

## Data Checking: Which Method is Superior?

Theresa M. Ross, Ernesto H. Bedoy, Ashley A. Anderson, & Kimberly A. Barchard  
University of Nevada, Las Vegas



**Reference:** Ross, T.M., Bedoy, E.H., Anderson, A.A., & Barchard, K.A. (2014, April). *Data checking: Which method is superior?* Poster presented at the Western Psychological Association annual convention in Portland, OR.

**Contact Information:** Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, P.O. Box 455030, Las Vegas, NV, 89154-5030, USA, barchard@unlv.nevada.edu

### Abstract

After research data are input into the computer, researchers use various data checking methods to find and correct possible data entry errors. Not all data checking methods are equally effective. The purpose of this study was to determine which data checking method is the most effective at finding and correcting errors. We examined four data checking methods: visual checking, solo read-aloud, partner read-aloud and double entry. Before participants arrived for the study, we created a dataset that they would check. This Excel dataset was deliberately created to contain 32 errors. Participants' job was to locate and correct those errors. Previous research has shown that double entry is more effective than visual checking (Barchard & Pace, 2011) or partner read-aloud (Kawado, Hinotsu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003). Although no research has examined solo read-aloud (where the same person reads the data sheet out loud as looks at the computer screen to ensure it was entered correctly), we hypothesized that double entry would be more effective than all three of the remaining methods.

Twenty-seven undergraduates participated in this study in return for course credit. These participants were randomly assigned to one of the four data checking methods. Participants checked the entries in the Excel file against the data on twenty data sheets. After each participant finished data checking, we calculated the number of errors that remained in the dataset. Double entry had the fewest errors, while partner read-aloud had the most errors. These results supported our hypothesis that double entry is the most effective. Future research should replicate these results using a larger sample size, so that we can distinguish between all four of the data checking methods. In addition, future research should use more experienced data entry personnel, such as graduate students and paid professionals, to extend the generalizability of these results.

### Introduction

Have you ever calculated a complex mathematical problem only to get the answer wrong because you typed a number wrong? It can be quite frustrating because you would have gotten the answer correct if you had only went back to check the data you entered. When researchers or doctors input incorrect data, their results could be catastrophic to their practice. Consider what happens when a researcher enters participant information incorrectly. . When the researcher runs the statistical analysis, the researcher will be using incorrect data which increases the possibility of incorrect significance. Moreover, having incorrect data increases to the chance of having an outlier which will skew results significantly (Wilcox, 1998). Consider what happens when a doctor enters patient information incorrectly. This increases the chance of a misdiagnosis. Thus, in both research and other work-related settings, data checking increases the accuracy of results and conclusions.

There are several different methods of checking data. These include partner read-aloud, solo read-aloud, double entry, and visual checking. Previous research has shown that double entry is more effective than visual checking. One study found that visual checking yielded 2958% more errors than double entry (Barchard & Pace, 2011). Double entry is also more accurate than a read aloud method where the computer reads the data out loud and the operator visually checks the original data sheet (Kawado et al., 2003). Double entry detected 69.03% of errors when the same operator did the data checking as did the original entries, whereas read aloud only detected 39.9% of the errors (Kawado et al., 2003). Similarly, double entry detected 88.3% of the errors when a different operator did the checking, whereas read aloud only detected 59.5% of the errors (Kawado et al., 2003). Thus, past research has unanimously concluded that double entry is the most accurate. However, no study has compared all four data entry techniques to each other. The purpose this study is to compare all four data checking techniques – double entry, visual checking, solo read aloud, and partner read aloud – to determine which method is the most accurate. We hypothesize that our results will support previous research and show that double entry is the most accurate.

### Methods

#### *Participants*

Participants were 27 undergraduate students (18 female, 9 male). Participants identified as the following: African American (14.8%), Asian (14.8%), Caucasian (40.7%), Hispanic (18.5%), Pacific Islander (7.4%), and Other (3.7%). The ages ranged from 18 to 46 years old ( $M = 21.48$ ,  $SD = 6.216$ ).

#### *Procedures*

Participants were tested individually during 90-minute supervised sessions. After reading and agreeing to the informed consent, participants watched a video tutorial about Microsoft Excel. Next, participants were randomly assigned to the data checking methods and watched a video tutorial explaining their assigned methods. Participants practiced their assigned method by checking the data for five data sheets. Finally, participants completed the main portion of the study, in which they checked the data on 20 data sheets.

Each data sheet contained 26 pieces of data. These data sheets had already been entered into the computer. Participants' job was to check that the data were entered correctly. When the data were entered, however, 32 errors were deliberately introduced into the Excel datasheet. Participants' job was to find and correct these errors.

Four data checking methods were used to identify errors. For visual checking, participants visually compared the Excel file with the original paper data sheet. For double entry, participants entered the data a second time. The Excel program highlighted mismatches and entries that were too high or too low. For solo read-aloud, participants read aloud the data that was on the original data sheet, and then visually compared that with the data that was shown in the Excel file. Finally, for partner read-aloud, administrators read the data sheet out loud while participants looked at the Excel file. If participants heard something that did not match what was shown in the Excel file, participants asked the administrator to repeat the data, by saying the word "verify". If the re-read data did not match, then participants knew this was an error.

#### *Data Analysis*

To compare the effectiveness of the four data checking methods, we used a one-way ANOVA. The independent variable was the group each participant belonged to (visual checking, double entry, solo read aloud, and partner read aloud). The dependent variable was the number of errors left in the data set after the participant checked it.

### **Results**

There were significant differences between the four data checking techniques ( $F(3,23) = 3.18, p = .043$ ). Double entry had the fewest errors and partner read-aloud had the most. See Table 1.

Table 1  
*Average Number of Errors Left in the Data Set after Data Checking*

Data Method	Checking	Mean	Standard Deviation
Visual Checking		1.33	1.86
Partner Read Aloud		2.33	1.52
Double Entry		0.00	0.00
Solo Read Aloud		2.00	1.73

### **Discussion**

The results supported our hypothesis by indicating double entry as the most effective method. In this study, double entry was shown to be the most accurate method for checking data. In double entry, the Excel file shows the number of discrepancies and out of range errors. Moreover, the person knows exactly which entry to correct because it is highlighted in the Excel file. Perhaps these user friendly features partially explain why double entry is the most effective data checking method. On the other hand, these results should be viewed with caution. Because participants were randomly assigned to groups and because there were very few participants, only a single participant was assigned to the double entry condition. These results should therefore be replicated with a larger sample size. With a larger sample size, we would be more confident that the results would generalize to other data checkers.

This study also found that partner read aloud resulted in the most errors being left in the data set after the participants finished data checking. There are many possible reasons for this poor performance. Perhaps the administrators read the data sheets too fast. Perhaps the participants did not want to interrupt the administrators, because interrupting is considered rude. Or perhaps the participants noticed the discrepancy between the Excel sheet and what they heard, but they were not confident that they have read the entry correctly; the next datum was read before they had made their decision, and their attention was torn away from the discrepant entry. These possibilities may suggest why solo read-aloud contained fewer errors. In solo read-aloud, the participant did not have to think about interrupting the administrator, and they were able to read aloud at their own pace.

Future research should examine differences between data checking methods with more experienced data entry personnel, such as graduate students and paid professionals. Undergraduate students may have little experience or motivation with data entry. In real research studies, the people who are entering and checking data would usually have more experience than the introductory psychology students used in this study. Including experienced data entry personnel in a future study would allow us to generalize results more widely.

### **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*, 834-1839.
- 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*, 560-569.
- Wilcox, R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*, 300-314. doi:10.1037/0003-066X.53.3.300.