Double Entry: Fun, Enjoyable, and Accurate

Ashley A. Anderson, Anum Warsi, Loise M. Ladrazo, and Kimberly A. Barchard
University of Nevada, Las Vegas

ABSTRACT

Whenever people enter data into the computer, it is possible for them to make data entry errors. Finding and correcting those errors is important. Previous research has found that double entry is the most accurate method of checking data (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado, Hinotsu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003). However, little research has examined the subjective experiences that data checkers have when they are using these methods. The purpose of this study was to compare subjective opinions of four data checking techniques: double entry, visual checking, solo read aloud, and partner read aloud. This will allow us to recommend the data checking method that researchers are most likely to enjoy. Moreover, if this study reveals negative opinions about some of the data checking methods, then future researchers could modify these methods to fix the problems, thus facilitating their use.

A total of 27 undergraduates (18 females and 9 males) participated in this study in return for course credit. They were randomly assigned to one of the four data checking methods: double entry, visual checking, solo read aloud, and partner read aloud. After individual training, each participant checked 20 data sheets. At the end of the study, participants completed a 16-item measure to evaluate the data checking method that they had used. To determine which data checking method was preferred, we compared the four groups using ANOVAs.

In this study, participants rated double entry as significantly more fun and enjoyable than the other techniques. Furthermore, previous research has found that double entry is the most accurate data checking method (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003). By using the method that is the most accurate and the most enjoyable, researchers (and research assistants) are likely to find data checking less onerous, and will continue to do data checking. We therefore recommend that researchers use the double entry method. A free double entry program will be available during the poster session

INTRODUCTION

When entering data, people often make errors. These errors can have drastic effects on the conclusions of research studies (Barchard & Pace, 2008). For example, a single error can turn a moderate correlation to zero or make a significant t-test non-significant (Barchard & Pace, 2011). It is therefore imperative that researchers locate and fix data entry errors. There are various methods to check data. These include double entry, visual checking, solo read aloud, and partner read aloud. Previous research has found that double entry is the most accurate (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado, Hinotsu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003). However, only a single study has been done on the subjective experiences of people using different data checking methods. In that study, participants rated double entry as significantly more accurate (F (2, 47) = 5.734, p = .006) and more reliable (F (2, 47) = 7.91, p = .001) than the other techniques, which included double entry, partner read aloud, and visual checking (Anang, Grob, Johnson & Barchard, 2011).

The purpose of this study is to compare subjective opinions of four data checking techniques: solo read aloud, partner read aloud, visual checking, and double entry. This study goes beyond previous research on this topic by examining a greater number of data checking techniques and by measuring opinions about a range of attributes. This research will allow us to identify the data checking techniques that are viewed most favorably by data enterers and will also allow us to pin-point weaknesses so that these data checking techniques can be improved.

METHOD

Participants

Participants included 27 undergraduates (18 females and 9 males), who ranged in age from 18 to 46 (mean 21.48, SD 6.22). Participants identified themselves as follows: Caucasian 40.7%, Hispanic 18.5%, African American 14.8%, Asian 14.8%, Pacific Islander 7.4%, and other 3.7%.

Measures

Subjective opinions of the data checking methods were measured using a 16-item self-report scale. The items were measured on a Likert-type scale that ranged from 1 (strongly disagree) to 5 (strongly agree).

Procedure

Participants completed the study during individual testing sessions that took approximately 90 minutes. First, the participants watched a 5-minute instructional video on how to use Excel. Then, the participants were randomly assigned to one of four data checking techniques (described below). The participants had no knowledge that there was more than one technique possible for them to be assigned to. Next, a short video explained the assigned data checking technique. To ensure that the participants understood the assigned technique, they checked five data sheets and were instructed to ask any questions necessary. Then the participants checked an additional 20 data sheets with no assistance from the researcher. Finally, the participants completed a survey based on their experiences during the study and answered a few questions about their computer use, data checking experience, and research experience.

This study used four data checking methods. Most of these techniques were done by a single person (the participant), but one of the techniques requires two people (the participant and the research assistant) to work together to check the data. The solo read aloud technique was completed by one person. The participant read the items off the paper data sheet out loud to themselves, and visually checked that these matched the data that was shown in the Excel file. If the participant noticed any discrepancies, they corrected the errors on the Excel file. The partner read aloud technique was done with two people. The trained researcher read the data out loud, while the participant checked the data that was in the Excel file. As each item was read, the participant said "check" if it matched what they saw on the Excel file. If the participant ran into any discrepancies between the two data sets, then they said "verify". The researcher then re-read that item so that the participant could correct the error. The visual checking technique was done with one person. The participant looked back and forth between the original paper data sheets and the Excel file and fixed any discrepancies on the Excel file. The double entry technique was completed with one person. The participant entered data a second time from the paper data sheets. The computer then compared the first and second entries, and highlighted any discrepancies or values that were outside the allowable range. The participant then fixed any errors.

Data Analysis

To compare subjective opinions of the four data checking methods, we conducted 16 one-way ANOVAs. The 16 dependent variables were the 16 adjectives on the self-report scale. The independent variable was the technique the participants were assigned to.

RESULTS

Participants rated double entry as significantly more fun (F(3, 23) = 3.87, p = .022) and more enjoyable (F(3, 23) = 3.53, p = .031) than the other techniques. No other differences were significant at the .05 level. See Table 1.

DISCUSSION

This research compared subjective opinions of four data checking methods. We examined 16 adjectives, but found significant differences on only two of them. Double entry was perceived as more fun and enjoyable. It is likely that most of the tests were non-significant because this study lacked adequate power due to the small sample size. As data collection continues and sample size increases, additional differences will likely be found between the data checking methods.

Statistical power might also be increased by using a less heterogeneous sample. This study was limited to university students who were largely in their early 20s and receiving course credit. Future research might consider including paid and unpaid research assistants, paid data entry staff, graduate students, and faculty. In general, greater heterogeneity in the sample will result in larger correlations, and thus might increase statistical power.

The statistical power of this study might also be improved by using a within-subject design. The study would take several days. Each day, participants would use a different data checking method and would complete a survey to indicate how they felt about that data checking method. Once participants had used all four types of data checking, they could state which method they preferred and why. The advantage of a within-subjects design is that it usually has higher statistical power (Maxwell & Delaney, 2004). The disadvantages are that there would be carry-over effects from one day to the next, due to

learning, that it would be difficult to recruit participants to come to a study for several days in a row, and that it would be impossible to keep participants blind to the purpose of the study.

Although our results should be considered tentative due to the small sample size, we did find that double entry is more fun and enjoyable than the other methods. Previous research has also found that double entry is the most accurate method (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003). We therefore recommend that researchers use double entry. By using the method that is the most accurate and the most enjoyable, researchers (and research assistants) are likely to find data checking less onerous, and will continue to do data checking. A free double-entry system is available from the following website http://faculty.unlv.edu/barchard/doubleentry/ and will be handed out during this poster session.

Table 1
Means of the Evaluation Items for the Four Data Checking Techniques

Items	Double Entry	Visual Checking	Solo Read Aloud	Partner Read Aloud
Satisfying	4.00	3.72	3.29	3.62
Comfortable	4.00	3.73	2.29	3.50
Pleasant	3.00	2.91	2.43	3.13
Relaxing	2.00	2.45	2.57	2.57
Accurate	4.00	3.45	3.14	3.75
Enjoyable	4.00	2.36	2.00	2.75
Fun	4.00	2.00	2.28	2.63
Calming	2.00	2.91	3.00	2.75
Reliable	3.00	2.00	2.85	3.38
Frustrating	4.00	3.64	3.00	3.75
Painful	3.00	4.09	3.14	4.00
Tedious	2.00	2.73	2.29	2.50
Depressing	5.00	4.18	3.89	3.88
Uncomfortable	3.00	3.73	2.43	3.13
Annoying	5.00	3.18	2.57	3.25
Boring	3.00	2.27	1.75	2.63

REFERENCES

Anang, C., Grob, K. E., Johnson, H. C., & Barchard, K. A. (2012, April). Subjective evaluations of data checking techniques. Paper presented at Western Psychological Association Annual Convention 2012, San Francisco, CA.

Barchard, K. A., & Pace, L. A. (2008). Meeting the challenge of high quality data entry: A free double-entry system. International Journal of Services and Standards, 4, 359-376.

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

Kawado, M., Hinotsu, S., Matsuyama, Y., 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

Maxwell, S. E., & Delaney, H. D. (2004). Designing experiments and analyzing data: A model comparison perspective (2nd ed.) New York: Taylor & Francis.