Checking Again vs. Checking Out Loud: A Surprising Second Best

Kimberly A. Barchard, Amber K. Stephens, and Elizabeth Ochoa
University of Nevada, Las Vegas

Reference: Barchard, K. A., Stephens, A. K., & Ochoa, E. (2017, April). *Checking again vs. checking out loud: A surprising second best.* Poster presented at the Western Psychological Association convention, Sacramento, CA.
Contact Information: Kimberly A. Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, P.O. Box 455030, Las Vegas, NV, 89154-5030, USA, barchard@unlv.nevada.edu

Abstract

For data checking, previous research showed double entry is out-performs visual checking and partner read aloud. No research has examined solo read aloud. Among 322 participants without data entry experience, double entry had the fewest errors and was most likely to produce perfect data. Next best was solo read aloud.

Introduction

In the last 20 years, technological advances such as optical mark recognition and online surveys have allowed much data entry to be computerized, which increases both efficiency and accuracy. However, not all data entry can be automated. Manual data entry is still common and inevitably leads to data entry errors. In psychological (Barchard & Pace, 2011) and life sciences research (Kozak, Kranowski, Cichocka, & Hartley, 2015), error rates are roughly 1%. In clinical research, error rates as high as 26.9% have been found (Goldberg, Niemierko, & Turchin, 2008). Even if only 1% of entries are errors, if a study contains just 200 items, manually entering the data is likely to result in data entry errors for almost every participant.

Simple data entry errors, such as typing an incorrect number or skipping over a line, can drastically change the results of a study (Barchard & Pace, 2008, 2011; Hoaglin & Velleman, 1995; Kruskal, 1960; Wilcox, 1998). For example, data entry errors can reverse the direction of a correlation or make a significant t-test non-significant (Barchard & Verenikina, 2013). Researchers therefore use a variety of strategies to prevent data entry errors, such as numbering the items, using data entry software that shows all items simultaneously, and entering data exactly as shown on the page (Schneider & Deenan, 2004). However, whatever steps are taken to ensure data are initially entered correctly, researchers cannot know that entries are correct unless they check them.

The Learning Study

ID: 346525

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.	ACOMADATE	165
2.	AMETURE	234
3.	CALANDAR	450
4.	CEMETARY	473
5.	CONSCIENCE	360
6.	EMBARRAS	350
7.	EXILARATE	230
8.	MAINTAINANCE	324

Figure 1. Example Data Sheet.

Some researchers have advocated holistic data checking methods, such as examining scatterplots and histograms (Tukey, 1977) and calculating univariate and multivariate statistics to detect outliers and other influential data points (Osborne & Overbay, 2004; Tabachnick & Fidell, 2013). However, these methods may not detect errors that fall within the allowable ranges for the variables. Therefore, item-by-item data checking is preferred. A variety of item-based data checking methods can be used (Barchard & Verenikina, 2013). In visual checking, the researcher visually compares the original paper data sheets with the entries on the computer screen. In solo read aloud, the researcher reads the original paper data sheets aloud and visually checks that the entries on the screen match. In partner read aloud, one researcher reads the data sheets aloud, while another researcher checks that the entries on the computer screen match. Finally, in double entry, the data are entered into the computer a second time and the computer compares the two entries and flags any discrepancies; the researcher then checks the original data sheet to determine which entry is correct. The purpose of all these data checking procedures is to identify and correct data entry errors. It is therefore important to study the effectiveness and efficiency of these methods.

Double entry is more accurate than visual checking and partner read aloud. Among experienced data entry operators checking renal transplant data, double entry detected 73% more errors than partner read aloud (Kawado et al., 2003). Among university students checking psychological data, visual checking resulted in 2958% more errors than double entry (Barchard

& Pace, 2011), and visual checking and partner read aloud were only one-third as likely to correct every data entry error as double entry (Barchard & Verenikina, 2013). No previous research has examined the accuracy of solo read aloud or compared its accuracy to that of other data checking methods.

The purpose of our study is to compare double entry to partner read aloud and visual checking, to replicate previous findings that double entry is more accurate than visual checking and partner read aloud. In addition to these methods, our study will also examine solo read aloud, which has never before been empirically tested. Like the previous studies by Barchard and colleagues, our study will use data and participants that are similar to those used in psychology research studies.

Method

Participants

A total of 322 undergraduates (195 female, 123 male), with no previous data entry experience, participated in return for course credit. They ranged in age from 18 to 50 ($M = 20.32$, $SD = 3.88$). They identified themselves as Caucasian (30.7%), Hispanic (27.3%), Asian (23.3%), African-American (10.2%), Pacific Islander (2.2%), and other (5.3%).

Because participants were assigned to the four conditions completely at random, there were slightly more participants in some conditions than others: double entry 75, visual checking 71, partner read aloud 95, solo read aloud 81.

Procedures

Participants completed this study individually in a single 90-min. session. Because Excel 2007 is used for data checking in this study, participants began the study by watching a video on how to use Excel. Next they were randomly assigned to one of the four data checking methods, and they watched a second video on how to use that data checking method. During the training phase, participants checked the data on 5 data sheets. During this time, the researcher answered any questions participants had and corrected any procedural errors they were making. During the testing phase, participants check the data on 20 additional data sheets, without further interactions with the researcher.

The goal of the data checking was to identify errors in the Excel file. Before participants arrived for the study, the 25 data sheets were entered into the computer. These data sheets contained six types of data: a six-digit ID code, a number to represent sex, five-point numerical rating scales, five point alphabetical scales (SD D N A SA), words in capital letters (often with spelling errors), and three-digit whole number (see Figure 1). When entering the data, we deliberately introduced 32 data entry errors. About half (13) were extreme errors that would be easy for researchers to identify later on (for example, entering a word where a number was expected). The other half (19) were less extreme errors which would be difficult for researchers to identify from a superficial examination of the data (for example, entering a number that is within the allowable range for that variable, but which is incorrect).

After participants complete the data checking, they evaluated their data checking method using a short survey.

Results

Number of Errors

Perfect Data Checking

Originally, there were 32 errors in the Excel file. Most of the participants who used double entry corrected every one of these errors, whereas only a third of the other participants corrected every error. See Figure 2. The probability of correcting every error was .83 for double entry, .32 for visual checking, .19 for partner read aloud, and .44 for solo read aloud (Wald statistic(3) = 59.68, $p < .001$). Tukey's follow-up tests showed that double entry was significantly more likely than any of the other methods to result in perfect data, and that solo read aloud was more likely than partner read aloud to result in perfect data.

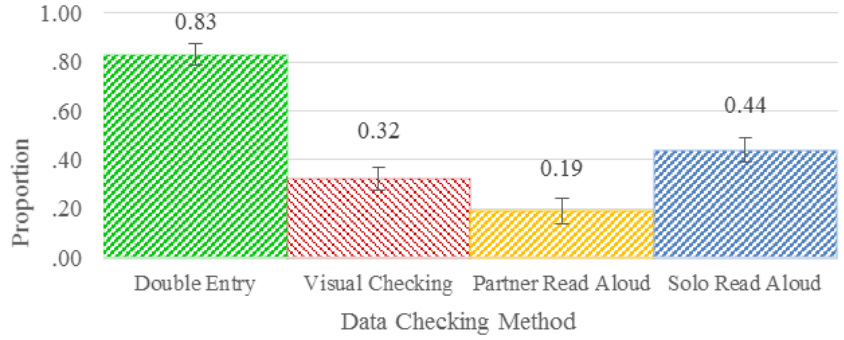

Figure 2. Proportion of Participants with Perfect Data (with Standard Error for the Proportion).

Number of Errors

Double entry had the fewest errors. See Table 1 and Figure 3. For five of the six types of data, it had fewer errors than the other methods, and for three of them, at least one of these comparisons was statistically significant. Overall, double entry had significantly fewer errors than partner read aloud, which had five times as many errors.

Visual checking was the second worse method. It was significantly better than partner read aloud for all data and for the Spelling Test (the hardest type of data to check). However, visual checking was significantly worse than double entry when checking IDs.

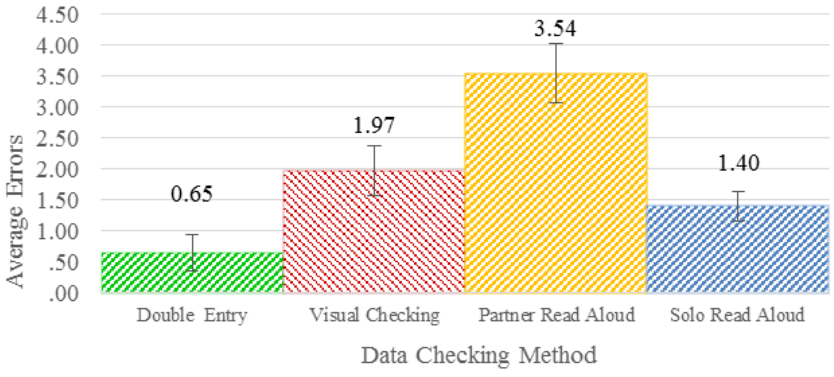

Figure 3. Mean Number of Errors (with Standard Error for the Mean).

Partner read aloud was the worst data checking method. It was significantly worse than the remaining three techniques for all data and for the Spelling Test specifically, and was worse than both the double entry and solo read aloud for the Math Test. Thus, the differences between the data checking methods are most apparent for the most difficult data checking tasks.

Solo read aloud worked better than partner read aloud. It was significantly better than partner read aloud across all data and for the Spelling Test and Math Test. Solo read aloud had more errors than double entry for almost every type of data; however, these differences were never statistically significant.

Table 1

Number of Errors by Data Type for Each of the Four Data Checking Methods

Data Type	Double Entry	Visual Checking	Partner Read Aloud	Solo Read Aloud	ANOVA
ID 6 digit	**.08^{a} [-.02, .18]**	.34^{b} [.18, .50]	.22^{ab} [.13, .32]	.20^{ab} [.10, .29]	$F(3, 318) = 3.23$, $p = .023$
Sex Convert letter to number	**.00^{a} [.00, .00]**	.01^{a} [-.01, .04]	.03^{a} [.00, .07]	**.00^{a} [.00, .00]**	$F(3, 318) = 1.61$, $p = .187$
Learning Style 5-point numerical scale	.21^{a} [.07, .36]	**.18^{a} [.02, .34]**	.24^{a} [.07, .42]	.20^{a} [.07, .33]	$F(3, 318) = 0.11$, $p = .955$
Study Habits 5-point letter scale convert to numbers	**.15^{a} [-.04, .33]**	.23^{ab} [.02, .43]	.92^{b} [.59, 1.24]	.46^{ab} [.20, .72]	$F(3, 318) = 7.21$, $p < .001$
Spelling Test Words (often misspelled)	**.08^{a} [-.01, .17]**	.69^{a} [.29, 1.09]	1.33^{b} [.86, 1.79]	.20^{a} [.09, .30]	$F(3, 318) = 12.37$, $p < .001$
Math Test 3 digit numbers	**.13^{a} [-.04, .30]**	.52^{ab} [.31, .73]	.80^{b} [.49, 1.11]	.35^{a} [.19, .50]	$F(3, 318) = 6.01$, $p = .001$
All Data	**.65^{a} [.06, 1.24]**	1.97^{a} [1.17, 2.77]	3.54^{b} [2.58, 4.50]	1.40^{a} [.92, 1.87]	$F(3, 318) = 11.07$, $p < .001$

Note. Within rows, means with different superscripted letters are significantly different at $p < .05$. In each row, the method with the lowest number of errors is in boldface.

Discussion

Every data checking method corrected the vast majority of errors. Even the worst data checking method (partner read aloud) left only 3.54 errors in the Excel file. Because the Excel file only contained 5% errors to start, 99.5% of the entries were correct when the data checking was complete. The best data checking method (double entry) resulted in 99.9% accuracy. In

absolute terms, the difference between 99.5% accuracy and 99.9% accuracy is small, which may explain why most researchers think that their particular data checking method is excellent. With all methods having accuracy rates greater than 99%, it is difficult for researchers to discern the differences between data checking methods without doing systematic research like the current study.

The reader is reminded, however, that an accuracy rate of 99.4% is far from optimal. If each participant in a study completes 200 items, then the records for almost every participant will contain errors. Such errors can reverse the sign of a correlation coefficient or make a significant *t*-test non-significant (Barchard & Verenikina, 2013). It is essential that researchers use high quality data checking methods to avoid such mistakes.

Double entry was the most accurate method. This is consistent with previous research, which has shown that double entry results in significantly fewer errors than single entry (Barchard & Pace, 2011), visual checking (Barchard & Pace, 2011; Barchard & Verenikina, 2013), and partner read aloud (Barchard & Verenikina, 2013; Kawado et al., 2003). This study replicated these findings by showing that double entry was substantially and significantly more likely than the other methods to result in perfect data entry, and that it had less than half as many errors as the next best technique.

Solo read aloud was the second-best data checking method. It was better than partner read aloud, being significantly more likely to result in perfect data and sometimes having significantly fewer errors. Solo read aloud never produced significantly more errors than double entry, although it cannot be considered equivalent to double entry because it was significantly less likely to result in perfect data. As well, solo read aloud was not significantly better than visual checking. Ours was the first study in any discipline to examine solo read aloud. We consider it to be a promising technique. Solo read aloud would likely be inappropriate in some types of environments and for some types, because reading data out loud would be distracting to others or would violate confidentiality. Where practical, though, solo read aloud may be a viable speedier alternative to double entry. Future research should examine this method with other types of data and other types of participants.

In examining the results of these analyses, readers might note that visual checking had the fewest number of errors when participants were checking the Learning Style data. These data consist of single digit numbers, from 1 to 5, which are typed directly into the computer. This is a common type of data in psychological research studies, and readers might be inclined to think that visual checking can be justified for this type of data. However, that conclusion cannot be justified from our results. There were no significant differences between any of the data checking methods for the Learning Style items (p = .955). Moreover, previous research does not support that conclusion. Barchard and Pace (2011) found that visual checking resulted in significantly (p = .013) and substantially more errors than double entry for this same type of data.

At this point, we conclude that double entry remains the most effective data checking method. Double entry software is available in commercial statistical packages, such as SPSS and SAS. Several free double entry programs are also available. These include web-based systems (Harris, Taylor, Thielke, Payne, Gonzalez, & Conde, 2009), stand-alone programs (Gao et al., 2008; Lauritsen, 2000-2008), and add-ons for Excel (Barchard, Bedoy, Verenikina, & Pace, 2016). A free double-entry system will be available during the poster session.

References

Barchard**,** K. A., Bedoy, E. H., Verenikina, Y., & Pace, L. A. (2016, May). *Poka-Yoke Double Entry System Version 3.0.76.* Excel 2013 file that allows double entry, checking for mismatches, and checking for out of range values. Available at http://faculty.unlv.edu/barchard/doubleentry/ and from kim.barchard@unlv.edu

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*(4), 359-376. doi:10.1504/IJSS.2008.020053

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

Gao, Q-B., Kong, Y., Fu, Z., Lu, J., Wu, C., Jin, Z-C., & He, J. (2008). EZ-Entry: A clinical data management system. *Computers in Biology and Medicine, 38,* 1042-1044. doi:10.1016%2Fj.compbiomed.2008.07.008

Goldberg, S. I., Niemierko, A., & Turchin, A. (2008). Analysis of data errors in clinical research databases. *AMIA Symposium Proceedings*, *6*, 242-246.

Harris, P. A., Taylor, R. Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. doi:10.1016/j.jbi.2008.08.010

Hoaglin, D. C. & Velleman, P. F. (1995). A critical look at some analyses of major league baseball salaries. *American Statistician, 49*(3), 277-285. doi:0.1080/00031305.1995.10476165

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

Kozak, M., Krzanowski, W., Cichocka, I., & Hartley, J. (2015). The effects of data input errors on subsequent statistical inference, *Journal of Applied Statistics, 42*, 2030-2037. doi:10.1080/02664763.2015.1016410

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics, 2.* Available at http://www.tufts.edu/~gdallal/out.htm

Lauritsen J. M. (Ed.) (2000-2008). *EpiData Data Entry, Data Management and Basic Statistical Analysis System*. EpiData Association: Odense Denmark. http://www.epidata.dk Accessed January 14, 2013.

Ohmann C., Kuchinke W., Canham S., Lauritsen J., Salas N., Schade-Brittinger, C., Wittenberg, M., et al. (2011). Standard requirements for GCP-compliant data management in multinational clinical trials. *Trials 12:85.* doi:10.1186/1745-6215-12-85

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, 9(6), 1-12. Schneider, J. K., & Deenan, A. (2004). Reducing quantitative data errors: Tips for clinical researchers. *Applied Nursing Research, 17*(2), 125-129. doi:10.1016/j.apnr.2004.02.001

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics, 6th ed.* Boston, MA: Pearson.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*(3), 300-314. doi:10.1037/0003-066X.53.3.300