Subjective Evaluations of Data Checking Techniques

Cyndy Anang, Kelly E. Grob, Heather C. Johnson, and Kimberly A. Barchard

University of Nevada, Las Vegas

ABSTRACT

The aim of this research was to analyze people's subjective opinions about the data checking techniques double entry, visual checking, and read aloud. Previous research has shown that entering data twice is more effective in reducing the data entry errors. We therefore hypothesize that participants would perceive the double entry technique as most accurate and reliable. Forty-eight undergraduate students answered a set of 16 items which was used to gather participants' opinions on the three techniques. The results showed that the double entry technique was perceived as significantly more accurate (F (2, 47) = 5.734, p = .006) and reliable (F (2, 47) = 7.91, p = .001). No other technique received better ratings than double entry on any of the other items. Based on our results and the information found by previous research, we recommend that researchers use the double entry technique to check data.

INTRODUCTION

The purpose of our research is to distinguish between double entry, visual checking, and read aloud data checking techniques by considering people's opinions about each one. In general, these three data checking techniques are used to ensure that errors found in data are reduced drastically so that in the long-run results based on the data used are as accurate as possible.

Several data-checking techniques are commonly used to detect errors and correct them. This paper will focus on three data-checking techniques: double entry, visual checking, and read aloud. In the double entry technique, the user enters the data a second time, and the computer compares the two data entries. If there is a difference between the two data entries, the user corrects the error. The visual checking technique involves the user visually comparing the data on the paper data sheet with the entries on the computer screen and correcting any errors found. For the read aloud technique, another person reads the data aloud from the paper data sheet, and the user visually checks the data on the computer, correcting any errors found.

This research in particular is important for two reasons. First, if data-checking techniques are equally effective, we can recommend the one that is preferred. Second, it allows us to determine the strengths and weaknesses of each technique. This allows us to modify the techniques, so that we can make them more user-friendly.

LITERATURE REVIEW

Research data help us examine many issues in life, and help us determine whether certain hypotheses are correct. When data are not correct, our conclusions can be affected drastically (Burchinal & Neebe, 2006). With just one data entry error, a significant t-test or correlation can be made non-significant (Barchard, Scott, Weintraub, & Pace, 2008). Therefore, it is imperative that we check data in the most efficient way possible.

For a data checking technique to be the most efficient, it needs to be both accurate and user-friendly. A particular technique may be ignored because of the discomfort it causes the user, even if that technique is the most effective in reducing errors. If researchers do not like or do not have faith in a certain technique, then they probably will not use it. Therefore, it is important to consider what people think about the techniques, so that we understand what they prefer and what they do not prefer.

Some researchers have compared these three techniques, to determine which is most effective. When the techniques double entry, read aloud, and visual checking have been compared, the double entry technique has consistently been found to be more accurate than the other two techniques (Barchard, Pace, & Burns, 2009; Barchard et al., 2008). Medical research that used small samples also found double entry to be more efficient than visual checking (Kawado et al., 2003). However, research has also shown that the double entry technique is the most tiring technique (Beaty, 1999).

Among the three techniques being discussed, the visual checking technique results in more errors than the double entry and read aloud techniques (Verenikina, Anang, Jenkin, Grob, & Barchard, 2012; Scott, Thompson, Wright-Thomas, Xu, & Barchard, 2008). In fact, visual checking doesn't even reduce more errors than single entry (Barchard et al., 2008) which takes place when researchers enter data a single time, without checking the data afterwards to make sure it is correct. This means that visual checking isn't even worth using, and may waste time and resources.

Unlike the other two data checking techniques, the read aloud technique involves two people, for example, the administrator and the participant. When two people check data, the process is more effective because it allows users to detect errors that may be missed by a single person (Nihei, Terashima, Suzuki, & Morikawa, 2002). It also helps eliminate boredom and mental fatigue because the work load is shared by two people. When users are bored or experience mental fatigue during data checking, there is a chance that more errors may be ignored (Kole, Healy, & Bourne, 2008).

Even though double entry is most efficient at checking data, and visual checking is the worst, data checking still isn't easy, no matter what technique someone uses. When people look at data on a computer for long amounts of time, they can be affected in one of two ways. They may either perfect the art of data checking, or they may become bored and tired (Healy, Kole, Buck-Gengler, & Bourne, 2004). Because data checking is already challenging enough, it is important that we consider people's opinions about the techniques. Considering people's opinions will help with choosing the data checking technique that will suit each person.

In our research, we have considered people's opinions about the three data checking techniques double entry, visual checking, and read aloud. Our hypotheses, which stem from both previous research and our experiences with data checking are as follows. We hypothesize that participants who use the double entry technique would perceive the technique as the most accurate and reliable.

METHOD

Participants

A total of 48 participants (26 females and 22 males) participated in this study for course credit. Their ages ranged from 18 to 39 (mean 22, standard deviation 5.26). The participants included African American (12.50%), Asian (22.92%), Caucasian (41.67%), Hispanic (14.58%), Pacific Islander (6.25%), and Other (2.08%).

Measures

This study used a self-report questionnaire that includes 16 items. Each of the items is measured on a 5-point Likert scale, which ranges from (1) "Strongly Disagree" to (5) "Strongly Agree". See Figure 1

Procedure

The participants used the computer for the entire study. First, they watched a video that explained how to use Excel. Second, they were randomly assigned to a technique. They did not know that there were other data checking techniques involved in the research. They only learned about the one that they were assigned. Third, they watched a video that explained how to use the particular technique. The participants were then given a set of data so they could practice their assigned technique. This set included five fake participants. After this, the participants checked data from twenty fake participants. Finally, after they completed checking the data, the participants were asked to complete the subjective evaluation of the technique they used. This evaluation took two to five minutes. See Figure 1.

Data Analysis

We performed a one-way Analysis of Variance (ANOVA). Our dependent variables were the 16 items from the evaluation form. Our independent variable was the group to which participants were assigned. This variable had three levels: double entry, visual checking, and read aloud.

RESULTS

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. No other differences were significant. See Table 1

It is interesting to note how each technique was described. Participants rated the double entry technique as satisfying and depressing, in addition to describing it as accurate and reliable. The read aloud technique was rated as painful and depressing. For the visual checking technique, none of the average ratings exceeded 4 on the 5-point scale.

DISCUSSION

In this paper, we have examined what the participants thought about each data checking technique. While there wasn't much significance in the difference in opinions between the three data checking techniques, it is still important to discuss the techniques' trends in ratings. We believe that these trends may become significant with a larger sample size.

Double entry was rated as significantly more accurate and reliable than visual checking and read aloud. This result relates to previous research, as double entry has been shown to be the most accurate of the three techniques (Barchard, Pace, & Burns, 2009; Barchard et al., 2008). In data checking, accuracy is very important. Therefore if a technique is perceived to be most accurate when being used for data checking, it is important that the technique is chosen over other techniques that are not perceived as most accurate. This means that even though the double entry technique is a lengthy technique because participants have to enter all of the data a second time and correct all errors on both sides, it is advisable to use the double entry technique instead of the visual checking or the read aloud techniques.

Double entry was also described as satisfying and depressing. Participants may have rated this technique as satisfying because much effort is put into ensuring that errors are eliminated by checking the data a second time. However, because data has to be c hecked a second time, the double entry may be considered as depressing because it involves so much time and energy.

Visual checking had the highest average ratings for fun, enjoyable, and pleasant. These ratings for visual checking may be because this technique takes less time and therefore is less stressful to use. However, these differences were not significant, and none of these average ratings was more than 4 on the 5 point scale.

Read aloud was rated as most painful and depressing. Some possible reasons why this technique had been rated as painful may be because of the speed at which the administrator was reading the items to the participant and how fast the participant was able to type the items. It could also be more frustrating and painful for individuals who do not enjoy interacting with others, or who prefer to work alone.

Our findings support our hypothesis which was that participants would perceive double entry as most accurate and reliable. Unlike the other two techniques, the chance of errors being detected is higher when using the double entry technique because the double entry technique compares two data sets and errors are highlighted for users to correct them. Participants also spend a lot more time on checking data when using the double entry (Gibson, Harvey, Everett & Parmar, 1994) than when using the read aloud or visual checking technique. This is because unlike the other two techniques which require listening to the data being read and looking to find errors using a data set respectively, the double entry requires that data are entered a second time. Spending more time checking errors for a data set may help reduce the number of errors.

Based on our findings and results shown by previous research, we recommend the double entry technique for data checking. Our sample showed that most participants found the double entry technique to be the most accurate and reliable. It is very important that the participants are confident in these two factors. This is because if people do not think that the technique being used is efficient at data checking, then there is no point in using the technique since the main purpose of using data checking techniques in the first place is to ensure that errors are reduced or better yet completely eliminated. Specifically, why would researchers go through the trouble of checking data, when they do not have confidence in the technique being used? There would be no point of checking the data in this case, because resources (time and people) would be wasted for no sufficient reason.

Conclusion

Our study examined the subjective evaluation of participants who used the three data checking techniques double entry, read aloud and visual checking to correct errors in a data set. Based on the data we collected and analyzed, we concluded that the double entry technique is perceived to be the most accurate and reliable. These are the only significant results. However, data collection is ongoing and we expect additional differences between the techniques to emerge once we have a larger sample size. Because of our results and what past research has found, we recommend that researchers use the double entry technique to check data.

REFERENCES

Barchard, K. A., Scott, J., Weintraub, D., & Pace, L. A. (2008, August). Better data entry: Double entry is superior to visual checking. Paper presented at the American Psychological Association, Boston, MA.

Barchard, K. A., Pace, L. A., & Burns, S. S. (2009, May). Double entry: Accurate results f om accurate data. Poster presented at the Association for Psychological Science convention, San Francisco, CA.

Beaty, J. (1999). The PowerChecker: A Visual Basic program for ensuring data integrity. ehavior Research Methods, Instruments & Computers, 31(4), 737-740.

Burchinal, M., & Neebe, E. (2006). Best practices in quantitative methods for developmentalists: I. Data Mangement: Recommended practices. Monographs of the Society for Research in Child Development, 71 (3), 9-23. doi:10.1111/j.1540-5834.2006.00354.x

Czaja, S., & Sharit, J. (1998). Ability–performance relationships as a function of age and t ask experience for a data entry task. Journal of Experimental Psychology: Applied, 4(4), 332-351. doi:10.1037/1076-898X.4.4.332.

Gibson D, Harvey A, Everett V & Parmar M. (1994). Is double data entry necessary? The CHART trials. CHART Steering Committee. Continuous, hyperfractionated, accelerated radiotherapy. Control Clinical Trials 1994;15:482-488.

Healy, A. F., Kole, J. A., Buck-Gengler, C., & Bourne, L. E. (2004). Effects of prolonged work on data entry speed and accuracy. Journal of Experimental Psychology: Applied, 10(3), 188-199. doi:10.1037/1076-898X.10.3.188

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

Kole, J., Healy, A., & Bourne, L. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. Applied Cognitive Psychology, 22(7), 917-937. doi:10.1002/acp.1401

Nihei, Y., Terashima, M., Suzuki, I., & Morikawa, S. (2002). Why are four eyes better than t wo? Effects of collaboration on the detection of errors in proofreading. Japanese Psychological Research, 44(3), 173-179. doi:10.1111/1468-5884.00020

Scott, J. R., Thompson, A. R., Wright-Thomas, D., Xu, X., & Barchard, K. A (April 2008). Data entry method: Is double entry the way to go? Poster presented at the Western Psychology Association Annual Convention, Irvine, CA.

Verenikina, Y., Anang, C., Jenkin, T., Grob, K.E., Barchard, K.A. (2012, April). Visual checking – the road to disaster. Poster presented at the Western Psychological Association Annual Convention, San Francisco, CA.