

Actual Data Checking Accuracy versus Perceived Accuracy  
 Yevgeniya Verenikina, Cyndy Anang, Elizabeth A. Craun, and Kimberly A. Barchard  
 University of Nevada, Las Vegas

**Reference:** Verenikina, Y., Anang, C., Craun, E. A., Barchard, K.A. (2012, April). *Misperceptions of data checking*. Poster presented at the Western Psychological Association Annual Convention, San Francisco, CA.

**Contact Information:** Kimberly A. 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

Data entry errors can substantially change research results and conclusions. Therefore, researchers use a variety of techniques to identify and eliminate data entry errors. In visual checking, the data in the computer are visually compared to the original paper data sheet. In read aloud, the data are read aloud by one person while a second person looks at the data in the computer. In double entry, the data are entered a second time and the computer checks their accuracy: The computer identifies mismatches between the two entries, and also identifies values that are outside the allowable range for each variable. In all data checking techniques, the person corrects any errors that have been identified.

Previous research has shown that double entry results in higher accuracy than visual checking or read aloud (Barchard & Pace, 2010). However, few researchers use double entry. Perhaps researchers continue to use other data checking techniques because they think those techniques work better than they actually do. The purpose of this study was to examine the relationship between perceived and actual accuracy. A total of 117 participants were randomly assigned to three data checking techniques: visual checking, read aloud, and double entry. Next participants watched a video that explained their assigned data checking technique. To ensure participants were using their technique properly, they received feedback while they checked five data sheets. Finally, participants completed the main section of the study by checking the data for 20 additional sheets. To increase the difficulty of the data checking, these data sheets contain both letters and numbers.

After checking the data, participants rated the perceived accuracy of their data checking method. As expected, the correlation between perceived and actual accuracy was small and non-significant in each of the data checking conditions and overall. The largest correlation was  $-.14$ . These results show that people misjudge the accuracy of their data checking techniques. Researchers should not select data checking techniques based upon subjective perceptions. They should instead select techniques based upon empirical evidence. Empirical evidence shows that double entry is more accurate than visual checking or read aloud.

### Introduction

Data entry errors can drastically affect research results and conclusions. In order to ensure the accuracy of their data entry, researchers use various data checking techniques. Some of these commonly used data checking techniques are visual checking, read aloud and double entry.

In visual checking data are entered once and visually compared to the original data. In read aloud the data are read aloud by a person or a computer while another person compares the data being read to the entered data. In double entry data are entered twice; then, the computer identifies any mismatches and values outside the allowable range and the researcher corrects the indicated errors.

Not every data checking technique results in the same accuracy rate. To ensure the most accurate results, researchers should use the most effective data checking technique. Previous studies have found that the double entry results in significantly fewer errors than visual checking (Barchard & Pace, 2010; Reynolds-Haertle & McBride, 1992). Furthermore, double entry corrects more errors than the read aloud technique (Kawado et al, 2003).

However, few researchers choose to use double entry. Perhaps, researchers perceive their preferred data checking technique to be more accurate than it actually is. The purpose of this study is to examine the relationship between the perceived and actual accuracy of these three data checking techniques.

Figure 1  
Example Data Sheet

**The Learning Study**

ID: 739925  
Sex: (M) F

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

  

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

## Method

### Participants

A total of 117 participants (64 females and 53 males) participated in this study for course credit. Participants' ages ranged from 18 to 67 (mean 21.9, standard deviation 6.9). Participants included African American (14.5%), Asian (24.8%), Caucasian (36.8%), Hispanic (16.2%), Pacific Islander (6.8%), and Other (0.9%).

### Measures

The perceived accuracy of the data checking techniques was measured using a five-point scale: Strongly Disagree = 1, Disagree = 2, Neutral = 3, Agree = 4, and Strongly Agree = 5.

Actual accuracy was measured by the number of correct entries. An entry was considered correct if it was identical to the original data.

### Procedure

Participants checked data during a 90-minute supervised session. First, participants watched a short instructional video about Microsoft Excel. Second, participants watched an instructional video about one of the three randomly assigned data checking techniques: visual checking, read aloud, or double entry. Next, participants practiced the assigned data checking technique by checking five data sheets (see Figure 1) and correcting any errors. The data sheets contain both numbers and letters. To increase the difficulty of the data checking task, some of these data had to be converted from letters to numbers. Specifically, Sex and Study Habits sections were converted from letters to numbers during data checking. Finally, participants checked data for 20 additional sheets. At the end of the study, participants answered a brief survey about their demographic information, computer skills, experience during the study, and their perception of the accuracy of the data checking technique they used.

### Analysis

We calculated the correlations between the perceived accuracy of the data checking techniques and the actual accuracy.

## Results

The correlation between perceived and actual accuracy was small, negative and non-significant in each of the data checking conditions and overall. See Table 1.

Table 1

*Correlations between perceived and actual accuracy*

Condition	Correlation
All participants	-.08
Double Entry	-.11
Visual Checking	-.14
Read Aloud	-.08

*Note.* All p values were greater than .39.

### Discussion

The purpose of this study was to examine the relationship between perceived and actual accuracy of three data checking techniques. There were no relationships between perceived and actual accuracy for any of the data checking techniques. All correlations were small, negative, and non-significant.

We conclude that people misjudge the accuracy of their data checking techniques. Researchers should not rely on their subjective perceptions of data accuracy. Instead, they should choose data checking techniques based on empirical evidence regarding their accuracy. Empirical evidence has shown that double entry is the most accurate. Therefore, we recommend that researchers use double entry.

## References

- Barchard, K A. & Pace, L.A. (2008). Meeting the challenge of high quality data entry: Free double entry system. *International Journal of Services and Standards*, 4, 359-376.
- 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
- Reynolds-Haertle, R. A., & McBride, R. (1992). Single versus double data entry in CAST *Controlled Clinical Trials*, 13, 487-494. doi:10.1016/0197-2456(92)90205-E