Double Entry: Slow and Steady Wins the Race

Brianna R. Maxim, Michael G. Curtis, Marielle Leo, and Kimberly A. Barchard
University of Nevada, Las Vegas

ABSTRACT

The smallest data entry error can completely change and invalidate research results (Barchard & Pace, 2011; Barchard & Verenikina, 2013). Researchers use a variety of methods to identify and correct data entry errors. Previous research has consistently shown that double entry finds the most errors (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003), but there have been inconsistent findings regarding speed: Some studies have found it to be faster than other methods (Kawamoto et al., 2014), whereas other studies have found it to be slower (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003). The purpose of this study was to compare four data checking methods – double entry, visual checking, solo read aloud, and partner read aloud – to determine which is the fastest and which is the most accurate.

A total of 101 undergraduates (36 males, 65 females) participated in this study in return for research credit. The study was individually administered during a single 90-minute session. During the study, participants checked the data for 20 fake data sheets, each of which contained 26 pieces of data. This fake data had already been entered into the computer, and we deliberately introduced errors into the data file. The participant's task was to identify and correct these errors.

We examined the data file after the participant had finished checking it, to determine how many errors were left. Double entry had far fewer errors than the other methods. For example, partner read aloud had almost five times as many errors as double entry. However, double entry also took longer than the other data checking methods. Given the magnitude of the differences in accuracy, we recommend double entry: It's worth the extra time.

Although the undergraduates who participated in this study are similar to the research assistants that typically enter data in psychological studies, future research should examine the accuracy and speed of other types of data checking personnel, such as graduate students and paid professionals.

INTRODUCTION

Researchers must be extremely careful about the way they enter data. Even small errors can destroy the validity of the study results (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Day et al., 1988; Kawado et al., 2003; King & Lashley, 2000). To avoid data entry errors, researchers have used many different data checking techniques. Of these, double entry results in the fewest errors. Double entry detected 69.0% of errors compared to the 39.9% that were detected by partner read aloud (Kawado et al., 2003), and visual checking left 2958% more errors than double entry (Barchard & Pace, 2011). Barchard and Verenikina (2013) compared double entry, visual checking, and partner read aloud, and double entry and demonstrated that double entry was the most accurate of the three. Specifically, participants who used visual checking and read aloud left 20% more errors in the data than double entry. No research has compared double entry to solo read aloud. Thus, the primary purpose of this paper was to expand upon previous studies by comparing accuracy of all four techniques: double entry, visual checking, partner read aloud, and solo read aloud.

Previous studies have found conflicting results regarding the relative speed of double entry and other data checking techniques. Kawamoto et al. (2014) found that double entry was the fastest of the four methods, although the average times were not significantly different. In contrast, two other studies found double entry to be the slowest data checking method (Barchard & Pace, 2011; Barchard & Verenikina, 2013).

Taking into account prior research, we hypothesize that double entry will be the most accurate and slowest data checking method. In addition, solo read aloud will detect the same number of errors and be as fast as partner read aloud.

The Learning Study

ID: 739925

Sex: (M) F

	Learning Style	Study Habits
1.	1 2 3 (4) 5	SD D N A (SA)
2.	1 2 3 4 (5)	SD D (N) A SA
3.	1 2 3 4 (5)	SD D N A (SA)
4.	1 2 (3) 4 5	SD D N A (SA)
5.	1 2 3 4 (5)	SD D N (A) SA
6.	1 2 3 (4) 5	SD D N (A) SA
7.	1 2 3 4 (5)	SD (D) N A SA
8.	1 2 (3) 4 5	SD (D) N A SA

	Spelling Test	Math Test
1.	ACCOMMODATE	156
2.	AMATEUR	235
3.	CALENDAR	485
4.	CEMETERY	493
5.	CONSHENCE	364
6.	EMBARRASS	327
7.	EXHILARATE	203
8.	MAINTAINANCE	347

Figure 1: Example of Data Sheet

METHOD

Participants

A total of 101 undergraduates (36 males, 65 females) participated in our study for course credit. Participants ranged in age from 18 to 50 years (mean 21.7, SD 6.4). Participants identified themselves ethnically as follows: 37.6% Caucasian, 23.8% Hispanic, 13.9% African American, 16.8% Asian, 3% Pacific Islander, and 5% as other. The education levels of participants were as follows: 37.6% Freshmen, 21.8% Sophomore, 13.9% Junior, 12.9% Senior, and 13.9% as other. Participants listed their lab experience as follows: 59.4% none, 14.9% 1 semester of research experience, 5.9% 2 semesters of research experience, and 19.8% as other.

Procedure

The study was run by trained administrators during one 90-minute session. Participants watched two training videos: the first was a tutorial introducing Excel because the study was completed using this software, and the second was a randomly assigned video explaining one of the four data checking methods (double entry, visual checking, partner read aloud, and solo read aloud).

After watching the training video, participants practiced their assigned data checking method using five practice data sheets. Each of the data sheets contained 26 pieces of data. See Figure 1 for an example of the data sheet. These data sheets had already been entered into Excel. However, we deliberately introduced errors into the data file. Participants were instructed to find and correct these data entry error. During this training period, the administrator carefully watched the participants and corrected any procedural mistakes that the participants made. Finally, participants were given an additional 20 data sheets to check on their own. The administrator remained in the room to answer any questions. This is the data that were analyzed for our paper.

Participants were randomly assigned to one of four data checking methods. In the double entry method, the data were entered into the computer a second time by the participant. Once the participant finished entering the data, a computer program identified any mismatches between the two data sets, as well as any values that were outside of an allowable range. In the visual checking method, the participant visually compared the data on the computer screen to the paper data sheet. In the partner read aloud method, the administrator read the paper data sheets aloud while the participant visually checked the data shown on the computer screen. Finally, in the solo read aloud method, participants read the data on the paper out loud and compared that to what they could see on the computer screen. Regardless of which of the four methods participants used to identify errors, participants corrected any errors they found in the computer data.

Data Analysis

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

We used another ANOVA to compare how long it took participants to complete the data checking. Time was calculated by taking the difference between the start and end time for the main part of the study, during which participants checked 20 data sheets.

RESULTS

Double entry had far fewer errors than the other methods ($F(3, 97) = 4.20$, $p < .008$). See Table 1. On average, solo read aloud resulted in twice as many errors in the final data set as double entry, visual checking had 2.4 times as many errors, and partner read aloud had 4.7 times as many errors.

However, double entry also took longer than the other data checking methods ($F(3, 97) = 18.39$, $p < .001$). See Table 2. On average, double entry took 34% longer then solo read aloud, 45% longer than visual checking, and 60% longer than partner read aloud.

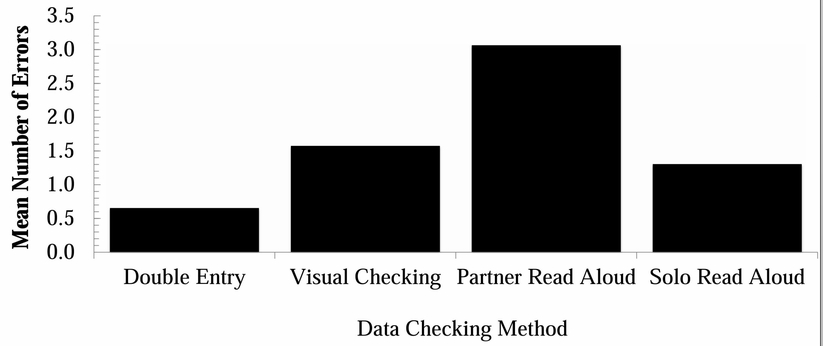

*Figure 2:*Number of Errors in the Data Set after Checking

Table 1
Number of Errors in the Data Set after Checking

Method	Mean	Standard Deviation
Double Entry	0.65	2.08
Visual Checking	1.57	1.92
Partner Read Aloud	3.06	3.71
Solo Read Aloud	1.30	1.34

Table 2
Time Taken to Complete Data Checking (in min)

Method	Mean	Standard Deviation
Double Entry	45.77	13.97
Visual Checking	31.49	7.03
Partner Read Aloud	28.68	4.36
Solo Read Aloud	34.18	7.39

DISCUSSION

The purpose of this study was to determine which of four data checking methods (double entry, visual checking, partner read aloud, or solo read aloud) was the fastest and which was the most accurate. We found that double entry was the most accurate. Visual checking and solo read aloud both left at least twice as many errors in the data sheet compared to double entry, and partner read aloud left about five times as many errors. These findings duplicate previous research and is consistent with our hypotheses that double entry would be the most accurate and that solo read aloud would have similar results to visual checking.

This study showed that double entry was the slowest. This is consistent with the results of three studies (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003) but inconsistent with a fourth (Kawamoto, et al. 2014), which found that double entry was the fastest method. This discrepancy may be due to the fact that the last study only had one participant use double entry method. Our study had a larger sample size and so many participants used the double entry method.

The majority of the undergraduate participants in our study (59.4%) had no research experience. Many participants also identified themselves as Freshmen (37.6%) or Sophomore (21.8%). Participants with no research experience and lower educational level are less likely to have entered large amounts of data in the past. This limited experience may have hindered some participants from entering data more accurately. Future research should examine the speed and accuracy of various data checking methods when they are being used by graduate students and paid professionals. This would allow researchers to generalize the findings to more experienced data entry personnel, to determine how speed and accuracy are affected by experience, and to determine if double entry is still the most accurate but slowed data checking method.

Given that double entry is the most accurate data checking method, we recommend that researchers use double entry computer programs to enter and check data for mistakes. In the case of data entry, slow and steady wins the race. A free double entry program will be distributed during the poster session. This double entry system is also available at the following website: http://faculty.unlv.edu/barchard/doubleentry/

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. 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,* 1917–1922. 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.

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. doi:10.1016/s0197-2456(03)00089-8

Kawamoto, R., Nagata, M., Roe, M., Ladrazo, L. M., & Barchard, K. A. (2014, April). *Need for speed: What is the fastest data checking method?* Poster presented at the Western Psychological Association annual convention, Portland, OR.

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.