.doi.org/10.1577/T08-075.1

Kawado, M., Hinotsu, S., Matsuyama, Y., Yamaguchi, T., Hashimoto, S., & Ohashi, Y. (2003). A comparison of error detection rates between the reading aloud method and the double data entry method. *Controlled Clinical Trials, 24*, doi:10.1016/S0197 2456(03)00089-8

King, D., & Lashley, R. (2000). A quantifiable alternative to double data entry. *Controlled Clinical Trials, 21*, 94–102. doi:10.1016/S0197-2456(00)00042-8

Hawley, C. J., Gale, T. M., Sivakumaran, T., & Littlechild, B. (2010). Risk assessment in mental health: Staff attitudes and an estimate of time cost. *Journal of Mental Health, 19*, 88-98. doi:10.3109/09638230802523005

Heilbrun, K., DeMatteo, D., & Marczyk, G. (2004). Pragmatic psychology, forensic mental health assessment, and the case of Thomas Johnson: Applying principles to promote quality. *Psychology, Public Policy, and Law, 10*, 31. doi:0.1037/1076-8971.10.1-2.31