Full speed ahead: Data checking methods and accuracy
Joanne Angosta, William R. Ebmeyer, Deirdre L. Bonner, Megan Holly, & Kimberly A. Barchard
University of Nevada, Las Vegas

UNLV

Reference: Angosta, J., Ebmeyer, W., Bonner, D. L., Holly, M., & Barchard, K. A. (2016, February). *Full speed ahead: Data checking methods and accuracy*. Paper presented at American Association of Behavioral and Social Sciences convention, Las Vegas, NV.

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, kim.barchard@unlv.edu

Abstract

When researchers enter data into computers, data entry errors are common. Therefore, researchers work hard to identify and correct these errors. Data checking accuracy can be impacted by speed. In this study, 154 undergraduates checked 20 data sheets, each containing 34 items. Participants were randomly assigned to one of four data checking methods: solo read aloud, partner read aloud, visual checking, and double entry. The correlation between number of correct entries and time spent on data checking was non-significant. However, the scatterplot shows that the people who spent the most time had the highest accuracy and those who spent the least time had the lowest accuracy. A large number of people had perfect accuracy. This study used university students, who may have stronger computers skills, on average, than the general population. This may have contributed to the ceiling effect. Therefore, future research should use a more heterogeneous sample.

Introduction

It is important to have accurate data. Errors in data influence statistical results (Levitt, Aeplli, Potish, Lee, & Nierengarten, 2006). Therefore, it is important to check data. However, the quality of data checking accuracy is impacted by speed. When speed is emphasized, work quality often decreases (Galinsky, Schiefer, & Pan, 1995). The inverse relationship between speed and accuracy is called the speed-accuracy tradeoff. The speed-accuracy tradeoff is a complex process that involves multiple cognitive processes. Discovering errors (even false errors) results in slower responses and improved performance, in order to prevent future errors (Saunder & Jentzch, 2012). Tasks that require more cognitive processes create a slowed-down effect which results in fewer errors (Wetter, Wegge, Jonas, & Schmidt, 2012). Slowing down creates cautious data entry and increases accuracy (Beilock, Bertenthal, Hoerger, & Carr, 2008).

Barchard and Verenikina (2013) compared the accuracy of three data checking methods: visual checking, read aloud, and double entry. Prior research has shown that taking more time to check data positively impacts the accuracy of data checking methods. Visual checking produced 2958% more errors than double entry, but is the quickest method (Barchard & Pace, 2011). Read aloud takes less time than double entry; however, read aloud is less accurate (Kawado, Hinostu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003). Double entry is the slowest method out of the three. Double entry increases the time taken to enter data, and as a result, there are more correct data entries (Cummings & Masten, 1994). Based on this research, we hypothesize that by spending more time checking data, results will be more accurate.

Method

Participants

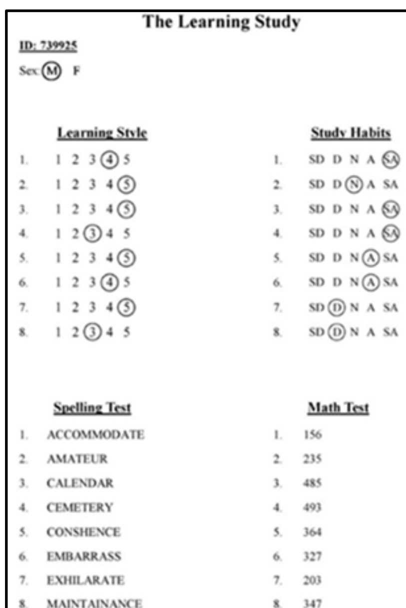
A total of 154 undergraduate students (103 females, 49 males, and 2 undisclosed) participated in this study in return for course credit. Ages ranged from 18 to 50 years (mean 21.56, SD 6.42). Participants identified themselves as follows: 10.4 % African American, 20.1% Asian, 36.4 % Caucasian, 26.0 % Hispanic, .6 % Native American, 1.9 % Pacific Islander, and 3.9 % Other.

Procedures

Participants were part of a larger study under the supervision of a research assistant. After participants consented to participate, they watched a short video on how to use Microsoft Excel, the program used during the study. After the video, they then watched a second video which explained their randomly-assigned data checking method.

Participants were randomly assigned to one of four possible data checking methods: visual checking, solo read aloud, partner read aloud, and double entry. Visual checking involved looking at a data sheet and comparing it to what was on the Excel file. Solo read aloud involved reading what was on the data sheet aloud as they compared it to what was already on the Excel file. Partner read aloud involved having the study administrator read the data sheet while the participant checked the Excel file. Double entry involved entering the data into Excel and then having the computer compare that data to the previously entered data. Whenever participants found data entry errors, they were asked to correct them.

Over the 90-minute session, participants checked two sets of data. The first set had five data sheets. Each data sheet had four different types of data to be checked. The participant was able to ask any question they had of the study administrator while checking the data sheets in this first set. After they completed these five data sheets, they were able to take a short break before beginning the second set of data sheets. The second set of data sheets had twenty data sheets. The participant was not provided any assistance while checking these data sheets. We used these data sheets to calculate the amount of correct entries for our data analysis. See Figure 2 for an example data sheet.


The Learning Study

ID: 739925

Sex: M F

Learning Style		Study Habits	
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

Spelling Test		Math Test	
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

Figure 1. *Example Data Sheet.*

Each data sheet had six different types of data: ID, Sex, five-point numerical response, five-point letter-responses (which had to be converted into numbers when entered), single words (which were often misspelled), and three digit numbers.

To calculate how long it took participants to enter the data, the time was recorded twice: first when participants opened the Excel files that contained the data they were checking, and second when they advanced to the next page of the study. To calculate the time it took participants to complete the data entry, we subtracted the times.

Data Analysis

Using a Pearson's correlation, we found the correlation between the amount of correct entries after data checking and the time it took to complete the data checking.

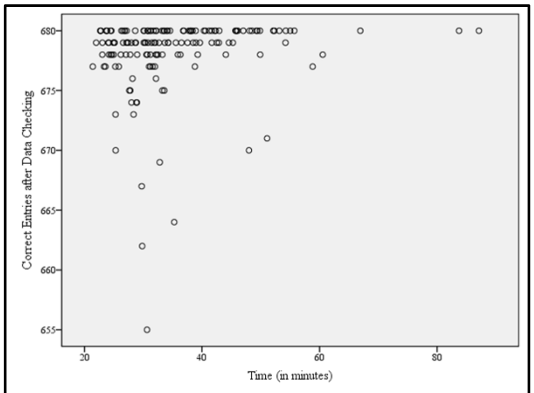


Figure 2. Scatterplot of time spent completing the study and amount of correct entries.

Results

The correlation between number of correct entries and time spent on the data checking was not statistically significant ($r(152) = .14$, $p = .075$). However, the scatterplot clearly shows that the people who were the least accurate spent very little time on the data checking and the people who spent the most time on the data checking were very accurate.

Discussion

Data checking is an important part of the research process; without it, results can become inaccurate (Kawado et al., 2003). Accuracy is affected by speed. This study analyzed the impact of speed on the accuracy of data checking. We found the correlation between time spent completing the data checking and the number of errors on the spreadsheet checked by the participant. We hypothesized that spending more time checking data improves the accuracy of results. Our results, however, showed that speed and accuracy did not significantly correlate with one another. We did find that there was a general positive trend in our scatterplot. Our scatterplot showed that most participants did not take a lot of time to complete the data checking. Those who did, however, had accurate data. On the other hand, those who did not take a lot of time, tended to have more errors in their data.

Many participants were able to fix every error. This could be because the data checking task was fairly easy. There are data entry tasks that are harder than research data. For example, the data that nurses enter, including medications (which are not common everyday words), addresses (which combine words and numbers), and names (which the enterer may never have heard before in their lives). If the data checking was a harder task, there would not be as much of a ceiling effect, and the correlation would likely be larger.

The study was short and was free of distractions. The data being checked was entirely legible. However, data checking may not always have these qualities. The setting may be loud and the data poorly written, which makes errors more common and checking harder.

Future research should use a larger and more heterogeneous subject pool. Current research for data checking mainly consists of undergraduates as participants (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Castellar, Notebaert, & Houtman, 2012; Healy et al., 2004; Wetter, Wegge, & Schmidt, 2012). College students may have more experience working with computers, which would increase accuracy. This might explain the ceiling effect we found in our data. Ceiling effects reduce correlations. If we repeated this study in the general population, our correlation would likely be larger.

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*(5), 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*(5), 1917 – 1922. doi:10.1016/j.chb.2013.02.021

Beilock, S. L., Bertenthal, B. I., Hoerger, M., & Carr, T. H. (2008). When does haste make waste? Speed-accuracy tradeoff, skill level, and the tools of the trade. J*ournal Of Experimental Psychology: Applied, 14*(4), 340-352. doi:10.1037/a0012859

Cummings, J. & Masten, J. (1994). Customized dual data entry for computerized data analysis. *Quality Assurance*, *3*, 300-303.

Galinsky, T. L., Schleifer, L. M., & Pan, C. S. (1995). The influence of performance standards and feedback on speed and accuracy in an electronically monitored data-entry task. *International Journal of Human-Computer Interaction, 7*(1), 25 – 36. doi: 10.1080/10447319509526108

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*(5), 560-569. doi:10.1016/S0197-2456(03)00089-8

Levitt, S. H., Aeplli, D. M., Potish, R. A., Lee, C. K.,& Nierengarten, M. E. (2006). Influences on inferences: Effect of errors in data on statistical evaluation. *Cancer*, *72*(7).

Saunders, B. & Jentzsch, I. (2012). False external feedback modulates posterror slowing and the f-P300: Implications for theories of posterror adjustment. *PsychonomicBulletin & Review, 19*(6), 1210 – 1216. doi: 10.3758/s13423-012-0314-y

Wetter, O. E., Wegge, J., Jonas, K., & Schmidt, K. (2012). Dual goals for speed and accuracy on the same performance task: Can they prevent speed-accuracy trade-offs? *Journal of Personnel Psychology, 11*(3). doi:10.1027/1866-5888/a000063