Which Data Checking Method is More Accurate?

UNLV

Sarah Cobb, McNair Scholar, Psychology Major
Dr. Kimberly A. Barchard, Faculty Mentor, Department of Psychology

The Center
CENTER FOR ACADEMIC ENRICHMENT & OUTREACH

Abstract

Researchers use multiple methods for data checking. Each method can help identify and fix errors that were introduced during the data entry process. Fixing the errors that were introduced during the data entry process increases the accuracy of the research results. Accuracy is important because if a researcher publishes inaccurate results other researchers would not be able to replicate those results and draw the same conclusions. The purpose of this study is to compare the accuracy of four different data checking methods: double entry with one person, double entry with two people, visual checking, and solo read aloud. So far, previous research has shown that double entry is more accurate than visual checking (Barchard & Pace, 2011) and partner read aloud (Kawado, Hinotsu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003). Although there has not been many studies done on the comparison of these four methods and only one study has used solo read aloud, double entry has been shown to produce the highest quality data. I therefore hypothesize that the two double entry methods will have the highest accuracy.

Introduction

There are four common data checking methods: single person double entry, two person double entry, read aloud, and visual checking. The single person double entry method consists of the one person entering and checking data. The two person double entry method has one person entering the data and a second person entering the data a second time and checking that they match. The read aloud method has one person entering the data and either the same person (solo read aloud) or a different person (partner read aloud) checking the data by reading it aloud. The visual checking method consists of one person entering the data and checking the data visually. One study that compared three different data checking methods found that two-person double entry produces fewer errors but takes longer than other data checking methods (Barchard & Verenikina, 2013). Through comparing the accuracy of the four different data checking methods this study will be able to identify which method produces the fewest errors.

Methods

Participants

There will be 100 participants for each data checking method, giving a total of 400 participants. Participants will be undergraduates at the University of Nevada, Las Vegas. Participants will be recruited from the Department of Psychology subject pool.

Materials

Measures

Accuracy will be measured by the numbers of errors on the participant's final Excel sheet. An error is defined as a discrepancy between the Excel sheet and what was actually on the data sheets. The data checking method that produces the greatest number of errors will be considered the least accurate. The method that produces the least number of errors will be considered the most accurate data checking method.

Data Analysis

To compare the accuracy of the four data checking methods, an ANOVA will be calculated. The independent variable will be the group each participant belongs to (one-person double entry, two-person double entry, solo read aloud, or visual checking). The dependent variable will be the number of errors left in the Excel sheet after the participant has completed entering and checking data.

Literature Review

There have been many different data checking methods. These include solo read aloud, partner read aloud, visual checking, and double entry with one or two people. Double entry has consistently been shown to be the most accurate (Barchard & Pace, 2011; Gibson, Harvey, Everett, & Parmar, 1994; Paulsen, Overgaard, & Lauritsen, 2012) and has been defined as the definitive gold standard of good clinical practice (Paulsen et al., 2012). However, even though double entry may find the most errors, not all researchers have concluded that it benefited their research (Gibson et al., 2012). The biggest problem with the double entry method is that it takes up a greater amount of time. Double entry requires up to 37% more time than other data checking methods (Reynolds-Haertle & McBride, 1992).

Regardless of which data checking method is used, accuracy rates increase when the data checking person is someone different from the data entry person. One study found that read aloud detected about 60% of the errors when a different person did the checking, but only 39.9% of errors when the same person did the checking and the original entering. Similarly, double entry detected 88.3% of errors using different operators, but only 69% of the errors when the data checking person was the same as the original data entry person (Kawado et al., 2003). Because of this, we hypothesize that double entry will be more accurate than read aloud or visual checking, and that double entry with two people will be more accurate than double entry with one person in the present study.

Discussion

Unlike previous studies, this study will compare four data checking methods simultaneously. One method that this study includes is solo read aloud method, for which there has been very little published research. Only a single study has examined solo read aloud and in that study, it was only compared to only one other data checking method: double entry (Kawado et al., 2003). Moreover, that study used only two participants. In our study, we will be comparing 100 participants in solo read aloud to 300 participants in the other three data checking methods.

One weakness of our study is that we are not including every possible data checking method: We are excluding partner read aloud. Partner read aloud is a data checking method similar to solo read aloud. In partner read aloud, there are two people checking the data. One person reads from the original data sheet while the other person visually checks the data in the Excel file. Partner read aloud has been excluded in order to simplify administration procedures and reduce the time it will take to complete the study. However, because we are excluding partner read aloud, we are not comparing all data checking methods that are available to researchers. A better study would be to compare the accuracy of all possible data checking methods.

We are still in the process of designing this study. So far, we have almost finished designing the data sheets that participants will enter and check. We have created scripts for the Adobe Captivate videos that will be used to train the participants in each method. Next, we need to finalize all data sheets and all Captivate videos, and create a Qualtrics website that includes the consent form, links to the relevant videos and Excel files, and the evaluation form. Then we will be able to write an IRB proposal, print the 50 data sheets and place them in the two testing rooms, and train research assistants to administer the study. We expect to begin data collection this fall. We hope to finish data collection and calculate our results within the next four semesters.

Acknowledgements

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., & Pace, L. A. (2013, Aug). Poka-Yoke Double Entry System Version 2.1.43. Excel file that allows double-entry data entry with checking for mismatches, out-of-range values, and out-of-list values. Available at http://faculty.unlv.edu/barchard/doubleentry/ or from Kimberly A. Barchard, Department of Psychology, University of Nevada, Las Vegas, Barchard@unlv.nevada.edu.

Barchard, K. A., & Verenikina, Y. (2013). Improving data accuracy: Selecting the best data checking technique. Computers in Human Behavior, 29, 1917-1922. doi:10.1016/j.chb.2013.02.021

Gibson, D., Harvey, A. J., Everett, V., & Parmar, M. K. B. (1994). Is double data entry necessary? The CHART trials. Controlled Clinical Trials, 15, 582-488.

Kawado, M., Hinotsu, M.D., 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. doi:10.1016/S0197-2456(03)00089-8

Paulsen, A., Overgaard, S., & Lauritsen, J. (2012). Quality of data entry using single entry, double entry and automated forms processing--an example based on a study of patient-reported outcomes. Plos ONE, 7, 1-6. doi:10.1371/journal.pone.0035087

Reynolds-Haertle, R. A., & McBride, R., (1992). Single versus double data entry in CAST Controlled Clinical Trials, 13, 487-494. doi:10.1016/01972456(92)90205-E