Esc'ing Errors: Pick the Best Method, Not the Best Person

Amber K. Stephens, Carmela Tayag, Claudia Chiang-Lopez, Megan Holly, & Kimberly A. Barchard
University of Nevada, Las Vegas

Reference: Ruggiero, A., Tayag, C., Chiang-Lopez, C., Holly, M., & Barchard, K. A. (2016, Feb). *Esc'ing errors: Pick the best method, not the best person*. Paper presented at Western Psychological Association convention, Long Beach, 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, kim.barchard@unlv.edu

Abstract

Researchers are likely to make the wrong conclusions if their databases contain data entry errors. Therefore, data checking is an important step in the data analysis process. Many factors influence the accuracy of data checking. The purpose of this study was to determine whether computer skills influence data checking accuracy. We created 25 data sheets that contain the kinds of variables that are used in psychological studies. Each data sheet had 34 items. When we entered this data into the computer, we deliberately introduced data entry errors. The participants' job was to identify and correct these data entry errors. A total of 154 undergraduate students participated in this study for course credit. They were randomly assigned to one of four data checking methods: visual checking, solo read aloud, partner read aloud, or double entry. After completing the data checking, participants rated their computer skills on four different variables: computer ability, spreadsheet ability, typing ability, and number pad ability. Surprisingly, none of the four computer skills had significant correlations with accuracy, despite the high sample size. Confidence intervals show that the population correlations are likely small. These small correlations might be due to the strong ceiling effect. Many participants had perfect or near perfect accuracy, perhaps because of the strong computer skills that most university students have. Future research could explore the relationship between data checking accuracy and computer skills in a more heterogeneous population. This would likely result in higher correlations. For researchers who are recruiting undergraduate research assistants to complete data entry tasks, it is helpful to know that computer skills have only small relationships to accuracy. To improve data checking accuracy, researchers should select the most accurate data checking methods. Previous research has shown that double entry is the most accurate technique (Barchard & Verenikina, 2013), and we therefore recommend it. A free double entry program will be available during the poster session.

Introduction

Researchers enter data into computers to preserve vital information while conducting a study. Much data entry is becoming computerized, but there is still a need for manual data entry in research (Atkinson, 2012; Barchard & Pace, 2011; Davidson & Skinner, 2010; Fletcher, Erickson, Toomey, & Wagenaar, 2003; Grieve, Witteveen, & Tolan, 2014; Koubek, Salvendy, & Noland, 1994). Whenever data are entered by hand, data entry errors can occur. Data entry errors can have detrimental effects on the integrity of study results (Barchard & Verenikina, 2013). In order to help eliminate data entry errors, data checking must be utilized. For the purpose of this paper we will focus on four methods of data checking: visual checking, solo read aloud, partner read aloud, and double entry.

To check the accuracy of the data that have been entered into the computer, basic knowledge of computers is essential. For example, the ability to use Microsoft Excel to create spreadsheets and the ability to use a keyboard and mouse efficiently are considered some of the basic skills required for data entry (Kay, 2007) and may be equally important for data checking. Similarly, poor typing skills might introduce errors into the data. The purpose of this study is to determine which computer skills are associated with higher data checking accuracy. More specifically we examined four key variables: computer ability, spreadsheet ability, typing ability, and number pad ability. We hypothesized that the largest relationship would be for typing ability.

Knowing which computer skills are important could be beneficial in two ways. First, data entry personnel could be selected based upon whether they have these skills. Second, data entry personnel could be trained on the specific skills they need.

Method

Participants

A total of 154 undergraduate students (103 female, 49 male, 2 undisclosed) participated in this study in return for course credit. Their ages ranged from 18 to 50 (mean 21.56, SD 6.42). These participants identified their ethnicities as follows: 10.4% African American, 20.1% Asian, 36.4% Caucasian, 26.0% Hispanic, 0.6% Native American, 1.9% Pacific Islander, and 3.9% other.

Measure

Evaluation Survey

The evaluation survey was a self-reported questionnaire administered at the end of the study that asked participants to rate their ability to use different computer functions. Using a five-point scale, the participant rated their ability from *very low* to *expert* on four variables: computer ability, spreadsheet ability, typing ability, and number pad ability. For this study, the variables were defined as follows: Computer ability was the participant's overall knowledge and familiarity with using computers; Spreadsheet ability was considered the participant's overall knowledge and familiarity with using programs such as Excel or Works Spreadsheet. Typing ability was the participant's ability to type without looking at the keyboard. Number pad ability was the participant's ability to use the numeric keypad, or number pad, without looking at the keys.

Procedures

The study took approximately 90 minutes to complete and was conducted on a computer inside a psychology lab under the supervision of an administrator. After being greeted and shown to the computer, participants were instructed to watch two tutorial videos: one that introduced them to the basic functions of Excel and one that explained the data checking method to which they had been randomly assigned.

There were four data checking methods. During visual checking, the participant held the paper data sheet and visually compared it to the data on the screen. During solo read-aloud, the participant read the data out loud from the data sheet, while simultaneously checking the entries on the screen. During partner read-aloud, the administrator read the information from the data sheet out loud and the participant verified the entries on the screen; when the participant hear something that did not match what they saw on the screen, they asked the administrator to repeat it to make sure it was an error. During double entry, the participant entered the data into a blank sheet in the Excel file, and the computer compared it to the existing entry to identify mis-matches between the entries and values that were outside the allowable range for that variable. Regardless of the data checking method to which they were assigned, participants were asked to correct any data entry errors that they found.

Participants checked the data on 25 data sheets. We had entered these data sheets into the computer before participants arrived. When we entered the data, we deliberately introduced errors for the participants to find. The first 5 sheets were completed as practice while the administrator closely watched. The participants then completed the next 20 sheets without help from the administrator.

Data Analysis

We calculated Pearson correlations to determine if there were significant relationships between data checking accuracy and the four computer skills variables.

Results

Data checking accuracy had no significant relationship with any of any of the four computer abilities we tested (see Table 1 and Figure 1). Ninety-five percent confidence intervals for these correlations demonstrated that the relationships between computer skills and data checking accuracy are typically small.

Table 1

The Relationship of Data Checking Accuracy to Computer Abilities

Computer Skills	Correlation	95% Confidence Interval
Computer ability	$r(150) = -.05, p = .540$	[-.21, .11]
Spreadsheet ability	$r(151) = -.01, p = .917$	[-.17, .15]
Typing ability	$r(151) = -.12, p = .147$	[-.27, .04]
Number pad ability	$r(150) = .04, p = .668$	[-.12, .19]

Interestingly, both data checking accuracy and the four computer abilities demonstrated ceiling effects: Most participants had perfect or nearly perfect data (see Figure 2) and most participants reported very strong computer skills (see Table 2).

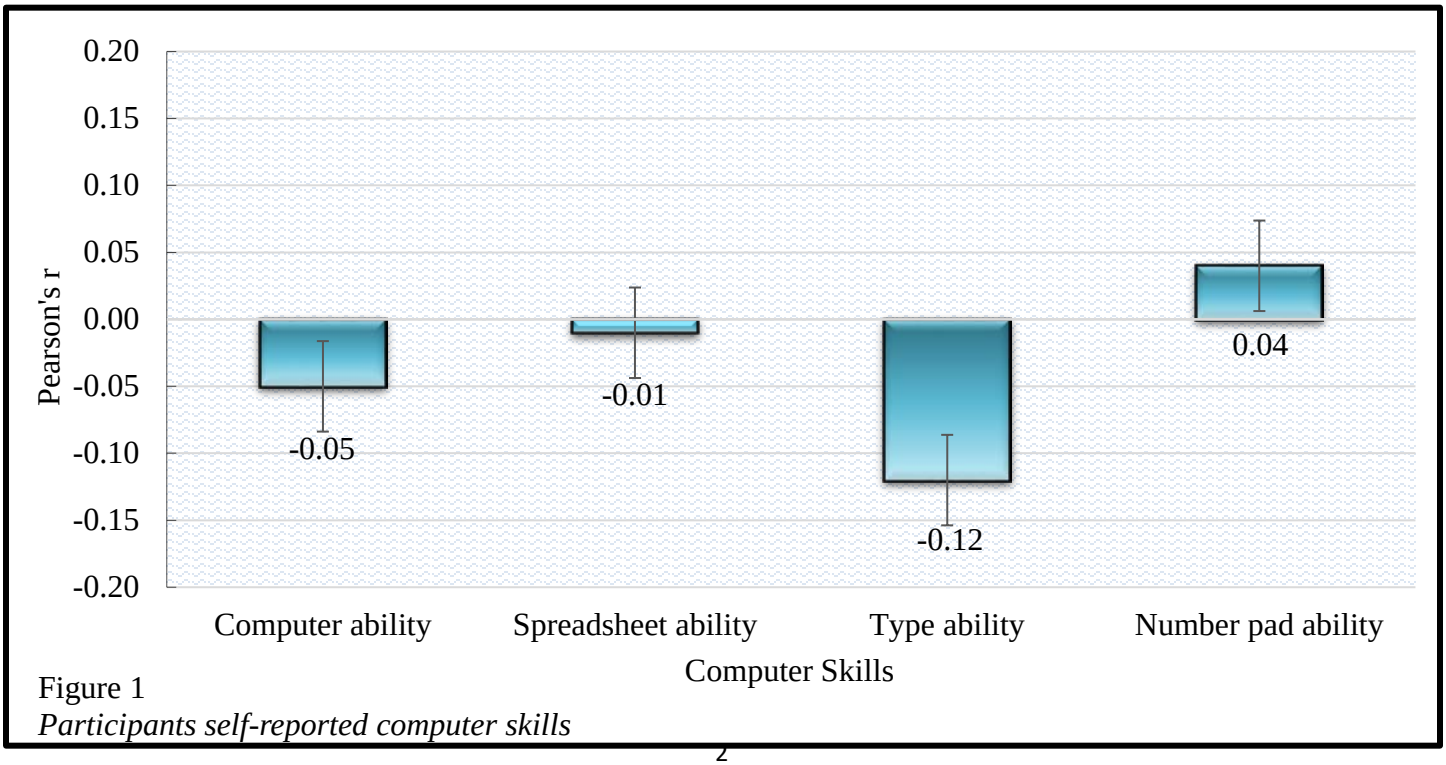

Figure 1

Participants self-reported computer skills

Table 2
Frequency of Responses

Rating of Skills	Computer Ability	Type Ability	Spreadsheet Ability	Number Pad Ability
1. Very low	0	5	2	2
2. Low	2	12	23	22
3. Moderate	46	54	64	48
4. High	89	52	57	58
5. Expert	15	26	7	22

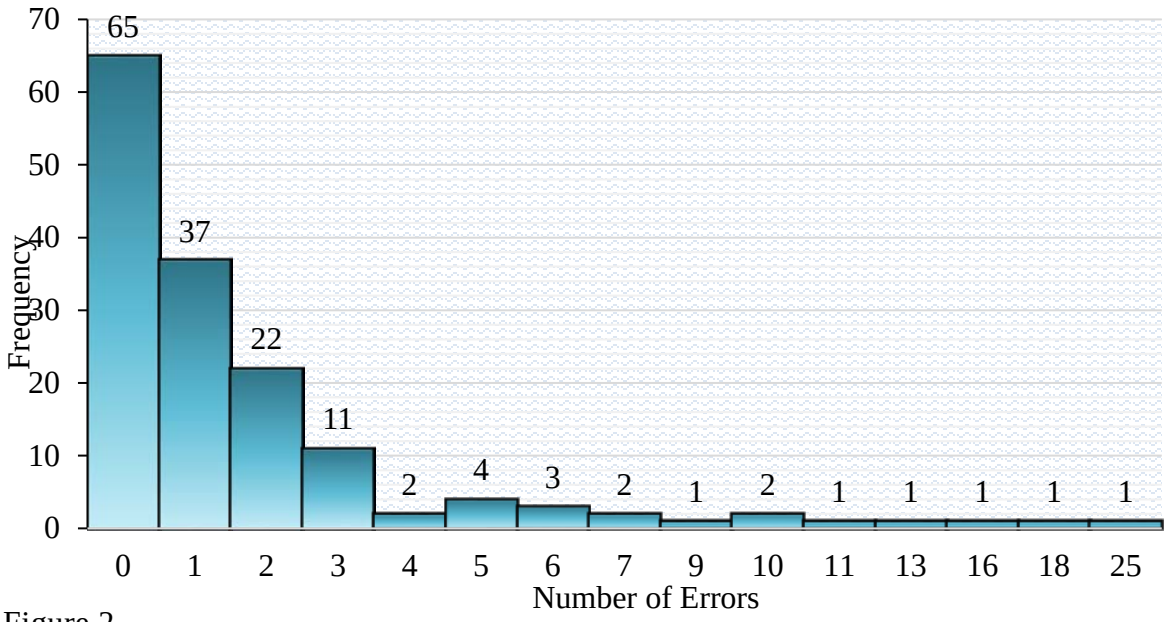

Figure 2
Frequency of Errors

Discussion

The purpose of this study was to determine if there were correlations between computer skills and data checking accuracy. This paper examined four self-reported computer skills: computer ability, spreadsheet ability, typing ability, and number pad ability. No significant relationships were found between computers skills and data checking accuracy.

One limitation of our study is that we used self-reported computer skills. If computer skills were measured using maximum-performance tests, they might have had stronger relationships with data checking accuracy. However, we suspect that the relationships are indeed small in the population of university students. University students generally have good computer skills, which leads to restriction of range on these variables. Moreover, most university students are young adults. Age predicts performance on computer tasks (Czaja & Sharit, 1993), which may explain the high accuracy rates for our participants. The combination of ceiling effects for the predictor variables and ceiling effects for the dependent variable explains the low correlations.

Although undergraduate students are some of the most commonly used data entry personnel in academic research, future research should examine the relationship between computer skills and accuracy among other data entry personnel, such as graduate students and professional data entry personnel. Previous studies found that data entry errors and data checking errors decreased as experience increased (Braa, Heywood, & Sahay, 2012; Czaja & Sharit, 1998; Kinoshita, 2003). We predict that the relationships between computers skills and data checking accuracy will be even smaller in populations that have more data entry experience, because the ceiling effects will be at least as strong as they were here.

Given this, it seems that researchers should not focus on selecting data entry personnel who have strong computer skills. They should instead train their research assistants carefully and use a data checking method that has high accuracy. Previous research has shown that double entry results in fewer errors than other data checking methods (Barchard & Verenikina, 2013). A free double entry program will be available during the poster session.

References

Atkinson, I. (2012). Accuracy of data transfer: Double data entry and estimating levels of error. *Journal of Clinical Nursing, 21*(19-20), 2730-2735. doi:10.1111/j.1365-2702.2012.04353.x

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

Braa, J., Heywood, A., & Sahay, S. (2012). Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania. *Bulletin of the World Health Organization, 90*(5), 379-384. doi:10.2471/BLT.11.099580

Czaja, S. J., & Sharit, J. (1993). Age differences in the performance of computer-based work. *Psychology and Aging, 8*(1), 59-67. doi:10.1037/0882-7974.8.1.59

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

Davidson, L., & Skinner, H. (2010). I spy with my little eye: A comparison of manual versus computer-aided analysis of data gathered by projective techniques. *Qualitative Market Research: An International Journal, 13*(4), 441-459. doi:10.1108/13522751011078845

Fletcher, L. A., Erickson, D. J., Toomey, T. L., & Wagenaar, A. C. (2003). Handheld computers: A feasible alternative to paper forms for field data collection. *Evaluation Review, 27*(2), 165-178. doi:10.1177/0193841X02250527

Grieve, R., Witteveen, K., & Tolan, G. A. (2014). Social media as a tool for data collection: Examining equivalence of socially value-laden constructs. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 33*(4), 532-544. doi:10.1007/s12144-014-9227-4

Kay, R. H. (2007). The role of errors in learning computer software. *Computers & Education, 49*(2), 441-459. doi:10.1016/j.compedu.2005.09.006

Kinoshita, K., Teshima, T., Ohno, Y., Inoue, T., Yamashita, T., Hiraoka, M., & Japanese PCS Working Group. (2003). Logical checking function increases the accuracy of data entry in the patterns of care study. *Strahlentherapie und Onkologie, 179*(2), 107- 112. doi:10.1007/s00066-003-1011-6

Koubek, R. J., Salvendy, G., & Noland, S. (1994). The use of protocol analysis for determining ability requirements for personnel selection on a computer-based task. *Ergonomics, 37*(11), 1787-1800. doi:10.1080/00140139408964951