eefdergi
Cilt 21 Sayı 1 Erzincan Üniversitesi Eğitim Fakültesi Dergisi
e-ISSN 2148-7510
http://eefdergi.erzincan.edu.tr | eefdergi@erzincan.edu.tr
Doi numarası: 10.17556/erziefd.448559

Missing Data Management Practices in L2 Research: The Good, The Bad and The Ugly

Talip GONULAL*

Received date: 27.07.2018 **Accepted date: 26.02.2019**

Abstract

Missing data are one of the frequently encountered problems in quantitative research. When neglected or handled improperly, this problem can have adverse impact on research results. However, the issue of missing data in quantitative second language (L2) research has largely been ignored when compared to the other sister disciplines such as education and psychology. The purpose of this methodological synthesis was, therefore, to investigate the issue of missing data in L2 research, with a particular focus on L2 researchers' current missing data management practices. A total of 143 studies published in six leading L2 journals were reviewed in this synthesis. The results indicated that missing data were indeed quite common in L2 research in that 41% of the studies indicated evidence of missing data, but L2 researchers' management and reporting of missing data was often less than optimal. In light of the results, several directed suggestions were made to improve the rigor and quality of L2 research.

Keywords: Missing data, quantitative research methods, statistical literacy, L2.

* Erzincan Binali Yildirim University, English Language Teaching Department, Erzincan, Turkey; talip.gonulal@erzincan.edu.tr

İkinci Dil Araştırmalarında Kayıp Veri Yönetim Uygulamaları: İyi, Kötü ve Çirkin

Talip GONULAL*

Geliş tarihi: 27.07.2018 **Kabul tarihi: 26.02.2019**

Öz

Kayıp veriler nicel araştırmalarda sıklıkla karşılaşılan sorunlardan biridir. İhmal edildiğinde ya da yanlış şekilde ele alındığında, kayıp veriler araştırma sonuçları üzerinde olumsuz etki yaratabilir. Ancak, eğitim ve psikoloji gibi diğer yakın alanlarla karşılaştırıldığında, ikinci dil araştırmalarında kayıp verilerin durumu göz ardı edilmiştir. Bu nedenle, bu metodolojik sentezin amacı ikinci dil araştırmalarındaki mevcut kayıp veri yönetim uygulamalarını araştırmaktır. Bu sentezde ikinci dil araştırmaları dergisinde yayınlanan toplam 143 çalışma ele alındı. Sonuçlar, ikinci dil araştırmalarında kayıp verilerin gerçekten oldukça yaygın olduğunu gösterdi. İncelenem çalışmaların %41'inde kayıp veri bulgusuna rastlanmıştır. Ancak, ikinci dil araştırmacılarının kayıp veri yönetimi ve sunumu genel olarak çok yetersiz. Sonuçların ışığında, ikinci dil araştırmalarının kalitesini artırmak için çözüm odaklı bazı öneriler sunulmuştur.

Anahtar kelimeler: Kayıp veri, nicel araştırma yöntemleri, istatistiksel okur-yazarlık, ikinci dil.

* Erzincan Binali Yıldırım Üniversitesi, İngilizce Öğretmenliği Bölümü, Erzincan, Turkey; talip.gonulal@erzincan.edu.tr

1. Introduction

Missing data are one of the most ubiquitous issues in data analysis that can occur in almost any discipline. Although this issue is virtually guaranteed in quantitative research, little is still known about why data are missing, how they influence the results and how this problem can be properly handled (McKnight, McKnight, Sidani & Figueredo, 2007). While a number of scholars in different fields such as education (Cheema, 2014; Peugh & Enders, 2004; Rousseau, Simon, Bertrand & Hachey, 2012), counseling psychology (Schlomer, Bauman & Card, 2010) and management information systems (Karanja, Zaveri & Ahmed, 2013) have addressed the problem of missing data, there is a paucity of missing data research in the field of second language acquisition (SLA[1]). Further, considering the recent scholarly work (e.g., Gonulal, 2016, 2018, Gonulal, Loewen & Plonsky, 2017; Loewen et al., 2014) on promoting statistical knowledge and statistical practices in the field, examining this issue in second language (L2) research is a logical and timely step. Given that, the purposes of this methodological synthesis are: (a) to provide a quick snapshot of the prevalence of missing data in L2 research (b) to reveal the current missing data management and reporting practices, and (c) to make directed suggestions towards improving missing data analytic practices in the field of SLA.

1.1. Missing Data

Missing data can be broadly defined as the absence or lack of some kind of information about the phenomena under investigation (McKnight et al., 2007). There can be various reasons triggering this issue. Some of these are fatigue (e.g., failing to respond the last questions on a long test), carelessness (e.g., forgetting to complete the items on the back-side of a test or survey), item difficulty, unwillingness to answer certain items (e.g., what's your recent TOEFL score?), unclear items/questions and limited test time (Enders, 2010; Karanja et al., 2013; Schafer & Graham, 2002). Taking a more systematic approach, McKnigt et al. (2007) highlighted that there are three potential sources of missing data: "missing cases, missing variables and missing occasions" (p. 17). Missing cases refer to the situation in which respondents do not take the test, for instance, because of not showing up for the test, whereas missing variables refer to the situation in which respondents take the test and answer some questions but skip some other questions. As for missing occasions which are likely to occur in longitudinal studies, respondents participate in some parts of the data collection process but do not attend the remaining parts or sessions.

Regardless of how they occur, missing data warrant further attention and explanation because they can have serious impact on study results. First and foremost, according to Cohen and Cohen (1983), missing data lead to a reduction of the available sample size, which can sequentially result in reduced statistical power and increased standard errors. Additionally, McKnight et al. (2007) argue that missing data can affect construct validity, internal validity as well as the generalizability of results. Further, most statistical techniques (e.g., factorial analysis of variance) do not produce optimal results when used with datasets with missing values largely because these techniques are principally designed for complete datasets (Peng, Harwell, Liou & Ehman, 2006).

Pertinent to the potential impact of missing data on research results are the amount and mechanisms of missing data. According to Schafer (1999) and Tabachnick and Fidell (2013), when the amount of missing data is less than 5%, the consequences might be negligible. However, Bennett (2001) stressed that a missing rate of 10% or more can be quite consequential on the results. Another equally or maybe even more important factor than the amount of missing data is the missing data mechanism. Three types of missing data mechanisms, with slightly different names, frequently appear in the missing data literature: *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR). These missing data mechanisms can be of practical assistance to researchers for understanding the nature of missing data they are dealing with before they take any remedial steps.

In the MCAR condition, the missing data represent an arbitrary subset of the hypothesized complete data. Missing data occur by chance in this condition. Put another way, there is no clear association between the missing values and the non-missing or rather observed values (Peng et al., 2006). An example of MCAR would be when data are missing for participants whose survey responses are lost in the mail. In the MAR condition, although the term 'missing at random' may seem misleading and confusing, missing data are not random and indeed are somehow associated with the observed values and do not depend on the missing data themselves (McKnight et al., 2007; Schafer & Graham, 2002). For instance, the missing data would be MAR when only elder respondents have missing values for an IQ test. In this case, the missingness is associated with age but not with what is measured. Although MCAR and MAR conditions have the potential to cause statistical power problems, they are not likely to bias the results (Osborne, 2013).

The MNAR mechanism, which has potentially the most serious influence on the study results, refers to the condition in which the probability of missing data is systematically related to missing data themselves (Osborne, 2013). As an example of MNAR, consider a researcher investigating the statistical knowledge of a group of graduate students by using a comprehensive statistics survey. Students with quantitative research orientation try to answer all the questions whereas students with weak quantitative research orientation or having taken fewer statistics courses appear to skip most questions. In such a case, the missing values are MNAR because only those with low statistical knowledge have missing observations. Therefore, it is likely that the results of this study will be biased since the available data come only from students with strong quantitative research orientation.

Overall, missing data mechanisms require a good understanding on the parts of the researchers. However, a thorough examination of missing data mechanisms is beyond the scope of this synthesis. Readers who are interested in learning more about the technical aspects might refer to Enders (2010), McKnight et al. (2007), and Schafer and Graham (2002).

1.2. Missing Data Handling Methods

The best missing data handling method is not to have any missing data. However, the occurrence of missing data is often inevitable and mostly out of the control of researchers. Even though missing data are usually unintended, researchers can take several remedial steps to effectively deal with the issue of missing data. Missing data management methods can be broadly categorized in two main groups: deletion methods and imputation methods (Cheema, 2014; Enders, 2010; McKnight et al., 2007; Osborne, 2013; Peugh & Enders, 2004). Deletion methods include the omission of the cases or variables with missing values whereas imputation methods rely on filling in the missing values with the imputed ones.

Listwise and pairwise deletion methods are among the commonly used deletion methods. In the listwise deletion method, any cases with missing information are excluded from the analysis. The analysis is then carried out with the remaining complete cases. As an example, imagine that a researcher conducts a study on EFL learners' beliefs about written corrective feedback with 300 Turkish EFL learners using a 25-item survey. However, let say, 100 learners do not answer some of the items on the survey. If the researcher decides to use the listwise deletion method, s/he has to discard all the cases with missing values, which will reduce the sample size to 200. As for the pairwise deletion, it only discards the missing data at the level of variable, not at the level of case. Referring back to the same example above, the pairwise deletion would result in different sample sizes for different variables. For instance, Item 25 might have 240 responses whereas Item 15 might have 200 responses. These two methods are often the default options for certain statistical techniques (e.g., listwise deletion in factor and regression analyses and pairwise deletion in correlation analysis) in some statistical software programs such as SPSS. However, the disadvantages of the deletion methods can outweigh the advantages (Enders, 2010). Listwise deletion can drastically reduce the sample size, which in turn will adversely affect the statistical power. Similarly, pairwise deletion can lead to different sample size for each variable which might result in serious issues especially when a covariance or correlation matrix is analyzed (McKnight et al., 2007). Providing that, these two methods are often considered as 'unwise deletion' methods. Indeed, Wilkinson and Task Force on Statistical Inference (1999) noted that "the two popular methods for dealing with missing data that are found in basic statistics packages—listwise and pairwise deletion of missing values—are among the worst methods available for practical applications" (p. 598).

When it comes to data replacement and data imputation methods, mean substitution appears as one of the most frequent ones. This method relies on the replacement of missing value on a variable with the mean for that variable. The mean substitution method is not without issues. This method is highly likely to yield biased parameter estimates (i.e., sample statistics; see Peugh & Enders, 2004), and to increase the risk of Type I error (Cheema, 2014). The second imputation method is regression imputation or estimation. In this method, a regression equation is computed to come up with predicted values for the missing values using the non-missing values. Although this method is considered superior to mean substitution, it is not highly recommended presumably because it can produce biased estimates (Enders, 2010).

Modern and probably more robust missing data handling methods include maximum likelihood and multiple imputation. Compared to multiple imputation, maximum likelihood method is not a method designed for dealing with missing data per se, but a method commonly used for parameter estimation (McKnight et al., 2007). According to Newman (2014), maximum likelihood method can "directly estimate parameters of interest from incomplete data matrix*…[or]* compute summary estimates [means, SDs, correlations]…then proceed with analysis based on these summary estimates" (p. 383). Similar to maximum likelihood, multiple imputation method is also based on simulating parameter estimates. However, as the name suggests, in the multiple imputation, multiple (usually 3 to 5) imputed data sets are created. Then, the parameter estimates and standard errors are calculated for each imputed data set. In the final step, a single set of estimates is created by averaging the parameter estimates in the imputed data sets, which results in unbiased parameter estimates (Cheema, 2014; Schlomer et al., 2010)

Despite the availability of the various missing data handling methods in the missing data management literature, there are issues attached to more traditional methods (e.g., listwise and pairwise deletion methods) and some recent methods (e.g., maximum likelihood and multiple imputation) are not common practices among researchers mainly due to their complex nature. Yet researchers should be aware of the available missing handling methods and be able to choose and apply the most appropriate one to their data sets. Although this section can serve as a fundamental introduction to missing data handling methods, readers may want to consult other voluminous sources (e.g., Enders, 2010; McKnight et al., 2007; Peugh & Enders, 2004; Schafer & Graham, 2002) for an expanded understanding of the methods.

1.3. Research on Missing Data

As has become apparent in the previous sections, missing data, a ubiquitous problem of quantitative research, have the potential to unfavorably alter research results and therefore require further attention on the part of the researchers. Given the prevalence of and probable consequences attached to missing data, researchers in a variety of fields, mostly in education and psychology, have investigated the missing data issue to reveal how missing data are managed by researchers and to provide suggestions for more rigorous research practices (e.g., Cheema, 2014; Karanja et al., 2013; Peugh & Enders, 2004; Peng et al., 2006; Rousseau et al., 2012; Schlomer et al., 2010). For instance, in their comprehensive and well-written review, Peugh and Enders (2004) attempted to provide a state-of-the-art analysis of missing data in educational research and to report two methodological reviews which were conducted in 1999 and 2003. The results indicated that although missing data were inevitable part of educational research, researchers' missing data management and reporting practices were often less than optimal. More specifically, traditional missing data handling methods were remarkably popular among educational researchers. Further, limited journal space was spent on reporting missing data. In a more recent review, Rousseau et al. (2012) focused on the frequency, handling and reporting of missing data in educational research. The review of the 68 articles published in a well-known educational journal from 2003 to 2007 painted a similar picture in that approximately a two-third of the studies showed evidence of missing data. In addition, simple deletion methods were researchers' first go-to methods of missing data handling. In a similar vein, Karanja et al. (2013) looked at the missing data problem in management information systems (MIS) research to address how MIS researchers dealt with this common issue. Karanja et al. reviewed 749 articles publish in nine MIS journals between 1990 and 2010. When compared to other reviews, this study did not draw different conclusions. That is, approximately 42% of the articles reviewed had evidence of missing data but only 22% explicitly stated the presence of missing data. Similarly, listwise and pairwise deletion methods were again frequently used by MIS researchers whereas more modern and robust methods such as multiple imputation were hardly ever used.

However, alhtough there is a growing body of research on missing data, this has not unfortunately been reflected in the field of SLA. The only study that examined the missing data issue in language research is Pichette et al.'s (2015) review. Pichette et al. investigated the missing binary data issue (i.e., missing responses to dichotomous items/questions such as yes/no questions or agree/disagree items) and what kinds of methods language researchers commonly employed to deal with such missing issues in binary data. Pichette et al. also compared the five commonly-used data insertion methods by focusing on the Cronbach's alpha changes. Although this study plays a pioneering role in missing data in L2 research, it has a very narrow scope because the primary focus was on dichotomous data, excluding studies that include other types of numerical data. Given that, further research in this area is definitely needed to better capture the current situation of missing data issue in L2 research. Such research is important and necessary, especially in light of the "methodological and statistical reform movement" taking place in the field of SLA (Plonsky, 2015, p. 4). Much scholarly attention (e.g., Gonulal, 2016, 2018, Gonulal et al., 2017; Loewen et al., 2014, forthcoming; Norris, Ross & Schoonen, 2015; Plonsky, 2013, 2014, 2015; Plonsky & Gonulal, 2015) has recently been placed on methodological quality in L2 research accentuating the need for increased rigor in statistical analysis and transparent reporting practices. Proper and transparent missing data management practice appears to be the gateway to the current quantitative reform movement.

Overall, the importance of missing data management in quantitative research and the scarcity of discipline-specific research on missing data, taken together with the quantitative turn cropping up in the field of SLA, gave impetus to the present study. The following research questions guided this study:

1. To what extent are missing data common in L2 research?
2. What are L2 researchers' current missing data management practices?

2. Method

In order to address the research questions, a methodological synthesis approach was adopted in this study. Methodological synthesis, part of the meta-analytic tradition, is the systematic review of methodological aspects of quantitative research (Plonsky, 2011). Although this approach has a slightly short history in L2 research, an increasing number of syntheses (e.g., Derrick, 2016; Plonsky, 2013; Plonsky & Gonulal, 2015; Winke, 2014) have been conducted in the recent years. Contrary to meta-analysis or research synthesis, in methodological synthesis, the main focus "is not so much on aggregating substantive findings but, rather, on the methods that have produced them" (Marsden, Thompson & Plonsky, 2018, p. 6). In particular, the current methodological synthesis aims to draw a broad picture of the amount, nature and reporting practices of missing data in quantitative L2 research.

2.1. Study Selection

To investigate the frequency of the occurrence of missing data in L2 research and how L2 researchers handle missing data, quantitative L2 studies published in top-tier SLA journals from 2015 to 2016 were reviewed. Six highly-rated SLA-oriented journals were selected: *Applied Linguistics*, *Language Learning*, *Language Teaching Research*, *Studies in Second Language Acquisition*, *The Modern Language Journal*, and *TESOL Quarterly*. These journals were chosen simply because of their wide L2 research scope, slightly strict publication policy and higher impact factors. Meta-analyses, qualitative studies, literature reviews, reports, forum papers and opinion essays were not the focus of this study, and were therefore excluded.

Table 1. SLA journals from which the primary studies were selected

Journals (2015-2016)	N	%
Applied Linguistics	14	9.8
Language Learning	27	18.8
Language Teaching Research	25	17.5
Studies in Second Language Acquisition	35	24.5
The Modern Language Journal	28	19.6
TESOL Quarterly	14	9.8
TOTAL	143	100

More than 350 articles were manually reviewed within the review period. Of these 350 articles, 130 were quantitative-oriented articles and met the criteria of this synthesis.

Since several articles included more than one study (e.g., two different experiments with different sample sizes in the same article), studies rather than articles were chosen as the unit of analysis. Given that, the total number of studies reviewed in this synthesis was 143 (see Table 1 for the frequency of the studies included from six SLA journals).

2.2. Coding Procedure

After the study identification step, a coding scheme was created to code each study. An initial coding scheme was designed based on the previous missing data reviews in other fields (e.g., Karanja et al., 2013; Peugh & Enders, 2004). The initial coding scheme went through several iterations to have clear and comprehensive coding categories. The final version of the coding scheme included categories such as amount of missing data, missing data handling methods, and missing data reporting practices. Table 2 shows the coding categories in detail.

Table 2. Missing data coding scheme

Variable	Values
Study Identification	
Author(s)	Open
Journal	Applied Linguistics, Language Learning, Language Teaching Research, Studies in Second Language Acquisition, The Modern Language Journal, TESOL
Year	2015-2016
Amount of Missing Data	
Sample size	Open
Missing sample size	Open
Missing rate	Open
Complete sample	0,1
Type of Missingness	
Explicitly	0,1
Implicitly	0,1
Unknown	0,1 + open
Missing Data Handling Methods	
Listwise deletion	0,1
Pairwise deletion	0,1
Mean substitution	0,1
Regression estimation	0,1
Estimation-maximization	0,1
Maximum-likelihood	0,1
Multiple imputation	0,1
Other	0,1 + open
Unknown	0,1 + open
Software	
Software used	Open

Note. 0 = variable is not reported; 1 = variable is reported; open = variable can take any value.

When reviewing the studies, special emphasis was given to methodology and results sections. Further, several key words (i.e., *missing data*, *listwise*, *pairwise*, *imputation*, *substitution*, *exclude*, *remove*) were searched through the studies to identify the prevalence of missing data. However, it was not always easy to locate the missing data because very few researchers explicitly acknowledged the presence of missing data or used missing data treatment methods. In case where missing data were not reported, the degrees of freedom or the sample sizes across different analyses were carefully examined to see if there were any discrepancies (e.g., when there are differences between the reported sample size in the method section and the returned sample size in the results section). Such variations in degrees of freedom or sample size indicated that the data had some missing values.

All the studies were first coded by the researcher. Then, a second quantitatively-oriented coder coded a random sample of 9 studies. A simple percent agreement was calculated and a good inter-coder agreement was found (i.e., 89%; see Landis & Koch, 1977).

2.3. Data Analaysis

The analysis of the selected studies was quite straightforward and similar to previous missing data reviews conducted in other fields (e.g., Karanja et al., 2013; Peugh & Enders, 2004; Rousseau et al., 2012). To answer the research questions, raw frequencies and simple percentages were calculated for each category on the coding scheme. In addition, basic descriptive statistics and confidence intervals were provided when necessary.

3. Results

Results for Research Question 1: To what extent are missing data common in L2 research?

The results indicated that out of the 143 studies reviewed, 57% had no missing data whereas approximately 41% had instances of missing data (see Table 3). In a couple of studies, the prevalence of missing data could not be determined from the information reported. For example, it was not possible to detect the missing data in case where the authors used listwise deletion and did not report it explicitly in the study.

Table 3. Prevalence of Missing Data (N = 143)

Categories	N	%
Studies with complete data	81	56.6
Studies with missing data	58	40.6
Explicitly reported	(42)	(72.4)
Implicitly reported	(16)	(27.6)
Not determined	4	2.8

In addition, in cases where there were instances of missing data, almost 28% of the authors did not explicitly acknowledge the presence of missing data even though there was some incongruity in the initial reported sample sizes and the returned sample sizes in the analyses. It is likely that the authors might have used some types of missing data handling methods (e.g., listwise or pairwise deletion) intentionally but forgot to report it in the study or they might have used unintentionally because certain missing data handling methods (i.e., listwise and pairwise deletion) are default options in the statistical packages.

Table 4. Descriptive statistics

	Min	Max	Median	M	SD	95% CIs
Sample size						
Studies with complete data	24	1270	73	118.87	160.72	[92.11, 145.63]
Studies with missing data	24	1270	76	169.12	234.94	[106.78, 231.46]
Rate of missing data (%)	.63	62	7.15	12.45	14.2	[8.49, 16.40]

As can be seen in Table 4, the average sample size for the collected sample in this methodological synthesis was 118.87 (*SD* = 160.72, *Median* = 73, 95% CIs [92.11, 145.63]). When it comes to the average sample size for the studies that had missing data, it was 169.12 (*SD* = 234.94, *Median* = 76, 95% CIs [106.78, 231.46]). As for the proportion of missing values, the missingness rate ranged from less than 1% to 62%. Additionally, the average missing data rate were 12.45 (*SD* = 14.2). This means that approximately 13 values were missing in a study with 100 values.

Results for Research Question 2: What are L2 researchers' current missing data management practices?

The 58 studies that were identified as having missing data were further investigated to reveal whether and what kinds of methods were employed to remedy the missing data issue. As presented in Table 5, the listwise deletion method was the most frequently used one, followed by the pairwise deletion method. In looking at the results more closely, approximately 89% (*N* = 52) studies employed listwise deletion, pairwise deletion or both. Apart from these old-school deletion methods, two studies used basic mean substitution method and three studies used some other forms of imputation. More specifically, one study used regression estimation, one study estimation-maximization and one study maximum-likelihood. Unfortunately, multiple imputation method, which is considered as one of the most robust and powerful missing data handling techniques, was not employed in any of the studies.

Table 5. Missing Data Handling Methods (N = 58)

Methods	N	%
Listwise deletion	40	69
Pairwise deletion	21	36.2
Mean substitution	2	3.4
Regression estimation	1	1.7
Estimation-maximization	1	1.7
Maximum-likelihood	1	1.7
Multiple imputation	0	0

Note. The percentage does not add up to 100 because several studies employed more than one method.

Although approximately 72% of the 58 studies that showed evidence of missing data issue explicitly stated the presence of the missing data, the level and amount of details given to the missing data reporting varied. To be more specific, there were bad, ugly and good exemplars of missing data treatment and reporting practices in L2 studies, with bad and ugly exemplars outnumbering the good ones.

First, most studies just acknowledged the presence of missing data in one or two sentences either as a footnote or a table note. Second, a number of L2 studies attempted to be more transparent in their management and reporting of missing data, but most often the researchers did not explicitly name the missing data handling methods employed. Third, only few studies showcased good practice in that they explicitly mentioned the missing data and then provided an in-depth treatment to remedy the problem. The following four excerpts show some exemplars of good reporting practices:

> *Excerpt 1:*
>
> First, we used data screening to examine missing data, outliers, and normality. To deal with missing data, we used an expectation–maximization algorithm in which a missing score is replaced by a predictive distribution (Khajavy, Ghonsooly, Hosseini Fatemi & Choi, 2016, p. 12).
>
> *Excerpt 2:*
>
> Missing data were present for three participants on the Spanish Passage Comprehension assessment and therefore full information maximum likelihood was used in all analyses. Seven students had missing data on the language of instruction variable and therefore were not included in the multiple-group analyses (Goodwin, August & Calderon, 2015, p. 610).
>
> *Excerpt 3:*
>
> Data collected from the WSSRLQ were screened and cleaned first. Missing responses, normality, and homogeneity for multivariate analyses were examined thoroughly…In addition, six cases with missing values were removed without imputation because the total proportion of missing values was far less than the cutoff value of 5% (Teng & Zhang, 2016, p. 12).
>
> *Excerpt 4:*
>
> …this reduced the sample size to 1,270, a data reduction of 9%. Next, listwise deletion had to be applied to 114 records because not all relevant questionnaire items were filled out by the student. Listwise deletion was deemed a suitable solution because the data were determined to be missing completely at random ($\chi 2 = 148.46$, df = 143, $p = .36$) using Little's (1988) test and the overall prevalence of missing values was low: Per questionnaire item, fewer than 1% of the responses were missing. Based on Mahalanobis distance ($p \leq .001$), 39 more records were removed because they represented multivariate outliers. The final dataset contained 1,117 records (Denies, Yashima & Janssen, 2015, p. 727).

As reflected in Excerpts 1 through 4, some L2 researchers not only pay attention to the issue of missing data and accordingly take remedial steps but also attempt to model good reporting practice in dealing with missing data. However, there are obviously some discrepancies in the amount of details given to the treatment of missing data.

4. Discussion and Conclusions

The purposes of the current methodological synthesis were three-fold: (a) to examine the prevalence of missing data in quantitative L2 research, (b) to uncover the current state of the missing data management practices among L2 researchers, and (c) to draw further attention to the issue of missing data and make directed suggestions for best missing data analytic practices in the field of SLA.

This synthesis revealed that missing data are quite ubiquitous in quantitative L2 research in that almost 41% studies showed evidence of missing data. This finding is in line with other missing data reviews conducted in different fields such counseling psychology (45%, Schlomer et al., 2010) and educational research (42%, Peugh & Enders, 2004). Given that the studies chosen in this synthesis were published in relatively well-known journals (e.g., *Language Learning, Studies in Second Language Acquisition* and so forth) with rigorous review process and strict publication policy, it would not be wrong to assume that the picture of the whole field would not be essentially different, if not worse.

In addition to such prevalence, the proportion of missing data was not at a trivial level either. In fact, roughly 13% of the data in quantitative L2 research was missing due to various reasons. This rate of missing data is considerably larger than the suggested threshold level (i.e., above 5%; see Schafer, 1999; Tabachnick & Fidell, 2013). This finding implies that L2 researchers should be extra cautious during the data analysis process because the proportion of missing data are often at non-negligible levels.

In spite of the pervasiveness and substantial rate of missing data, there were some variations in acknowledging, treating and reporting missing data. For instance, although some studies acknowledged the issue of missing data, very few studies allotted a reasonable amount of journal space for missing data treatment. Additionally, the missingness issue was not even explicitly stated in about one-fourth of the studies even though there was clear evidence of missing data.

When looking more closely at how L2 researchers treated missing data, the old-fashioned and less robust missing data treatment methods (e.g., listwise deletion and pairwise deletion) appeared to be L2 researchers' first go-to approach. Indeed, approximately 90% the studies that had missing data issues used listwise deletion, pair-wise deletion methods or both. This finding is consistent with Peugh & Enders' (2004) and Peng et al.'s (2006) reviews in which almost all the studies (around 96%) that showed evidence of missing data employed these traditional methods. However, a more recent review on missing data in educational psychology (Dong & Peng, 2013) reported that the rate of the employment of these two deletion methods has decreased to less than 30%. One potential explanation for such high use of traditional data handling methods in L2 research might be related to L2 researchers' use of default options in statistical packages. For instance, the reliability analysis, factor analysis, and regression analysis on SPSS use listwise deletion method by default. Similarly, pairwise deletion method is the default option in correlation analysis on SPSS. However, the Wilkinson and Task Force on Statistical Inference (1999) advised against using them since these are considered as the 'unwise' deletion methods.

Many statistical software programs are now allowing researchers to use superior and more effective missing data analytic methods (e.g., maximum-likelihood and multiple imputation). However, L2 researchers have hardly ever applied these sophisticated and robust methods even when the proportion of missing data was considerably high. When it comes to missing data mechanisms (e.g., MCAR, MAR and MNAR), just a single study took the missing data mechanisms into consideration when dealing with the missing data issue.

When reviewed in its entirety, this methodological synthesis revealed that although the issue of missing data is inevitable in quantitative L2 research, it is often neglected or handled in a relatively superficial manner. There are several potential reasons for the current state of the missing data practices and why L2 researchers appear to vary in their acknowledging, treating and reporting missing data. First, it seems that not many L2 researchers are aware that the consequences of missing data on the results of the study can be profound (e.g., ameliorating the statistical power or biasing the parameter estimates; Enders, 2010; McKnight et al., 2007; Peugh & Enders, 2004) especially when the proportion of missing data is large. Consequently, L2 researchers tend to either ignore or put less emphasis on the missing data problem during data-screening and data analysis stages. This might, to a great extent, be related to the current level of statistical literacy in the field in that L2 researchers' and applied linguists' statistical training and knowledge of statistics is mostly limited to basic descriptive and common inferential statistics (Gonulal, 2016, 2018; Gonulal et al., 2017; Loewen et al., forthcoming). It is, thus, not surprising to see that contemporary missing data handling techniques and missing data mechanisms are not familiar to many L2 researchers. Second, the issue of missing data has drawn less editorial and scholarly attention in the field of SLA. Although, in light of the quantitative turn taking place in the field, some editorial work (e.g., journal guidelines on publishing quantitative research) has recently been undertaken to highlight transparency and improve the quality of reporting in L2 research, the problem of missing data has been overlooked. To illustrate, among the six SLA journals included in this synthesis, only one journal (i.e., TESOL Quarterly) has emphasized missing data in their guidelines for prospective authors with the following words: "Describe how missing cases were addressed (e.g., if an expected participant was absent, what sort of follow-up was conducted to ensure adequate sampling)" (Mahboob et al., 2016, p. 50). Given that, there are obviously limited guidelines on how to properly handle and report missing data in L2 research. Third, it is also likely that L2 researchers are aware of the missing data in their data, knowledgeable about the potential consequences of missing data on the results and familiar with the contemporary missing data handling methods, but precious journal space might limit the information they want to report regarding missing data. However, considering that most journals now accept supplementary materials wherein further details about the studies can be provided, this explanation seems to hold less true.

Although the current synthesis reveals the frequency of the occurrence of missing data in L2 research and raises several questions regarding the missing data analytic practices of L2 researchers, the findings should be handled with care due to a number of limitations. First and foremost, this is a small-scale methodological synthesis in that the number of the primary studies included in the analysis was relatively small. Further, the selection of the studies was restricted to a specific time span (i.e., 2015-2016) and certain journals. Taken together, these factors might have inflated or deflated the results. Therefore, future research might expand on this research area by including studies published in a variety of journals and in additional venues as well (e.g., books, theses, and dissertations) to get a more complete picture of the missing data issue in L2 research. Similarly, future studies might also focus on whether there is a change in the missing data analytic practices over time. Additionally, contrary to the present synthesis which adopted a slightly broad scope, future research might take a narrow focus in investigating the missing data problem (e.g., missingness issue in binary data; see Pichette et al., 2015). For instance, a methodological synthesis looking at the studies that employed surveys and questionnaires might tell us more about the state of the art of missing data in L2 research because surveys and questionnaires are notoriously known for their missing data rate.

As an initial foray into the issue of missing data in L2 research, this methodological synthesis attempted to showcase the situation of the problem in the field and to provide a snapshot of the missing data analytic practices of L2 researchers. Although there were some 'good' exemplars of proper missing data treatment and reporting practices, 'bad' and 'ugly' exemplars mostly exceeded in quantity. In light of the results, the current study suggests that missing data analysis should be added to the statistical repertoire of L2 researchers and be a routine part of data screening and data analysis. For this purpose, the dedicated and continued efforts of journal editors, reviewers and slatisticians[2] are needed to make missing data analysis a best practice in the field.

References

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, *25*(5), 464-469.

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research*, *84*(4), 487-508.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/ correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Denies, K., Yashima, T., & Janssen, R. (2015). Classroom versus societal willingness to communicate: Investigating French as a second language in Flanders. *The Modern Language Journal*, *99*(4), 718-739.

Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly*, *50*(1), 132-153.

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus, 2*(1), 1-17.

Enders, C. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Gonulal, T. (2016). *Statistical literacy among second language acquisition graduate students*. Unpublished dissertation. Michigan State University, East Lansing.

Gonulal, T. (2018). An investigation of the predictors of statistical literacy in second language acquisition. *Eurasian Journal of Applied Linguistics, 4*(1), 49-70.

Gonulal, T., Loewen, S.., & Plonsky, L. (2017). The development of statistical literacy in applied linguistics graduate students. *ITL – International Journal of Applied Linguistics, 168*(1)*, 4-32*.

Goodwin, A. P., August, D., & Calderon, M. (2015). Reading in multiple orthographies: Differences and similarities in reading in Spanish and English for English Learners. *Language Learning*, *65*(3), 596-630.

Karanja, E., Zaveri, J., & Ahmed, A. (2013). How do MIS researchers handle missing data in survey-based research: A content analysis approach. *International Journal of Information Management*, *33*(5), 734-751.

Khajavy, G. H., Ghonsooly, B., Hosseini Fatemi, A., & Choi, C. W. (2016). Willingness to communicate in English: A microsystem model in the Iranian EFL classroom context. *TESOL Quarterly*, *50*(1), 154-180.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Loewen, S., Gonulal, T., Isbell, D. R., Ballard, L., Crowther, D., Lim, J., Maloney, J., & Tigchelaar, M. (forthcoming). How knowledgeable are SLA researchers about basic statistics? Data from North America and Europe. *Studies in Second Language Acquisition.*

Loewen, S., Lavolette, E., Spino, L. A., Papi, M., Schmidtke, J., Sterling, S., & Wolff, D. (2014). Statistical literacy among applied linguists and second language acquisition researchers. *TESOL Quarterly*, *48*, 360–388.

Mahboob, A., Paltridge, B., Phakiti, A., Wagner, E., Starfield, S., Burns, A., Jones, R.H. & De Costa, P. I. (2016). TESOL Quarterly research guidelines. *TESOL Quarterly*, *50*(1), 42-65.

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861-904.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: Guilford Press.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, *17*(4), 372-411.

Norris, J. M., Ross, S. J., & Schoonen, R. (2015). Improving second language quantitative research. *Language Learning, 65*(S1), 1-8.

Peng, C., Harwell, M., Liou, S., & Ehman, L. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real data analysis* (pp. 31–78). Charlotte, NC: New Information Age.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of educational research*, *74*(4), 525-556.

Plonsky, L. (2011). *Study quality in SLA: A cumulative and developmental assessment of designs, analyses, reporting practices, and outcomes in quantitative L2 research* (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*(4), 655-687.

Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal*, *98*(1), 450-470.

Plonsky, L. (2015). *Advancing quantitative methods in second language research*. New York, NY: Routledge.

Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning, 65*, (S1), 9-36.

Pichette, F., Béland, S., Jolani, S., & Lesniewska, J. (2015). The handling of missing binary data in language research. *Studies in Second Language Learning and Teaching, 5*(1), 153–169.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*(3), 537-560.

Osborne, J. W. (2013). *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Thousand Oak, CA: Sage.

Rousseau, M., Simon, M., Bertrand, R., & Hachey, K. (2012). Reporting missing data: a study of selected articles published from 2003–2007. *Quality & Quantity*, *46*(5), 1393-1406.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, *8*(1), 3-15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling psychology*, *57*(1), 1-10.

Tabachnick B., & Fidell, L. (2013). *Using multivariate statistics*. Boston: Pearson Education Inc.

Teng, L. S., & Zhang, L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal*, *100*(3), 674-701.

Wilkinson, L. & Task Force on Statistical Inference, (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594-604.

Winke, P. (2014). Testing hypotheses about language learning using structural equation modeling. *Annual Review of Applied Linguistics, 34*, 102-122.

Notes

[1]In this study, ,the whole field was referred to as SLA, which in this paper encompasses SLA, applied linguistics, language assessment and testing. Further, SLA and L2 research were used interchangeably in this study.

[2]This term was coined by the researcher to describe SLA researchers who are highly knowledgeable in applied statistics and well-trained to properly use an array of statistical techniques within L2 research.

Research article: Gönülal, T. (2019). Missing data management practices in L2 research: The good, the bad and the ugly. *Erzincan University Journal of Education Faculty, 21* (1), 56-73.